Em Questão
ISSN: 1807-8893
ISSN: 1808-5245
emquestao@ufrgs.br
Universidade Federal do Rio Grande do Sul
Brasil

# Aplicação da folksonomia assistida na construção de *corpus* de referência em Ciência da Informação

**Silva, Bruno Felipe de Melo; Correa, Renato Fernandes**
Aplicação da folksonomia assistida na construção de *corpus* de referência em Ciência da Informação
Em Questão, vol. 26, núm. 2, 2021
Universidade Federal do Rio Grande do Sul, Brasil
**Disponível em:** https://www.redalyc.org/articulo.oa?id=465662940018
**DOI:** https://doi.org/10.19132/1808-5245262.413-436

# Aplicação da folksonomia assistida na construção de *corpus* de referência em Ciência da Informação

The application of assisted folksonomy in the construction of Information Science reference corpus

*Bruno Felipe de Melo Silva* 1
*Universidade Federal de Alagoas, Brasil*
bruno.fms545@gmail.com

*Renato Fernandes Correa* 2
*Universidade Federal de Pernambuco, Brasil*
renato.correa@ufpe.br

DOI: https://doi.org/10.19132/1808-5245262.413-436
Redalyc: https://www.redalyc.org/articulo.oa?id=465662940018

Recepção: 05 Fevereiro 2019
Aprovação: 15 Julho 2019

## Resumo:

O presente trabalho propõe e discute a aplicação da folksonomia assistida na construção de *corpus* de referência de artigos científicos da área de Ciência da Informação. A hipótese levantada é que tal aplicação pode garantir maior qualidade na indexação de artigos científicos e uma melhor avaliação dos sistemas de indexação automática através do *corpus* compilado. Para a pesquisa foi delimitado o uso do *corpus* composto por 60 artigos escritos em língua portuguesa selecionados por Souza (2005). A plataforma colaborativa de indexação social assistida do *corpus* foi configurada usando o *software* de gerenciamento de coleção denominado *Tainacan*. As etapas da pesquisa envolveram a configuração e preparação da coleção no *Tainacan*, a realização da indexação social assistida por grupos de usuários e análise dos resultados do processo de indexação. A análise da folksonomia assistida ocorreu mediante comparação daquilo que consta disponibilizado nos campos de metadados Assuntos e *tags* dos artigos. Como indicadores da qualidade da indexação obtiveram-se média de 28% do coeficiente de consistência, 32% de precisão, 68% de revocação, e 41% de medida F. As médias alcançadas representam bons níveis de consistência e revocação, e níveis satisfatórios de precisão e medida F, dando a entender que o uso da folksonomia assistida é útil no aperfeiçoamento da indexação do *corpus* de referência.

Palavras-chave: Indexação social, Folksonomia assistida, Indexação Automática, Avaliação de sistemas de indexação automática, *Corpus* de Referência.

## Abstract:

This paper proposes and discusses the application of assisted folksonomy in the construction of reference corpus of scientific articles of Information Science area. The hypothesis is that the result of the application can guarantee a better quality in the indexing of scientific articles and a better evaluation of the automatic indexing systems through the compiled corpus. The outlined research uses the corpus composed by 60 articles written in Portuguese selected by Souza (2005). The collaborative platform for assisted social indexing of the corpus consisted of the collection management software called *Tainacan*. The stages of the research involved the configuration and preparation of the collection in *Tainacan*, the accomplishment of the assisted social indexing by groups of users and analysis of the results of the indexing process. The analysis of the assisted folksonomy consists of comparing what is available in the Subject and tags metadata fields of the articles. The indicators of the quality of indexation were a mean of 28% of the consistency coefficient, 32% of precision, 68% of recall, and 41% of the F measure. The indicators represent good levels of consistency and recall and satisfactory levels of precision and F measure, implying that the use of assisted folksonomy is useful in improving the indexing of the reference corpus.

Keywords: Automatic Indexing, Evaluation of automatic indexing systems, Reference Corpus, Social indexing, Assisted Folksonomy10.19132/1808-5245262.413-436, Assisted Folksonomy.

## Autor notes

1 Mestre; Universidade Federal de Alagoas, Maceió, AL, Brasil
bruno.fms545@gmail.com

2 Doutor; Universidade Federal de Pernambuco, Recife, PE, Brasil;
renato.correa@ufpe.br

## 1 Introdução

Pensar no próprio conhecimento como sendo fruto da junção de um conhecimento prévio adquirido com um conhecimento extraído de uma informação, faz com que se entenda que todo indivíduo que se encontra num meio colaborativo não pode ser notado isoladamente. O indivíduo como ator ativo potencializa o surgimento de modelos de indexação social.

Baseando-se em Santos (2016), modelos de indexação social podem ser pensados como compostos por atividades colaborativas de indexação que serão efetuadas pelos usuários em sistemas documentais e que geram a folksonomia como principal resultado. Tais modelos auxiliam no desenvolvimento da linguagem de indexação dos próprios sistemas, dando a flexibilidade de serem adaptados ou aperfeiçoados em diversos contextos.

O uso da folksonomia apresenta questões que estão diretamente ligadas à polissemia, quando um significante tem vários significados, e sinonímia, quando o mesmo significado pode ser representado por vários significantes.

Além disso, até por característica, os modelos de indexação social não apresentam preocupação semântica ou hierárquica com o que é produzido no processo de interação dos usuários.

A busca por alternativas que preservem o efeito positivo da folksonomia, eliminando os aspectos negativos, permite se pensar acerca de um modelo híbrido de representação, a folksonomia assistida.

A premissa básica da folksonomia assistida gira em torno de garantir uma liberdade na escolha de termos de uma linguagem controlada por usuários na atribuição de *tags* a conteúdos informacionais, principalmente de caráter científico, fazendo uso de vocabulários controlados como fonte de termos de indexação.

É importante mencionar que as aplicações dos conceitos apresentados por Santarém Segundo (2010), Silva (2013) e Santos (2016) acerca da folksonomia assistida trazem consigo a possibilidade de surgimento de uma nova perspectiva também para o campo da recuperação da informação em bases de dados científicas.

Em consonância de ideias, acredita-se que a aplicação da folksonomia assistida pode permitir uma melhor indexação de artigos de periódicos científicos, e posteriormente dar condições de compilar um *corpus* de referência que possa servir como padrão para avaliar a qualidade da indexação automática no corpus resultante.

O bom entendimento acerca das temáticas como a indexação automática e avaliação dos sistemas de indexação automática dá base para a proposta da pesquisa aqui desenvolvida, que perpassa a aplicação da folksonomia assistida na construção de corpus de referência de artigos científicos da área de Ciência da Informação para avaliação de sistemas de indexação automática.

O conjunto de artigos de periódicos científicos que constitui o *corpus* foi escolhido levando em consideração o uso por pesquisas anteriores na temática indexação automática para assim determinar sua relevância.

A avaliação da qualidade da folksonomia resultante da indexação social, neste contexto de compilação de um *corpus* de referência, dar-se-á diante da comparação a ser feita entre o que é utilizado pelo **indexador oficial**, sendo ele o próprio autor no ato de escolha das palavras-chave que representarão seu artigo, e o que é entendido e representado pelo **indexador-leitor** através das tags.

A partir dessa relação será possível pensar a respeito sobre em que grau se encontra o elo entre produtor e consumidor da informação, principalmente em ambientes formais como os de revistas científicas, com um nível de especificidade maior. Com isso, é possível estabelecer questões como: a indexação do autor corresponde ao que o indexador-leitor representou?

Como problema de pesquisa fica a pergunta: Na construção do *corpus* de referência para fins de avaliação de sistemas de indexação automática, a aplicação da folksonomia assistida traz uma melhor qualidade na indexação de artigos científicos?

Destarte, esta pesquisa tem por objetivo propor e discutir a aplicação da folksonomia assistida na construção de *corpus* de referência de artigos científicos em Ciência da Informação, uma vez que o seu resultado pode garantir maior especificidade e exaustividade na indexação de artigos científicos, permitindo uma melhor avaliação dos sistemas de indexação automática através do *corpus* compilado.

## 2 Avaliação de sistemas de indexação automática

Realizar avaliações não é uma das tarefas mais simples, envolve a subjetividade de um julgamento, bem como um juízo de valor acerca de um processo e seus resultados. O processo de indexação, como atividade importante na organização e recuperação da informação, traz consigo a necessidade de avaliar se seu objetivo fim está sendo atingido.

Segundo Hlava (2002), para que um sistema automatizado seja considerado bom precisará fornecer inicialmente, a partir de uma boa lista de palavras-chave, uma taxa de precisão de 60%, e com treinamento ou regras de construção um percentual de 85% ou mais.

Adicionalmente, Gil Leiva; Rubi e Fujita (2008) discorrem que os procedimentos de avaliação da indexação podem ser categorizados como avaliação intrínseca, que tem como objetivo medir o grau de consistência da indexação, e como avaliação extrínseca, que objetiva alcançar através de medição o grau de revocação, precisão e medida F na recuperação da informação.

O autor ainda apresenta uma fórmula conhecida como **coeficiente de consistência**, sendo ela proposta num primeiro momento por Salton e McGill (2003). A consistência entre a indexação automática e manual é calculada por meio da fórmula:

Na qual:

· CI = a coerência entre dois sistemas ou dois indexadores;

· T∞ = o número de termos comuns atribuídos pelos dois sistemas ou dois indexadores;

· A = o número de termos atribuídos pelo sistema 1 ou indexador 1;

· B = o número de termos propostos pelo sistema 2 ou indexador 2.

Gil Leiva (1997) explica que a fórmula pode ser utilizada na perspectiva de avaliação de sistemas de indexação automática, e não no sentido de oposição entre a indexação manual e automática. O intuito é permitir a verificação da consistência para posterior melhora ou correção dos parâmetros do sistema de indexação, em um trabalho conjunto da indexação manual e automática.

Outra maneira de avaliar a indexação automática ocorre por meio de cálculos dos coeficientes de precisão e revocação no ato de recuperação de documentos em um sistema. Os resultados obtidos denotam os índices de exaustividade e precisão na recuperação da informação. Segundo Narukawa, Gil Leiva e Fujita (2009), o procedimento consiste em: consultar através de um conjunto de expressões de busca duas bases de dados que contêm os mesmos campos e idênticos conteúdos, salvo os campos que armazenam a indexação; realizar o julgamento de relevância dos documentos retornados para cada expressão de busca; e cálculo das métricas de precisão, revocação e medida F.

De forma análoga, a revocação (R) e precisão (P) podem ser calculadas quanto à atribuição ou recuperação de termos relevantes por um sistema de indexação a um documento:

No intuito de apresentar um único valor, surge a possibilidade de combinar os valores de revocação e precisão através da medida F. A medida F é a média harmônica ponderada entre o índice de revocação (R) e o índice de precisão (P). No caso de nenhum termo relevante ser recuperado a medida F assume o valor zero. A medida F assumirá o valor 1 quando todos os termos recuperados forem relevantes e forem exaustivamente recuperados todos os termos relevantes.

Sabendo disso, é necessário então atentar que a medida F só assume altos valores quando os índices de revocação e precisão são igualmente altos. A busca por determinar um valor para medida F é vista como uma tentativa de se encontrar equilíbrio entre estes dois índices.

Celerino (2018) afirma que na indexação são utilizados diversos critérios e indicadores que possibilitam avaliar sua qualidade. Os critérios são: revocação, que é relacionado à exaustividade; precisão, que está relacionada à especificidade; coerência, que está relacionada à consistência; e relevância, que está relacionada à pertinência.

Para efetivar a avaliação de sistemas de indexação automática, Hasan e Ng (2014) dividem o processo em duas etapas: (1) Mapeamento das palavras-chave dos *corpora*, tanto da indexação manual quanto da automática; (2) Uso de métricas de avaliação como precisão (P), revocação (R) e medida F (F), confrontando as duas formas de indexação.

Fato importante é que os sistemas automáticos ainda não conseguem lidar de forma satisfatória com a linguagem natural, oriunda do autor e do usuário, ao ponto de a indexação de documentos textuais terem um alto grau de qualidade.

Assim, no processo de avaliação da indexação automática é preciso ter em foco que a indexação é a ligação existente entre o que é disponibilizado através do sistema e os anseios do usuário.

## 3 Corpus de referência

A seleção de um *corpus* consiste na escolha manual de documentos seguindo testes de agrupamento. Tem assim como propósito, caracterizar publicações a partir dos objetivos e aspectos metodológicos traçados para uma pesquisa.

É importante destacar uma sutil diferença existente entre o que vem a ser *corpus* para o contexto da indexação automática e o da linguística. Isso se torna importante para que fique evidente que os conceitos aqui adotados estão fundamentados dentro do campo da indexação automática.

Na indexação automática, a relevância dada à utilização de *corpus* vai muito além da compilação de documentos significativos. Hasan e Ng (2014) apresentam um estudo de revisão acerca da extração de frases-chave (do inglês *automatic keyphrase extraction*), onde examinam os principais erros dos sistemas de indexação automática, bem como discutem os desafios futuros em torno da temática. Diante disso, os autores descrevem que nos trabalhos científicos os sistemas foram avaliados em *corpora* de diversas de fontes, que vão desde longas publicações científicas, a resumos e mensagens de *e-mail*.

Assim, entende-se que a proposta dos autores através do uso de vários *corpora* é a avaliação de sistemas de indexação automática. A justificativa para a aplicação de tal método de avaliação vai de encontro ao fato da avaliação humana ter valor significativo. Acrescenta-se a isso a ideia de que em diversos momentos a utilização de mais de um *corpus* tem a intenção de avaliar e perceber como cada sistema se comporta frente ao *corpus* de domínio especifico ao qual foi aplicado.

Levando em conta a indexação automática de artigos científicos da área de Ciência da Informação em português, Souza (2005) utiliza em sua tese dois *corpora* de documentos visando à aplicação de uma metodologia. O objetivo era estudar a indexação automática através da identificação, extração e seleção de sintagmas nominais (SNs) dos textos completos de documentos digitais.

O primeiro *corpus* já havia sido utilizado por Kuramoto (1995), e era a compilação de 15 documentos, que tinha como intuito validar a extração automática dos SNs. O segundo *corpus* foi construído pelo próprio Souza (2005) e era composto por 60 documentos.

A proposta era ter dois *corpora*, onde no primeiro teria a aplicação de uma metodologia prospectiva e, no segundo, uma metodologia consolidada. A utilização dos *corpora* permitiu assim apresentar avaliações e percepções de como a metodologia se comportava em cada *corpus*.

Os usos do *corpus* por Souza (2005) e no presente trabalho se assemelham no intuito de definir uma metodologia que auxilie o processo de escolha de termos de indexação para documentos, guiando-se por estruturas linguísticas pré-estabelecidas. No caso de Souza (2005), a estrutura utilizada foram os sintagmas

nominais, e no trabalho aqui desenvolvido, um tesauro. Nota-se que o principal objetivo em ambas as pesquisas gira em torno de definir um *corpus* significativo para avaliar a indexação automática.

Em sua dissertação, Bandim (2017) utiliza o *corpus* de 60 artigos de Souza (2005) no intuito de realizar um estudo de caso onde compara as palavras-chave desses artigos com os termos atribuídos pelo sistema Sistema de Indexação Semi-Automático (SISA) usando o Tesauro Brasileiro de Ciência da Informação (TBCI). A proposta de Bandim (2017) era analisar a indexação automática por atribuição na representação de artigos científicos da área de Ciência da Informação, fazendo uso de descritores definidos no tesauro.

O uso do *corpus* nos trabalhos de Bandim (2017), Bandim e Corrêa (2018) e Bandim e Corrêa (2019) visou validar mecanismos de atribuição de termos por sistema automático de indexação, bem como verificar se no contexto de uma base de dados de artigos científicos em Ciência da Informação, a utilização da indexação automática por atribuição do sistema SISA com base no TBCI gera uma indexação de qualidade.

Outro trabalho que também utilizou do corpus de 60 artigos de Souza (2005) foi o de Celerino (2018). Neste último trabalho, o *corpus* foi utilizado para avaliar uma metodologia de indexação automática baseada na normalização de sintagmas nominais extraídos do título e resumo dos documentos (CELERINO; CORRÊA, 2017; CORRÊA; CELERINO, 2019).

Já na presente pesquisa, o âmbito de aplicação deixa de ser sistema automático voltado aos profissionais indexadores e passa a ser o usuário-indexador. Nesse ponto o *corpus* auxilia a pesquisa a alcançar parâmetros que visem determinar como a aplicação de um controle terminológico na indexação social, em um espaço *web* aberto, pode gerar uma indexação de qualidade, aumentando assim a possibilidade de garantir melhores resultados na indexação e recuperação da informação. Vale lembrar que nesse ponto, o usuário, mesmo diante do controle terminológico, ainda terá liberdade sobre a produção de registro sobre o conteúdo a ser indexado.

## 4 Folksonomia assistida

O fato dos usuários participarem de forma ativa do processo de construção de um cenário informacional, gerando como resultado a folksonomia, faz com que diversas questões sobre a validade do que é produzido sejam levantadas.

Nascimento (2008) enfatiza que a folksonomia tendo emergido no cenário do software social, permite "[...] que as pessoas compartilhassem entre si não apenas os objetos digitais, mas também seus pensamentos, reflexões, críticas, e suas formas de indexar esses objetos." (NASCIMENTO, 2008, p. 42).

Corrêa e Santos (2018) enfatizam que a folksonomia determina um alto grau de liberdade para a categorização dos recursos informacionais e acentua a descentralização no processo de representação da informação, pois quem classifica o conteúdo são as próprias pessoas interessadas nele.

Diante da liberdade dada aos usuários na prática de descrição de conteúdos diversos, constata-se com frequência o encadeamento de pontos positivos e negativos da folksonomia no ato de representar a informação em sistemas colaborativos.

Nessa perspectiva, surge então na literatura uma abordagem híbrida no intuito de sanar alguns desses pontos negativos. Santarém Segundo (2010) e Santarém Segundo e Vidotti (2011) visam apresentar uma proposta de modelo de Representação Iterativa, que tem seu resultado denominado folksonomia assistida, e que tem como propósito estabelecer um método de recuperação semântica da informação em repositórios digitais.

A folksonomia assistida se concretiza através de um processo de apoio ao usuário no momento de definir os termos ou tags que representarão o conteúdo depositado em um repositório digital (SANTARÉM SEGUNDO, 2010).

De forma cooperativa, a folksonomia assistida tem um papel valioso no gerenciamento do vocabulário utilizado na representação informacional, em aspectos que contribuem para a não redundância de

informação, e na confiabilidade da pesquisa e recuperação dos artigos pelos usuários do repositório digital (SANTARÉM SEGUNDO; SIQUEIRA, 2013).

Assim, no momento em que o usuário vai informar a palavra-chave, acontece uma intervenção do sistema apresentando sugestões de *tags* similares presentes em uma estrutura de representação do conhecimento, que pode ser um tesauro ou ontologia por exemplo.

Nesse sentido, o modelo de indexação em questão oferece ao usuário uma quantidade de termos presentes em um vocabulário controlado, no intuito de que ele possa usar a base informacional do instrumento de controle terminológico para qualificar a descrição de um recurso. Resumindo, é uma espécie de **folksonomia controlada** que auxilia o usuário no tageamento do assunto do seu objeto informacional.

Santarém Segundo (2010) afirma que a folksonomia assistida visa a 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." (SANTARÉM SEGUNDO, 2010, p. 181).

Como resultado dessa intervenção, alguns dos principais problemas encontrados na folksonomia passam a ser evitados, tais como a polissemia, erros de grafia, entre outros.

Santos (2016) reforça que o modelo de indexação social em questão contempla atividades a serem realizadas por dois perfis de usuários (atores) no sistema, são eles: **o usuário administrador**, que deve ser um profissional da informação ou até mesmo uma equipe multidisciplinar composta por bibliotecário, arquivista e/ou cientista da informação responsável pela manutenção da estrutura de representação do conhecimento associada ao repositório digital; e, o **usuário autor do trabalho**, a ser depositado e indexado no repositório digital.

Mediante tais considerações, observa-se que nessa perspectiva, o método a ser utilizado opta por usar um modelo colaborativo: por meio de permissão, tendo em vista que somente o usuário autor do documento pode arquivar e descrever o seu recurso no âmbito do repositório digital.

Silva (2013) propõe um modelo colaborativo que tem como objetivo a construção de um catálogo *web* facetado, chamado por ele de *Facetlog*. Tal modelo oferece uma estratégia complementar à atribuição livre de etiquetas através da seleção de termos provenientes de uma taxonomia facetada. Neste cenário, a taxonomia facetada descrita pelo autor vem a ser um conjunto de facetas com conceitos subordinados de forma hierárquica.

Ainda segundo Silva (2013), tal modelo apresenta a necessidade específica do administrador do ambiente 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.

## 5 Procedimentos metodológicos

Para classificar essa pesquisa foi levado em consideração inicialmente, dentre seus diversos aspectos, a conceituação dos fundamentos da metodologia científica descritas por Gil (2002), Marconi e Lakatos (2006), no qual trabalham no intuito de explicitar os procedimentos sistemáticos e racionais, condensando a metodologia científica, técnicas de pesquisa e metodologia do trabalho científico.

Sendo assim, seguindo o primeiro passo nessa direção, a pesquisa se apresenta classificada como exploratória e empírica, envolvendo um estudo de caso. Para tanto, é salutar que o caráter exploratório se justifique em decorrência do pouco conhecimento acumulado e sistematizado sobre diversos aspectos em torno da folksonomia assistida.

A parte empírica da pesquisa envolve a realização de experimento que tem a finalidade de maximizar a qualidade do processo de representação da informação através da construção da folksonomia assistida por grupos de indexadores para os documentos que compõem o *corpus* de Souza (2005) e a posterior análise da folksonomia gerada.

A pesquisa tem como estudo de caso a aplicação de uma adequação do modelo colaborativo de indexação social desenvolvido enquanto protótipo na tese de Silva (2013), sendo disponibilizado através da plataforma *Tainacan* na versão *web*. As adequações principais foram: (1) a fonte de termos para construção da folksonomia assistida se dará através de consulta ao TBCCI em um ambiente externo ao *Tainacan*, ao invés de ser na própria plataforma; (2) não será permitido um usuário contestar a indexação feita por outro usuário, sendo a validação das *tags* inseridas pelos usuários-indexadores realizada posteriormente de forma manual pelo administrador.

Para alcançar o objetivo traçado foram estabelecidas etapas contemplando os procedimentos a serem realizados. Para isso, tomou-se como base a fundamentação teórico-metodológica, e assim o percurso metodológico segue a seguinte estrutura:

**Etapa 1:** Construção da folksonomia assistida por meio da indexação social do corpus de artigos de periódicos em Ciência da Informação:

· Configuração e prepararação de uma coleção no *Tainacan*, levando em conta a adaptação do modelo colaborativo de indexação social de Silva (2013) quanto a papeis, atividades e regras de interação entre usuários, bem como inserção na coleção dos artigos do *corpus* de Souza (2005) com respectivos metadados descritivos;

· Desenvolver uma política de indexação social assistida adotando o Tesauro brasileiro de Ciência da Informação no processo de indexação, direcionando o registro de descritores no campo de metadado *tags* da coleção;

· Instruir os usuários indexadores e guiá-los no processo de indexação social assistida in lócus na plataforma *Tainacan*.

**Etapa 2:** Analisar a folksonomia assistida resultante do processo de indexação social assistida do *corpus* de Souza (2005) por grupos de indexadores.

· Analisar o resultado do processo de indexação colaborativa realizada pelos grupos de usuários, mediante comparação daquilo que consta disponibilizado na plataforma *Tainacan* no campo *tags* e no campo de metadado Assuntos a partir das métricas de avaliação da qualidade da indexação.

Seguindo essas etapas, a proposta perpassa por deixar evidente o caminho para a realização do estudo de caso estabelecido para a pesquisa.

## 6 Análise dos resultados

A folksonomia assistida resultante da indexação social dos 60 artigos que compõem o *corpus*, da exata forma que foi gerada, incluindo os problemas no uso de termos que não constam no TBCI, de termos com caracteres especiais e letras maiúsculas indevidas, contabilizou 261 *tags* diferentes, totalizando 583 registros totais.

É salutar que a folksonomia assistida gerada não apresentou erros de escrita, contudo, resultaram algumas inconsistências no que diz respeito ao uso incorreto de caracteres e letras maiúsculas. Para isso, algumas regras de correção e exclusão foram levadas em consideração, para que de fato se chegasse ao resultado final. Foram elas:

a) Não ser um termo autorizado no TBCI (**Quadro 1**)– regra de exclusão;

b) Uso indevido de caracteres (**Quadro 2**)– regra de correção;

c) Uso indevido de letras maiúsculas (**Quadro 3**) – regra de correção.

Os quadros a seguir apresentam separadamente os problemas identificados na folksonomia assistida gerada.

QUADRO 1
Termos descartados da folksonomia assistida gerada

| Termos | |
|---|---|
| Autonomia | Informação |
| ciência e tecnologia | informação e sociedade |
| ciências sociais | informação histórica |
| Código | memória social |
| Conhecimento | métodos quantitativos de avaliação |
| conhecimento científico | pós-graduação |
| Contexto | produtividade em pesquisa |
| contrato social | profissão e mercado de trabalho |
| Documento | Rio de Janeiro |
| espaço prisional | sociologia da leitura |
| estoques informacionais | teorias da leitura |
| estrutura informacional | tipo de usuário |
| estudo da leitura | vannevar bush |
| estudos mnemônicos | |

Fonte: Elaborado pelos autores.

QUADRO 2
Erros no uso de caracteres na folksonomia assistida gerada

| Como foi inserido | Forma correta |
|---|---|
| Gestão da informação. | gestão da informação |
| Informação tecnológica. | informação tecnológica |
| Inteligência Competitiva; | inteligência Competitiva |
| Inteligência empresarial. | inteligência empresarial |
| Projetos de pesquisa. | projetos de pesquisa |
| Recuperação da informação. | recuperação da informação |
| Serviços de informação. | serviços de informação |
| Teoria da informação. | teoria da informação |
| Transdisciplinaridade. | transdisciplinaridade |

Fonte: Elaborado pelos autores.

QUADRO 3
Erros no uso de letras maiúsculas e minúsculas na folksonomia assistida gerada

| Como foi inserido | Forma correta |
|---|---|
| Agentes Inteligentes | agentes inteligentes |
| Análise de dados | análise de dados |
| Análise quantitativa | análise quantitativa |
| Áreas do conhecimento | áreas do conhecimento |
| Artigos de periódico | artigos de periódico |
| Automação | automação |
| Avaliação | avaliação |
| Bases de Dados Cadastrais | bases de dados cadastrais |
| Bibliometria | bibliometria |
| Bibliotecários | bibliotecários |
| Bibliotecas | bibliotecas |
| Bibliotecas digitais | bibliotecas digitais |
| Bibliotecas virtuais | bibliotecas virtuais |
| Bolsas de Pesquisa | bolsas de pesquisa |
| Categorias | categorias |
| Ciberespaço | ciberespaço |
| Cibermetria | cibermetria |
| Cienciometria | cienciometria |
| Comércio eletrônico | comércio eletrônico |
| Complexidade | complexidade |
| Compressão de Dados | compressão de dados |
| Comunicação científica | comunicação científica |
| Conhecimento organizacional | conhecimento organizacional |
| Direito à informação | direito à informação |
| Disseminação da informação. | disseminação da informação. |
| Documentação | documentação |
| Educação à distância | educação distância |
| Educação em ciência da informação | educação em ciência da informação |
| Ensino de biblioteconomia | ensino de biblioteconomia |
| Ensino e Pesquisa em Ciência da Informação e Áreas Afins | ensino e pesquisa em ciência da informação e áreas afins |
| Epistemologia | epistemologia |
| Epistemologia da Ciência da Informação | epistemologia da ciência da informação |
| Estudos de Usuários | estudos de usuários |
| Governo eletrônico | governo eletrônico |
| Hipertextos | hipertextos |
| Indicadores de C&T | indicadores de C&T |
| Informetria | informetria |
| Inovação | inovação |
| Inovação Tecnológica | inovação tecnológica |
| Internet | internet |
| Intranetes | intranetes |
| Lei de lotka | lei de Lotka |
| Levantamentos | levantamentos |
| Literatura cinzenta | literatura cinzenta |
| Mediadores da informação | mediadores da informação |
| Métodos quantitativos de avaliação | métodos quantitativos de avaliação |
| Notícias | notícias |
| Organização da informação | organização da informação |
| Organização do conhecimento | organização do conhecimento |
| Periódicos | periódicos |
| Pesquisa | pesquisa |
| Pesquisa em ciência da informação. | pesquisa em ciência da informação. |
| Políticas públicas | políticas públicas |
| Produtividade cientifica | produtividade cientifica |
| Programas de pós-graduação | programas de pós-graduação |
| Publicações científicas: periódicos | publicações científicas: periódicos |
| Semântica | semântica |
| Sociedade da Informação | sociedade da informação |
| Tecnologia da informação | tecnologia da informação |
| TRANSFERÊNCIA DA INFORMAÇÃO | transferência da informação |
| Unidades de informação | unidades de informação |
| Usabilidade | usabilidade |

Fonte: Elaborado pelos autores.

É importante destacar que as inconsistências em ambos os casos, erro indevido de caracteres e uso de letras maiúsculas ou minúsculas, foram conferidas pelo administrador de acordo com o que é apresentado no TBCI, reiterando seu papel de mediador da plataforma. Além disso, é necessário mencionar que mesmo com a inserção da forma indevida, o primeiro registro passa a ser o padrão para todos os demais, impedindo que mais de uma forma tenha sido inserida para o mesmo termo.

Nesse cenário, mesmo com os erros destacados, a folksonomia assistida desempenhou um papel importante, pois suas características negativas, que são oriundas da folksonomia, foram reduzidas, garantindo que o resultado possa ser utilizado para avaliar sistemas de indexação automática, proposta da presente pesquisa.

Partindo do ponto que foi gerada uma folksonomia assistida, a sua revisão se fez necessária. É importante frisar que cabe ao administrador o papel de revisar todas as questões e problemas levantados, gerando assim o que aqui será chamado de folksonomia assistida revisada. Isto para garantir a correta avaliação da qualidade da indexação automática a partir da compilação do *corpus* de referência.

Na **Tabela 1**, são apresentados indicadores de qualidade da indexação do *corpus* a partir da análise da folksonomia assistida revisada.

Com base nos dados apresentados na **Tabela 1** no item Assunto, à primeira vista não existe uma determinação por parte das revistas analisadas quanto à quantidade obrigatória de palavras-chave a serem utilizadas nos documentos, causando assim uma grande variação de um artigo para outro. Para o campo Assuntos foi identificado um mínimo de 2 e um máximo de 9 termos atribuídos para indexação, gerando uma média de 4 a 5 descritores ou palavras-chaves por documento.

Referente ao número de termos gerados da folksonomia assistida, por ter sido gerada por três grupos de usuários se poderiam encontrar 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.

A quantidade de termos comuns foi determinada confrontando as palavras-chave, geradas pelos autores, e *tags* geradas pelos usuários-indexadores, levando em consideração não só a mesma escrita, mas os conceitos e respectivos termos autorizados no TBCI.

TABELA 1
Coeficientes de consistência, precisão, revocação e medida F

| ID | ASSUNTO | TAGS | TERMOS COMUNS | CONSISTÊNCIA | P | R | MEDIDA F |
|---|---|---|---|---|---|---|---|
| Artigo 01 | 3 | 10 | 2 | 18% | 20% | 67% | 31% |
| Artigo 02 | 2 | 10 | 2 | 20% | 20% | 100% | 33% |
| Artigo 03 | 5 | 11 | 1 | 7% | 9% | 20% | 13% |
| Artigo 04 | 5 | 9 | 2 | 17% | 22% | 40% | 29% |
| Artigo 05 | 5 | 11 | 2 | 14% | 18% | 40% | 25% |
| Artigo 06 | 7 | 10 | 5 | 42% | 50% | 71% | 59% |
| Artigo 07 | 5 | 10 | 5 | 50% | 50% | 100% | 67% |
| Artigo 08 | 5 | 12 | 2 | 13% | 17% | 40% | 24% |
| Artigo 09 | 5 | 10 | 2 | 15% | 20% | 40% | 27% |
| Artigo 10 | 5 | 10 | 2 | 15% | 20% | 40% | 27% |
| Artigo 11 | 4 | 10 | 3 | 27% | 30% | 75% | 43% |
| Artigo 12 | 6 | 10 | 6 | 60% | 60% | 100% | 75% |
| Artigo 13 | 3 | 11 | 2 | 17% | 18% | 67% | 29% |
| Artigo 14 | 6 | 10 | 3 | 23% | 30% | 50% | 38% |
| Artigo 15 | 2 | 10 | 3 | 33% | 30% | 100% | 46% |
| Artigo 16 | 5 | 9 | 4 | 40% | 44% | 80% | 57% |
| Artigo 17 | 3 | 12 | 4 | 36% | 33% | 133% | 53% |
| Artigo 18 | 4 | 14 | 2 | 13% | 14% | 50% | 22% |
| Artigo 19 | 5 | 11 | 4 | 33% | 36% | 80% | 50% |
| Artigo 20 | 4 | 11 | 4 | 36% | 36% | 100% | 53% |
| Artigo 21 | 5 | 6 | 1 | 10% | 17% | 20% | 18% |
| Artigo 22 | 9 | 8 | 5 | 42% | 63% | 56% | 59% |
| Artigo 23 | 3 | 9 | 0 | 0% | 0% | 0% | 0% |
| Artigo 24 | 6 | 8 | 1 | 8% | 13% | 17% | 14% |
| Artigo 25 | 2 | 11 | 2 | 18% | 18% | 100% | 31% |
| Artigo 26 | 4 | 12 | 2 | 14% | 17% | 50% | 25% |
| Artigo 27 | 3 | 12 | 1 | 7% | 8% | 33% | 13% |
| Artigo 28 | 7 | 11 | 3 | 20% | 27% | 43% | 33% |
| Artigo 29 | 3 | 7 | 3 | 43% | 43% | 100% | 60% |
| Artigo 30 | 3 | 13 | 2 | 14% | 15% | 67% | 25% |
| Artigo 31 | 5 | 9 | 3 | 27% | 33% | 60% | 43% |
| Artigo 32 | 6 | 8 | 4 | 40% | 50% | 67% | 57% |
| Artigo 33 | 5 | 11 | 3 | 23% | 27% | 60% | 38% |
| Artigo 34 | 3 | 12 | 2 | 15% | 17% | 67% | 27% |
| Artigo 35 | 5 | 6 | 5 | 83% | 83% | 100% | 91% |
| Artigo 36 | 4 | 11 | 3 | 25% | 27% | 75% | 40% |
| Artigo 37 | 2 | 10 | 2 | 20% | 20% | 100% | 33% |
| Artigo 38 | 3 | 12 | 2 | 15% | 17% | 67% | 27% |
| Artigo 39 | 5 | 9 | 4 | 40% | 44% | 80% | 57% |
| Artigo 40 | 2 | 11 | 2 | 18% | 18% | 100% | 31% |
| Artigo 41 | 4 | 12 | 3 | 23% | 25% | 75% | 38% |
| Artigo 42 | 5 | 10 | 3 | 25% | 30% | 60% | 40% |
| Artigo 43 | 4 | 10 | 2 | 17% | 20% | 50% | 29% |
| Artigo 44 | 3 | 11 | 3 | 27% | 27% | 100% | 43% |
| Artigo 45 | 5 | 8 | 4 | 44% | 50% | 80% | 62% |
| Artigo 46 | 6 | 6 | 4 | 50% | 67% | 67% | 67% |
| Artigo 47 | 3 | 10 | 2 | 18% | 20% | 67% | 31% |
| Artigo 48 | 3 | 9 | 2 | 20% | 22% | 67% | 33% |
| Artigo 49 | 3 | 12 | 3 | 25% | 25% | 100% | 40% |
| Artigo 50 | 2 | 10 | 2 | 20% | 20% | 100% | 33% |
| Artigo 51 | 5 | 8 | 2 | 18% | 25% | 40% | 31% |
| Artigo 52 | 5 | 6 | 3 | 38% | 50% | 60% | 55% |
| Artigo 53 | 5 | 6 | 4 | 57% | 67% | 80% | 73% |
| Artigo 54 | 7 | 8 | 3 | 25% | 38% | 43% | 40% |
| Artigo 55 | 4 | 6 | 3 | 43% | 50% | 75% | 60% |
| Artigo 56 | 5 | 9 | 3 | 27% | 33% | 60% | 43% |
| Artigo 57 | 7 | 9 | 2 | 14% | 22% | 29% | 25% |
| Artigo 58 | 7 | 7 | 6 | 75% | 86% | 86% | 86% |
| Artigo 59 | 5 | 9 | 3 | 27% | 33% | 60% | 43% |
| Artigo 60 | 4 | 9 | 4 | 44% | 44% | 100% | 62% |
| Média | 4,4 | 9,7 | 2,8 | 28% | 32% | 68% | 41% |

Fonte: Elaborado pelos autores.

A quantidade de termos comuns alternou de um mínimo de 0 (zero), em apenas um caso onde nenhum termo coincidiu (**Artigo 23**), a um máximo de 6 (seis) termos comuns, presentes em dois casos (**Artigo 12** e **58**). A média de termos comuns variou entre dois e três termos, tendendo a três termos comuns por documento.

Na avaliação da qualidade da indexação foram utilizadas as formulas dos índices de consistência, revocação, precisão e medida F. A avaliação foi realizada levando em consideração os termos presentes no campo Assuntos, que consistem das palavras-chave de cada documento, e os termos presentes no campo *tags*, consistindo dos termos da folksonomia assistida revisada.

Ao realizar de imediato uma avaliação generalista dos resultados alcançados, é possível afirmar, confrontando as palavras-chave com as *tags*, que a grande quantidade de *tags* influenciou diretamente nos cálculos dos índices de coeficiente de consistência e os índices de precisão, revocação e medida F.

Bandim e Corrêa (2018) categorizaram em níveis de desempenho faixas de valores do índice de consistência, de acordo com a fórmula do índice de consistência:

a) 0 a 11% - desempenho insatisfatório (corresponde a uma média de 0 a 1 termo em comum);
b) 11 a 25% - desempenho satisfatório (corresponde a uma média de 1 a 2 termos em comum);
c) 25 a 43% - desempenho bom (corresponde a uma média de 2 a 3 termos em comum);
d) 43 a 67% - desempenho ótimo (corresponde a uma média de 3 a 4 termos em comum);
e) 67 a 100% - desempenho excelente (corresponde a uma média de 4 a 5 termos em comum).

Sendo que para cálculo dos limites dos intervalos apresentados, foi estipulado o número médio de termos de indexação atribuídos pelas duas indexações iguais a cinco.

Com base nas faixas de valores apresentadas, o resultado médio de 28% para o coeficiente de consistência permite entender que o desempenho da folksonomia assistida é bom, ou seja, o nível de consistência entre as tags dos usuários e as palavras-chaves dos autores é bom.

Na sequência, a aplicação das fórmulas gerou um índice de precisão médio de 32%, número muito abaixo do que foi descrito por Hlava (2002) que seria por volta de 60% para ser considerado bom, logo, inversamente proporcional ao número de revocação médio maior do que o esperado, 68%. Assim, percebe-se que a folksonomia alcança um bom nível de revocação das palavras-chaves, mas que por atribuir muitos termos não-coincidentes com as palavras-chaves acaba alcançando um nível de precisão apenas satisfatório.

Como a medida F é a média harmônica entre precisão e revocação, entende-se que na busca por uma melhor harmonia entre revocação e precisão seria possível obter ambos os índices iguais a 41%, caso o entendimento comum do autor e do usuário acerca do mesmo documento, expresso através de termos comuns de indexação, girasse em torno de duas a cada cinco palavras-chaves do documento. Entretanto, uma alternativa para aumentar a precisão, sem necessariamente diminuir a revocação, seria atribuir menos *tags*, buscando uma indexação mais seletiva através de critérios para esta seleção.

Diante do cenário apresentado, alguns pontos precisam ser destacados individualmente no que se refere à relação do número de palavras-chave, o número de termos da folksonomia assistida e o número de termos em comum.

É interessante observar que o fato de possuir maior número de termos da folksonomia assistida, não necessariamente garante que vão existir termos em comum, sendo os casos mais emblemáticos os documentos **Artigo 58** e **Artigo 23**.

No **Artigo 58**, é possível notar que a quantidade de termos é idêntica em ambos os campos de metadados, e ainda assim o resultado de termos em comum entre eles é de quase 100%. Já no caso do **Artigo 23**, mesmo tendo um número de nove *tags* e três palavras-chave, o número de termos em comum é zero. Nesse caso em específico, isso pode ocorrer por falta de consenso terminológico entre autor e leitores na representação da publicação.

Nesse ponto, é importante acrescentar à discussão o fato de que muitas das palavras-chave utilizadas pelos autores não constam no TBCI, resultando numa falta de controle terminológico e, consequentemente, problemas de representação e recuperação da informação. Ao relatar tal aspecto, é importante destacar que o uso de um tesauro permite limitar de forma semântica a aplicação de termos a serem utilizados numa base de dados, dando maior respaldo para melhores resultados nas buscas por informações.

Na **Figura 1** é apresentada uma nuvem de *tags* que permite visualizar a folksonomia assistida. Nela é possível visualizar os termos mais usados pelos usuários. Vale lembrar que o tamanho no qual o termo se apresenta evidencia sua frequência de uso, ou seja, quanto maior sua fonte na nuvem de *tags*, mais vezes o termo foi utilizado. Além disso, a visualização em forma de nuvem, onde palavras como Ciência da Informação, Sociedade da Informação e acesso à informação são recorrentes, permite conhecer os termos representativos do *corpus* segundo a perspectiva do usuário-indexador.

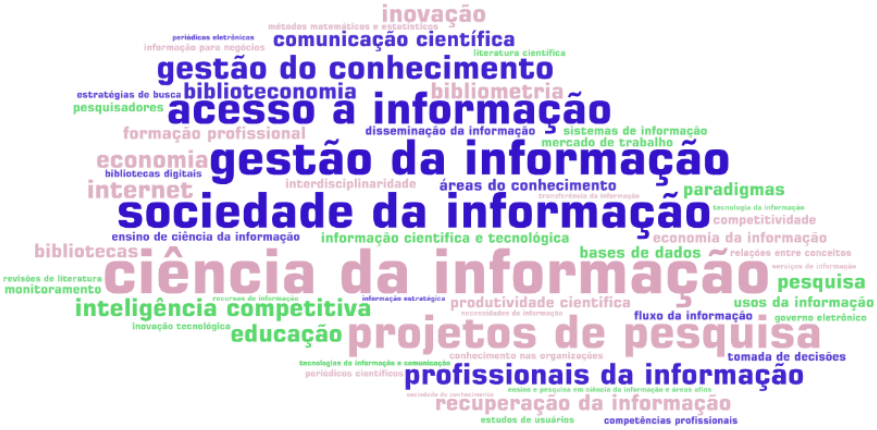


FIGURA 1
Nuvem de *tags* do *corpus*
Fonte: Elaborado pelos autores.

Diante de uma breve comparação dos resultados obtidos por Bandim (2017) e Celerino (2018), onde ambos utilizaram o mesmo *corpus* em suas respectivas pesquisas, é possível observar que de uma forma geral, é preciso se ter ciência de que para alcançar melhores índices de consistência, revocação, precisão e medida F, devem ser levadas em consideração as necessidades semânticas na seleção dos termos de indexação.

Bandim e Corrêa (2018) realizam uma análise acerca da consistência na indexação automática por atribuição de artigos científicos na área de Ciência da Informação. Os autores, que se utilizam dos textos completos dos artigos que compõem o *corpus* de Souza (2005), chegam à conclusão de que a utilização da indexação automática, com base no Tesauro Brasileiro de Ciência da Informação (TBCI) e o uso do *software* SISA, proporciona uma consistência média satisfatória na indexação de artigos científicos em português da área de Ciência da Informação. Em contraponto, o presente trabalho apresenta um índice de consistência média de nível bom quando se leva em consideração as palavras-chaves e *tags* com base no TBCI.

Ao buscar mensurar o índice de revocação na indexação automática por sintagmas nominais de artigos de periódicos em Ciência da Informação, Corrêa e Celerino (2019) utilizaram os títulos e resumos do mesmo *corpus*. O índice de revocação médio obtido foi de 56% das palavras-chaves. Os autores apontaram que é viável a utilização da indexação automática por sintagmas nominais do título e do resumo na construção de bases de dados científicas na área de Ciência da Informação. Na presente pesquisa, obteve-se um nível médio de revocação de palavras-chaves maior, que foi de 68%, isto devido principalmente à liberdade de atribuição de termos do TBCI que podem não aparecer no título ou resumo dos artigos.

Celerino e Corrêa (2017) descrevem que a maior adoção pela comunidade científica das normas existentes para elaboração de resumo e atribuição de termos de tesauros especializados pode minimizar a limitação

quanto à uniformidade, expressividade e coerência da representação dos trabalhos científicos via palavras-chave. A presente pesquisa contribui com a indexação retrospectiva do *corpus* utilizado, aperfeiçoando a indexação dos artigos.

Por fim, é importante observar que em ambos os trabalhos, de alguma forma a necessidade de normalização dos termos de indexação trazem resultados positivos, o que foi contemplado na indexação social assistida do *corpus*.

Diante do que foi discutido nesta seção, aponta-se que a folksonomia assistida gerada se apresenta como alternativa ao uso das palavras-chave dos autores na avaliação de sistemas de indexação automática, agregando termos especializados no processo de representação e recuperação dos artigos.

## 7 Considerações finais

A proposta de utilização da folksonomia assistida na construção de *corpus* de referência de artigos científicos em Ciência da Informação permitiu que os principais aspectos em torno da temática fossem validados.

No que se refere aos resultados da aplicação do modelo de indexação social assistida, foi possível tomar conhecimento sobre os desdobramentos da participação ativa e coordenada dos usuários frente à construção de uma linguagem para representação da informação em ambientes colaborativos. A garantia de uma maior qualidade na indexação de artigos científicos, assim como a melhor avaliação dos sistemas de indexação automática através do *corpus* compilado foi maximizada pelo papel desempenhado pelos usuários-indexadores e administrador.

No uso da plataforma *Tainacan* como sistema colaborativo, a configuração da coleção de acordo com o modelo de indexação social almejado e a política de indexação social assistida foram essenciais para a correta construção da folksonomia, ressaltando que ambas tiveram como elemento central a adoção de um tesauro da área de Ciência da Informação na indexação social assistida de artigos científicos do *corpus* de Souza (2005).

Nesse ponto, é válido destacar que o uso do TBCI se apresentou coerente ao *corpus* escolhido, apesar de não abranger de forma satisfatória o que os usuários entendiam como propício para representar o conteúdo de alguns artigos, reforçando a necessidade de atualização (aperfeiçoamento) do tesauro em questão.

Quanto à qualidade da indexação da folksonomia assistida, os índices verificados para determinar o grau de qualidade da indexação apresentaram médias de 28% do coeficiente de consistência, 32% de precisão, 68% de revocação, e 41% de medida F. Quando avaliados conjuntamente, esses índices médios nos levam a entender que o nível de harmonização entre as palavras-chave produzidas pelos autores e as tags produzidas pelos indexadores, alcançam um resultado bom.

A folksonomia assistida aplicada na construção de *corpus* de referência efetiva a indexação de melhor qualidade dos documentos e possibilita a compilação do corpus para fins de avaliação de sistemas de indexação automática. A partir do *corpus* compilado será possível avaliar a qualidade da indexação automática levando em conta agora as *tags* definidas pelos indexadores com base no TBCI.

Adicionalmente, como contribuições originais deste trabalho, podem ser apontadas o uso da plataforma *Tainacan* na construção de folksonomia assistida, e a apresentação de resultados práticos e substanciais para a temática e para a evolução da plataforma na constituição de coleções digitais.

corpus

## Referências

BANDIM, M. A. S. **Indexação automática por atribuição de artigos científicos da área de Ciência da Informação.** 2017. 144 f. Dissertação (Mestrado em Ciência da Informação) Universidade Federal de Pernambuco, Recife, 2017.

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

BANDIM, M. A. S.; CORRÊA, R. F. Indexação automática por atribuição de artigos científicos em português da área de Ciência da Informação. **Transinformação**, Camppinas, v. 31, p. 1-12, 2019.

CELERINO, V. G. **Proposta de normalização dos sintagmas nominais em termos para indexação automática.** 2018. 176 f. Dissertação (Mestrado em Ciência da Informação) Universidade Federal de Pernambuco, Recife, 2018.

CELERINO, V. G.; CORRÊA, R. F. A revocação na indexação automática por sintagmas nominais de artigos de periódicos em Ciência da Informação. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2017, Marília. **Anais... .**Marília: Ancib, 2017. p. 1 - 20.

CORRÊA, R. F.; CELERINO, V. G. Método de normalização de sintagmas nominais na indexação automática. **Em questão**, Porto Alegre, v. 25, n. 1, p. 321-344, 2019.

CORRÊA, 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.

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

GIL LEIVA, I. **La automatización de la indización, propuesta teórico-metodológica**: aplicación al área de Biblioteconomía y Documentación.1997. 268f. Tese (Doutorado em Filosofia e Letras) – Universidad de Murcia, Murcia, España, 1997.

GIL LEIVA, I.; RUBI, M. P.; FUJITA, M. S. L. Consistência na indexação em bibliotecas universitárias brasileiras. **Transiformação,** Campinas,v. 20, n. 3, p. 233-253, set./dez. 2008.

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.

HLAVA, M. M. Automatic indexing: a matter of degree. **Bulletin of the American Society for Information Science and Technology**, Maryland, v. 29, n. 1, p. 12-15, out./nov. 2002.

KURAMOTO, H. Uma abordagem alternativa para o tratamento e a recuperação de informação textual: os sintagmas nominais. **Ciência da Informação**, Brasília, v. 25, n. 2, p. 1-18, 1995.

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

NARUKAWA, C. M.; GIL LEIVA, I.; FUJITA, M. S. L. Indexação automatizada de artigos de periódicos científicos: análise da aplicação do software SISA com uso da terminologia DeCS na área de odontologia. **Informação. & Sociedade**: Estudos, João Pessoa, v. 19, n. 2, p. 99-118, maio/ago. 2009.

NASCIMENTO, G. F. C. **Folksonomia como estratégia de indexação dos bibliotecários no Del.icio.us**. 2008. 82f. Dissertação (Mestrado em Ciência da Informação) - Universidade Federal da Paraíba, João Pessoa, 2008.

SALTON, G.; MCGILL, M. J. **Introduction to modern information retrieval**. 2. ed. New York: McGraw-Hill, 2003.

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

SANTARÉM SEGUNDO, J. E.; SIQUEIRA, C. S. Aplicación teórico-conceptual de folksonomías asistidas para la recuperación de información. **Scire**: Representación y organización del conocimento, 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.

SANTOS, R. F. dos. **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 em Ciência da Informação) - Universidade Federal de Pernambuco, Recife, 2016.

SANTOS, R. F. dos; CORRÊA, R. F. A Folksonomia e a representação colaborativa da informação em ambientes digitais. **Tendências da Pesquisa Brasileira em Ciência da Informação**, João Pessoa, v. 8, n. 1, p. 69-84, 2015.

SILVA, M. F. **Proposta de modelo de colaboração para catálogo web facetado**. Belo Horizonte, 2013. 269 f. Tese (Doutorado em 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 em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2005.