




O PROCESSO DE CONSTRUÇÃO DO CORPUS DE REFERÊNCIA EM CIÊNCIA DA INFORMAÇÃO

The process of construction of the reference corpus in Information Science

Bruno Felipe de Melo **SILVA**
Biblioteca Central, Universidade Federal de Alagoas,
Maceió, Brasil
bruno.fms545@gmail.com
<https://orcid.org/0000-0002-6697-3564> 

Renato Fernandes **CORREA**
Departamento de Ciência da Informação, Universidade
Federal de Pernambuco, Recife, Brasil
renato.correa@ufpe.br
<https://orcid.org/0000-0002-9880-8678> 

Mais informações da obra no final do artigo 

RESUMO

Objetivo: O presente trabalho propõe e aplica um percurso metodológico para a construção de corpus de referência na área de Ciência da Informação.

Método: A metodologia da pesquisa pode ser classificada como exploratória, empírica e estudo de caso. Para a pesquisa foi delimitado o uso do corpus composto por 60 artigos selecionados por Souza (2005) e a adoção do Tesauro Brasileiro de Ciência da Informação (TBCI) como linguagem de indexação. O percurso metodológico proposto consistiu da adaptação de um modelo colaborativo de indexação social, assim como a elaboração de uma política de indexação social e a configuração e uso de um software para a construção de uma folksonomia assistida por grupos de usuários. A folksonomia assistida foi então analisada visando à compilação do corpus de referência.

Resultado: O modelo colaborativo de indexação social foi obtido por meio da configuração da plataforma de gerenciamento de coleção denominada Tainacan, mais especificamente na configuração e preparação de uma coleção no software. Foi desenvolvida uma política de indexação social assistida e houve instrução aos grupos de usuários indexadores de forma a guiá-los no processo de indexação. A análise da folksonomia resultante se deu mediante comparação dos termos com as palavras-chave dos autores e com termos do TBCI, permitindo a posterior correção e exclusão de termos inadequados na compilação do corpus.

Conclusões: Conclui-se que o percurso metodológico proposto se mostra eficaz na construção de corpus de referência no âmbito da Ciência da Informação.

PALAVRAS-CHAVE: Folksonomia assistida. Indexação social. Modelo colaborativo de indexação social. Política de indexação social. Ciência da Informação.

ABSTRACT

Objective: This work aims to propose and apply a methodological approach for the construction of a reference corpus for Information Science area.

Methods: The research methodology is exploratory, empirical and case study. For the research was delimited the use of the corpus of 60 articles selected by Souza (2005), and the *Tesauro Brasileiro de Ciência da Informação* (TBCI) as indexing language. The methodological approach consists of the adaptation of a collaborative model of social indexing, the elaboration of a policy of social indexing, and the configuration and use of software as steps for the construction of an assisted folksonomy. After this, the analyses of the constructed assisted folksonomy and the compilation of the reference corpus are the last steps.

Results: The collaborative model of social indexing results of the configuration of the collection management platform called Tainacan, more specifically in the configuration and preparation of a collection in the software. An assisted social indexing policy was developed and user groups receive instructions to guide them through the indexing process. The analysis of the resulting folksonomy was made by comparing the terms with the authors' keywords and with terms from the TBCI, allowing the correction and exclusion of inappropriate terms in the compilation of the corpus.

Conclusions: It concludes that the proposed methodological approach is effective in the construction of reference corpus of Information Science area.

KEYWORDS: Assisted folksonomy. Social indexing. Social indexing collaborative model. Social indexing policy. Information Science.

1 INTRODUÇÃO

A busca por melhores resultados no processo de recuperação da informação exige uma melhor qualidade na indexação de documentos. Segundo Gil Leiva, Rubi e Fujita (2008, p. 234) “a consistência da indexação é um elemento característico tanto do processo quanto do resultado do tratamento temático da informação”.

Visando uma maior consistência e agilidade no processo de indexação, a indexação automática tem sido pesquisada na área da Ciência da Informação no âmbito nacional e internacional desde a década de 1950 (CORRÊA; LAPA, 2013).

Segundo Hjørland (2008), a indexação automática é realizada por meio de procedimentos algorítmicos. Adicionalmente, Lancaster (2004) a define como “um processo que ocorre quando o computador é utilizado para substituir a indexação manual realizada por um indexador”. Este tipo de indexação acontece automaticamente em diversos sistemas informacionais.

Os estudos em torno da indexação automática no Brasil ocorreram mediante esforços individuais realizados nos cursos de pós-graduação, e que tinham como foco a análise de frequência, análise semântica e sintática do termo (CORRÊA; LAPA, 2013).

Bandim e Corrêa (2018) apresentam os sistemas de indexação automática mais citados na literatura brasileira de Ciência da Informação, sendo eles: 1) o sistema PRECIS - usado na construção automática de índices de assunto; 2) o AUTOMINDEX/II - utilizado para gerar tabelas de frequência de aparecimento dos descritores encontrados em cada documento; e 3) o SISA - no uso de linguagens documentárias para indexação semiautomática e indexação automática por atribuição.

No aperfeiçoamento dos sistemas de indexação automática, avaliar a qualidade do resultado da indexação realizada é tão importante quanto a própria indexação. Assim, tais sistemas precisam ser avaliados na indexação de *corpora* de documentos cuja qualidade na indexação intelectual é a melhor possível, visando a correta avaliação e a comparação do desempenho dos mesmos. Tais *corpora* de documentos indexados são denominados corpus de referência (HASAN; NG, 2014) ou *gold standard* (Golub et al. 2016), e permitem avaliar o quão próximo está a qualidade da indexação automática em relação à indexação intelectual (KIM et. al., 2013).

Neste contexto, o presente trabalho propõe e aplica um percurso metodológico para a construção de corpus de referência de artigos científicos escritos em português da área

de Ciência da Informação, visando a posterior aplicação do corpus na avaliação de sistemas de indexação automática.

Na literatura internacional de Ciência da Informação, os trabalhos mais recentes sobre indexação automática fazem uso de *corpora* de referência para avaliação de sistemas de indexação automática e análises comparativas entre eles. Entre esses trabalhos, destacamos os que realizam a indexação automática de textos escritos em português, tais como (SOUZA; LEIVA, 2016) e (LEIVA, 2017).

O corpus de documentos utilizado no presente trabalho consiste de 60 artigos de periódicos em Ciência da Informação selecionados por Souza (2005). Este corpus foi escolhido levando em consideração o idioma e domínio dos documentos, bem como seu uso em pesquisas anteriores na temática indexação automática, tais como: (SOUZA, 2006), (SOUZA; RAGHAVAN, 2006) e (SOUZA; RAGHAVAN, 2014) – na extração e seleção de sintagmas nominais do texto completo; (BANDIM; CORRÊA, 2018) e (BANDIM; CORRÊA, 2019) – na indexação por atribuição do texto completo; (CORRÊA; CELERINO, 2019) – na normalização de sintagmas nominais relevantes presentes no título ou resumo do artigo.

O percurso metodológico proposto visa aperfeiçoar a indexação feita pelos autores via palavras-chave nos respectivos documentos, por meio da indexação social de leitores via atribuição de termos de um tesauro da área, ou mais especificamente, na indexação social assistida.

Levando em conta que o processo de indexação social assistida resulta em uma folksonomia assistida, tem-se a folksonomia assistida como elemento chave para a construção do corpus de referência, visando uma melhor qualidade na indexação dos artigos científicos do corpus.

Assim, como problema de pesquisa tem-se a pergunta: Qual percurso metodológico possibilita construção do corpus de referência via folksonomia assistida?

O objetivo deste artigo é propor e aplicar um percurso metodológico para a construção de corpus de referência na área de Ciência da Informação para avaliação de sistemas de indexação automática.

O percurso metodológico conta com as seguintes etapas: 1) Configuração do modelo colaborativo de indexação social na plataforma Tainacan; 2) Configuração da coleção no Tainacan; 3) Desenvolvimento de uma política de indexação social assistida; 4) Indexação social assistida; 5) Análise da folksonomia assistida. Tais etapas serão discutidas na seção 4.

Porém, visando um melhor entendimento do leitor, são apresentados na seção 2 os conceitos e recursos principais que compõem o percurso metodológico e na seção 3 é apresentada a metodologia da presente pesquisa.

2 FUNDAMENTAÇÃO TEÓRICA

A presente seção tem como intuito discorrer acerca dos aspectos teóricos necessários para entendimento da pesquisa. Serão abordados aspectos ligados a folksonomia assistida, modelos colaborativos de indexação social, política de indexação social e o software Tainacan.

2.1 Folksonomia assistida

De acordo com Roschelle e Teasley (1995), a colaboração consiste no empenho mútuo de indivíduos em um esforço coordenado para auxiliar/ajudar na realização de uma tarefa.

A colaboração na indexação de documentos em ambientes digitais dá origem à folksonomia. A folksonomia pode ser entendida como resultado da indexação dos usuários em ambientes colaborativos por meio da etiquetagem (CORREA; SANTOS, 2018). A folksonomia em sua essência traz consigo diversos fatores positivos e negativos na representação da informação.

Implementada em algumas bibliotecas nos Estados Unidos da América e recomendada pela Biblioteca do Congresso (YEDID, 2013), a etiquetagem colaborativa em catálogo tem trazido proveitos para essas instituições e para seus usuários, pois permite que os catálogos se apresentem mais moldados aos hábitos dos utilizadores e se melhore o acesso aos recursos (YEDID, 2013).

Segundo Guy e Tonkin (2006), a falha nos sistemas de folksonomia decorre da imprecisão dos termos de marcação utilizados nesses sistemas. Os usuários dos sistemas que utilizam folksonomia inserem livremente as *tags*, o que significa que as *tags* são muitas vezes ambíguas, excessivamente personalizadas e inexatas.

Santarém Segundo (2010, p.209), um dos primeiros autores nacionais a abordar a temática folksonomia assistida, afirma que essa "vem ao encontro à necessidade de auxiliar o usuário na descrição do assunto do recurso a ser depositado, em relação ao domínio do conhecimento do qual o repositório faz parte". Nota-se então que a proposta

do autor visa apresentar soluções a questões tratadas como problemáticas na folksonomia.

Nessa proposta, o indivíduo mantém sua liberdade em definir quais termos irão representar o conteúdo. Essa questão é essa tratada como determinante para não descaracterizar a folksonomia, acrescentando e apresentando a eles possibilidades de escolha de termos de uma hierarquia pré-existente.

Santos (2016, p.75) descreve que a gênese da folksonomia assistida está na atividade de apoio ao usuário na definição das *tags* mais pertinentes para representação dos conteúdos informacionais. Como consequência dessa ação do sistema colaborativo, o usuário passa a evitar abreviações, problemas com plurais e singulares, ou ainda palavras que possam dificultar a recuperação da informação (SANTARÉM SEGUNDO, 2010, p.181).

Corroborando com o princípio da assistência ao usuário na construção da folksonomia assistida, entre os estudos sobre a distribuição de *tags* em folksonomias, Zubiaga *et al.* (2012) revelam que os utilizadores são influenciáveis pelos sistemas de sugestão de etiquetas, tendendo a selecionar etiquetas sugeridas mais do que a colocar as suas próprias.

Entre os aspectos positivos da folksonomia assistida pode-se apontar o surgimento de um relacionamento horizontal entre as etiquetas que representam o mesmo conteúdo, além da padronização dos termos atribuídos tanto pelo usuário quanto pelo autor do documento. A folksonomia assistida tende a apresentar menos termos isolados e apresentar mais termos específicos, além de termos com maior significado em relação ao conteúdo indexado.

O sistema colaborativo utilizado na construção da folksonomia assistida favorece a colaboração através da sugestão de termos que já foram inseridos anteriormente pelos usuários ou através de termos oriundos de estruturas já existentes, como por exemplo, os tesouros. As regras de funcionamento dos sistemas serão determinadas pelo modelo colaborativo de indexação social adotado, tema que será discutido na próxima subseção.

2.2 Modelo colaborativo de indexação social

Os modelos colaborativos de indexação social são aplicados com o intuito de dar maior autonomia no que se refere à organização da informação por parte dos usuários dentro dos ambientes digitais.

Santos e Corrêa (2015) acrescentam que tais modelos podem ser úteis no desenvolvimento de metadados com o intuito de melhorar a representação das informações nos espaços web, bem como para o aperfeiçoamento dos objetos, processos e produtos construídos no âmbito da ciência da informação, como os tesauros, as ontologias e as taxonomias.

Santarém Segundo (2010) apresenta o modelo de representação iterativa para gerar uma “folksonomia assistida” e estabelecer um processo de recuperação semântica da informação dentro dos repositórios digitais. Alinhando os recursos da Web 2.0 e Web Semântica, tal modelo colaborativo de indexação social se baseia na coexistência da folksonomia e dos vocabulários controlados para a retroalimentação e a descrição da informação (SANTARÉM SEGUNDO; VIDOTTI, 2011).

O surgimento desse modelo tem como premissa inicial agregar funcionalidades aos repositórios digitais, bem como ambientes que se desenvolvam a partir de estruturas de representação do conhecimento, como as taxonomias, tesauros e ontologias, oferecendo ao usuário a chance de atribuir suas próprias *tags*, mantendo-as com uma ligação semântica a um vocabulário especializado (SANTAREM SEGUNDO; SIQUEIRA, 2013).

Um segundo modelo colaborativo de indexação social é apresentado por Silva (2013) visando a busca de registros em um catálogo web facetado, chamado pelo autor de Facetlog. O modelo busca oferecer uma estratégia complementar à atribuição de etiquetas através da seleção de termos provenientes de uma taxonomia facetada. A ideia é permitir que o usuário possa realizar de forma colaborativa a descrição dos objetos informacionais usando a etiquetagem, tendo como característica a linguagem natural, ou, como segunda opção, selecionando as categorias da taxonomia facetada. Vale apenas reforçar que a taxonomia facetada descrita pelo autor consiste de um conjunto de facetas com conceitos organizados de forma hierárquica.

Segundo Silva (2013), o modelo web facetado apresenta a necessidade específica do gerenciador ou administrador da coleção deter conhecimentos em torno dos princípios de representação da informação para a modelagem dos termos, assim como para a aprovação/reprovação das contribuições dos usuários que venham a ocorrer.

Nesse modelo a taxonomia facetada e a navegação facetada complementam a indexação e a recuperação da informação. Isso traz ganhos no momento em que o usuário não souber o descritor correto a ser digitado na busca, ou quando a busca por termos livres não demonstrar resultados adequados à necessidade do usuário.

Santos (2016) destaca que os modelos colaborativos estão agregados aos sistemas colaborativos, tendo em vista que antes da implementação ou utilização de um sistema colaborativo é necessária a especificação e adesão a um modelo de colaboração.

Assim, um modelo colaborativo de indexação social deve ser especificado para a construção de uma folksonomia assistida, delimitando regras de interação e colaboração entre usuários, visando alcançar uma maior qualidade semântica na descrição dos recursos em ambiente digital.

2.3 Política de indexação social

Para a execução da atividade de indexação, a definição de diretrizes que visam direcioná-la corretamente deve ser considerada na fase de desenvolvimento dos sistemas de recuperação de informação. Nesse sentido, Lancaster (2004) reforça que se o sistema de recuperação de informação tiver uma política de indexação bem descrita, levando em conta o perfil da comunidade de usuários, haverá maiores probabilidades de efetividade em relação à realização dos seus objetivos.

Carneiro (1985, p.221) afirma que uma política de indexação deve servir como guia para tomada de decisões, bem como levar em conta alguns fatores: características e objetivos da organização, determinantes do tipo de serviço a ser oferecido; identificação dos usuários, para atendimento de suas necessidades de informação; e recursos humanos, materiais e financeiros, que delimitam o funcionamento de um sistema de recuperação de informação.

Atualmente, nos ambientes digitais onde existe a participação dos indivíduos, têm ocorrido problemas referentes à escrita e até contextualização do que vem a ser indexado. Assim, existe a necessidade de especificação de uma política de indexação social, no intuito de amenizar as questões problemáticas relacionadas com a utilização da linguagem natural na prática da indexação social. Por conta disso, apresentar uma política com diretrizes, pode vir a ser o primeiro passo para potencializar a produção de termos por parte dos indexadores.

A busca por melhor qualificar e estruturar a representação do conhecimento em ambientes colaborativos, possibilitando que as *tags* tenham um maior significado em relação ao objeto informacional tratado, faz com que surja a possibilidade de construção da folksonomia assistida atrelada a uma política de indexação social.

Para isso, bem como Fujita e Rubi (2006, p.49) afirmam, é preciso ter em mente que a política de indexação dentro de um sistema de informação deve ser entendida como uma filosofia pertinente aos objetivos de recuperação da informação e não somente como uma lista de procedimentos a serem seguidos durante a realização da indexação.

Essa afirmação garante uma dimensão à política de indexação que vai muito além de questões técnicas, e que segundo as próprias autoras, foi à falta desse entendimento que determinou a defasagem nos estudos acerca da temática (FUJITA; RUBI, 2006).

Sobre o desenvolvimento de políticas de indexação, quando se refere à aplicação dessas políticas em ambientes de indexação social, pouco se encontra na literatura científica, fazendo com que seja necessária para sequencia desse estudo uma adequação dos conceitos considerados tradicionais para tal conjuntura.

Guimarães (2000) descreve que o estabelecimento de uma política de indexação contribuirá para que o usuário e o documento deixem de ser sujeito e objeto para se tornarem dois sujeitos que interagem, uma vez que o usuário estará sempre recriando o documento e, por consequência, alimentando novamente o sistema. Assim, a partir do entendimento dessa afirmação, a política de indexação é um instrumento importante para delinear a participação do indivíduo que tem acesso à descrição do conteúdo.

Ao estabelecer uma política de indexação social, deve ser levada em consideração a importância que a construção coletiva passa a ter, e como o indivíduo contribui dentro da perspectiva da indexação.

2.4 Tainacan

O Tainacan¹ é um projeto que surge no ano de 2014, a partir de uma parceria entre o MediaLab da Universidade Federal de Goiás e o Ministério da Cultura, tendo como objetivo se tornar a referência técnica para a Política Nacional de Acervos Digitais.

O software oferece soluções técnicas para que o usuário possa classificar e contextualizar os conteúdos lá disponibilizados, dando visibilidade às coleções publicadas, para ainda permitir a preservação e divulgação de parte da memória. Seu desenvolvimento teve como intuito efetivar ao máximo a inteligência coletiva por meio da colaboração, se apresentando de fácil configuração e personalização quanto ao processo de gestão de coleções digitais.

¹ <http://tainacan.org/>

A ideia de coletividade no qual se baseia a web 2.0 pode ser visualizada de forma efetiva na maneira amigável e responsiva dada ao usuário final quando utiliza a ferramenta.

Para isso, o Tainacan apresenta requisitos que podem ser observados na sua organização a partir dos componentes da arquitetura da informação fornecidos (como busca, navegação, organização e rotulagem) e posteriormente, na execução das funções essenciais de um repositório digital, como por exemplo: administração, armazenamento, preservação e disponibilização de conteúdos informacionais nos diversos formatos eletrônicos.

Dentre os eventos (ações) que visam efetivar o caráter colaborativo do Tainacan, estão: adicionar um item, editar uma *tag*, invalidar um item, remover uma *tag*, validar um item, adicionar um metadado, adicionar uma classificação, editar um metadado, invalidar uma classificação, remover um metadado, validar uma classificação, adicionar uma coleção, adicionar uma categoria, remover uma coleção, mover uma categoria, adicionar um comentário, remover uma categoria, editar um comentário, adicionar uma *tag* e remover um comentário.

A base de desenvolvimento do Tainacan é o software livre WordPress, propício ao desenvolvimento de soluções para a Web. Por essa razão, engloba em seu cerne diversas facilidades e funcionalidades de caráter técnico que o tornam compatível com tendências das tecnologias da web. A plataforma consiste em quatro módulos que podem ajudar na gestão de repositórios, ontologias, documentos e museus.

Como diferencial, o software tem por objetivo agregar características que facilitem a integração com as mídias sociais, permita um alto grau de participação social de usuários e possibilite a criação de repositórios de fácil utilização, configuração e efetivação, além de livre acesso.

3 METODOLOGIA

O desenvolvimento dessa pesquisa levou em consideração o conceito de fundamento metodológico descrito por Gil (2002), Marconi e Lakatos (2006), no intuito de deixar claros os procedimentos sistemáticos e racionais, condensando a metodologia científica, técnicas de pesquisa e metodologia do trabalho científico.

A pesquisa aqui apresentada se encontra classificada como exploratória e empírica, sendo construída a partir de um estudo de caso.

O aspecto exploratório é justificado pelo pouco conhecimento acumulado acerca da folksonomia assistida e dos modelos colaborativos de indexação social, bem como do percurso metodológico para a construção de corpus de referência utilizando esses elementos.

Já no que se refere à parte empírica, tem-se o planejamento e realização de um experimento visando a construção de corpus de referência com base em folksonomia assistida gerada por grupos de indexadores para um corpus de documentos, bem como a posterior análise da folksonomia gerada.

O estudo de caso consiste da construção de corpus de referência para a área de Ciência da Informação, levando em conta os 60 artigos que compõem o *corpus* de Souza (2005) e o Tesouro Brasileiro de Ciência da Informação (TBCI) como linguagem de indexação (PINHEIRO; FERREZ, 2014).

À medida que for sendo planejado e realizado o experimento envolvendo o estudo de caso, será delineado o percurso metodológico que determina o processo de construção de corpus de referência em Ciência da Informação.

Nos parágrafos a seguir, são descritas sequencialmente as etapas da pesquisa.

Inicialmente, foi realizada uma adequação no modelo colaborativo de indexação social de Silva (2013). Tal modelo será especificado e aplicado através da configuração da plataforma Tainacan para uso por grupos de usuários indexadores.

Como o domínio do corpus de referência foi delimitado como sendo a área de Ciência da Informação, foi escolhido o corpus de documentos de Souza (2005) para ser alvo da indexação social assistida. A escolha desse corpus se deu inicialmente por ser redigido no idioma português, ser do domínio da Ciência da Informação e pela sua utilização em outras pesquisas acerca da indexação automática, dando condições dos resultados dos estudos serem comparados em outra oportunidade.

O corpus é constituído por publicações da revista DataGramZero, sendo um total de 29 documentos no formato HTML, e da revista Ciência da Informação, com 31 documentos no formato PDF.

Portanto, a configuração do Tainacan foi realizada por meio da criação de uma coleção de documentos. Na configuração e preparação da coleção no Tainacan, levou-se em conta a adaptação do modelo colaborativo de indexação social de Silva (2013) (quanto a papéis, atividades e regras de interação entre usuários), bem como a inserção na coleção dos artigos do corpus de Souza (2005) com respectivos metadados descritivos. Foi utilizada nesta pesquisa a versão web do Tainacan.

Posteriormente foi desenvolvida uma política de indexação social para guiar os usuários indexadores. Tal política adota o TBCI como base terminológica.

O experimento foi então realizado com três grupos de usuários distintos, sendo todos estudantes de graduação de cursos da área da Ciência da Informação que já haviam cursado as disciplinas de indexação dos seus respectivos cursos. Foram eles:

- Grupo 1: Turma de Práticas em Biblioteconomia 2018.1;
- Grupo 2: Turma de Estruturas e Linguagens de Informação 2018.2;
- Grupo 3: Turma de Recuperação da informação 2018.2.

A indexação social foi realizada por cada grupo de forma independente dos demais grupos, de forma que o resultado da indexação social de um não fosse visível aos outros.

Os 60 artigos foram divididos entre os usuários de cada grupo de forma que cada usuário ficou responsável por realizar a indexação de uma fração dos artigos, sendo obrigatória a inserção de cinco termos do TBCI por artigo, conforme especificado na política de indexação social. Essa ação de indexação social foi repetida para cada grupo, totalizando no máximo 15 termos inseridos por documento.

Finalizando o experimento, foi realizada a análise da folksonomia assistida visando a compilação do corpus de referência.

Por fim, o percurso metodológico envolvendo todas as etapas até então desenvolvidas, foi então delineado.

4 ANÁLISE DE RESULTADOS

Nesta seção são discutidas as etapas do percurso metodológico para construção do corpus de referência de Ciência da Informação. Para isso, foram definidas subseções, que discutem os elementos que compõem o percurso, conforme seu enquadramento nas etapas da pesquisa.

4.1 Modelo colaborativo de indexação social assistida

O modelo de colaboração para catálogo web facetado de Silva (2013), aqui categorizado como modelo de indexação colaborativa, foi adaptado para fins desta pesquisa.

Visando estabelecer aos papéis, atividades e regras de interação dos usuários, Silva (2013) define as categorias de usuários e suas respectivas atribuições. Visando a simplificação do modelo, para esta pesquisa foi definido que nem todas as modalidades

de atores seriam contempladas e atribuições foram alteradas como pode ser visto no Quadro 1.

Quadro 1 – Comparativo dos atores do sistema

Silva (2013)	Atual
<ul style="list-style-type: none">• Super administrador: usuário responsável pela gestão das configurações do sistema;• Administrador: usuário responsável pela criação, configuração do catálogo, taxonomia facetada e aprovação de contribuições dos membros;• Membro: usuário cadastrado e logado no sistema;• Visitante: usuário que não está cadastrado ou não está logado no sistema. Também conhecido como usuário anônimo.	<ul style="list-style-type: none">• Super administrador: usuário responsável pela gestão das configurações do sistema e pela criação, configuração do catálogo, taxonomia facetada e aprovação de contribuições dos membros;• Membro: usuário cadastrado e logado no sistema que é responsável pela indexação social assistida de itens através do registro de termos;• Visitante: usuário que não está cadastrado ou não está logado no sistema. Também conhecido como usuário anônimo.

Fonte: Elaborado pelos autores.

Os resultados a serem aqui discutidos giram em torno do papel desempenhado por cada ator do sistema apresentado no Quadro 1. Comparando os atores do modelo desenvolvido por Silva (2013) em relação aos desta pesquisa, percebe-se que os atores e funções passaram por uma adequação de acordo com as condições encontradas para aplicação do modelo.

É possível observar que para a pesquisa foi removido o ator administrador, fazendo com que o usuário super administrador assumisse as responsabilidades de ambos. O super administrador tem como responsabilidade a gestão das configurações do sistema como criação, configuração da coleção, taxonomia facetada e aprovação de contribuições dos membros.

Aqui, percebe-se que a moderação por aprovação na figura de um administrador é até certo ponto descomedido quanto à centralização das atribuições, indo na contramão do que vem sendo praticada em ambientes digitais web, principalmente com características Web 2.0, no qual a participação do usuário é considerável. Contudo, o mesmo aspecto traz como benefício a garantia de uma qualidade no que vem a ser registrado para as coleções, isso quando se leva em consideração que toda contribuição passa por um processo de aprovação.

A atribuição dos usuários membro foi alterada para ser mais específica, limitando a sua contribuição à indexação social. Nesse ponto é válido reforçar que a folksonomia assistida será gerada a partir dos registros produzidos pelos usuários membros, ou como chamado até aqui, usuário-indexador.

O Tainacan, plataforma que dará condições de aplicação do modelo, permite acesso à coleção livremente aos internautas, sendo esses categorizados como visitantes.

Para se tornar membro, o visitante deve efetuar seu registro de log in e senha. Para efetuar qualquer intervenção, o usuário membro deve estar logado com seu log in e senha no sistema.

No Quadro 2 é apresentado um comparativo dos requisitos funcionais para contribuição no catálogo (para todos os membros) dos modelos de Silva (2013) e o utilizado nesta pesquisa. A necessidade de apresentar esse item é auxiliar no devido entendimento das adaptações realizadas no modelo de Silva (2013).

Quadro 2 – Comparativo dos requisitos funcionais para a contribuição no catálogo (para todos os membros)

Silva (2013)	Atual
<ul style="list-style-type: none"> • Adicionar um registro, utilizando os termos das facetas, através da etiquetagem livre ou ambas as estratégias para classificação; • Contestar a indexação de um registro. Permite ao usuário revisar a indexação do mesmo adicionando novas categorias das facetas, ou informar eventual classificação erroneamente aplicada (deve enviar um e-mail para o administrador do catálogo). Assim, o usuário pode contribuir com o catálogo, melhorando a qualidade da indexação. O administrador deve aprovar a contestação realizada para que seja efetivada; • Denunciar abuso de um registro no catálogo. Voltado principalmente para questões de vandalismo ou conteúdo impróprio, permite ao usuário informar se um registro cadastrado livremente deve ser retirado do ar. O administrador ou moderador pode aprovar ou rejeitar a contestação; • Adicionar comentários aos registros; • Criar um ranking dos membros com maior quantidade de registros cadastrados. 	<ul style="list-style-type: none"> • Adicionar um registro, utilizando os termos autorizados do Tesouro Brasileiro de Ciência da Informação (TBCI) na etiquetagem livre para classificação; • A estrutura hierárquica levada em consideração para a inserção de um registro foi o TBCI, acessado através do site http://www.vocabularyserver.com/tbci/. • Os termos devem ser grafados como especificado no TBCI. Nesse ponto a própria plataforma Tainacan assume para si a responsabilidade de assumir a condição de manter padronizados os registros inseridos, mantendo a qualidade da indexação. Eventuais erros na grafia dos termos serão corrigidos posteriormente fora da plataforma Tainacan. • O registro de termos na indexação dos documentos pelos usuários membros será uma atividade pré-aprovada. • Não haverá contestação da indexação de um registro.

Fonte: Elaborado pelos autores.

Os fatos mais importantes a serem mencionados a partir da análise do Quadro 2 são referentes aos seguintes pontos:

1. Os termos usados na indexação social serão os termos autorizados do TBCI, visando manter a liberdade de escolha dos termos do vocabulário controlado;
2. A condição de apoio à construção da folksonomia assistida se dará através de um ambiente externo, ao invés de ser na própria plataforma Tainacan. Partindo da premissa que a folksonomia assistida é o resultado do apoio ao usuário no ato da etiquetagem, a falta da ferramenta “autocomplete” fez com que o site do TBCI no software TemaTres fosse a alternativa viável para registro dos termos pelos usuários;
3. O Tainacan mantém padronizados os registros inseridos. O primeiro registro inserido na coleção, independente do item, é assumido como formato padrão para os demais que vierem a ser inseridos posteriormente. Por exemplo, o termo “transferência de informação” ao ser inserido “TRANSFERÊNCIA DA INFORMAÇÃO” (todo em letras maiúsculas), assume assim a condição padrão. Eventuais correções na grafia dos termos são realizadas posteriormente fora da plataforma Tainacan.
4. Buscando a simplicidade na realização do experimento, foi estabelecido que a atividade de registro de *tags* no ato da indexação de documentos será uma atividade pré-aprovada, bem como não haveria contestação da indexação de um registro de um usuário por outro usuário. A validação dos registros, *tags*, inseridas pelos usuários-indexadores será realizada posteriormente de forma manual pelo usuário administrador.

Uma dificuldade identificada na aplicação do Modelo foi observada no relato de alguns usuários que não encontraram termos capazes de representar da forma como gostariam o item (documento) a qual foram destinados a indexar. Tendo em vista que a base terminológica foi o TBCI, a falta de termos adequados à necessidade dos usuários-indexadores fez com que eles sugerissem a criação de novos termos ou, como orientado na instrução de como proceder na indexação social assistida, escolhessem um termo alternativo do TBCI, capaz de se aproximar da representação que se desejava inicialmente. Vale ressaltar que a ferramenta autocomplete disponibilizada através do site <http://www.vocabularyserver.com/tbci/>, principalmente para casos como esses, possibilita ao usuário identificar alternativas, servindo assim como apoio.

Para esta pesquisa, o papel de administrador coube a um único usuário, podendo em outras aplicações ter mais de um, trabalhando de forma a amenizar a questão de centralização. Com a ressalva de que dar condições aos usuários realizarem ações de editar e deletar, ainda que seja o desfazer de uma ação, pode acarretar certas dificuldades, como a exclusão de termos equivocadamente.

Fora observado que a aplicação de um modelo de colaboração no qual o nível de moderação pode ser considerado alto e mais centralizador gira em torno de um menor dinamismo, e como consequência o fluxo de trabalho ocorre de forma mais lenta. Como efeito desse aspecto, pode vir a ocorrer um afastamento natural e desinteresse dos usuários.

Para de fato funcionar, o modelo colaborativo de indexação social assistida necessita garantir que não haja diminuição da cooperação ou do compartilhamento de conteúdo por parte dos usuários.

Em alguns modelos, o visitante, usuário que não está cadastrado ou não está logado no sistema, podem atuar realizando algumas ações. Contudo, para esta pesquisa, o papel do visitante se restringiu exclusivamente a possibilidade consultar e verificar a coleção, isso porque foi entendido que em ambientes como bases de dados científicas, por exemplo, mesmo com a indexação sendo feita livremente, é necessário garantir o mínimo de identificação daqueles que estão realizando essa atividade.

Sendo assim, diversos são os pontos que merecem ser destacados enquanto favoráveis ao uso do modelo colaborativo de indexação social assistida. O primeiro está relacionado à presença de um tesouro da área de Ciência da Informação associado ao sistema, pois possibilita uma melhor qualidade da representação e recuperação da informação.

Adiante, caso seja bem construído, assim como devidamente atualizado, o tesouro, que servirá de apoio, pode proporcionar um aumento na confiança dos usuários na seleção de tags para a representação e recuperação dos documentos. Aliado a isso, a possibilidade de visualização de uma lista de potenciais tags, disponível no sistema para o usuário no momento da indexação ou recuperação, surge como uma estratégia interessante, principalmente no momento em que o usuário não sabe a etiqueta a ser empregada.

Como outra possível vantagem, a participação colaborativa dos usuários na indexação de qualquer produção científica aumenta de forma significativa a memória coletiva acerca da temática.

Por fim, como uma das principais questões atreladas a aplicação de um modelo colaborativo, está o fato da pesquisa baseada em tags da folksonomia assistida proporcionarem uma busca semântica por documentos (orientada a conceitos), superando as limitações da busca sintática por termos (orientada a expressões em linguagem natural).

4.2 A configuração da coleção

Como forma de propor um modelo de indexação colaborativa e, sobretudo aplicável, é preciso relatar algumas características acerca do processo de configuração da coleção para entender de forma direta, como o modelo pode vir a ser útil em ambientes de caráter científico, por exemplo.

O Tainacan, espaço utilizado para desenvolver as ações desta pesquisa, apresenta diversos recursos que possibilitaram contribuir na construção de um modelo colaborativo de indexação social assistida, e, além disso, trabalhar de forma integrada com outras ferramentas Web 2.0.

Como premissa para a escolha do Tainacan está a possibilidade de uso de recursos, tendo uma interface voltada para usuários sem conhecimento técnico em programação. Dessa forma, apresenta-se uma série de características para este fim, das quais como destaque se tem: a interface de busca facetada, o recurso de autocompletar na caixa de busca, a interface simplificada para criação de coleções, a possibilidade de submeter arquivos em massa, dentre outros recursos.

Partindo das possibilidades propostas pela ferramenta, vários são os apontamentos a serem feitos sobre sua utilização no decorrer da pesquisa. Para tanto, inicialmente será discutido como ocorreu a configuração da coleção.

No primeiro momento, para incluir uma nova coleção o assinante, que passará a ser administrador, conta com uma série de possibilidades de configuração, e precisa fazer seu cadastro na página inicial. Como nome para a Coleção foi definido que ela se chamaria **Corpus de Referência em CI** e que só haveria um administrador.

Na sequência, antes de inserir qualquer tipo de item na coleção é necessário definir as permissões dos usuários que nortearão sua construção. Como permissões, definiu-se que todos os membros (usuários) poderiam criar categorias, classificação, comentários e *tags*, mas que para toda ação de edição e exclusão precisariam de aprovação do administrador. Isso impediu que diante da indexação pelos grupos de usuários, imprevistos pudessem vir a comprometer o que já havia sido inserido.

Um ponto crítico na configuração da coleção se deu na etapa de inserção das categorias, e que acabou implicando diretamente na adaptação da forma de inserção das *tags* ou apoio ao usuário-indexador no ato da escolha das *tags*. O Tainacan permite duas

formas de adição de Categorias, a primeira de forma manual, onde se acrescenta cada termo da hierarquia, e a segunda através de importação da taxonomia.

Definido que o TBCI seria o instrumento terminológico a ter os termos inseridos nas categorias, a forma de importação descrita pelo Tainacan seria por meio de arquivo no formato XML. Contudo, em nenhum momento a importação funcionou como deveria, fazendo com que, por conta de tempo hábil, a inserção manual de todos os termos que constam no TBCI não fosse uma alternativa viável.

Diante da problemática apresentada, à saída foi à busca de uma alternativa para o andamento da pesquisa, de forma a não comprometer os objetivos. Tal alternativa consistiu do uso do site <http://www.vocabularyserver.com/tbci/> já descrito na seção anterior.

A próxima etapa da configuração da Coleção passa pela definição de quais são os metadados que irão constar na descrição dos itens. Para esta pesquisa definiu-se como metadados os campos Dublin Core, acrescentando apenas o metadado “tags”, que receberá o registro dos termos atribuídos pelos usuários indexadores na indexação social assistida.

Na sequência, o Tainacan permite ao administrador definir o Layout da Coleção, para por fim permitir a inserção dos itens que irão fazer parte da Coleção, podendo ser de vários formatos. No caso da **Coleção Corpus de Referência em CI** foram inseridos os 60 artigos que constam no Corpus de Souza (2005) em formato PDF.

Com todos os artigos inseridos e catalogados pelo administrador, os usuários têm a sua disposição a possibilidade de interagir como toda ferramenta Web 2.0, incluindo fazer buscas personalizadas, dentre elas por Assunto, após a indexação social pelos usuários. Finalizado todas as etapas de configuração da Coleção, o Tainacan permite que todas as ações sejam revistas caso necessárias, incluindo o espaço para aprovação das ações realizadas.

No processo de configuração é importante ter em mente o papel do administrador e o quanto as atividades serão centralizadas. No que depender do contexto de aplicação, a necessidade de vários administradores para acompanhar todas as ações passa a ser determinante para um melhor andamento das atividades.

4.3 Política de indexação social assistida

Na perspectiva de direcionar os usuários a executarem a atividade de indexação social assistida de forma homogênea, foi criada como instrumento a política de indexação social assistida.

Por se tratar de um modelo colaborativo que contempla características híbridas da junção de taxonomia e folksonomia (KIU; TSUI 2010), observaram-se diversos aspectos que reiteram a necessidade da aplicação de uma política para o alcance dos objetivos traçados.

O primeiro, diz respeito ao fato de se tratar de um modelo até então novo quando se pensando na aplicação ao usuário. Isso é reforçado quando observado que algumas das recomendações feitas na política não são seguidas, mostrando assim a necessidade de direcionamento ou instrução com as ações a serem executadas.

Inicialmente é importante tomar nota que no intuito de congregiar as categorias e manter uma sequência lógica para as análises, foi solicitado que a indexação seguisse alguns direcionamentos, e que independentemente do conhecimento técnico ou da área na qual fosse aplicada, pudessem ser alcançados visando uma indexação de qualidade. O resultado disso é a harmonização do acesso por assunto na recuperação da informação na coleção.

Assim, para que a indexação transcorresse com qualidade, foi solicitado que os usuários levassem em consideração os seguintes aspectos: 1. Compreensão do conteúdo do documento como um todo; 2. Identificação dos conceitos que representam este conteúdo e 3. Seleção de cinco conceitos do TBCI para a recuperação do documento, sendo cada conceito representado pelo respectivo termo autorizado no tesouro.

Quanto à forma de escrita, para evitar problemas de erros de escrita, bem como o uso indiscriminado de plural e letras maiúsculas ou minúsculas erroneamente, foi orientado que quando escolhido a *tag* que representaria um conceito do artigo, o usuário executasse a simples ação de copiar do site <http://www.vocabularyserver.com/tbci/> o termo autorizado e colasse no campo *tags* na página do artigo que se encontra no Tainacan, da exata forma que consta no site em questão.

Ainda diante desse contexto, é válido acrescentar que o processo de acompanhamento dos termos, ou avaliação da linguagem de indexação inserida, precisa ser realizada. Tudo isso de forma a garantir que esse cenário continue a se manter propício e vantajoso, servido como registros úteis para a recuperação da informação.

Por fim, a utilização de uma Política de indexação social assistida, se apresentou de forma a permitir a elaboração de uma linguagem comum e possivelmente compreensiva a todos os usuários. Fato esse garantido através do processo de cogestão estabelecida na participação dos grupos de atores (administrador e membros) em torno do modelo de indexação social.

4.4 Análise da folksonomia assistida

Dentro do processo de desenvolvimento da folksonomia assistida, a funcionalidade que caracteriza a sua construção é o apoio ao usuário no exato momento de inserção de uma *tag* que vise descrever um conteúdo. Além disso, bem como descreve Santarem Segundo (2010, p.181) a “folksonomia assistida prima pela consistência das *tags*, de forma que o usuário do sistema evite abreviações, plurais/singulares ou ainda palavras que possam dificultar a recuperação da informação, posteriormente”.

Foi observado que não existe uma determinação por parte das revistas, cujos artigos compõem o corpus, quanto à quantidade “obrigatória” de palavras-chave a serem utilizadas nos documentos, causando assim uma grande variação de um artigo para outro do número de palavras-chave informadas no campo assunto da coleção, como pode ser observada na **Tabela 1**.

Tabela 1 – Quantitativo de palavras-chave dos autores nos documentos do corpus

ID	Nº. DE PALAVRAS-CHAVE	ID	Nº. DE PALAVRAS -CHAVE	ID	Nº. DE PALAVRAS -CHAVE	ID	Nº. DE PALAVRAS -CHAVE
Artigo 01	3	Artigo 16	5	Artigo 31	5	Artigo 46	6
Artigo 02	2	Artigo 17	3	Artigo 32	6	Artigo 47	3
Artigo 03	5	Artigo 18	4	Artigo 33	5	Artigo 48	3
Artigo 04	5	Artigo 19	5	Artigo 34	3	Artigo 49	3
Artigo 05	5	Artigo 20	4	Artigo 35	5	Artigo 50	2
Artigo 06	7	Artigo 21	5	Artigo 36	4	Artigo 51	5
Artigo 07	5	Artigo 22	9	Artigo 37	2	Artigo 52	5
Artigo 08	5	Artigo 23	3	Artigo 38	3	Artigo 53	5
Artigo 09	5	Artigo 24	6	Artigo 39	5	Artigo 54	7
Artigo 10	5	Artigo 25	2	Artigo 40	2	Artigo 55	4
Artigo 11	4	Artigo 26	4	Artigo 41	4	Artigo 56	5
Artigo 12	6	Artigo 27	3	Artigo 42	5	Artigo 57	7
Artigo 13	3	Artigo 28	7	Artigo 43	4	Artigo 58	7

Artigo 14	6	Artigo 29	3	Artigo 44	3	Artigo 59	5
Artigo 15	2	Artigo 30	3	Artigo 45	5	Artigo 60	4

Fonte: Dados da Pesquisa (2018)

Na folksonomia assistida gerada, foram contabilizadas 261 *tags* diferentes, totalizando 583 registros totais. Como foram envolvidos três grupos de usuários na sua construção, poderia conter até 15 *tags* por documento, podendo variar para menos quando houvesse coincidência dentre as *tags* escolhidas. Por essa razão, foi verificado que o número máximo de *tags* encontrada foi de 14, e o mínimo 6. Esse cenário gerou uma média de 9 a 10 *tags* por documento e que pode ser vista na Tabela 2.

Tabela 2 - Quantitativo de *tags* geradas pelos usuários nos documentos do corpus

ID	Nº. DE TAGS	ID	Nº. DE TAGS	ID	Nº. DE TAGS	ID	Nº. DE TAGS
Artigo 01	10	Artigo 16	9	Artigo 31	9	Artigo 46	6
Artigo 02	10	Artigo 17	12	Artigo 32	8	Artigo 47	10
Artigo 03	11	Artigo 18	14	Artigo 33	11	Artigo 48	9
Artigo 04	9	Artigo 19	11	Artigo 34	12	Artigo 49	12
Artigo 05	11	Artigo 20	11	Artigo 35	6	Artigo 50	10
Artigo 06	10	Artigo 21	6	Artigo 36	11	Artigo 51	8
Artigo 07	10	Artigo 22	8	Artigo 37	10	Artigo 52	6
Artigo 08	12	Artigo 23	9	Artigo 38	12	Artigo 53	6
Artigo 09	10	Artigo 24	8	Artigo 39	9	Artigo 54	8
Artigo 10	10	Artigo 25	11	Artigo 40	11	Artigo 55	6
Artigo 11	10	Artigo 26	12	Artigo 41	12	Artigo 56	9
Artigo 12	10	Artigo 27	12	Artigo 42	10	Artigo 57	9
Artigo 13	11	Artigo 28	11	Artigo 43	10	Artigo 58	7
Artigo 14	10	Artigo 29	7	Artigo 44	11	Artigo 59	9
Artigo 15	10	Artigo 30	13	Artigo 45	9	Artigo 60	9

Fonte: Dados da Pesquisa (2018)

É importante frisar que coube ao usuário administrador o papel de revisar todas as questões e problemas oriundos da inserção de termos, gerando assim o que aqui será chamado de folksonomia assistida revisada. As correções consistiram da remoção de *tags* que não eram termos do TBCI (27 *tags* foram eliminadas) e correção da grafia de alguns termos (nove termos com presença de caractere ponto no final e 62 termos com caracteres capitalizados indevidamente).

Foi observado que as *tags* da folksonomia assistida revisada apresentam-se como alternativa às palavras-chave dos autores na avaliação de sistemas de indexação automática, agregando novas possibilidades no processo de recuperação dos artigos. Contudo tendo a devida atenção aplicada ao problema de inserção de *tags* não presentes no TBCI ou com grafia errada.

Comparando os termos gerados pelos autores com o que foi produzido enquanto *tags*, o quantitativo de termos comuns pode ser visto na Tabela 3.

A quantidade de termos comuns se deu quando confrontado palavras-chave, geradas pelos autores, e *tags*, geradas pelos usuários-indexadores, levando em consideração não só a mesma escrita, mas os conceitos contemplados e respectivos termos autorizados no TBCI. Como resultado disso, a quantidade de termo alternou de um mínimo de zero, em apenas um caso onde nenhum termo coincidiu, a um máximo de seis termos comuns entre eles.

Tabela 3 - Quantitativo de termos em comum nos documentos do corpus

ID	Nº. DE TERMOS EM COMUM	ID	Nº. DE TERMOS EM COMUM	ID	Nº. DE TERMOS EM COMUM	ID	Nº. DE TERMOS EM COMUM
Artigo 01	2	Artigo 16	4	Artigo 31	3	Artigo 46	4
Artigo 02	2	Artigo 17	4	Artigo 32	4	Artigo 47	2
Artigo 03	1	Artigo 18	2	Artigo 33	3	Artigo 48	2
Artigo 04	2	Artigo 19	4	Artigo 34	2	Artigo 49	3
Artigo 05	2	Artigo 20	4	Artigo 35	5	Artigo 50	2
Artigo 06	5	Artigo 21	1	Artigo 36	3	Artigo 51	2
Artigo 07	5	Artigo 22	5	Artigo 37	2	Artigo 52	3
Artigo 08	2	Artigo 23	0	Artigo 38	2	Artigo 53	4
Artigo 09	2	Artigo 24	1	Artigo 39	4	Artigo 54	3
Artigo 10	2	Artigo 25	2	Artigo 40	2	Artigo 55	3
Artigo 11	3	Artigo 26	2	Artigo 41	3	Artigo 56	3
Artigo 12	6	Artigo 27	1	Artigo 42	3	Artigo 57	2
Artigo 13	2	Artigo 28	3	Artigo 43	2	Artigo 58	6
Artigo 14	3	Artigo 29	3	Artigo 44	3	Artigo 59	3
Artigo 15	3	Artigo 30	2	Artigo 45	4	Artigo 60	4

Fonte: Dados da Pesquisa (2018)

Com o cenário apresentado na Tabela 3, é possível chegar à reunião dos termos mais representativos para o *Corpus de Referência em Ciência da Informação*, contudo, na perspectiva de agregar valor à indexação em bases de dados, também agregam valor os termos não coincidentes. Assim, é válido destacar que o número de termos incomuns é significativo, a ponto de gerar maior exaustividade num processo de busca.

5 CONSIDERAÇÕES FINAIS

A folksonomia assistida construída por meio da aplicação de modelo colaborativo de indexação social assistida possibilitou validar o que vem sendo estudado pelos pesquisadores da Ciência da Informação quanto ao assunto. Isso permitiu que os principais aspectos acerca da temática fossem enfatizados de acordo com as produções científicas, mas, sobretudo, aplicados de forma a verificar o grau de validade das proposições em torno da folksonomia assistida.

O percurso metodológico para construção de corpus de referência conta com as seguintes etapas: 1) Elaboração do modelo colaborativo de indexação social; 2) Configuração da coleção no Tainacan; 3) Desenvolvimento de uma política de indexação social assistida; 4) Indexação social assistida; 5) Análise da folksonomia assistida.

A primeira e segunda etapas do percurso metodológico foram concretizadas por meio da configuração da plataforma Tainacan, que oferece suporte aos aspectos principais da web 2.0. Isso permitiu que houvesse condições de configurar o processo de indexação social assistida por meio da definição dos papéis, atividades e regras de interação entre usuários, como previsto no modelo colaborativo de indexação social, bem como a configuração da coleção de itens ou documentos do corpus.

A terceira etapa do percurso consistiu do desenvolvimento de uma política de indexação social assistida, que permitiu delinear a indexação dos documentos para além do modelo aplicado, e principalmente apresentar a adoção de um tesouro da área de Ciência da Informação para a realização da indexação social assistida de artigos do corpus.

A quarta etapa consistiu do experimento com aplicação do modelo e política na indexação social da coleção por grupos de usuários-indexadores, permitindo que a folksonomia assistida fosse concretizada.

A análise da folksonomia assistida foi a quinta e última etapa do percurso metodológico. Poucas foram as *tags* removidas por não serem termos do TBCI e poucas

tags foram corrigidas quanto a ortografia dos termos. Foi percebido que, se bem construído e atualizado, o instrumento de controle terminológico pode aumentar a confiança do usuário-indexador na identificação de conceitos relacionados para a indexação e recuperação de documentos. O fato do usuário ter permissão para definir quais *tags* representam as produções científicas, torna possível que ele realize uma espécie de "lembrete", atribuindo pontos de acesso para futura leitura ou recuperação das produções que ele julga relevantes sobre os assuntos atribuídos como *tags*.

Com a aplicação do percurso metodológico proposto nesta pesquisa chegou-se a algumas vantagens, que são:

1. A produção significativa de metadados semânticos;
2. A padronização na grafia dos descritores;
3. A capacidade de associar a linguagem dos usuários à linguagem utilizada pelo sistema.

Tais vantagens reforçam que o percurso metodológico se mostrou eficaz para construção do corpus de referência em Ciência da Informação, e pode ser adaptado para construção de corpus de referência em outras áreas do conhecimento.

Como trabalho futuro aponta-se o uso de uma nova compilação por meio da folksonomia assistida com outros grupos de usuários no intuito de verificar o grau de confiabilidade da indexação realizada. Nessa perspectiva, o objetivo seria também mensurar a consistência e o grau de comprometimento do usuário no ato da indexação.

Outra possibilidade de trabalho futuro consiste da adaptação e aperfeiçoado do percurso metodológico para a construção de bases de dados de artigos de periódicos eletrônicos ou repositórios digitais com a possibilidade da indexação social assistida dos documentos pelos usuários.

6 REFERÊNCIAS

BANDIM, M. A. S.; CORRÊA, R. F. A consistência na indexação automática por atribuição de artigos científicos na área de Ciência da Informação. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 23, n. 53, p.64-77, set. 2018. DOI: <https://doi.org/10.5007/1518-2924.2018v23n53p64> .

BANDIM, M. A. S.; CORREA, R. F. Indexação automática por atribuição de artigos científicos em português da área de Ciência da Informação. **Transinformação**, v. 31, p. 1-12, 2019. DOI: <https://doi.org/10.1590/2318-0889201931e180004>

CARNEIRO, M. V. Diretrizes para uma política de indexação. **R. Esc. Biblioteconomia**. UFMG, Belo Horizonte, v.14, n.2, p.221-241, set. 1985.

CORRÊA, R. F.; CELERINO, V. G. Método de normalização de sintagmas nominais na indexação automática. **EM QUESTÃO**, v. 25, p. 321-344, 2019. DOI: <https://doi.org/10.19132/1808-5245251.321-344> .

CORRÊA, R. F.; LAPA, R. C. Panorama de Estudos sobre indexação automática no âmbito da Ciência da Informação no Brasil (1973-2012). **Ciência da Informação**. Brasília, DF, v. 42, n. 2, p. 255-273, 2013.

CORREA, R. F.; SANTOS, R. F. dos. Análise das definições de folksonomia: em busca de uma síntese. **Perspectivas em Ciência da Informação**, Belo Horizonte , v. 23, n. 2, p. 1-32, jun. 2018. DOI: <https://doi.org/10.1590/1981-5344/2571>.

KIU, Ching-chieh; TSUI, Eric. TaxoFolk: a hybrid taxonomy–folksonomy classification for enhanced knowledge navigation. **Knowledge Management Research & Practice**, v. 8, n. 1, p.24-32, mar. 2010. DOI: <https://doi.org/10.1057/kmrp.2009.33> .

FUJITA, M. S. L.; RUBI, M. P. Um modelo de leitura documentária para a indexação de artigos científicos: princípios de elaboração e uso para a formação de indexadores. **DataGramZero**, Rio de Janeiro, v. 7, n. 3, p. 1-19. jun. 2006.

GIL, A. C. **Como elaborar um projeto de pesquisa**. 4. ed. São Paulo: Atlas, 2002. 176 p.

GIL LEIVA, I.; RUBI, M. P.; FUJITA, M. S. L. Consistência na indexação em bibliotecas universitárias brasileiras. **Transinformação**, v. 20, n. 3, p. 233-253, set./dez. 2008. DOI: <https://doi.org/10.1590/S0103-37862008000300003>.

Golub, K.; Soergel, D.; Buchanan, G.; Tudhope, D.; Lykke, M.; Hiom, D. A Framework for Evaluating Automatic Indexing or Classification in the Context of Retrieval. **Journal of the Association for Information Science and Technology**, v.67, n.1, p. 3-16. 2016. DOI: <https://doi.org/10.1002/asi.23600> .

GUIMARÃES, J. A. C. Políticas de análisis y representación de contenido para La gestión del conocimiento en las organizaciones. **Scire**, Zaragoza, v. 6, n. 2, p. 48-58, jul./dez. 2000.

GUY, M.; TONKIN, E. Folksonomies: tidying up tags?. **D-Lib Magazine**, v. 12, n. 1, Jan. 2006. Disponível em: < <http://www.dlib.org/dlib/january06/guy/01guy.html> >. Acesso em: 13 fev. 2019

HASAN, K. S.; NG, V. Automatic Keyphrase Extraction: A Survey of the State of the Art. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 52., 2014, Baltimore. **Proceedings...** . Baltimore: Association For Computational Linguistics, 2014. p. 1262 - 1273.

HJØRLAND, Birger. Automatic Indexing. In: **Lifeboat for Knowledge Organization**, 2008.

KIM, S. N.; MEDELYAN, O.; KAN, M.Y.; BALDWIN, T. Automatic Keyphrase: extraction from Scientific Articles. In: **Language Resources and Evaluation** (2013) v. 47, n. 3, p. 723-742, Springer. December 2012.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. 2. ed. Brasília: Briquet de Lemos, 2004. 452 p.

LEIVA, I. G. SISA – Automatic Indexing System for Scientific Articles: Experiments with Location Heuristics Rules Versus TF-IDF Rules". **Knowledge Organization**, v.44, n. 3, p. 139-162, 2017.

MARCONI, M. A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 5.ed. São Paulo: Editora Atlas, 2006. 315 p.

PINHEIRO, L. V. R.; FERREZ, H. D. **Tesouro Brasileiro de Ciência da Informação**. Rio de Janeiro, Brasília: IBICT, 2014. 384 p.

ROSCELLE, J.; TEASLEY, S. D. The Construction of Shared Knowledge in Collaborative Problem Solving. **Computer Supported Collaborative Learning**, Berlim, v. 25, p.69-97, 1995.

SANTARÉM SEGUNDO, J. E. **Representação Iterativa**: um modelo para repositórios digitais. Marília, SP, 2010. 224 f. Tese (Doutorado) – Curso de Ciência da Informação, Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, SP, 2010.

SANTAREM SEGUNDO, J. E.; SIQUEIRA, C. S. Aplicación teórico-conceptual de folksonomías asistidas para la recuperación de información. **Scire**, Zaragoza, v. 19, n. 2, p.77-82, 2013.

SANTARÉM SEGUNDO, J. E.; VIDOTTI, S. A. B. G. Representação Iterativa e folksonomia assistida para repositórios digitais. **Liinc em Revista**, Rio de Janeiro, v.7, n.1, mar. 2011, p. 283-300. DOI: <https://doi.org/10.18617/liinc.v7i1.414> .

SANTAREM SEGUNDO, J. E.; VIDOTTI, S. A. B. G. Rede de tags para recuperação da informação no contexto da representação iterativa. **InCID: Revista de Ciência da Informação e Documentação**, v. 2 n. 1, n. 1, p. 86-109, 2011. DOI: DOI: <https://doi.org/10.11606/issn.2178-2075.v2i1p86-109> .

SANTOS, R. F. **Modelos colaborativos de indexação social e a sua aplicabilidade na base de dados referencial de artigos de periódicos em Ciência da Informação (BRAPCI)**. 2016. 185 f. Dissertação (Mestrado) - Curso de Ciência da Informação, Departamento de Ciência da Informação, Universidade Federal de Pernambuco, Recife, 2016.

SANTOS, R. F.; CORRÊA, R. F. Modelos colaborativos de indexação social e sua aplicabilidade em bibliotecas digitais. **Liinc em Revista**, v. 11, n. 1, 2015. DOI: <https://doi.org/10.18617/liinc.v11i1.768>

SILVA, M. F. **Proposta de modelo de colaboração para catálogo web facetado**. Belo Horizonte, 2013. 269 f. Tese (Doutorado) – Curso de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2013.

SOUZA, R. R. **Uma proposta de metodologia para escolha automática de descritores utilizando sintagmas nominais**. 2005. 215 f. Tese (Doutorado) - Curso de Ciência da Informação, Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

SOUZA, R. R. Uma proposta de metodologia para indexação automática utilizando sintagmas nominais. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, n. esp. 1. sem., p. 42-59, 2006. DOI: <https://doi.org/10.5007/1518-2924.2006v11nesp1p42>.

SOUZA, R. R.; LEIVA, I. G. Automatic Indexing of Scientific Texts: A Methodological Comparison. **Proc. of the Fourteenth International ISKO Conference**, Rio de Janeiro, Brazil. Würzburg: Ergon Verlag, p. 243-250, 2016.

SOUZA, R. R.; RAGHAVAN, K. S. A methodology for noun phrase-based automatic indexing. **Knowledge Organization**, v. 33, n. 1, p. 45-56, 2006.

SOUZA, R. R.; RAGHAVAN, K. S. Extraction of keywords from texts: an exploratory study using Noun Phrases. **Informação & Tecnologia (ITEC)**. Marília/ João Pessoa. v. 1, n. 1. p. 5-16, 2014.

YEDID, N. Introducción a las folksonomías: definición, características y diferencias con los modelos tradicionales de indización. **Información, Cultura Y Sociedad**, v. 29, p. 13-26, 2013. Disponível em: <http://revistascientificas.filo.uba.ar/index.php/ICS/article/view/673>. Acesso em: 03 jul. 2019.

ZUBIAGA, A. et al. Analyzing tag distributions in folksonomies for resource classification. 2012. p. 1-12. Disponível em: <http://arxiv.org/abs/1202.5477v1>. Acesso em: 03 jul. 2019.

NOTAS

AGRADECIMENTOS

Não se aplica.

CONTRIBUIÇÃO DE AUTORIA

Concepção e elaboração do manuscrito: B. F. M. Silva, R. F. Correa

Coleta de dados: B. F. M. Silva, R. F. Correa

Análise de dados: B. F. M. Silva, R. F. Correa

Discussão dos resultados: B. F. M. Silva, R. F. Correa

Revisão e aprovação: B. F. M. Silva, R. F. Correa

CONJUNTO DE DADOS DE PESQUISA

Todo o conjunto de dados que dá suporte aos resultados deste estudo foi publicado no próprio artigo.

FINANCIAMENTO

Não se aplica.

CONSENTIMENTO DE USO DE IMAGEM

Não se aplica.

APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA

Não se aplica.

CONFLITO DE INTERESSES

Não se aplica.

LICENÇA DE USO

Os autores cedem à **Encontros Bibli** os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a [Licença Creative Commons Attribution](#) (CC BY) 4.0 International. Esta licença permite que **terceiros** remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os **autores** têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

PUBLISHER

Universidade Federal de Santa Catarina. Programa de Pós-graduação em Ciência da Informação. Publicação no [Portal de Periódicos UFSC](#). As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

EDITORES

Enrique Muriel-Torrado, Edgar Bisset Alvarez, Camila Barros.

HISTÓRICO

Recebido em: 14-05-2019 – Aprovado em: 04-06-2019