



UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE ARTES E COMUNICAÇÃO  
DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

BRUNO FELIPE DE MELO SILVA

**A FOLKSONOMIA ASSISTIDA APLICADA NA CONSTRUÇÃO DE CORPUS DE  
REFERÊNCIA PARA AVALIAÇÃO DE SISTEMAS DE INDEXAÇÃO  
AUTOMÁTICA**

Recife  
2019

BRUNO FELIPE DE MELO SILVA

**A FOLKSONOMIA ASSISTIDA APLICADA NA CONSTRUÇÃO DE CORPUS DE  
REFERÊNCIA PARA AVALIAÇÃO DE SISTEMAS DE INDEXAÇÃO  
AUTOMÁTICA**

Dissertação apresentada ao Programa de Pós-graduação em Ciência da Informação da Universidade Federal de Pernambuco, como requisitos parcial para a obtenção do título de Mestre em Ciência da Informação.

**Área de concentração:** Informação, Memória e Tecnologia.

**Orientador:** Prof. Dr. Renato Fernandes Corrêa

Recife

2019

Catálogo na fonte  
Bibliotecária Jéssica Pereira de Oliveira, CRB-4/2223

S586f Silva, Bruno Felipe de Melo  
A folksonomia assistida aplicada na construção de Corpus de Referência para avaliação de sistemas de indexação automática / Bruno Felipe de Melo Silva. – Recife, 2019.  
188f.: il.

Orientador: Renato Fernandes Corrêa.  
Dissertação (Mestrado) – Universidade Federal de Pernambuco. Centro de Artes e Comunicação. Programa de Pós-Graduação em Ciência da Informação, 2019.

Inclui referências, anexo e apêndices.

1. Folksonomia Assistida. 2. Indexação Social. 3. Corpus de Referência. 4. Avaliação de Sistemas de Indexação Automática. 5. Ciência da Informação. I. Corrêa, Renato Fernandes (Orientador). II. Título.

020 CDD (22. ed.) UFPE (CAC 2019-108)

BRUNO FELIPE DE MELO SILVA

**A FOLKSONOMIA ASSISTIDA APLICADA NA CONSTRUÇÃO DE CORPUS DE  
REFERÊNCIA PARA AVALIAÇÃO DE SISTEMAS DE INDEXAÇÃO  
AUTOMÁTICA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Mestre em Ciência da Informação.

Aprovada em: 25/02/2019

**BANCA EXAMINADORA**

---

Prof. Dr. Renato Fernandes Corrêa (Orientador)  
Universidade Federal de Pernambuco

---

Prof. Dr. Fábio Assis Pinho (Examinador Interno)  
Universidade Federal de Pernambuco

---

Prof. Dr. Guilherme Ataíde Dias (Examinador Externo)  
Universidade Federal da Paraíba

À minha esposa Janyelle, aos meus pais, Marcos e Meire, e por fim, a todos aqueles que incansavelmente estão sempre olhando inquietos ao redor.

## AGRADECIMENTOS

A *Deus*, por ter guiado e abençoado todas as minhas idas e vindas no percurso Maceió-Recife-Maceió.

Ao orientador, *Professor Renato Fernandes Corrêa*, que sempre esteve disponível, e que por muito contribuiu durante toda a minha caminhada no mestrado.

Ao *Professor Fábio Assis Pinho* e ao *Professor Diego Salcedo* por todas as sugestões para avanço da pesquisa apresentadas durante o exame de qualificação dissertação.

Aos *Professores Fábio Assis Pinho* e *Guilherme Ataíde Dias*, por aceitarem participar desta construção. Por dispensar sua sabedoria e experiência para que esta pesquisa ganhasse em qualidade e fizesse sentido.

À minha esposa *Janyelle*, pelo amor, incentivo e carinho nos momentos mais tensos dessa trajetória.

Aos meus pais *Marcos* e *Meire*, e meu irmão *Paulinho*, por nunca deixarem de me lembrar o quanto sou querido.

Aos *amigos* que sempre estão comigo na presença e na ausência.

Aos *colegas* de turma *Ismael*, *Elanna*, *Geórgia*, *Sandryne*, *Patrícia*, *Rúbia*, *Wiwiam*, *Wilma*, *João*, *Salim*, *Mônica*, *Renata*, *Paula*, *Henry* e *Tatyane* pelos grandes debates dentro e fora da sala de aula.

Aos professores e funcionários do PPGCI-UFPE.

A *Universidade Federal de Alagoas* por ter me possibilitado um afastamento para cumprir com minhas obrigações junto ao PPGCI-UFPE.

A todos, e que não foram poucos, que de alguma forma confiaram, deram palavras de incentivo e me apoiaram no dia a dia dessa caminhada.

A todos os *caroneiros* que se aventuram nessas BRs semanalmente em busca de sonhos e conquistas. Compartilhar todas as idas e vindas fez com que tudo fosse menos desgastante.

Por fim, agradeço a todas as dificuldades que enfrentei, incluindo os momentos de solidão longe da família. Não fosse por elas, eu não teria saído ainda mais fortalecido dessa trajetória. As facilidades nos impedem de caminhar, e nesses dois anos eu aprendi e me tornei uma nova pessoa.

Não se trata do que você sabe, e não se trata nem mesmo de quem você conhece. Trata-se do volume de conhecimento que você transmite. O acúmulo de conhecimento diminui seu poder. (WEINBERGER, 2007, p. 45).

## RESUMO

O presente trabalho visa correlacionar os conceitos de folksonomia assistida e indexação automática, propondo meios para que a primeira possa ser aplicada na avaliação de sistemas de indexação automática a partir da compilação de Corpus de Referência. A perspectiva é ser capaz de responder ao seguinte problema de pesquisa: 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? Para isso, buscou-se alcançar os objetivos específicos que partem inicialmente da ideia de construção da Folksonomia Assistida por meio da adaptação do modelo colaborativo de indexação social assistida de Silva (2013), onde são contemplando papéis, atividades e regras de interação entre usuários, com base no desenvolvimento de uma política de indexação social assistida que adota o Tesouro Brasileiro de Ciência da Informação. O resultado da aplicação do modelo norteou o segundo objetivo que era o de avaliar a qualidade da indexação do corpus de Souza (2005). O terceiro objetivo visou à compilação do corpus de referência de artigos científicos na área de Ciência da Informação para avaliação de sistemas de indexação automática. A metodologia construída para o presente trabalho pode ser classificada como exploratória, empírica e de estudo de caso, tendo em vista o experimento realizado. Para a pesquisa fora delimitado o uso do *corpus de referência* composto por 60 artigos da tese de doutorado de Souza (2005). O modelo colaborativo de indexação social assistida do *corpus* ocorreu por meio da plataforma de gerenciamento de coleção, Tainacan. O Tainacan foi utilizado como instrumento para garantir a mediação da pesquisa, visando à construção da Folksonomia Assistida. Como forma de alcançar os objetivos propostos na pesquisa foi definida etapas a serem seguidas. Essas etapas envolveram: 1) Levantar e analisar a literatura a cerca da Folksonomia assistida e a construção de corpus de referência para avaliação da indexação automática; 2) Configurar e preparar uma coleção no Tainacan, levando em conta papéis, atividades e regras de interação entre usuários, bem como inserção na coleção dos artigos do corpus de artigos de periódicos em Ciência da Informação de Souza (2005) com respectivos metadados descritivos; 3) Desenvolver uma política de indexação social assistida adotando o Tesouro brasileiro em Ciência da Informação no processo de indexação colaborativa assistida, a ser registrado no

campo de metadado *Tags* da coleção; 4) Instruir os usuários indexadores e guiá-los no processo de indexação social assistida; 5) Analisar o resultado do processo de indexação colaborativa realizada pelos grupos de usuários, mediante comparação daquilo que consta disponibilizado no campo de metadado “assuntos”; 6) Definir e aplicar regras para compilação do corpus de referência na área de Ciência da Informação para avaliação de sistemas de indexação automática. O resultado das análises do grau de qualidade da indexação tem-se média de 32% de precisão, 69% de revocação, 28% do coeficiente de consistência e 41% de medida F. As médias alcançadas apresentam assim, algumas discrepâncias no que pode ser tratado como nível de harmonização entre as palavras-chave definidas pelos autores, quando comparada às dos usuários, acarretando problemas, dentre tantas, no ato da recuperação dos artigos. Por fim, chegou-se a conclusão que o uso da Folksonomia Assistida pode vir a ser útil no aperfeiçoamento dos objetos, processos e produtos construídos no âmbito da Ciência da Informação, principalmente no que se refere ao corpus compilado que permitiu chegar a um cenário adequado frente as correção e exclusão de termos da folksonomia realizadas.

**Palavras-chave:** Folksonomia Assistida. Indexação Social. Corpus de Referência. Avaliação de Sistemas de Indexação Automática. Ciência da Informação.

## ABSTRACT

The present work goal is to correlate Assisted Folksonomy and automated indexing, propositioning ways in which the first can be applied on the assessments of automated indexing systems that come from the compilation of Reference Corpus. The perspective is to be able to respond to the following search issue: In the construction of Reference Corpus with a goal of automated indexing systems assessments, the use of Assisted Folksonomy brings a better quality in the indexing of scientific articles? To achieve that, was sought specific goals that come initially from the idea of an Assisted Folksonomy through the adaptation of the collaborative model of assisted social indexing by Silva (2013), where are considered roles, activities and interaction rules between users, based in the development of an assisted indexing policy that shall adopt the Brazilian Thesaurus of Information Science. The results of applying that model guided the second goal which was to do an assessment on the quality of the Souza (2005) indexing corpus. The third goal wanted to achieve the compilation of reference corpus of Information Science scientific articles to assessments of automated indexing systems. The methodology that was constructed for this present work can be classified as exploratory, empiric and of case studies, when you consider the experiment carried out. To do the research was asked the use of the reference corpus composed by the 60 articles of the doctoral thesis by Souza (2005). The collaborative model of social assisted indexing of the corpus occurred through the Tainacan management platform. The Tainacan was used as an instrument to guarantee the mediation of the research, in order to construct the Assisted Folksonomy. As a way to achieve the proposed goals in the research steps to be followed were defined. The process in which those steps configured were: 1) To research and evaluate the written literature about assisted Folksonomy and the building of reference corpus to an automated indexing assessment; 2) Configure and prepare a Tainacan collection, taking into consideration roles, activities and interaction rules between users as the input of the collection of the corpus of periodicals articles by Souza (2005) Information Science with respective descriptive metadata; 3) Develop an assisted indexing policy adopting the Brazilian Thesaurus of Information Science in the process of the collaborative indexing, to be inserted in the metadata field Collection *Tags*; 4) Instruct the indexing users and guide in the process of the social assisted indexing; 5)

Annalise the results of the collaborative indexing process made by the users groups, comparing data that is found in the metadata "subjects" field; 6) Define and apply rules to do the compilation of reference corpus in the Information Science area to make assessments of the automated indexing systems. The results of analyses of the quality degree of the indexing have an average 32% of accuracy, 69% of retrieval, 28% of consistency coefficient and 41% of score-F. The average achieved, then, present some discrepancies which can be seen as a harmonization level between keywords defined by authors, when compared to users' keywords, bringing problems, among others, in the articles retrieval action. Lastly, the conclusion found was that the use of Assisted Folksonomy can useful in the improvement of the objects, processes and products build in the Information Science field, particularly in reference of the corpus compiled that allowed to achieve a proper setting facing the corrections and exclusions of terms of the applied Folksonomy.

**Keywords:** Assisted Folksonomy. Social Indexing. Reference Corpus. Evaluation of Automatic Indexing Systems. Information Science.

## LISTA DE FIGURAS

Figura 1 –	Ação de colaboração para Eventos e Usuários do Tainacan .....	42
Figura 2 –	Processo de indexação do usuário .....	43
Figura 3 –	Processo de contestação para colaboração .....	43
Figura 4 –	Cadastro de membros no Tainacan .....	75
Figura 5 –	Campos categoria da coleção .....	76
Figura 6 –	Permissões para Coleção Corpus de Referência em CI no Tainacan .....	85
Figura 7 –	Inserção de categorias para Coleção Corpus de Referência em CI no Tainacan .....	86
Figura 8 –	Itens da Coleção Corpus de Referência em CI no Tainacan .....	88
Figura 9 –	Configuração do Corpus de Referência em CI no Tainacan .....	89
Figura 10 –	Página de aprovação das ações do Corpus de Referência em CI no Tainacan .....	89
Figura 11 –	Nuvem de <i>tag</i> do Corpus .....	95

## LISTA DE QUADROS

Quadro 1 –	Fatores positivos e negativos da Indexação automática .....	48
Quadro 2 –	Critérios de indexação .....	57
Quadro 3 –	Planejamento das etapas com relação aos objetivos da pesquisa ..	67
Quadro 4 –	Etapas do desenvolvimento do protótipo .....	71
Quadro 5 –	Quadro comparativo dos requisitos funcionais para a contribuição no catálogo (para todos os membros) .....	72
Quadro 6 –	Atores do sistema .....	74
Quadro 7 –	Erros no uso de caracteres na Folksonomia Assistida gerada .....	92
Quadro 8 –	Erros no uso de letras maiúsculas e minúsculas na Folksonomia Assistida gerada .....	92
Quadro 9 –	Termos descartados da Folksonomia Assistida .....	105
Quadro 10 –	Folksonomia Assistida resultante da junção dos grupos .....	106
Quadro 11 –	Compilação da Folksonomia Assistida Revisada .....	110
Quadro 12 –	Termos representativos da Folksonomia Assistida Revisada .....	113

## LISTA DE TABELAS

Tabela 1 –	6 (seis) tags mais recorrentes na Folksonomia Assistida gerada.....	91
Tabela 2 –	Quantitativo de palavras-chave dos documentos do Corpus .....	96
Tabela 3 –	Quantitativo de tags dos documentos do Corpus .....	97
Tabela 4 –	Quantitativo de termos em comuns das palavras-chave dos autores e tags dos usuários-indexadores.....	98
Tabela 5 –	Índices médios de coeficiente de consistência, precisão, revocação e medida F entre as palavras-chave dos autores e tags dos usuários-indexadores.....	99
Tabela 6 –	Os 6 (seis) melhores casos de termos em comum entre as Palavras-chave e Folksonomia Assistida .....	101
Tabela 7 –	Os 5 (seis) piores casos de termos em comum entre as Palavras-chave e Folksonomia Assistida .....	101
Tabela 8 –	Os 6 (seis) melhores casos de índice de precisão (P) .....	102
Tabela 9 –	Os 6 (seis) melhores casos do índice de Medida F .....	103

## LISTA DE SIGLAS

ABNT	ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS
CI	CIÊNCIA DA INFORMAÇÃO
IBICT	INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA
LI	LINGUAGEM DE INDEXAÇÃO
P	PRECISÃO
R	REVOCAÇÃO
RI	RECUPERAÇÃO DA INFORMAÇÃO
TBCI	TESAURO BRASILEIRO DE CIÊNCIA DA INFORMAÇÃO

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>17</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b> .....	<b>25</b>
2.1	INDEXAÇÃO SOCIAL .....	25
<b>2.1.1</b>	<b>Folksonomia</b> .....	<b>29</b>
2.2	FOLKSONOMIA ASSISTIDA .....	32
<b>2.2.1</b>	<b>Origem e definição</b> .....	<b>34</b>
2.3	MODELOS COLABORATIVOS DE INDEXAÇÃO SOCIAL.....	35
2.4	POLÍTICA DE INDEXAÇÃO SOCIAL.....	38
2.5	TAINACAN .....	40
2.6	TESAUROS.....	43
<b>2.6.1</b>	<b>Tesouro brasileiro em ciência da informação (TBCI)</b> .....	<b>46</b>
2.7	SISTEMAS DE INDEXAÇÃO AUTOMÁTICA.....	47
<b>2.7.1</b>	<b>Avaliação de sistemas de indexação automática</b> .....	<b>53</b>
2.8	CORPUS DE REFERÊNCIA .....	59
<b>3</b>	<b>PROCEDIMENTO METODOLÓGICO</b> .....	<b>64</b>
3.1	CLASSIFICAÇÃO DA PESQUISA .....	64
3.2	INSTRUMENTO E UNIVERSO DA PESQUISA.....	66
3.3	ETAPAS DA PESQUISA .....	67
<b>4</b>	<b>ANÁLISE DOS RESULTADOS</b> .....	<b>81</b>
4.1	MODELO COLABORATIVO DE INDEXAÇÃO SOCIAL ASSISTIDA...81	
4.2	A CONFIGURAÇÃO DA COLEÇÃO .....	84
4.3	POLÍTICA DE INDEXAÇÃO SOCIAL ASSISTIDA .....	90
4.4	ANÁLISE DA FOLKSONOMIA ASSISTIDA GERADA .....	91
4.5	ANÁLISE DA FOLKSONOMIA ASSISTIDA REVISADA .....	98
4.6	COMPILAÇÃO DO CORPUS DE REFERÊNCIA.....	104
<b>5</b>	<b>CONSIDERAÇÕES FINAIS</b> .....	<b>116</b>
	<b>REFERÊNCIAS</b> .....	<b>122</b>
	<b>APÊNDICE A - POLÍTICA DE INDEXAÇÃO SOCIAL</b> .....	<b>130</b>
	<b>APÊNDICE B - MANUAL DE INDEXAÇÃO</b> .....	<b>133</b>
	<b>APÊNDICE C - TERMOS DE INDEXAÇÃO DOS AUTORES E</b> <b>USUÁRIOS-INDEXADORES</b> .....	<b>142</b>

<b>APÊNDICE D - TAGS POR FREQUÊNCIA DE UTILIZAÇÃO PARA DESCRIÇÃO DOS DOCUMENTOS DO CORPUS DE REFERÊNCIA.....</b>	<b>156</b>
<b>APÊNDICE E - ÍNDICES DE AVALIAÇÃO, PRECISÃO, REVOCAÇÃO E MEDIDA F .....</b>	<b>162</b>
<b>APÊNDICE F - FOLKSONOMIA ASSISTIDA REVISADA POR GRUPO .....</b>	<b>164</b>
<b>ANEXO A - CORPUS DE REFERÊNCIA DE SOUZA (2005).....</b>	<b>173</b>

## 1 INTRODUÇÃO

O ato de indexar, independente do indivíduo que a realiza, profissional ou não, traz consigo aspectos sociais, políticos, econômicos, entre outras, e que reforçam o valor cognitivo desempenhado no desenvolvimento de tal atividade.

Entender isso permite traçar uma linha que justifique a construção daquilo que se entende como indexação social, pois em ambientes de caráter colaborativos, principalmente na chamada Web 2.0, a indexação é realizada frequentemente através dos usuários. Esses perfazem o mesmo caminho da indexação controlada para representar um conteúdo, que vai desde identificação e seleção, a partir dos próprios recursos informacionais.

No processo de entendimento da proposta de indexação social é preciso entender como a Ciência da Informação (CI) passou a discutir questões ligadas a representação da informação. Trilhando esse caminho, a CI aqui será abordada numa perspectiva tecnológica e social, seguindo a linha apresentada por Borko (1968) e Saracevic (1995), que consideram determinantes nesse processo a explosão informacional e a implosão do tempo de alcance de toda informação produzida num período pós-guerra (LE COADIC, 2004).

Diante dessa perspectiva tecnológica e social, é possível então ver-se diante de uma visão positivista, que nos remete ao senso comum, seguido do empirismo, para que por fim possamos chegar ao pensamento racional. Pinheiro (1999, p. 156) sintetiza a CI como sendo interdisciplinar, de natureza social que está “relacionada à tecnologia da informação e do novo papel da informação na sociedade e na cultura contemporânea”.

A Representação da Informação (RI), sendo aqui apresentados na figura da Folksonomia, segundo alguns autores, é gerada pela expressão cognitiva humana. Para Lima *apud* Jerry Fordor (2003, p. 80), os sistemas cognitivos envolvem representações, e as operações cognitivas envolvem a manipulação de representações simbólicas que partem da – percepção, raciocínio, aprendizagem da língua e de valores semelhantes – que os seres humanos são capazes de fazer.

Saracevic (1995) diz que a RI, enquanto área de pesquisa é primordial no processo de disseminação e criação do conhecimento e coopera para o progresso científico no campo da CI. Para isso, pode-se pensar no próprio conhecimento como sendo fruto da junção de um conhecido prévio adquirido e um conhecimento

extraído de uma informação, e que mediante os paradigmas em questão, é possível entender que o sujeito que se encontra no meio de todo esse cenário não pode ser notado isoladamente.

Segundo Santos (2016, p.17) os modelos colaborativos de indexação social podem ser pensados como atividades colaborativas de indexação que serão efetuadas pelos usuários e que, como principal resultado, serve para desenvolvimento dos próprios sistemas colaborativos de indexação, dando a flexibilidade de ser adaptados ou aperfeiçoados em diversos contextos.

Os modelos colaborativos de indexação trazem consigo algumas problemáticas, e por conta disso é preciso ter clareza de quais elementos devem constar na sua concepção. Dessa forma, Kim et al. (2011) buscam definir que um modelo colaborativo de indexação que utiliza a Folksonomia para a descrição dos recursos deve ser composto por um conjunto de elementos, são eles:

- Objeto: recurso a ser etiquetado (livro, artigo, imagem, áudio, por exemplo)
- Tag(s): etiqueta(s) a ser(em) atribuída(s) ao recurso;
- Etiquetador: o usuário (uma pessoa) que estabelece o relacionamento entre o objeto e a etiqueta;
- Fonte: o ambiente onde a ação de atribuição de etiquetas será realizada (base de dados, catálogo, repositório, rede social, por exemplo).
- Polaridade: um voto contra ou a favor da *tag* elencada pelo usuário - com o objetivo de solucionar possíveis problemas de RI.

Santos (2016, p.71) afirma que os modelos esboçam um nível de interação e comunicação entre usuários no que se refere a cumprir atividades que compõem o processo de indexação social. Para tanto, essa perspectiva evidencia alguns problemas semânticos, principalmente pelo processo de interação dos usuários e que precisam ser analisados para encontrar meios de contorná-los.

Os modelos que utilizam a Folksonomia para descrição de recursos, por exemplo, apresentam problemas que como descritos por Silva (2014, p.30-31) estão diretamente ligados à polissemia, quando um significante tem vários significados, e sinonímia, quando o mesmo significado pode ser representado por vários significantes.

No intuito de amenizar esses possíveis efeitos negativos, que são em sua grande parte, decorrentes da produção de grandes volumes de informações, o

tratamento físico e de conteúdo dos documentos passam a deter um papel fundamental para a recuperação da informação.

Além disso, a busca por alternativas que preservem o efeito positivo da Folksonomia, eliminando os aspectos negativos, deu origem a um modelo híbrido de representação, a Folksonomia Assistida. Sua premissa básica gira em torno da possibilidade de garantir uma liberdade controlada para usuários que visam definir *tags* a conteúdos informacionais, principalmente de caráter científico.

Tal liberdade controlada tem como proposta garantir que o usuário mantenha seu direito à escolha de como deseja etiquetar o conteúdo. Isso se assemelha ao que a Folksonomia propõe, mas com a diferença de que uma pré-estrutura hierárquica será a base de escolha, tornando possível garantir qualidade sobre a indexação produzida.

Para a pesquisa que se segue, é preciso então levar em consideração diversos aspectos que visem dar condições de tal modelo ser validado, principalmente aqueles que se relacionem a representação da informação. Sendo assim, alguns conceitos precisam ser abordados para garantir uma fundamentação adequada.

Com isso, questões ligadas a Indexação Automática, bem como se apresenta ao campo da CI, e que desempenha papel fundamental para a RI, precisa ser amplamente discutido, mesmo que ainda não caminhe diretamente com o que se propõe a Folksonomia Assistida.

Robredo (2005), afirma que a indexação automática é “qualquer procedimento que permita identificar e selecionar os termos que representam o conteúdo dos documentos, sem a intervenção direta do indexador”.

Acrescenta-se a isso as palavras de Lapa (2014, p.97-98), que afirma o princípio geral da indexação automática sendo baseado na comparação de cada palavra do texto com uma relação de palavras vazias de significado pelo computador, previamente estabelecidas, que conduz, por eliminação, a considerar as palavras restantes do texto como palavras significativas.

Contudo, ainda sim, segundo Robredo (2005) o processo de indexação automática guarda semelhanças com a forma que se realiza a indexação manual, quem em muito, é similar ao processo de leitura e memorização humano. Isso ocorre uma vez que está baseado na comparação de cada palavra do texto com uma lista de palavras vazias de significado.

O importante diante do apresentado é destacar o valor da Indexação na garantia de uma boa recuperação do conteúdo. Com o atual cenário, onde a quantidade de informação produzida passa a ser considerável, avaliar o resultado dessa indexação passa a ser tão importante quanto à própria indexação.

Ao atuar na perspectiva de garantir qualidade na indexação, muito precisa ser levado em consideração. Atualmente, alguns autores apresentam formas de avaliar a indexação automática, e que para alguns espaços não são devidamente aplicados.

Por exemplo, a indexação é importante na recuperação de artigos científicos de diversas bases de dados. Contudo, algumas bases existentes fazem uso apenas das palavras-chaves designadas pelos próprios autores, ou em outros poucos casos, realizam a indexação por equipe de indexadores.

Lancaster (2004) reforça a necessidade de trazer à tona o que se conhece como avaliação. Segundo o mesmo é o único caminho que leva à melhoria dos serviços e produtos oferecidos aos usuários e à detecção dos pontos positivos e negativos da usabilidade de um catálogo em linha ou em fichas.

Sendo assim, é válido destacar que o importante dentro desse processo de busca numa base acontece quando a recuperação traz o maior número de documentos úteis com o menor número de ruídos. E que para isso efetivamente vir a acontecer, é preciso trabalhar na perspectiva de avaliar os sistemas de indexação automática.

A garantia de um bom entendimento acerca das temáticas apresentadas, Indexação automática e avaliação dos sistemas de indexação automática, servem como base para dar condições a um bom desenvolvimento da aplicação da Folksonomia Assistida na construção de corpus de referência de artigos científicos em CI.

Vale reforçar que a própria Folksonomia Assistida, enquanto resultado, é um modelo híbrido de conceitos existentes. Ao ser aplicado, atua com uma alternativa para questões essenciais dentro do processo de representação e consequentemente recuperação da informação, principalmente em ambientes digitais e, principalmente, pode vir a ser uma alternativa para a avaliação de sistemas de indexação.

A proposta da pesquisa aqui desenvolvida perpassa pela aplicação da folksonomia assistida na construção de corpus de referência de artigos científicos da área de Ciência da Informação escritos em português para avaliação de sistemas de

indexação automática. O corpus foi escolhido levando em consideração o uso por pesquisas anteriores na temática indexação automática para assim determinar sua relevância. A perspectiva é aperfeiçoar a indexação feita pelos autores nos respectivos documentos via palavras-chave, através da indexação social de leitores baseada na atribuição de termos de um tesouro da área.

A busca por uma relevância da indexação social neste contexto de compilação de um *corpus de referência* se dará diante da relação a ser feita entre o que é utilizado pelo “indexador oficial”. Ele é o próprio autor no ato de escolha das palavras que representarão seu documento, e que é entendido pelo “indexador-leitor”.

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 a que o indexador-leitor representou e até que ponto o tesouro, enquanto base hierárquica, é capaz de alcançar a representação desejada por esse mesmo indexador-leitor?.

Nesse ponto é importante compreender que a folksonomia assistida consiste no resultado do processo de indexação social assistida, apoiada pelo usuário. Isso ocorre haja vista que no processo ele é auxiliado no momento de definição das *tags* mais adequadas para a representação dos objetos informacionais a partir dos descritores atribuídos pelo próprio usuário ou disponibilizados no sistema (SANTOS, 2016, p.75).

Santos (2016, p.75) ainda afirma que no momento em que o usuário informa uma palavra-chave é preciso existir uma intervenção do sistema, dando a eles sugestões de *tags* similares presentes em uma estrutura de representação do conhecimento, que pode ser um tesouro ou ontologia por exemplo.

O **problema de pesquisa** será guiado pela 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 geral** analisar a aplicação da folksonomia assistida, resultante do processo de indexação social assistida, na construção de corpus de referência na área de Ciência da Informação para

avaliação de sistemas de indexação automática. Para isso, os **objetivos Específicos** são:

- a) Construção da folksonomia assistida - adaptar um modelo colaborativo de indexação social assistida, configurar um software para dar suporte a indexação social assistida contemplando papéis, atividades e regras de interação entre usuários previstas no modelo colaborativo;
- b) desenvolver uma política de indexação social assistida que adota um tesauro da área de Ciência da Informação, e realizar experimento de indexação social assistida de artigos científicos do corpus de Souza (2005) por um grupo de indexadores in locus na plataforma Tainacan;
- c) Avaliar a qualidade da indexação do corpus - analisar a folksonomia assistida resultante do processo de indexação social assistida quanto à qualidade na indexação, levando em conta as palavras-chaves dos autores e ocorrência dos termos em tesauro da área de Ciência da Informação e no próprio texto do documento;
- d) Compilar o corpus de referência de artigos científicos na área de Ciência da Informação para avaliação de sistemas de indexação automática.

A justificativa para a realização da pesquisa surge a partir da identificação de que, na área de Ciência da Informação, os estudos relacionados ao assunto folksonomia assistida se mostram, na sua grande maioria, apenas no campo da construção conceitual. Assim, vem a precisar ser devidamente testada em contextos práticos, como por instância, na indexação de artigos científicos.

A importância das indagações em torno do assunto decorre inicialmente da folksonomia se apresentar como tendência para repositórios institucionais em ambientes digitais. Sendo assim, é notório destacar as poucas publicações referentes à temática folksonomia assistida e, conseqüentemente, a real necessidade de se aprofundar nas pesquisas sobre o tema. Como principais trabalhos identificados na área estão os de Santarem Segundo (2010), Silva (2013) e Santos (2015).

A concepção de uma folksonomia assistida visando a posterior análise e compilação de um corpus de referência, apresenta uma nova dimensão para o que até então vem sendo discutido sobre a temática, podendo vir a efetivar, ou não, questões abordadas nos respectivos trabalhos. A aplicação dos conceitos desses

autores visa averiguar de forma empírica tendências, e ser capaz de fazer projeções futuras para pesquisas em torno da RI em ambientes digitais.

É importante mencionar que as aplicações desse conceito trazem consigo, enquanto valor científico, a possibilidade de surgimento de uma nova perspectiva para o campo da RI em bases de dados científicas. Segundo Sayão (1996) as bases de dados científicas se constituem como a metáfora da memória da ciência que se pratica nos dias atuais, objetivando agregar as produções intelectuais dos pesquisadores. Dessa forma, é válido destacar que a construção dessa memória, principalmente em ambientes digitais, depende de estudos como esses para auxiliar, especialmente na RI.

Ao longo dos capítulos, essa dissertação é estruturada para apresentar no segundo capítulo questões ligadas a *Indexação Social*, com o intuito de entender como funcionam dentro do contexto atual frente à sociedade, assim como a *Folksonomia* e seus prós e contras. Ainda no segundo capítulo, reforçasse que serão apresentados alguns modelos colaborativos de indexação sociais já existentes, dentre eles a proposta de *Folksonomia Assistida*, abordada por Santarem Segundo (2010), Silva (2013) e Santos (2015). Por fim, no capítulo ainda são abordados como subcapítulos, *Sistemas de Indexação automática*, *Avaliação de Sistemas de Indexação automática* e *Corpus de Referência*. O intuito é dar fundamentação e entender de que forma a Folksonomia assistida pode surgir como alternativa para avaliar sistemas de indexação automática, e nesse caso em específico a partir da construção do Corpus de Referência em CI.

O terceiro capítulo é responsável por apresentar a metodologia aplicada para alcançar os objetivos traçados na pesquisa. O mesmo se encontra subdividido em *Classificação da pesquisa*, *Instrumento e universo da pesquisa* e *Etapas da pesquisa*, sendo nesse último apresentado detalhadamente através de procedimentos.

Para o quarto capítulo foi decidido que a análise dos resultados será apresentada seguindo os tópicos, *O Modelo Colaborativo de Indexação Social Assistida*, no qual será apresentada a execução do modelo proposto, bem como características identificadas na sua construção e aplicação. *A Política de Indexação Social Assistida* e configuração da coleção aplicada ao grupo de usuários e como se deu o desenvolvimento das atividades propostas. Por fim, será possível visualizar a *Folksonomia assistida* e *Compilação do corpus de referência* gerada a partir da

aplicação da metodologia, bem como as inferências de acordo com o que foi abordado no referencial teórico.

## 2 REFERENCIAL TEÓRICO

O presente capítulo tem como objetivo discorrer acerca dos aspectos teóricos necessários para entendimento dos estudos e pesquisas relacionados a esta dissertação.

Serão abordados aspectos ligados a indexação social e a folksonomia, bem como o surgimento da folksonomia assistida a partir da visão de Santarem Segundo (2010), Silva (2013) e Santos (2015), chamando atenção à busca pela qualidade no resultado da indexação a partir da aplicação dos conceitos por eles apresentados. Além disso, neste capítulo serão apresentados modelos colaborativos de indexação cujas regras possam servir para identificar também melhores formas de indexação das produções científicas. Para por fim tratar do contexto de aplicação da folksonomia assistida neste trabalho, que é a avaliação de sistemas de indexação automática.

### 2.1 INDEXAÇÃO SOCIAL

Ao tratarmos de indexação social ou indexação colaborativa, o primeiro ponto a ser tocado diz respeito a algo que ocorre com frequência dentro da Ciência da Informação, o fato de falar de um mesmo objeto, em contextos distintos. Segundo Alvarenga (2006, p.97) isso se dá principalmente em decorrência da mudança dos meios, bem como a sofisticação dos instrumentos, fazendo com que surjam novos nomes para designar coisas novas e velhas, mas que em sua essência, permanece a mesma.

Partindo dessa premissa é notável, e necessário, relatar a quantidade de termos utilizados para descrever a indexação orientada pelo usuário, surgida a partir do advento da Web 2.0. O aumento significativo de espaços no qual os usuários passaram a interagir e contribuir diretamente na construção de conteúdos fez com que surgissem com elas termos que representavam a ação de indexação por meio desses usuários.

Dentre essas ações, podem-se observar as chamadas *Tags* (etiquetas), no qual os usuários definem rótulos para os conteúdos por eles disponibilizados. O intuito dessa rotulagem, é representar a partir das experiências, sejam elas políticas, econômicas, culturais ou sociais.

Na literatura, dentre as referências que se encontram para o ato de etiquetar, Tennis (2006) define a *Social Tagging* como sendo um tipo de indexação aberta que se manifesta na Web. Além disso, reforça que o melhor a ser feita para atribuição de etiquetas é olhar para os propósitos dos sistemas no qual está inserido, pois crescem em torno de comunidades que desejam compartilhar interesses em comum.

Destaca-se outras formas de designar a ação de indexar nesses espaços, algumas delas são: etiquetagem colaborativa, classificação social, indexação social e folksonomia (VOSS, 2007, p. 2).

Diante dos termos apresentados, a indexação social se apresenta como aquele que se enquadra mais efetivamente diante da dinâmica nesses espaços. Além de se apresentar melhor situado dentro do que se propõem os estudos sobre tratamento da informação na grande área da Biblioteconomia e Ciência da Informação.

Como mencionado, o caráter dinâmico dos ambientes Web traz consigo um cenário onde os tradicionais modelos de tratamento da informação são colocados a prova por uma série de fatores. O aumento exponencial, bem como a diminuição do tempo na produção da informação, torna complexa a maneira de quantificar e qualificar o que é veiculado nesses ambientes. Por essa razão tais fatores se apresentam como principais questões a serem lidadas dentro do tratamento informacional.

A importância desses pontos levantados pode ser observada pela maneira com que se é feito o gerenciamento da informação produzida nesses ambientes, que por muitas vezes é por meio de ferramentas ou softwares específicos. Diante disso, se torna necessário a existência de estudos que consigam atender ao novo fluxo demandado.

Na perspectiva de conseguir entender a indexação social, é preciso ter em mente a função e o que vem a ser a indexação. Para isso, Lancaster (2004) traz uma definição para a indexação como ela sendo um conjunto de procedimentos que objetivam expressar ou representar o conteúdo temático de documentos através de linguagens de indexação ou documentárias no intuito de recuperar posteriormente.

Tennis (2006, p.5) apresenta a ideia de que no processo de indexação se usa estruturas de representação, já na *social tagging* de assunto usam diferentes estruturas, permitindo assim ver diferenças entre um e outro

Lancaster (2004) ainda reforça que os termos ao serem atribuídos aos documentos, durante o processo de indexação, visam servir como ponto de acesso mediante os quais um item pode vir a ser localizado e recuperado.

No processo de indexação, é preciso a adoção de uma Linguagem de Indexação - LI, que segundo Guinchat e Menou (1994) são instrumentos convencionais de uso das unidades de informação para a descrição dos conteúdos temáticos dos documentos.

Rowley (2002, p.169) afirma que a LI pode ser classificada como linguagens controladas, linguagens naturais e linguagens livres.

- **Linguagens controladas de indexação** – definidas como um conjunto de termos autorizados para uso na indexação do assunto de documentos. É subdividida em dois tipos: as linguagens alfabéticas de indexação, como os tesouros e listas de cabeçalhos de assunto; e os sistemas de classificação, representado por código ou notação.
- **Linguagens naturais de indexação** – referem-se a quaisquer expressões que ocorram em alguma parte do documento. Todos os termos no corpo do documento são candidatos a serem termos de indexação.
- **Linguagens livres de indexação** – para esta linguagem não existem limitações quanto aos termos a serem utilizados no processo de indexação.

No que se refere às classificações apresentadas, a terceira se assemelha ao que ocorre nos espaços Web, e é realizada por um indexador humano. A qualidade dependerá única e exclusivamente do conhecimento daquele que vem a indexar sobre o(s) assunto(s) tratado no documento.

Hjørland (2001) apresenta uma nova dimensão, quando afirma que a indexação deve ser moldada para se ajustar às necessidades de determinado grupo, indo além da escolha de linguagem de indexação adequada.

Segundo Rafferty e Hilderley (2007, p. 399) existem três grupos de candidatos a atores dentro do processo de indexação, são eles:

- **Indexação orientada por especialistas** – baseia-se no tratamento da informação através da intervenção de intermediários (bibliotecários, indexadores, editores voluntários), é a indexação feita por especialistas sendo dispendiosa e cara.

- **Indexação orientada pelo autor** – esta abordagem pressupõe que o autor irá utilizar termos que são comumente compreendidos e geralmente aceitos. Um problema que essa abordagem enfrenta é o fato do autor não ser necessariamente um gestor de informação com os conhecimentos profissionais de um especialista.
- **Indexação orientada pelo usuário** – esse tipo de indexação possibilita um elevado nível de interação com a comunidade que, provavelmente, não seria possível se tivesse decisões a serem feitas sobre códigos, convenções e regras que regem qualquer taxonomia controlada.

Dentro dessas questões apresentadas anteriormente, o surgimento e evolução da *Web* veio com muitas transformações, indo bem além das questões tecnológicas, mas também agregando mudanças na posição da produção de conteúdo centradas no usuário. Quando *Web 1.0*, ainda uma rede explicitamente corporativa e comercial, ela passa a ser conhecida pelo seu poder de interação em sua segunda geração.

A nível de entendimento, a *Web 1.0* ou simplesmente *Web* foi criada pelo físico Tim Berners Lee, que deu início ao uso dos sistemas de hipertextos (HTML), sendo essa uma linguagem para representar o conteúdo em termos de páginas na *Web* e de expressá-las em links. Outra das criações de Tim Berners Lee é o Protocolo de transferência de Hipertextos (HTTP), sendo esse recurso o responsável por permitir que o usuário navegue na rede de forma não estática. Segundo Primo (2007, p.2) “na primeira geração da *Web* os sites eram trabalhados como unidades isoladas”.

Segundo Blattmann e Silva (2007, p. 197):

Esse cenário possibilitou um nível de interação onde as pessoas poderiam colaborar principalmente para a qualidade do conteúdo disponível, produzido e classificando, além de reformular tudo aquilo que já está disponível até então.

Como contrapontos surgiram questionamentos a que forma toda essa informação produzida passaria a ser tratada e, conseqüentemente organizado, mais ainda por se tratar de um novo contexto, o digital. Para isso algumas ferramentas começaram a surgir e ganhar forma, e entre as principais ações dessas ferramentas é possível visualizar a atribuição de etiquetas (*Tags*) por usuários que, como resultado, permite gerar “índices” de descritores para que posteriormente a recuperação dos itens etiquetados seja realizada

Esse cenário possibilitou que usuários dos mais variados lugares acessassem e manipulassem o mesmo documento em tempo real. Tal característica reforça o potencial descentralizador que a organização da informação passa a deter.

Brown *et al.* (1996) destaca a importância de se ter conversações entorno de uma indexação democrática, onde os usuários passem a auxiliar nos registros de termos de acordo com seu interesse.

Reforçando esse movimento, Lancaster (2004), traz que para certos tipos de materiais, a indexação orientada pelo usuário pode até ser mais importante do que o é no caso de artigos de periódicos, livros, ou relatórios técnicos.

Agregando uma nova conotação, a indexação social é definida por Hassan-Montero (2006, online) como sendo:

[...] um novo modelo de indexação, em que são os próprios usuários ou consumidores dos recursos os que levam a cabo sua descrição [...] A descrição de cada recurso se obteria por agregação, ou seja, um mesmo recurso seria indexado por inúmeros usuários, dando como resultado uma descrição intersubjetiva e portanto mais fiel que a realizada pelo autor do recurso (Tradução do autor).

No centro deste novo conceito, é possível colocar as folksonomias, que podem ser definidas como sistemas orgânicos baseados na atribuição livre e pessoal de etiquetas às informações ou objetos visando à organização e recuperação (WAL, 2006).

É importante salientar que a indexação social não se apresenta como um contraponto às outras formas de indexação (indexação realizada por profissionais, indexação automática, etc.), mas sim como uma alternativa de organização, adaptada a necessidade de tratar conteúdos nos ambientes digitais.

### **2.1.1 Folksonomia**

A organização e a interface entre o homem/informação, assunto esse diretamente ligado à Folksonomia, vem sendo debatido desde que Thomas Vander Wal criou o termo *folksonomy* para definir um conjunto de categorias que emerge da maneira de como as pessoas rotulam as informações por eles disponibilizadas.

Folksonomia é o resultado da atribuição livre e pessoal de etiquetas às informações dos recursos na Web, em um ambiente social, compartilhado e aberto a outros, pelos próprios usuários da informação, visando a sua recuperação. Destacam-se portanto três fatores essenciais: o primeiro é o resultado de uma indexação livre, feita pelo próprio usuário do recurso; o segundo objetiva a recuperação posterior da informação e o terceiro é desenvolvida num ambiente aberto que possibilita o compartilhamento e,

até, em alguns casos, a sua construção conjunta. (CATARINO; BAPTISTA, 2007).

Ao apresentar o conceito de folksonomia, Catarino e Baptista trazem consigo pontos ligados a usabilidade da sua aplicação frente à sociedade. A grande quantidade de informação gerada a partir da liberdade dada ao usuário possibilita descobertas frente à utilização do vocabulário próprio destes. Passando assim pela interpretação dos conteúdos disponibilizados, sendo a soma das ações ou das características de cada indivíduo.

Silva (2007, p. 10) afirma que “a Web ganhou caráter social, uma rede formada de pessoas participativas e influentes no processo de geração, distribuição e organização da informação”. A partir do momento em que a criação de palavras-chaves ou “*tags*” (etiquetas) passa a ser feita, os usuários começam a se ajudar, isso acontece devido o agrupamento de termos parecidos. A classificação por meio de etiquetas torna possível a organização dos recursos da web, criando uma espécie de *etnoclassificação*, isto é, uma classificação popular, capaz de causar conflitos por diversos fatores, como também apresentar diversas surpresas.

Com um grande número de pessoas diagnosticando, classificando, se torna inevitável a criação de novas definições para os mais diversos campos. Nessa perspectiva, é possível afirmar que as etiquetas são espaços de sociabilização onde as pessoas compartilham informação podendo formar com facilidade grupos em torno de interesses na medida em que, ao utilizar funcionalidades da Folksonomia, os usuários dos ambientes digitais têm acesso aos outros participantes que possuem os mesmos interesses identificados através das *tags*, formando as redes sociais.

Diante de algumas críticas existentes sobre a Folksonomia, muitas relacionadas com a falta de controle e organização que o grande número de informação gerada causa, Van Amstel (2007, p. 19) explica que na Folksonomia, o controle existe, contudo é descentralizado e espontâneo. Trata-se de movimentos reguladores da cultura no sentido horizontal, que se espraiam pelos laços da rede social do indivíduo. Vale ressaltar que esses ambientes permitem pessoas colaborarem para o enriquecimento do seu conteúdo por meio da participação coletiva.

Catarino e Baptista (2007) afirmam que aparenta existir um consenso de que a característica de liberdade na etiquetagem dos conteúdos acarreta aquilo que é o seu maior problema, a falta de controle do vocabulário.

Santos (2016, p. 33) discorre sobre a Folksonomia como sendo uma antítese ao termo taxonomia, tendo como principal característica a criação de *tags* com função de descritor a partir da linguagem dos usuários. Nesse ponto é importante destacar a contraposição entre fatores positivos e negativos de ambas as perspectivas, a primeira e mais claras dizem respeito à linguagem natural que resulta na folksonomia, e a linguagem controlada da taxonomia.

A folksonomia facilita a construção de um mapa dos desejos dos usuários e de seu universo semântico, e junto a isso traz consigo uma série de questões que causam reflexão, e que variam entre ambigüidades, inconsistência, polissemia, inflexões, sinônimas e, principalmente, erros de ortografia.

Dentre as vantagens observadas, Lévy (2000) *apud* Blattmann e Silva (2007, p. 191), é possível observar que [...] a existência de uma Internet colaborativa possibilita a disseminação da inteligência e memória coletiva. Além disso, seu pensamento nos conduz à reflexão de que a Internet é um canal pelo qual flui uma grande quantidade de práticas sociais, culturais, políticas e econômicas. Trata-se de um espaço interativo, de trocas, de criação e geração, além de armazenamento de informações, tornando-se uma importante ferramenta de colaboração entre os participantes do mundo digital on-line [...]

Moura (2009) discorre que as folksonomias contribuíram enormemente para a popularização de novas perspectivas de classificação de documentos digitais e ampliaram as possibilidades de compartilhamento de novas significações para termos e conceitos socialmente estabelecidos e debatidos em ambientes virtuais.

Dentre os tantos conceitos apresentados sobre o que vem a ser folksonomia, Santos apresenta um perspectiva sobre a falta de consenso entre o que vem a ser de fato a folksonomia. Sobre a análise de vários conceitos encontrados na literatura, a mesma relata que:

não há um consenso conceitual e terminológico acerca da Folksonomia tendo em vista que em alguns momentos ela é considerada pelos autores como um fenômeno; um sistema; uma classificação; um vocabulário; um método ou até mesmo um resultado de um processo (SANTOS, 2016, p. 35).

A definição do conceito a ser dado para a folksonomia se torna determinante para o direcionamento da pesquisa, e assim sendo, bem como Santos (2006)

trabalhou em sua pesquisa, o presente estudo entende que a folksonomia é resultado do processo de etiquetagem livre (atribuição de etiquetas, palavras-chave) realizada pelos usuários mediante o emprego de termos provenientes de linguagem natural - dispensando o auxílio de vocabulários controlados. Ocorrem em ambientes digitais colaborativos visando indexar recursos informacionais compartilhados em qualquer formato (textos, imagens, áudio, vídeo etc.) para fins de sua representação.

Desta forma, é importante destacar que a maneira no qual o conceito de folksonomia é empregado dentro das produções científicas está intimamente ligada a abordagem para qual ela estará sendo aplicada, tendo em vista a apresentação das várias discussões para defini-la pertinente as suas principais características. Levando a conclusão que a distinção básica entre o que vem a ser etiquetagem e o que é de fato Folksonomia está na que uma parte do processo de atribuir palavras-chaves e a outra ao resultado dessa atribuição respectivamente.

## 2.2 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 seja levantada.

Nascimento (2008, p.42) enfatiza que:

a folksonomia emergiu no cenário do software social, permitindo que as pessoas compartilhem entre si não apenas os objetos digitais, mas também seus pensamentos, reflexões, críticas, e suas formas de indexar esses objetos.

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 de colaboração.

Nessa perspectiva, surge então na literatura uma abordagem híbrida no intuito de sanar alguns desses pontos. Santarém Segundo (2010); Santarém Segundo e Vidotti (2011), por exemplo, em seus trabalhos 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.

Folksonomia Assistida consiste em um processo de apoio ao usuário, que atua no momento de definir os termos mais adequados para as *tags* que representarão o conteúdo depositado em um repositório digital. (SANTAREM 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 (SANTAREM SEGUNDO; SIQUEIRA, 2013).

Assim, no momento em que o usuário vai informar a palavra-chave, Santarém Segundo e Vidotti (2011) descreve acontece uma intervenção do sistema apresentando sugestões de *tags* similares presentes em uma estrutura de representação do conhecimento, que pode ser um tesouro ou ontologia por exemplo. É importante que para esse primeiro momento exista uma alteração na interface de comunicação do usuário com o repositório. O segundo passo, que é de caráter interno, o repositório ao receber os termos elencados pelo usuário, deve seguir com a pesquisa de relacionamento da informação dada pelo usuário em relação ao conjunto de informações internas que o ambiente já tem em posse.

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 seu recurso. Resumindo, é uma espécie de “Folksonomia Controlada” que auxilia o usuário no “tagueamento” do assunto do seu objeto informacional.

Santarem Segundo (2010, p. 176) afirma que:

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.

Como resultado dessa intervenção, alguns dos principais problemas encontrados no ato da folksonomia passam a ser evitados, tais como a polissemia, erros de grafia, entre outros. Santos (2016, p.76) reforça que o modelo colaborativo em questão também permite a criação de um relacionamento horizontal entre o conjunto de etiquetas que representam o mesmo recurso, ou seja, entre os termos atribuídos pelo usuário autor para descrever o seu documento.

Santos (2016) ainda reforça que o modelo 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.

### 2.2.1 Origem e definição

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. Reforçando isso, Barros (1994) afirma que colaborar significa empenhar-se em conjunto na intenção de agregar algo.

Com essa ideia de colaboração, e o processo de desenvolvimento tecnológico alcançando os ambientes digitais, muito se questiona de que forma é possível alcançar melhores formas de representar o conteúdo gerado nesses espaços. Surge então a folksonomia como resultado da indexação dos usuários em ambientes colaborativos por meio da etiquetagem, e que em sua essência estabelece diversos apontamentos por trazer consigo diversos fatores considerados negativos.

Pensando disso, é possível observar que alguns autores, preocupados com tais questões, sugerem o surgimento de uma forma híbrida de representação, utilizando o conceito de folksonomia aplicado juntamente a taxonomia. A ideia inicial visa possibilitar a aplicação em repositórios digitais de uma recuperação semântica da informação.

Santarém Segundo (2010, p.209), um dos primeiros autores nacionais a abordar a temática, afirma que o surgimento da folksonomia assistida "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 dentro de uma hierarquia pré-existente.

Reforçando esse cenário, Santos (2016, p.75) traz a Folksonomia Assistida como resultado do apoio ao usuário, onde existe o auxílio ao usuário na definição das *tags* mais pertinentes para representação dos conteúdos informacionais. Para tanto, é importante refletir sobre a possibilidade do surgimento de um relacionamento horizontal entre as etiquetas que representam o mesmo conteúdo, ou seja, padronizam os termos atribuídos tanto pelo usuário, quanto pelo autor para descrever o seu documento quando se tratar de repositórios digitais.

Diante de tudo isso, em sua origem a Folksonomia Assistida busca melhorar a estrutura de RI em ambientes colaborativos de ambientes digitais, abandonando a possibilidade de excesso de termos isolados, para a idéia de dar um maior significado em relação ao conteúdo indexado. Para isso, como consequência dessa característica, o usuário do sistema passa a evitar abreviações, problemas com plurais e singulares, ou ainda palavras que possam dificultar a recuperação da informação, posteriormente (SANTARÉM SEGUNDO, 2010, p.181).

Como proposta de funcionamento, o sistema que se vale da Folksonomia assistida trabalha de forma que ele permita 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 tesauros.

### 2.3 MODELOS COLABORATIVOS DE INDEXAÇÃO SOCIAL

Como primeiro ponto a ser abordado acerca dos modelos colaborativos de indexação, é importante reforçar que o papel de descrição dos conteúdos informacionais, não é uma atividade exclusiva dos seus gestores, mas principalmente dos usuários, sendo realizada através da etiquetagem.

Para tanto, os modelos colaborativos de indexação social são aplicados com o intuito de dar maior autonomia no que se refere a organização da informação por parte dos usuários dentro dos ambientes digitais. Santos (2015, p. 280) acrescenta

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

Santarém Segundo (2010) apresenta uma proposta do modelo de representação iterativa que permitiria gerar uma “folksonomia assistida”, e busca estabelecer um processo de recuperação semântica da informação dentro dos repositórios digitais. Para isso, o autor ainda apresenta questões referentes aos recursos da *Web 2.0* e *Web Semântica*, dando notoriedade à possibilidade de coexistência da folksonomia e dos vocabulários controlados para a retroalimentação e a descrição da informação.

O surgimento desse modelo tem como premissa inicial agregar funcionalidades aos repositórios digitais. Também aos 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*, tentando assim, manter uma ligação semântica.

Para o devido funcionamento do modelo em questão, a inserção das *tags*, ou descritores, são determinadas tanto pelas *tags* já inseridas no repositório, como de termos provenientes de um instrumento de controle terminológico associado a esse sistema, podendo ser um Tesauro.

Um segundo modelo colaborativo é apresentado por Silva (2013), e tem como objetivo, a busca de registros em um catálogo web facetado, chamado por ele de Facetlog, com a proposta de oferecer uma estratégia complementar à atribuição de etiquetas através de uma seleção de termos provenientes de uma taxonomia facetada.

Nesse modelo, a ideia é permitir que o usuário possa vir a 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 vem a ser um conjunto de facetas com conceitos distribuídos de forma hierárquica.

Segundo Silva (2013), o modelo web facetado apresenta a necessidade específica de o gerenciador 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.

Ainda segundo Silva (2013) é importante ressaltar a viabilidade tanto da taxonomia facetada quanto da navegação facetada como complemento à indexação, e para a ação de recuperação da informação. Isso vai impactar diretamente 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.

Para Santos (2016), antes de descrever e apresentar os modelos colaborativos de indexação identificados na literatura da área de Ciência da Informação torna-se necessário contextualizar e apresentar fundamentos teóricos. Tudo isso para ter clareza que o papel da colaboração vai de encontro ao um esforço conjunto para a realização de tarefas, e isso antes mesmo de se pensar em um modelo prático que viabilize essa conjuntura.

O avanço tecnológico propiciou o surgimento de espaços que facilitaram a concretização de sistemas colaborativos, favorecendo a execução de ações em grupo. Santos (2016, p.68) destacam que os modelos colaborativos, por sua vez, estão agregados ao uso dos sistemas colaborativos - tendo em vista que antes da utilização de um sistema com essas características é necessária a seleção e adesão de um modelo de colaboração.

Dentro desse panorama, a autora utiliza-se do conceito de modelo apresentado por Priberan (2015), onde é afirmado que modelo seriam uma imagem, desenho ou objeto que serve para ser imitado; [...] que serve ou pode servir de modelo ou de exemplo. Para Sayão (2001, p.83) modelos são “[...] representações simplificadas e inteligíveis do mundo, que permitem vislumbrar características essenciais de um domínio ou campo de estudo”.

Santos (2016, p.68) faz um esforço para alinhar os conceitos de colaboração e modelos destacando as palavras de Gerosa (2006) que afirma que os modelos colaborativos podem ser voltados para um tipo específico de trabalho em grupo ou para um elemento específico de colaboração presente em um sistema.

Diante das questões então levantadas, e entendendo que o uso das Folksonomias é bem aplicado ao contexto no qual se apresentou, é notória a possibilidade de uso delas também para a classificação e a organização de conteúdos em ambientes digitais, tais como repositórios. Para isso, o surgimento de

modelos colaborativos de indexação social que se valem exatamente da Folksonomia, com adequações, se tornam uma alternativa para o alcance de uma maior qualidade semântica na descrição dos recursos gerados em ambiente digital.

## 2.4 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. Maximiano (2000) enfatiza que as políticas têm por objetivo nortear as ações dos indivíduos/grupos em longo período no que se trata dos cenários que apresentam atividades repetitivas ou permanentes por meio da padronização e delimitação dos comportamentos em um determinado contexto.

A política de indexação pode ser definida como:

uma decisão gerencial tradicional, que divide as atividades informativas em gerenciais, de processamento e de interface como o usuário, passando a ver a gestão da unidade ou serviço de informação como um todo, no qual se inserem atividades de organização, disponibilização e uso da informação (CESARINO; PINTO, 1978)

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 informações.

A afirmação feita pelo autor se torna importante quando se pensa no contexto a qual a política será aplicada. Atualmente, os ambientes digitais, no qual existe a participação dos indivíduos, têm frequentemente lidado com problemas dos mais variados referentes à escrita e até contextualização do que vem a ser indexado.

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. A isso, acrescentamos a possibilidade de uma Política de Indexação Social, no qual ela seja construída no intuito de amenizar as questões problemáticas relatadas anteriormente.

Uma das principais características da Indexação Social é a definição de palavras que possuem como propósito representar os conteúdos disponibilizados. Com base nisso podemos validar essa prática, com ressalvas, e aplicar ao contexto apresentado por Lancaster (2004, p.27) que afirma que "quanto mais termos forem utilizados para indexar um documento mais acessível ele se tornará e, provavelmente, mais vezes serão recuperadas".

Entende-se a existência da preocupação com a utilização da linguagem natural para a 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 dinamicidade, bem como interatividade conduzida a partir da Web 2.0 possibilitou o surgimento de novos produtos e serviços, alterando profundamente a forma com que os usuários participam do processo de produção, organização e disseminação de informações no ciberespaço. Houve assim, o surgimento de uma nova forma de organização e representação das informações através do resultado da interação, a Folksonomia.

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 trabalhar a Folksonomia assistida, atrelada a proposta de construção de uma política de indexação.

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 e Rubi (2006, p.50) refletem sobre a questão informando que o fato de a indexação ser considerada por muitos um processo técnico, exime a construção de uma política seguindo procedimentos sistematizados para identificação de assuntos, tendo a certeza de que para muitos é necessária apenas uma leitura rápida para identificar e extrair os termos que representem o conteúdo.

Carneiro (1985) traz as primeiras discussões em torno da necessidade de se existir uma política de indexação, e para isso apresenta o que para ela são requisitos imprescindíveis ao planejamento de um sistema de recuperação de informação:

- a identificação da organização à qual estará vinculada ao sistema de indexação (contexto);
- a identificação da clientela a que se destina o sistema (destinatário);
- os recursos humanos, materiais e financeiros (infraestrutura).

Com a conclusão de que pouco se tem sobre o desenvolvimento de políticas de indexação, quando se refere à aplicação dessas políticas para ambientes de cunho social colaborativo, isso é ainda mais crítico. Nada até então é verificado no que possa direcionar a esse campo de estudo, 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, podem ser levantados pontos que são amplamente discutidos no que se trata da participação do indivíduo que tem acesso ao conteúdo.

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

## 2.5 TAINACAN

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

Nesse contexto de acompanhamento de tendência, é válido destacar a título de informação que o Tainacan surgiu em 2009, através do próprio Ministério da Cultura - Minc, a partir de discussões em torno da Cultura Digital no Fórum da

Cultura Digital Brasileira, reforçando a ideia de que é preciso encontrar formas de entender e atender o cenário que se desenha.

O Tainacan tem como proposta, soluções técnicas de integração que anseiam pela simplicidade e 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 científica. Seu desenvolvimento teve como intuito efetivar ao máximo a inteligência coletiva por meio da colaboração, se apresentando de fácil configuração e personalização, principalmente no processo de gestão das coleções digitais.

A ideia de coletividade no qual se baseia a web 2.0 pode ser visualizada de forma efetiva na maneira amigável e responsiva dada pelo usuário final quando utiliza a ferramenta. Assim, se tem uma importante ferramenta para análises de comportamento do que pode ser desenvolvido em repositórios digitais.

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, por exemplo, como administração, armazenamento, preservação e disponibilização de conteúdos informacionais nos diversos formatos eletrônicos.

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.

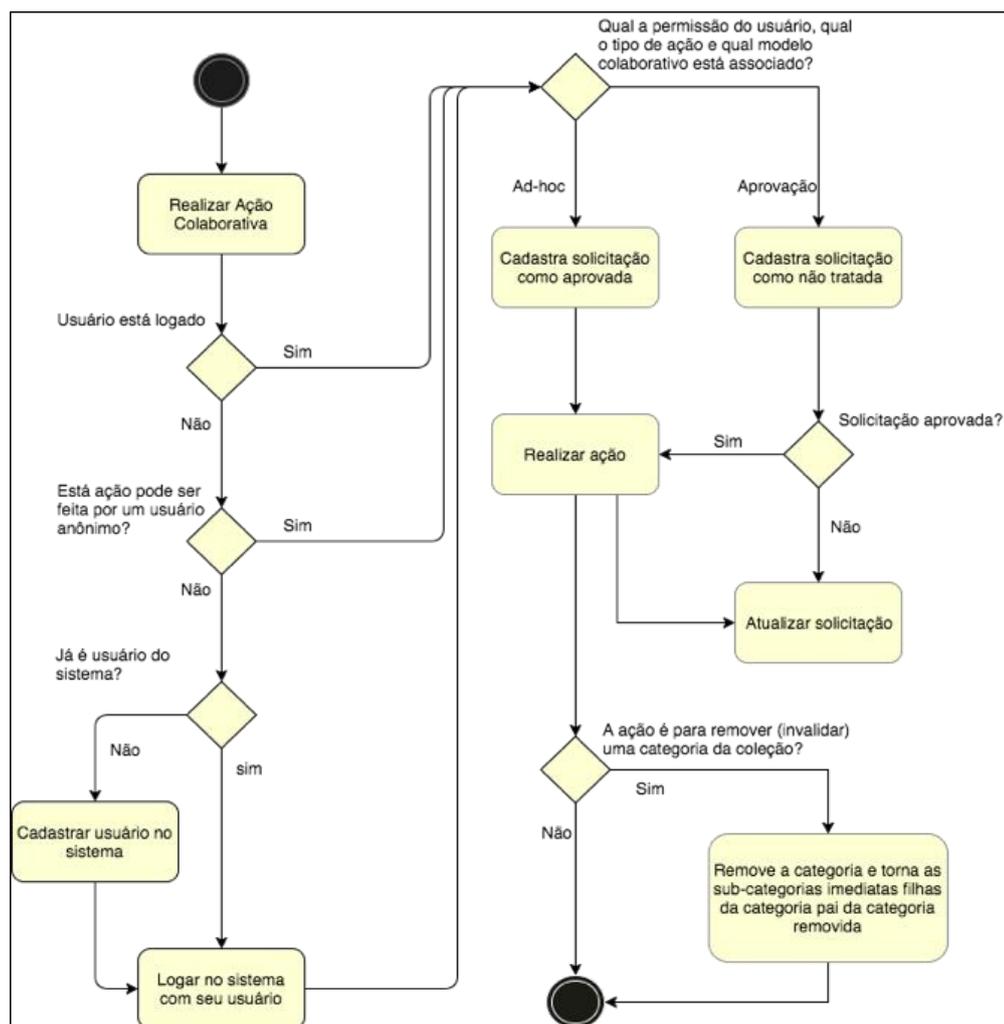
Como diferencial, o projeto tem por objetivo agregar características que facilitem a integração com as mídias sociais e, principalmente, permita um alto grau de participação social de usuários. Por fim, é notório que o Tainacan, enquanto ferramenta tem por objetivo dar condições à criação de repositórios de fácil utilização, configuração e efetivação, e tudo isso, sendo uma alternativa de livre acesso.

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 figura a seguir tem o intuito de apresentar o fluxo de trabalho para desenvolvimento das ações descritas como eventos dentro do Tainacan.

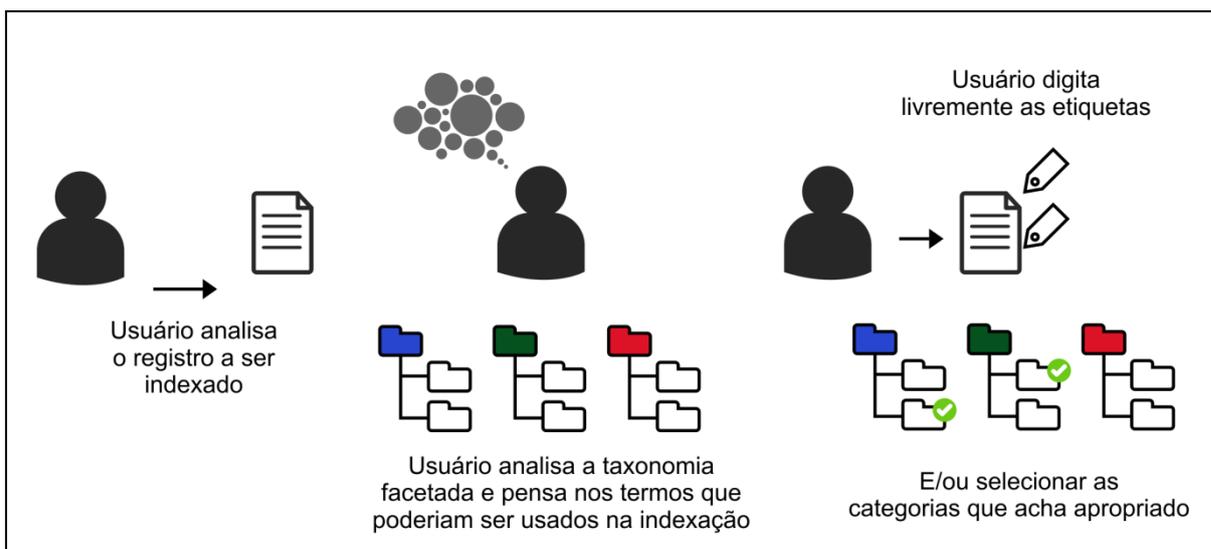
**Figura 1 – Ação de colaboração para Eventos e Usuários do Tainacan**



Fonte: Wiki do Tainacan (2018, *online*)

Além disso, o próprio Tainacan informa que existe o processo de indexação de itens (documentos, vídeos e arquivos sonoros) da coleção (catálogo), que pode ser realizados através do usuário. A representação desse processo pode ser visto na figura a seguir:

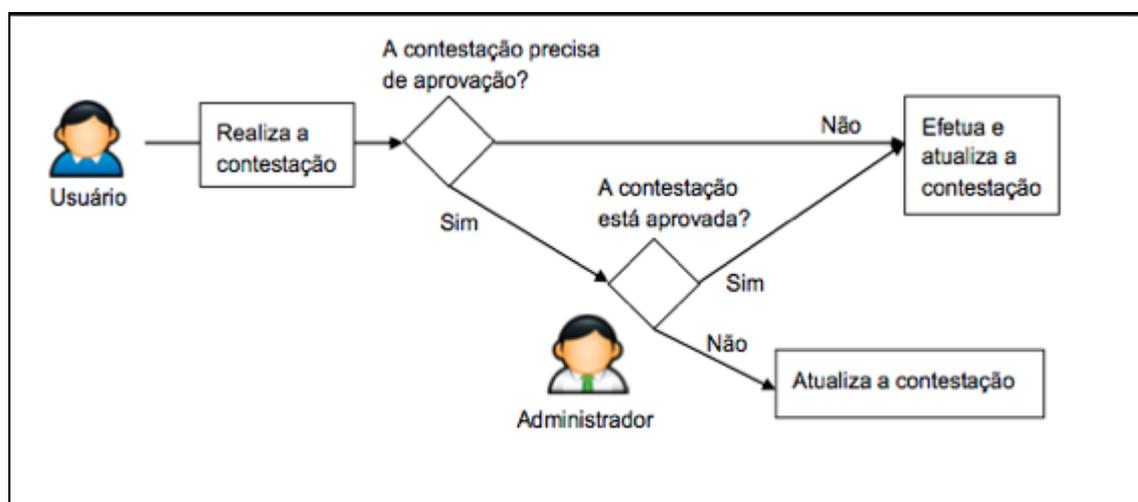
**Figura 2 – Processo de indexação do usuário**



Fonte: Wiki do Tainacan (2018, *online*, adaptada)

É importante destacar a também existência do processo de contestação da indexação realizada, ou simplesmente contribuição do usuário.

**Figura 3 – Processo de contestação para colaboração**



Fonte: Wiki do Tainacan (2018, *online*)

O fluxo de trabalho apresentado tem o intuito detalhar como ocorre a colaboração dentro das coleções disponibilizadas através do Tainacan, e como a plataforma funciona enquanto ferramenta de interoperabilidade.

## 2.6 TESAUROS

Decorrente à explosão informacional ocorrida no período pós-segunda guerra, muitas questões que passaram a infligir diretamente a recuperação da informação

começaram a serem objetos de estudo. Os tesouros surgiram então como ferramenta que visava auxiliar o processo recuperação de informação diante do cenário caótico que se apresentava.

Um tesouro é uma linguagem que tem como aspecto traçar ligações entre os termos que o compõem. Os tesouros assim constituem uma ferramenta de indexação aplicada às atividades de organização da informação.

O conceito de tesouro para análise de um assunto em um determinado documento derivou em última análise do *Thesaurus of english words and phrases* de Peter Mark Roget, que primeiramente constitui numa numeração de termos da língua inglesa, ordenando esses termos, segundo um número específico de categorias: espaço, matéria, intelecto, volição, afeições. Segundo Dodebei (2002, p. 64), essas categorias foram elaboradas seguindo o pensamento de Peter acerca dos significados das palavras. Segundo ele o seu tesouro foi organizado seguindo a teoria em que tendo uma ideia, qual palavra ou frase representaria essa mesma ideia.

Foi a partir da elaboração do tesouro de Peter Mark Roget, que em 1940 a Ciência da Informação passa a vislumbrar o tesouro como mais um processo de organização e recuperação da informação, que segundo aborda Dodebei (2002, p. 66) “[...] como sendo um instrumento capaz de transportar conceitos e suas relações mútuas, tal como expressos na linguagem dos documentos, em uma língua regular, com controle de sinônimos e estruturas sintáticas simplificadas”.

Diante disso, podemos afirmar que o tesouro é tido como um tipo de linguagem documentária, sendo um instrumento que permite o estabelecimento de relações de termos de determinado campo que estejam relacionados entre si com a finalidade de indexar e recuperar em um sistema de recuperação de informação. (CAMPOS; GOMES, 2006)

Desse modo, temos que os tesouros consistem em uma linguagem documentária específica cujo objetivo é a representação e recuperação da informação. Por conseguinte, os tesouros se caracterizam como instrumentos de organização objetivando a recuperação de informações a partir da utilização de termos.

A finalidade do uso do tesouro em bibliotecas e centros de documentação é para organizar o conteúdo dos assuntos apresentados em documentos ou em outros meios, para que seja eventualmente consultado. No entanto cabe ao bibliotecário ou

documentalista, desenvolver, organizar e sistematizar os termos que vão fazer parte da estrutura de um tesauro, a fim de garantir à qualidade das terminologias empregadas fornecendo desse modo a garantia de uso na recuperação da informação.

Para isso é necessário buscar na terminologia os conceitos teóricos para fundamentar o trabalho de construção de uma linguagem documentária, nas áreas especializadas no conhecimento humano.

Segundo Kobashi, Lara e Tálamo (1992, p. 197) discorrem que:

A linguagem documentária são instrumentos de controle terminológico que atuam em dois níveis: a) na representação da informação obtida pela análise e síntese de textos; b) na formulação de equações de busca de informação. Isso nos leva a entender que a terminologia vai fornecer a metodologia para a busca e uso dos termos a serem utilizados na estruturação do tesauro.

Nesse instante é necessário ressaltar que o trabalho de construção de um tesauro não se resume em uma tarefa apenas maquinal. Não apenas conta com a distribuição de termos em um determinado sistema de recuperação de informação, esse processo exige um desempenho também intelectual de toda a equipe envolvida.

Como afirma Kobashi, Lara e Tálamo (1992, p. 197) quando discorrem sobre os dois conjuntos referenciais os quais são básicos para a construção dos tesauros, são eles: o conhecimento categorizado em assuntos e em segundo os documentos confiáveis dos quais possam ser retirados os termos tomados como relevantes.

Esses dois conjuntos apresentados correspondem as fontes primárias e secundárias as quais auxiliam na elaboração do universo conceitual dos tesauros. Ambos são processos significantes na elaboração dos tesauros.

Entretanto para a determinação do universo terminológico, Dodebei (2002, p. 69) dá ênfase ao uso das fontes primárias por estas apresentarem maior confiabilidade no que diz respeito ao sentido da comunicação anunciada nos documentos ou ainda na formulação de uma necessidade de informação.

Em se tratando do primeiro conjunto este se resume em uma garantia classificatória da organização de todo o universo, isto é, no objeto representado. No segundo, podemos entendê-lo como uma garantia literária que visa a determinação das linguagens de indexação mais frequentes para o processo de representação da informação.

Após a apresentação desses dois conjuntos podemos intentar uma segurança significativa quanto a eficácia dos tesouros e sua funcionalidade enquanto um instrumento de controle terminológico objetivado para a representação e recuperação da informação.

Sendo assim, é preciso atenção na elaboração de tesouros principalmente no que concernem as suas limitações terminológicas e na RI, pois a definição precisa dos descritores na apresentação dos tesouros é de grande utilidade na compreensão dos termos apresentados. Assim, verifica-se que o uso de glossário se torna eficaz na constituição dos tesouros.

### **2.6.1 Tesouro Brasileiro em Ciência da Informação (TBCI)**

A Ciência da Informação sempre se caracterizou por manter estudos de tendência multidisciplinar e interdisciplinar. Como resultado dessa característica, a sua classificação alterna entre as especificidades do campo e as mesclas conceituais corporizadas nos limites do seu domínio.

A construção de um tesouro para a Ciência da Informação, como o Tesouro Brasileiro de Ciência da Informação -TBCI, levou em consideração essa questão, sob a pena de manter uma estrutura deficiente de consistência e um facetamento que não conseguisse refletir o real entrelaçamento conceitual, preocupação essa, presente em todo o processo.

Segundo o próprio Instituto Brasileiro de Informação em Ciência e Tecnologia -IBICT, responsável pelo TBCI, o Tesouro enquanto instrumento é fundamental como meio para alcançar a devida consistência terminológica e de vocabulário dentro de determinados campos do conhecimento. Sua aplicação tem funcionalidade não apenas para os profissionais indexadores, mas também para pesquisadores, professores e profissionais de informação em geral.

A construção do Tesouro em Ciência da Informação iniciou-se em 1989, e dentre as informações importantes em torno da construção do TBCI, está a do apoio da FINEP e do próprio IBICT, mas que de fato, é resultado do estudo iniciado há mais de 20 anos na linha de pesquisa em Ciência da Informação, do Curso de Doutorado em Comunicação e Cultura da Universidade Federal do Rio de Janeiro. Lena Vania Ribeiro Pinheiro, até então pesquisadora e professora do IBICT, com colaboração da Helena Dodd Ferrez (que já era autora principal do tesouro para

acervos museológicos) foram às responsáveis por darem início ao projeto. Contudo, vale ressaltar que o Tesouro não chegou a ser publicado de imediato, ficando restrita e destinada apenas ao uso interno.

Para a construção do TBCI, importantes também foram às fontes às quais recorreram, como outros tesouros e dicionários, todos identificados nesta publicação. Por essa razão, é possível visualizar como sendo a maior contribuição, do IBICT à ampliação dos estudos de questões ligadas a linguagem documentária.

O TBCI conta com cerca de 1.800 termos, sendo a maioria contemplada também com versão inglesa e espanhola, e por definições. O IBICT reforça que a expectativa é a de que o Tesouro Brasileiro de Ciência da Informação tenha um papel central na recuperação da informação no Brasil e para países lusófonos.

Como um dos principais resultados, está a definição de grandes categorias e as disciplinas de cada conceito. Pinheiro e Ferrez (2010, p.10) apresentam:

as disciplinas estruturais foram reunidas em Fundamentos da Ciência da Informação; as instrumentais passaram a ser representadas por Organização e Processamento da Informação; as disciplinas gerenciais ganharam o nome de Gestão da Informação; as tecnológicas foram agrupadas pelas Tecnologias da Informação; e as disciplinas socioculturais ficaram identificadas por Transferência da Informação.

Diante desse cenário, a consulta a tesouros de diferentes décadas se tornou importante por deixar mais evidente a evolução terminológica ao longo dos anos. Por outro lado, o fato de ser em diferentes idiomas e culturas, além de enriquecedor do ponto de vista do trabalho em si, motivou a expansão do tesouro brasileiro, razão pelo qual é importante estar atento e flexível com as mudanças que podem vir a ocorrer.

## 2.7 SISTEMAS DE INDEXAÇÃO AUTOMÁTICA

A consistência aferida ao assunto que será indexado depende de alguns fatores a serem considerados. 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”, e para isso eles acrescentam que o resultado da indexação descende da combinação de diversos elementos, e que são eles:

- 1) A formação, os conhecimentos na matéria, a profissão e a motivação do indexador;

- 2) As características do objeto indexado; e
- 3) As condições em que se dá a indexação.

No que se refere à dinâmica apresentada a esse conhecimento, a necessidade de atualizações constantes por parte de quem indexa, se torna ainda mais importantes para atingir um determinado grau de qualidade.

Na literatura de Ciência da informação é possível encontrar muitas discussões referentes a pontos positivos e negativos da aplicação de um sistema de indexação. Com o propósito de sistematizar os argumentos identificados, Gil Leiva (2008, p. 320) apresenta o seguinte quadro: aquilo que foi verificado como argumentos a favor da automatização da indexação:

**Quadro 1 – Fatores positivos e negativos da Indexação automática**

Fatores Positivos	Fatores Negativos
<ul style="list-style-type: none"> <li>• Caracteriza a indexação humana como lenta, subjetiva e de alto custo;</li> <li>• Afirmam que com a automatização da indexação tem-se diminuição de erros que repercutirá positivamente na recuperação das informações em bases de dados;</li> <li>• A indexação parece ser mais precisa, permitindo uma recuperação dos documentos mais rica.</li> </ul>	<ul style="list-style-type: none"> <li>• Alegam sobre a incapacidade dos sistemas automáticos de indexação reconhecerem diferentes significados em diferentes contextos, relacionar e selecionar conceitos implícitos dos documentos;</li> <li>• Afirmam que a indexação automática reconhece palavras e não conceitos, defendem que se deve perseguir a captação de terminologias dos textos, porque esta cumpre a função representativa, cognitiva e comunicativa que apresentam os conceitos e, por tanto, o conhecimento;</li> <li>• Alegam que na maioria das ocasiões a automatização da indexação restringe-se às áreas</li> </ul>

	<p>específicas do conhecimento;</p> <ul style="list-style-type: none"> <li>• Uma das principais razões abordadas contra a automatização do processo é a impossibilidade, no estado atual da investigação, conseguir indexação totalmente automática.</li> </ul>
--	---

Fonte: Gil Lopes (2008, p. 320)

Outro ponto a ser considerado, segundo Borko (1977 apud GUEDES, 1994), diz respeito à inconsistência interindexadores (diferentes indexadores atribuindo diferentes termos-índice a um mesmo conceito/documento) e intraindexador (o mesmo indexador atribuindo diferentes termos-índice a um mesmo conceito/documento, em diferentes momentos). Ainda como possíveis entraves, a possibilidade do indexador não ter domínio sobre o idioma e o vocabulário da linguagem de especialidade do documento, é um ponto determinante para uma baixa qualidade da indexação.

Frente às questões apresentadas, pesquisas sobre indexação automática passaram a ter uma atenção maior dos pesquisadores da Ciência da Informação. Como resultados desses estudos, surgiram algumas alternativas para solucionar tais problemas, sobretudo em ambientes digitais.

Para Robredo (1982), o processo de indexação automática é semelhante ao processo de leitura-memorização humano, tendo seu princípio geral baseado na relação feita de cada palavra do texto com a de palavras vazias de significado. O que nos leva a conduzir, por eliminação, as palavras que serão descartadas, e as que possuem algum significado para o todo.

No processo de leitura-memorização, não interessa ao indivíduo, as letras, ou palavras soltas, mas a ideia que elas passam a formar quando organizadas. O olho, janela do cérebro, reconhece as palavras significativas e suas associações fixando-se nelas um tempo necessário para assegurar a memorização das ideias, pulando, praticamente, as palavras não significativas (ROBREDO, 1982).

O processo de memorização humana pode ser visualizado a partir de duas etapas: 1) **memorização temporária e inconsciente**, no qual há a conservação das

palavras significativas passando por uma modificação ou aperfeiçoamento das mesmas a partir da detecção de novos conceitos significativos; e, 2) **memorização permanente**, onde se reforça os conceitos assim trabalhados, onde pode ser atribuído o nome de memória. Efetivada as duas etapas, o que ocorrem enquanto processo é a fixação na memória de uma série de palavras-conceitos-descritores que passa a representar as ideias básicas do documento. Em decorrência desses pontos, através de um processo de análise-indexação feita na leitura, pode se observar a armazenagem dos descritores que representam o conteúdo dos documentos (ROBREDO, 1982).

Em similaridade como ocorre no processo de indexação automática acontece na indexação intelectual à escolha das palavras que representarão o conteúdo analisado. Assim como um sistema define, por exemplo, as palavras mais representativas através da frequência que ela se apresenta, o cérebro humano, de forma intuitiva cria mecanismos para definir sobre o que se trata o conteúdo que está sendo lido.

Em seus estudos Lancaster (2004) demonstra que era necessário o processamento apenas da primeira e da última frase de cada parágrafo, pois, em 85% das vezes a primeira frase era o tópico frasal e em 7% dos casos a última frase o era. Considera-se como tópico frasal a parte do texto que provia o máximo de informações relativas ao conteúdo do texto.

Sendo assim, o Lancaster (2004, p.285) ainda destaca a existência de dois processos de indexação automática: 1) a **Indexação por Extração Automática**, onde as palavras ou expressões que aparecem no texto são extraídas para representar seu conteúdo como um todo. Quando considerado a versão eletrônica de documentos, é possível se valer de programas computacionais para a extração de termos, tendo os mesmos princípios utilizados pelos indivíduos, como: frequência da palavra dentro do texto; posição da palavra no texto (no título, nas legendas, no resumo etc.) e por seu próprio contexto; e 2) a **Indexação automática baseada em frequência** sugere que em substituição ao processo que analisa todo o texto, sejam analisadas apenas o “Tópico Frasal” e “Palavras Sugestivas”.

Outro processo de indexação automática é o de **Indexação por Atribuição Automática** no qual consiste na criação para cada termo do vocabulário controlado, de um “perfil” de palavras ou expressões que costumam verifica-se nos documentos aos qual um indexador humano atribuiria esses termos. Nesse ponto, Lancaster

(2004) atenta que a indexação acontece em duas fases: 1) extraem-se palavras ou expressões do texto por meio de técnicas estatísticas; e 2) partindo desse conjunto de palavras/expressões, seleciona-se no vocabulário controlado o termo cujo perfil possui certo nível de conhecimento.

Efetivamente, pode ser considerado o mais complexo de ser realizado com um maior grau de eficiência quando comparado ao de indexação por extração automática. De forma geral, a dificuldade se dá no ato da representação de conteúdo temático, pois é necessário um controle terminológico mais efetivo.

A história da indexação automática pode ser associada ao uso de programas computacionais que visam gerar índices pré-coordenados. Segundo Lancaster (2004, p. 50), “a flexibilidade inerente aos sistemas pós-coordenados deixa de existir quando os termos de indexação são impressos em papel ou fichas catalográficas convencionais”. Ainda segundo o autor, “vários programas de computador foram desenvolvidos para gerar, automaticamente um conjunto de entradas de índices a partir de uma sequência de termos”. Dois exemplos de índices pré-coordenados são os índices impressos e os catálogos, sendo caracterizados por: 1) dificuldade de representação de multidimensionalidade das relações entre os termos; 2) possibilidade de listagem dos termos somente em uma sequência, o que implica que o primeiro termo é mais importante que os demais; e, 3) dificuldade, ou impossibilidade, de combinação de termos no momento da busca (LANCASTER, 2004, p. 50).

Para Robredo (1982, p. 238), a primeira aplicação da indexação automática de documentos técnicos, a partir de palavras significativas dos títulos, aconteceu com KWIC, quando foi capaz de produzir índices de títulos permutados. O mesmo KWIC, que foi desenvolvido por H. P. Luhn em 1959 corresponde a um índice rotativo onde cada palavra-chave que aparece nos títulos dos documentos acaba se tornando uma entrada do índice.

Além do relatado acima, o KWIC é considerado um método simples e barato, que alcança em certo nível, acesso temático ao conteúdo disponibilizado em uma coleção. É válido, porém destacar, que sua qualidade está intimamente relacionada à qualidade dos títulos, no sentido em que estes devem ser bons indicadores do conteúdo dos textos (LANCASTER, 2004, p.54 e 55).

Dentre os métodos de indexação automática, ainda pode ser descrito o Nested Phrase Indexing System - NEPHIS (Sistema de Indexação de Frase

Encaixada), que segundo Lancaster (2004, p.56-59) é um índice articulado de assunto, que foi criado por T. C. Craven em 1977. Dentro desse modelo, os termos de entrada são reordenados de tal modo que cada um deles se liga a seu vizinho original por meio de uma palavra funcional ou pontuação especial, conservando-se, assim, a estrutura similar à de uma frase; mesmo que muitas vezes disposta em ordem diferente. Então, a sintaxe do texto original pode ser mantida de modo que o significado do enunciado não se perca.

Vieira (1988) *apud* Lapa (2014, p. 30) em sua dissertação traz um panorama acerca da indexação automática no Brasil, e assim relata a aplicação da indexação automática teve seu ponto de partida no país nos anos 60. Iniciou-se através do programa KWIC (Keyword In Context), e tinha como objetivo elaborar os índices das bibliografias especializadas publicados pelo Instituto Brasileiro de Bibliografia e Documentação (IBBD), atual Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).

Os estudos em torno da indexação automática no Brasil, já na década seguinte, 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.

Referente a isso, Lapa (2014) *apud* Bandim (2017, p. 44-45) apresenta os sistemas de indexação mais citados na literatura, a partir de uma análise em sua dissertação, são eles:

- 1) O sistema PRECIS que é usado na construção automática de índices de assunto (FUJITA, 1989 *apud* LAPA, 2014);
- 2) Já Robredo (1982) propõe a utilização do sistema de indexação automática, nomeado como AUTOMINDEX/II, para gerar uma tabela de frequências de aparecimento dos descritores encontrados em cada arquivo;
- 3) A proposta de Narukawa (2011) é a de investigar o uso e a adequação de linguagens documentárias alfabéticas e vocabulários controlados (tesauros) em indexação automatizada e indexação automática por atribuição na aplicação do (SISA).

Alguns autores relatam a importância da semântica para a indexação automática. Rector e Yunes (1980, p. 14) afirmam que,

uma explicação de propriedades semânticas requer mais do que a análise do sentido das palavras apenas, isto é, para que se entenda o sentido de uma sentença e suas relações semânticas com outras expressões, é

preciso saber não só o significado de suas unidades léxicas, mas, também, como estas se relacionam – a dependência da estrutura sintática da sentença.

Dessa forma a semântica passa a desempenhar papel importante na indexação automática ao passo que permite o trabalho de *softwares* para a formação das estruturas lexicais das frases, bem como o sentido dos termos que representam o conteúdo informacional.

### **2.7.1 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.

Quando se fala em indexação, as primeiras questões que vêm à mente perpassam pela definição dos termos que virão a representar o conteúdo informacional. O aumento exponencial, não só na produção, mas nas buscas por informações apresenta um cenário de desafio no qual a necessidade de organizar as informações de forma sistemática foi amplificada.

Com algumas características oriundas dos espaços digitais, essa questão se tornou ainda mais complexa, contudo traz consigo a preocupação ainda maior em avaliar a indexação. Diante disso, é possível fazer uma clara relação ao que foi dito por Lancaster (2004), no qual uma mesma publicação pode vir a apresentar conjuntos diferentes de termos de indexação dependendo do grupo de usuários ao qual se destina e aos interesses particulares desse grupo, ou seja, não há uma forma ‘correta’ única de indexar.

Essa citação amplia as discussões em torno da indexação e tudo aquilo que trata de critérios de avaliação, tornando tudo ainda mais complexo. Nessa conjuntura, Lancaster (2004) reforça a necessidade de trazer à tona o que se conhece como avaliação, pois segundo o mesmo é o único caminho que leva à melhoria dos serviços e produtos oferecidos aos usuários e à detecção dos pontos positivos e negativos da usabilidade de um catálogo em linha ou em fichas. Segundo o autor a realização da avaliação se torna necessária por:

- Estabelecer uma espécie de “escala” para mostrar em que nível de desempenho o serviço está funcionando no momento;
- Comparar o desempenho de várias bibliotecas ou serviços;
- Justificar a existência do serviço ou produto;
- Identificar as possíveis causas de malogro ou ineficiência do serviço, visando a elevar o nível de desempenho no futuro.

Para definir o que vem a ser uma indexação de qualidade é então necessário compreender que a mensuração através da quantidade de descritores representativos de determinado documento não é preponderante para definir uma boa indexação. O mais importante dentro desse processo de busca acontece quando a recuperação traz o maior número de documentos úteis com o menor número de ruídos.

Segundo Lapa (2014, p. 89) dois são os processos de avaliação da indexação:

- 1) Avaliar pela recuperação: nessa é observado através de medidas definidas os resultados apresentados na recuperação da informação de determinado documento;
- 2) Avaliar pela linguagem: o processo de avaliação é levado em consideração se os termos recuperados satisfazem o processo de busca.

Lancaster (2004) destaca todo esse cenário numa relação direta com o processo decisório na forma de execução da indexação. Através da avaliação é possível obter dados necessários para determinar quais dentre várias estratégias alternativas parecem ter mais probabilidade de obter um resultado almejado. A avaliação da indexação automática, segundo Lapa (2014, p. 117) pode ser realizado por meio dos cálculos dos coeficientes de precisão e revocação na recuperação de documentos de um sistema.

Para Hlava (2002) a avaliação dos sistemas de indexação automática parte do princípio de análise dos índices selecionados automaticamente. Para isso, ele define que as medidas da precisão de um sistema baseiam-se em três categorias:

- Acertos – resultados que coincidem exatamente com o que o indexador humano teria aplicado ao sistema;
- Erros (*misses*) – palavras-chave que o indexador humano teria selecionado e, o sistema automatizado não o fez;

- Ruídos (*noises*) – palavras-chave selecionadas pelo computador que um indexador humano preferiria terem ignorado.

Nessa perspectiva, o estudo estatístico, dos valores dos acertos, erros e ruídos são a medida de avaliar a qualidade do sistema. Segundo Hlava (2002), O percentual mínimo de acerto deve ser de 85% de um total de 100% de precisão de indexação humana para que o sistema seja considerado aceitável. Isso significa que a quantidade combinada dos índices, Erros e Ruídos, precisam ser inferiores a 15%.

Ainda 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.

Diante disso, torna-se evidente que uma das maneiras de avaliar a Indexação Automática surge por meio de cálculos dos coeficientes de precisão e revocação no ato de recuperação de documentos em um sistema. Antes disso, Gil Leiva (2008, p. 38) discorre que a avaliação da indexação pode ser realizada de duas formas:

- Intrínseca: sendo o conjunto de tarefas centradas no resultado da indexação (descritores, cabeçalhos, subcabeçalhos ou identificadores) com a finalidade de conhecer sua qualidade. A avaliação intrínseca da indexação pode ser qualitativa, isto é, por meio de valoração e consensos entre especialistas ou quantitativa através de fórmulas;
- Extrínseca: realiza uma análise interior enquanto que na avaliação extrínseca o resultado da indexação é utilizado para compará-lo com a indexação de outra unidade de informação que também indexou os mesmos documentos (interconsistência) ou para testar a função da indexação na recuperação (exaustividade e precisão na recuperação)

Baseado nisso, os procedimentos de avaliação de indexação definidos no que se refere à avaliação intrínseca, tem como objetivo medir o grau de consistência da indexação. Já a avaliação extrínseca, objetiva alcançar através de medição o grau de revocação, precisão e medida F (média harmônica ponderada entre a precisão e revocação) na indexação automática.

O autor ainda apresenta um formula conhecida como “coeficiente de consistência”, sendo ela proposta num primeiro momento por Salton e McGill (2003). A ideia consiste em definir a consistência entre a indexação automática e o manual a partir de uma fórmula, sendo ela:

$$CI = \frac{T^{\infty}}{(A + B) - T^{\infty}}$$

No qual:

- CI = a coerência entre dois sistemas ou dois indexadores;
- $T^{\infty}$  = 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 forma de avaliação segundo Narukawa, Gil Leiva e Fujita (2009, p. 110), consiste em consultar duas bases de dados que contém os mesmos campos e idênticos conteúdos salvo os campos que armazenam a indexação. Os resultados obtidos denotam os índices de exaustividade e precisão na recuperação. A preocupação com o uso deste método de avaliação ocorre frente ao fato dele ser efetivamente mais trabalhoso, contudo proporciona resultados claros, mediante o uso de fórmulas de exaustividade ou revocação (R) e precisão (P) na recuperação:

$$R = \frac{\text{Número de termos relevantes recuperados}}{\text{Número total de termos relevantes}}$$

$$P = \frac{\text{Número de termos relevantes recuperados}}{\text{Número total de termos recuperados}}$$

No intuito apresentar um único valor, surge à possibilidade de combinar os dois resultados gerados a partir das equações apresentados acima (Revocação e Precisão). Chamada de medida F, a equação 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 instantaneamente o valor 0. A medida F assumirá o valor 1 quando todos os termos recuperados forem relevantes e foram 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 máximo para medida F é visto como uma tentativa de se encontrar equilíbrio entre estes dois índices.

$$\text{Medida F} = \frac{2 \times (\text{Precisão} \times \text{Revocação})}{\text{Precisão} + \text{Revocação}}$$

Corroborando sobre a importância dos apontamentos realizados acima, Gil Leiva (2008, p. 76) *apud* Celerino (2018, p. 47) apresenta o seguinte quadro acerca dos critérios de indexação:

#### Quadro 2 – Critérios de indexação

Qualidades da Indexação	
1. Exaustividade	Conceitos caracterizadores do conteúdo presente no documento.
2. Especificidade	Relação exata entre a unidade conceitual e o termo escolhido para representar.
3. Coerência	Ausência de erros de inclusão e omissão.
4. Consistência	Grau de coincidência entre duas ou mais indicações.

Fonte: Gil Leiva (2008, p. 76) *apud* Celerino (2018, p. 47)

Com base no quadro apresentado, Celerino (2018, p.47) afirma que na indexação são utilizados diversos critérios e indicadores que possibilitam avaliar sua qualidade, sugerindo assim um acréscimo no que diz respeito aos critérios. Os critérios são: revocação que é relacionado à exaustividade, precisão à especificidade, coerência e consistência, para tanto o autor ainda aborda a relevância e pertinência.

Para Celerino (2018, p. 48) a **revocação, ou exautividade**, são usados como indicadores para a avaliação da qualidade e do desempenho, tanto do Sistema de Recuperação da Informação (SRI) como da indexação. Já o indicador de **precisão**,

**ou especificidade**, diz respeito ao número de documentos recuperados que atendem a necessidade do usuário.

Celerino (2018, p. 47), continua e explica que os critérios de **consistência e coerência**, onde para o mesmo estão relacionados, pois quando se busca uma indexação consistente, significa que está interessado em indexar com coerência. Nessa perspectiva, a coerência se apresenta relacionada aos termos atribuídos por diversos indexadores para representar o mesmo documento, quanto mais termos se coincidem, mais a indexação está coerente.

Uma questão que vale o destaque se trata dos critérios de relevância e pertinência, que no quadro de Gil Leiva (2008, p. 76) *apud* Celerino (2018, p. 47) não constam, mas que segundo Celerino (2018, p. 47) surgem como critérios que devem ser levados também em consideração dentro do processo de avaliação dos sistemas de indexação automática.

O autor explica que **relevância** na indexação está diretamente ligada à necessidade de informação do usuário, pois é através da sua satisfação que é possível indicar o nível de relevância da indexação. Já no que se refere a **pertinência**, Celerino (2018, p. 48) apresenta a existência de uma semelhança com a relevância, pois trata da relação da necessidade e do uso do documento. Como forma de acrescentar à discussão, Celerino amplia a discussão através do que fora mencionado por Harter (1992) *apud* Dias e Naves (2007, p. 19), onde o critério da pertinência é “[...] outro aspecto da relevância subjetiva, tem sido usado para se referir à relação entre documento e a necessidade de seu uso”.

Fato importante para os sistemas automáticos é que o grande volume de informação faz com que eles ainda não consigam lidar de forma satisfatória com a linguagem natural, oriunda do usuário, ao ponto da indexação de documentos textuais terem um alto grau de qualidade. Segundo Gil Leiva (1999, p.60) o indexador pode contribuir no desenvolvimento das atividades de construção e avaliação dos sistemas de indexação automática.

A avaliação dos sistemas de indexação automática segundo Farmer (2006, p. 99-100), traz uma nova perspectiva quando se refere às ferramentas de categorização automática de documentos digitais Ela confirma a necessidade da parceria Homem-Máquina para somar à capacidade de processamento de textos dessas ferramentas, a inteligência, julgamentos e experiência humanos.

Sendo assim, a visão da autora vai de encontro a melhoria na efetividade das taxonomias e no desempenho dos sistemas quando aplicadas as habilidades humanas à configuração das ferramentas, ao controle de qualidade da indexação (avaliação), à criação das taxonomias, em testes e treinamento dos sistemas e à criação de regras de classificação.

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. A indexação está passando por um processo de transformação em decorrência da multiplicação das publicações, ou informações geradas.

Narukawa, Gil Leiva e Fujita (2009) apresentam indícios da necessidade de verificar-se a qualidade da indexação, e para isso utiliza os métodos de indexação automática por atribuição, para assim investigar índices de consistência, revocação e precisão, sendo os dois últimos, avaliados na recuperação da informação. A ação ocorreu seguindo uma análise comparativa entre a indexação automática com o uso da linguagem dos descritores em Ciências da Saúde (DeCS) e a indexação manual do Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (BIREME), a proposta visava encontrar formas de contribuir com o aperfeiçoamento e aplicação do SISA.

Para Narukawa (2011) a análise trouxe como resultado a indispensabilidade de adequação do vocabulário controlado, essencialmente no que se refere à definição de relações entre os termos. As relações semânticas entre conceitos devem ser mantidas durante a indexação automática para garantir qualidade na recuperação da informação.

Bandim (2017) na perspectiva de avaliar a qualidade da indexação automática por atribuição, onde ele divide em Indexação automática com atribuição de termos gerais e Indexação automática sem atribuição de termos gerais, fez uso dos índices de consistência, revocação, precisão e medida F.

## 2.8 CORPUS DE REFERÊNCIA

A seleção de um Corpus consiste na escolha manual de documentos seguindo testes de agrupamento. Têm assim como propósito, caracterizar

publicações a partir dos objetivos e aspectos metodológicos traçados para uma pesquisa.

Para validar a construção de um corpus é preciso ter em vista que ele se constitui a partir de um quantitativo de documentos, que tendem a ser categorizados seguindo a natureza da investigação a ser desenvolvida. Tal quantitativo pode ser compilado, por exemplo, a partir de periódicos científicos, oriundos de diversas bibliotecas ou repositórios digitais.

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.

Trask (2004) afirma que corpus é “um conjunto de textos escritos ou falados numa língua, disponível para análise”. Contudo, é preciso ter cuidado no entendimento do que vem a formar esse grupo de textos que compõe um *Corpus*.

Souza (2005) utiliza em sua tese dois *corpora* de documentos visando à aplicação da metodologia proposta. O objetivo era estudar a indexação através da identificação, extração e seleção de SNs dos textos completos utilizados em 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 Sintagmas Nominais (SNs). O segundo *corpus* fora construído pelo próprio Souza (2005) e era composto por 60 documentos. A proposta era ter dois corpora, onde o 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 cada um dos dois se comportavam mediante as circunstâncias no qual foram aplicadas.

A utilização do Corpus por Souza (2005) e no presente trabalho se assemelha no intuito de definir uma metodologia que auxilie o processo de escolha automática de termos para documentos, se guiando por estruturas lingüísticas pré-estabelecidas. No caso de Souza (2005), a estrutura utilizada foi os sintagmas nominais, e no trabalho aqui desenvolvido, o Tesouro.

Nota-se que o principal objetivo em ambas as pesquisas gira em torno de definir um *corpus* significativo. Esse é capaz de dar respaldo e reconhecimento para validar a metodologia definida, bem como a devida contextualização.

Diante desse tipo de resultado, a relevância dada à utilização de *Corpus* para alguns autores vai muito além da compilação de documentos significativos. Hasan e Ng (2014) apresentam um estudo acerca *Automatic Keyphrase Extraction* para examinar os principais erros dos sistemas de indexação automática, bem como discutir os desafios futuros em torno da temática. Diante disso, os autores descrevem que os sistemas de *Automatic Keyphrase Extraction* foram avaliados em corpora de diversas 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 de definir como ele pode vir a servir no processo de avaliação de sistemas de *Automatic Keyphrase Extraction*. Isso possibilitaria uma melhora na qualidade em tarefas como o processamento de linguagem natural e recuperação de informações, sumarização de texto e categorização de texto, além da mineração de opinião e indexação de documentos.

Para efetivar a avaliação de sistemas de indexação automática, Hasan e Ng (2014, p. 1267) divide o processo em duas etapas: 1) diz respeito ao mapeamento das *Keyphrase* dos *Corpora*, tanto da indexação manual quanto da automática; 2) o uso de métricas de avaliação, como precisão (P), revocação (R) e medida-F (F) confrontando as duas formas de indexação.

A justificativa para a aplicação de tal método de avaliação vai de encontro ao fato da avaliação humana ter valor significativo, sendo sugerido cada vez mais como alternativa por diversos autores. A proposta diante desse cenário é a experimentação dos dois tipos de métricas de avaliação automática. (Matsuo e Ishizuka, 2004 *apud* Hasan e Ng, 2014).

Kuramoto (2006) se utiliza de um *corpus* para verificar a ocorrência de ambiguidades e de que forma pode ou não solucionar essas questões através do sistema Zstation. Para isso, o modelo foi elaborado a partir da experimentação de 15 artigos científicos da revista “Ciência da Informação” e que na sequência teve levada em consideração pelo autor o conjunto de dados registrados no dicionário, na gramática morfológica, na gramática de argumentos e na ontologia, para assim efetuar o tratamento sintático semântico de enunciados do corpus.

A proposta de Kuramoto (2006) tem como intuito permitir que o usuário num primeiro momento defina no sistema um termo. A partir daí o sistema funcionaria de forma a sugerir possíveis SN. O autor utiliza o exemplo do termo, “informação”, onde

a partir desse momento o sistema oferece todos os SN de primeiro nível que apresentam “informação” no seu centro, inclusive o próprio sintagma “informação”. Tendo as sugestões, o usuário seleciona o sintagma que atende a sua necessidade de informação. A ideia é que a possibilidade de uma interface como essa possa aumentar o nível de interação entre o usuário e o computador, assim como o processo de busca mediante interação com a máquina.

Em sua dissertação Bandim (2017) utiliza um *Corpus de Referência* no intuito de realizar um estudo de caso para comparar as palavras-chave extraídas desses artigos com os termos atribuídos pelo sistema SISA usando o Tesouro TBCI. A proposta do autor era analisar a indexação automática por atribuição na representação de artigos científicos da área de Ciência da Informação, fazendo uso de descritores definidos em tesauro.

Além disso, como valor a se atribuir à compilação de um *Corpus*, como o da pesquisa de Bandim (2017), por exemplo, tem-se o alcance, após as análises comparativas, de informações importantes principalmente no que diz respeito à indexação automática por atribuição, bem como embasamento para a construção de Sistemas de Recuperação da Informação em bases de dados científicas.

Dentre o que foi apresentado, é válido destacar que o uso do *Corpus* nos trabalhos de Kuramoto (2006) e Bandim (2017) podem ser justificados tendo em vista a necessidade de validar mecanismos de avaliação de sistemas automáticos de indexação. 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 corpus se comporta frente à metodologia a qual será aplicada.

A pesquisa aqui desenvolvida apresenta diversas semelhanças, assim como diferenças com o trabalho desenvolvido por Bandim (2017) sobre a utilização do *Corpus*. Em ambos existe a preocupação de aplicar meios de controle terminológico a partir do uso de vocabulário controlado, no qual possibilita a realização de avaliação do processo de indexação, em termos de qualidade, isso através dos índices de consistência, revocação, precisão e medida F.

A diferença se apresenta no foco de aplicação onde Bandim (2017) busca entender como dentro do contexto de uma base de dados de artigos científicos em Ciência da Informação, a utilização da indexação automática por atribuição com base no TBCI aplicado ao SISA gera uma indexação de qualidade. Já na pesquisa

que se segue, o âmbito de aplicação deixa de serem sistemas formais voltados aos profissionais indexadores e passa a ser o usuário-indexador.

Nesse ponto o *corpus* auxilia a pesquisa a alcançar parâmetro que visem entender como a aplicação de um controle terminológico a um espaço web aberto, com o desenvolvimento de características híbridas, pode gerar uma indexação de qualidade, aumentando assim a possibilidade de garantir melhores resultados na recuperação. 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.

### 3 PROCEDIMENTO METODOLÓGICO

Toda metodologia a ser desenvolvida tem papel determinante na execução de qualquer atividade científica. Ela consiste em uma série de etapas que tem como propósito dar condições para o entendimento do problema da pesquisa e alcançar os objetivos traçados. Tendo isso em vista, é na metodologia que se estabelece as abordagens, técnicas e processos sistematizados que serão utilizados para a investigação do problema exposto na pesquisa.

A importância da verificação de um objeto através de uma sistematização caracteriza-se como o ambiente próprio à efetivação da ciência. Isso reforça a ideia de que a sua principal função é compor ações relevantes e direcionadas a responder um problema. A formação de um conhecimento científico se diferencia, dentre as razões, pelas metodologias empregadas, por se tratar de um conhecimento real, sistemático e verificável (MARKONI; LAKATOS, 2006. p.80).

No intuito de descrever detalhadamente a metodologia da presente pesquisa, a seguir são expostos a sua classificação, o universo e a descrição das etapas a serem seguidas na pesquisa.

#### 3.1 CLASSIFICAÇÃO DA PESQUISA

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 justifica em decorrência do pouco conhecimento acumulado e sistematizado, tanto na literatura nacional quanto estrangeira, fazendo assim que seja necessidade apresentar diversos aspectos em torno da Folksonomia Assistida. Nesse sentido, Gil (2002, p.41) trabalha na perspectiva de que esse tipo de classificação para a pesquisa pode "proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a constituir hipóteses".

A parte empírica da pesquisa, envolve a realização de experimento que tem a finalidade de maximizar a qualidade do processo de RI através da construção da folksonomia assistida por um grupo de indexadores dos documentos que compõem o corpus de Souza(2005) e a posterior análise da folksonomia e compilação do corpus de referência.

A escolha pelo estudo de caso pode ser justificada levando em consideração a definição dada por Yin (2001, p. 32), onde define como sendo “uma investigação empírica que investiga um fenômeno contemporâneo dentro de seu contexto da vida real, especialmente quando os limites entre o fenômeno e contexto não estão claramente definidos”.

Nesse ponto, presume-se então que ao se valer do procedimento técnico estudo de caso, deseja-se trabalhar sob condições contextuais, por se acreditar que elas sejam significativas e oportunas ao fenômeno estudado.

A presente pesquisa terá como estudo de caso a aplicação de uma adequação do modelo colaborativo de indexação social desenvolvido enquanto protótipo na Tese de Silva (2013). Tal modelo será aplicado através da plataforma Tainacan, para uns grupos de usuários pré-estabelecidos. Vale reforçar nesse ponto que segundo Silva (2013, p.122) os estudos sobre a aplicação de modelos de colaboração na classificação e navegação facetada, e que aqui serão aplicados eram raros e que pelo estudo até aqui feito, não houve mudança nesse cenário. Por tal razão, e como forma de garantir o andamento do estudo seguindo o atual contexto de aplicação, achou-se por bem a adaptação do modelo já desenvolvido pelo autor e que será descrito nas etapas da pesquisa.

Assim, o intuito do estudo é ter condições de entender de que forma os possíveis fenômenos se apresentaram na construção de uma Folksonomia assistida, e de que maneira ela pode ser aplicada na construção efetiva de um corpus de referência. Podendo vir a atuar como mecanismo de avaliação de sistemas de indexação automática.

Para isso será configurado no Tainacan uma coleção de documentos de acordo com adaptação de um modelo de indexação social assistida, e desenvolvido uma política de indexação que adota um tesouro da área de Ciência da Informação, ambos determinantes para o cumprimento dos objetivos, incluindo também a aplicação de técnicas e instrumentos visando à compilação do corpus de referência.

### 3.2 INSTRUMENTO E UNIVERSO DA PESQUISA

Como instrumento para efetivação dos objetivos da pesquisa e para aplicabilidade dos conceitos apresentados nessa dissertação, será utilizado o Tainacan, plataforma desenvolvida para possibilitar a criação de coleções digitais facilitando ao máximo o reuso de objetos digitais já existentes em outros ambientes espalhados pela web.

A escolha da plataforma ocorreu por se tratar de um software livre nacional, desenvolvido pelo Laboratório de Políticas Públicas Participativas – L3P, após o lançamento do edital “Preservação e acesso aos bens do patrimônio Afro-brasileiro” financiado pelo Ministério da Cultura, que contou com a parceria com o Laboratório Liber – Tecnologia para conhecimento, da Universidade Federal de Pernambuco.

Dentre as características do Tainacan, vale destacar o fato dele ter sido construído como um espaço de convergência, permitindo uma maior interoperabilidade. Isso tornou a plataforma o ambiente ideal para atuar como instrumento da pesquisa.

Além disso, como forma de dar condições aos usuários que realizarão a indexação colaborativa social, foi desenvolvida a partir dos interesses da pesquisa uma Política de Indexação Social, no qual orienta acerca da atividade indexação (APÊNDICE A) e um Manual de Indexação, onde o intuito é direcionar o usuário-indexador dentro da plataforma selecionada (APÊNDICE B).

Como universo de pesquisa foi delimitado a área de concentração Ciência da Informação, e que para isso fora escolhido o *corpus* de Souza (2005) que pode ser verificado no Anexo A. A escolha se do *corpus* se deu inicialmente pela já utilização em outras pesquisas, dando condições de em outra oportunidade dos resultados servirem de parâmetros um para os outros.

Decorrente disso e a outros fatores, naturalmente é válido destacar o prestígio das revistas escolhidas junto ao QUALIS da CAPES, bem como a sua utilização em pesquisas acerca da indexação automática. Foram publicações das revistas **DataGramZero**, sendo um total de 29 documentos no qual se encontra disponibilizado no formato HTML, e a **Ciência da Informação**, com 31 documentos, que podem ser recuperados no formato PDF.

A pesquisa será desenvolvida a partir da aplicação com três grupos de usuários distintos, tendo todos eles relação direta à área da Ciência da Informação. Importante destacar ainda que em os grupos podem representar o perfil de indexadores iniciantes por terem passado pelas disciplinas voltadas à análise e tratamento da Informação. Assim, os grupos definidos foram:

- 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

Com a escolha dos grupos, eles serão subdivididos de forma que a partir da quantidade de usuários divididos pelos 60 artigos que compõe o *corpus*, cada um fique responsável por realizar a indexação de uma quantidade X de artigos, sendo obrigatória a inserção de 5 termos por artigo. Essa ação se repetirá para cada grupo, totalizando um total de 15 termos inseridos, desde que não exista coincidência nos termos escolhidos.

### 3.3 ETAPAS DA PESQUISA

A presente pesquisa no intuito de alcançar o objetivo traçado foi separada em etapas para melhor visualização dos procedimentos a serem realizados. Para isso, tomou-se como base a fundamentação teórico-metodológica descrita no capítulo anterior, assim o percurso metodológico segue a seguinte estrutura:

**Quadro 3 – Planejamento das etapas com relação aos objetivos da pesquisa**

<b>Objetivo A</b>	Construção da folksonomia assistida por meio da indexação social assistida do corpus de artigos de periódicos em Ciência da Informação
<b>Etapas</b>	<ol style="list-style-type: none"> <li>1. Levantar e analisar a literatura a cerca da Folksonomia assistida e a construção de corpus de referência para avaliação da indexação automática;</li> <li>2. Configurar e preparar 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</li> </ol>

	artigos de periódicos em Ciência da Informação de Souza (2005) com respectivos metadados descritivos;
<b>Objetivo B</b>	Desenvolver uma política de indexação social assistida que adota um tesauro da área de Ciência da Informação, e realizar experimento de indexação social assistida de artigos científicos do corpus de Souza (2005) por um grupo de indexadores in locus na plataforma Tainacan;
<b>Etapas</b>	<p>3. Desenvolver uma política de indexação social assistida adotando o Tesauro brasileiro em Ciência da Informação no processo de indexação colaborativa assistida, a ser registrado no campo de metadado <i>Tags</i> da coleção;</p> <p>4. Instruir os usuários indexadores e guiá-los no processo de indexação social assistida.</p>
<b>Objetivo C</b>	Analisar a folksonomia assistida resultante do processo de indexação social assistida do corpus de Souza (2005) por um grupo de indexadores in locus na plataforma Tainacan.
<b>Etapas</b>	5. Analisar o resultado do processo de indexação colaborativa realizada pelos grupos de usuários, mediante comparação daquilo que consta disponibilizado no campo de metadado “assuntos” a partir das métricas de avaliação da qualidade da indexação.
<b>Objetivo D</b>	Compilar o corpus de referência de artigos científicos na área de Ciência da Informação para avaliação de sistemas de indexação automática.
<b>Etapas</b>	6. Definir e aplicar regras para compilação do corpus de referência na área de Ciência da Informação para avaliação de sistemas de indexação automática.

Fonte: Elaborada pelo autor.

Como puderam ser observadas, foram definidas um total de 6 (seis) etapas, divididas de acordo com o objetivos específicos estabelecidas para a pesquisa. O quadro estrutural apresentado acima tem como propósito deixar evidente o caminho

para a realização do estudo de caso estabelecido para a pesquisa. As seções subsequentes da pesquisa irão apresentar detalhadamente cada etapa descrita, fazendo a devida relação com o objetivo a qual está atrelado.

### **Etapa 1 - Levantar e analisar a literatura a cerca da Folksonomia assistida e a construção de corpus de referência para avaliação da indexação automática**

Como primeira etapa do presente estudo, os principais aspectos trabalhados têm o intuito de identificar, descrever e averiguar modelos colaborativos de indexação sociais já propostos na literatura da Ciência da Informação, no intento de que as definições possam direcionar de forma adequada o caminho a ser seguido para promoção de um modelo colaborativo de indexação social assistida. Além disso, nessa etapa é possível visualizar a pesquisa bibliográfica fundamentada no segundo capítulo desta Dissertação, sendo ela responsável por dar embasamento teórico para estudo que se segue.

O desenvolvimento do capítulo 2 foi realizado a partir da busca por produções científicas acerca dos seguintes tópicos: indexação social, com ênfase na folksonomia; modelos colaborativos de indexação social e possíveis política de indexação social existentes; folksonomia assistida; representação da informação, usando tesouros como ferramenta; sistemas de indexação automática e formas de avaliação; e construção de corpus referência.

A busca por obras referente às temáticas relatadas acima foram realizadas em fontes impressas, como livros, glossários e periódicos, e em meio digital, como bases de dados de artigos de periódicos da área de Ciência da Informação dentre as quais podem ser destacados: Biblioteca Digital de Teses e Dissertações (BDTD), Repositórios Institucionais (RIs), Portal de Periódicos CAPES, periódicos eletrônicos, buscadores online, entre outros.

Como estratégia de busca foram definidos que os termos a serem utilizados seriam: Folksonomia Assistida, *Assistid Folksonomy*, Folksonomia, *Folksonomy* e *Folksonomie*. O recorte de tempo para a junção de referencias em torno da temática se deu de abril de 2017 a agosto de 2018, período esse condizente a formação e aplicação do Referencial Teórico e Metodologia da presente pesquisa.

### **Etapa 2 - Configurar e preparar uma coleção no Tainacan, levando 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 inserção na coleção dos artigos do corpus de artigos de periódicos em Ciência da Informação de Souza (2005) com respectivos metadados descritivos**

Um dos fatores que contribuíram para a escolha do Tainacan está na facilidade de seu manuseio. Dentre as possibilidades, a ferramenta oferece dois modos de uso, são elas:

- 1) **Instalação da plataforma** onde os requisitos mínimos de instalação do WordPress: servidor baseado em UNIX/Linux; PHP, versão 5.2.4 e o MySQL, versão 5.0 ter memória de 64MB para o PHP e Pacote PHP commons. Caso a intenção seja instalar recursos extras, torna-se necessário o mínimo de 256MB de memória para o PHP, a instalação do Apache ou Nginx, ter o módulo mod\_rewrite do Apache ativo e habilitar algumas extensões do PHP (UFG, 2017). Todas as opções de software são livres.
- 2) **Versão web** no qual é possível utilizar a própria base do Tainacan através do site: <http://afro.culturadigital.br/>. Essa opção é de fácil acesso, dispensando a necessidade de profissionais especializados, visto que suas configurações, básicas ou avançadas, podem ser realizadas pela própria interface. A partir da inserção do *Corpus* como coleção na plataforma, é definida um endereço próprio para acesso.

Em ambas as opções de criação de coleção a ferramenta oferece vários níveis de personalização. Para a pesquisa será utilizado o segundo modo de utilização da ferramenta, a **versão web**.

Como consequência e no plano de entender melhor o funcionamento da plataforma, se faz necessário conhecer os termos mais utilizados para desenvolver as ações de colaboração no Tainacan:

- **Coleção:** Conjunto de itens agrupados por uma determinada temática ou finalidade e organizados segundo algum esquema de metadados e sistema de organização.
- **Etiqueta:** É um termo digitado livremente pelo usuário para a descrição de um item sem um controle ou normalização.
- **Categoria:** É um termo que faz parte de um vocabulário controlado utilizado na descrição de um item. Pode estar ordenada em uma lista de termos ordenada alfabeticamente ou pode ser estruturado hierarquicamente, que

também é chamado de taxonomia. Nesse caso, os termos superiores da hierarquia da taxonomia são mais genéricos ou mais amplos e os subtermos são mais específicos ou de partes de um todo.

- **Item:** Um item é formado pelo objeto de conteúdo original e pelos metadados de descrição. Um item pode ser formado apenas pelos metadados de descrição apontando para o objeto de conteúdo original em outro repositório. Os metadados de descrição do item servem para representá-lo na coleção, para organizá-los e para recuperá-los. O padrão de metadados mais difundido na web é o Dublin Core, que define 15 elementos de descrição.

- **Metadado:** É um atributo ou relacionamento de um item ou objeto.

Entendido o funcionamento da plataforma Tainacan, é preciso tomar conhecimento que para a sequência da pesquisa foi diagnosticado a necessidade de ser realizada uma adequação no uso da plataforma, tudo isso no intuito de atender a proposta do protótipo apresentada por Silva (2013).

Para isso, antes de dar sequência ao relato acerca da adequação, é importante tomar ciência do modelo proposto por Silva (2013), sendo ele, o escolhido para desenvolvimento da pesquisa que se segue.

Em sua tese, Silva (2013) busca a criação de um protótipo de modelo de colaboração para catálogo web facetado, e para isso apresenta uma figura que ele chama de “Etapas do desenvolvimento do protótipo”, e que pode ser visto a seguir:

#### **Quadro 4 - Etapas do desenvolvimento do protótipo**

<b>Desenvolvimento do protótipo</b>
<p><b>Concepção</b></p> <ul style="list-style-type: none"> <li>• Objetivos e escopo do projeto</li> <li>• Atores</li> </ul>
<p><b>Elaboração</b></p> <ul style="list-style-type: none"> <li>• Requisitos funcionais</li> <li>• Requisitos não funcionais</li> <li>• Diagramas UML</li> <li>• Escolha da ferramenta</li> <li>• Customização do módulo</li> <li>• Modelo de dados</li> <li>• Modelagem de taxonomia facetada</li> </ul>

- Prototipação de baixa fidelidade da interface facetada

### **Construção**

- Criação do ambiente
- Módulos instalados
- Implementação da interface de busca
- Implementação do modelo de colaboração

Fonte: Silva (2013, p. 132)

Do quadro disponibilizado pelo autor, é importante reforçar que cada etapa corresponde a um item que precisa ser levado em consideração no anseio de aplicação do modelo de indexação colaborativa.

Como plano geral do que acontecerá na aplicação dessa pesquisa tendo como comparativo o que foi proposto por Silva (2013) em sua tese abaixo será apresentado um quadro correspondente ao que o autor especifica como sendo a etapa de elaboração, no item requisitos funcionais para a contribuição no catálogo (para todos os membros). A necessidade de apresentar esse item é auxiliar no devido entendimento das adequações realizadas do modelo de Silva (2013).

#### **Quadro 5 – Comparativo dos requisitos funcionais para a contribuição no catálogo (para todos os membros)**

Silva (2013)	Atual
<ul style="list-style-type: none"> <li>• Adicionar um registro, utilizando os termos das facetadas, através da etiquetagem livre ou ambas as estratégias para classificação;</li> <li>• Contestar a indexação de um registro. Permite ao usuário revisar a indexação do mesmo adicionando novas categorias das facetadas, ou informar eventual classificação erroneamente aplicada (deve enviar um email para o administrador do catálogo). Assim, o usuário pode contribuir com o catálogo,</li> </ul>	<ul style="list-style-type: none"> <li>• Adicionar um registro, utilizando os termos das facetadas, através da etiquetagem livre ou ambas as estratégias para classificação;</li> <li>• Não haverá contestação da indexação de um registro. 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.</li> </ul> <p>Obs.: O primeiro registro inserido</p>

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.

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.

- A estrutura hierárquica levada em consideração para a inserção de um registro foi o Tesouro Brasileiro de Ciência da Informação, e que deve ser acessado através do site <http://www.vocabularyserver.com/tbci/>.

Obs.: Nesse ponto a plataforma Tainacan não deu condições para que o ato de inserção do registro todo o processo foi exclusivamente nela. 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 apresentado acima fosse a alternativa de apoio necessária para a execução da ação por parte do usuário, isso por dar a condição do mesmo usufruir da condição de “autocomplete”.

Os fatos mais importantes a serem mencionados a partir da análise do Quadro 4 são referente a dois pontos:

1. A condição de apoio proposta enquanto folksonomia assistida se dará através de um ambiente externo, ao invés de ser na própria plataforma;
2. A validação dos registros, *tags*, inseridas pelos usuários-indexadores será realizada de forma manual, item esse que será descrito na sequência dentro do item que se refere aos papéis, atividades e regras de interação no Tainacan.

No que corresponde aos papéis, atividades e regras de interação, é válido destacar que nesse ponto a aplicação do Modelo de Indexação Colaborativa sugerido por Silva (2015) define especificamente como sendo o item *Atores*. Assim, o autor define as categorias de usuários e suas respectivas atribuições e que para esta pesquisa ficou definido que nem todas as modalidades de atores seriam contempladas, como pode ser visto no quadro a seguir:

**Quadro 6 - Atores do sistema**

Silva (2013, p. 136)	Atual
<ul style="list-style-type: none"> <li>• <b>Super administrador:</b> usuário responsável pela gestão das configurações do sistema;</li> <li>• <b>Administrador:</b> usuário responsável pela criação, configuração do catálogo, taxonomia facetada e aprovação de contribuições dos membros;</li> <li>• <b>Membro:</b> usuário cadastrado e logado no sistema;</li> <li>• <b>Visitante:</b> usuário que não está cadastrado ou não está logado no sistema. Também conhecido com usuário anônimo.</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Super administrador:</b> usuário responsável pela gestão das configurações do sistema e usuário responsável pela criação, configuração do catálogo, taxonomia facetada e aprovação de contribuições dos membros;</li> <li>• <b>Membro:</b> usuário cadastrado e logado no sistema;</li> <li>• <b>Visitante:</b> usuário que não está cadastrado ou não está logado no sistema. Também conhecido com usuário anônimo.</li> </ul>

Fonte: Elaborado pelo autor.

É possível observar que para a pesquisa foi excluído o ator **administrador**, fazendo com que o cargo do **super administrador** agregue as duas funções.

Quanto às duas outras modalidades de *atores*, **Membros e Visitantes**, mantiveram para si as mesmas atribuições. Nesse ponto é válido reforçar que a pesquisa tem como finalidade focar na Folksonomia Assistida gerada a partir dos registros produzidos pelos Membros, ou como chamado até aqui, usuário-indexador.

O Tainacan, plataforma que dará condições de aplicação do modelo, permite acesso a todos que tem acesso à internet. No primeiro momento para ser considerado membro, todo visitante deve efetuar seu registro como apresentado na Figura 4.

**Figura 4 – Cadastro de membros no Tainacan**



Fonte: Tainacan (2018, *online*)

No intuito de efetuar qualquer intervenção é indispensável que o visitante, em posse do seu login (registro) entre no sistema. Dessa forma, o visitante deixa essa condição e passa a ser considerado membro.

Fora definido que o *corpus* para validação da metodologia a ser aplicada nessa pesquisa seria o mesmo já utilizado na Tese de Souza (2005), sendo também utilizado em outras pesquisas. Cabe informar que o *Corpus*, como Souza (2005) evidencia, consta com um total de 60 documentos textuais, sendo todos eles da língua portuguesa e publicados entre os anos 2002 e 2003. Para conhecimento, os documentos foram referenciados seguindo o local no qual foram publicados, o título, o resumo e a autoria, assim como consta no anexo da dissertação de Souza (2005).

O *Corpus* foi construído com publicações das revistas **DataGramZero** (B3), sendo um total de 29 documentos, e **Ciência da Informação** (B1), com 31 documentos. Vale destacar o reconhecimento que ambas revistas possuem dentro da área da Ciência da Informação segundo o programa *Qualis*<sup>1</sup> da CAPES.

<sup>1</sup> *Qualis* é o conjunto de procedimentos utilizados pela Capes para estratificação da qualidade da produção intelectual dos programas de pós-graduação. (Ministério de Educação, 2018, *online*)

Partindo da escolha do *Corpus* a ser utilizado, ele será disponibilizado como coleção de Ciência da Informação na plataforma Tainacan.

**Etapa 3: Desenvolver uma política de indexação social assistida adotando o Tesouro brasileiro em Ciência da Informação no processo de indexação colaborativa assistida, a ser registrado no campo de metadado *Tags* da coleção:**

Dentre as opções de personalização do Tainacan, uma delas é a inserção de vocabulários controlados, no que é chamado categorias. Para a presente dissertação, essa opção seria útil para auxiliar o usuário na definição dos termos descritores dos documentos disponibilizados.

Segundo o manual do Tainacan na área de criação de categorias ela “possibilita a criação de categorias de forma intuitiva, bastando dar um clique duplo e escrever o nome da categoria desejada”. Para isso, existe duas formas de inserção: a manual e a importação de taxonomia:

- 1) Manual: Neste item existem algumas opções como a de adicionar uma subcategoria, além de ser possível manipular a ordem das categorias apenas clicando no nome criado e acionando algumas das funções desse item.
- 2) Importação de Taxonomia: Este item possibilita a importação de uma taxonomia em formato xml.

**Figura 5 – Campos categoria da coleção**



Fonte: Wiki do Tainacan, 2018, *online*

Na figura 2 está enumerado de forma que apresenta:

- 1) As categorias e subcategorias;
- 2) Opções de inserção para categorias e subcategorias; e
- 3) Importar Taxonomia no formato XML.

Entretanto, devido a dificuldades técnicas não foi possível a tempo fazer a inserção dos termos do tesouro como categorias no Tainacan, lembrando que o TBCI possui cerca de 1.800 termos.

Como solução e alternativa para garantir o andamento da pesquisa, os usuários-indexadores devem buscar no site <http://www.vocabularyserver.com/tbci/> os termos pertinentes no processo de representação de cada item da coleção.

A escolha do site ocorreu após a verificação de existência da funcionalidade “autocomplete” que ele oferece, bem como a garantia de que todos os termos do TBCI constavam no site.

A importância dessa funcionalidade vai de encontro ao aspecto determinante na construção de uma folksonomia assistida, e que foi discutido ao longo do referencial teórico. É ela a ideia de dar sugestões ao usuário-indexador de termos que já constem na base para representar o conteúdo que está indexando.

Como instrumento de auxílio para a execução da escolha dos termos que deverão ser escolhidos para descrever os artigos que compõe o *corpus* e disponibilizados através de uma coleção no Tainacan, foi desenvolvido uma breve política de indexação social no intuito nortear as ações do usuário-indexador.

No ato da aplicação da metodologia aos grupos escolhidos, a primeira ação a ser realizada será a apresentação da Política de Indexação, antes mesmo das instruções de manuseio do Tainacan. A partir desse ponto eles terão condições de iniciar as ações.

#### **Etapa 4: Instruir os usuários indexadores e guiá-los no processo de indexação social assistida:**

Criada a coleção, e tendo em vista sua devida personalização, a etapa que se segue é a utilização da mesma por parte dos usuários-indexador. Aqui, é importante destacar que o pré-requisito definido para a escolha do grupo de usuários é a de possuir afinidade com a área de construção do *Corpus*, sendo então a Ciência da Informação e suas respectivas disciplinas. Nesta etapa serão apresentadas aos usuários as devidas orientações de utilização da plataforma Tainacan tendo em vista a realização da indexação colaborativa.

Para tanto, os usuários precisarão seguir algumas etapas, para possuir acesso à plataforma, e conseqüentemente a Coleção e realizar a indexação. São elas:

- 1) Acessar a página da coleção *Corpus de referência em Ciência da Informação*<sup>2</sup>: <http://afro.culturadigital.br/colecao/corpus-de-referencia-em-ci/>;
- 2) Efetuar cadastro;
- 3) Acessar através de login e senha cadastrada;
- 4) Encontrar o item a ser indexado através do filtro Identificação no menu à esquerda, informando o identificador composto por dois dígitos (de 01 a 60).
- 5) Acessar a página do item;
- 6) Na página do item preencher o campo de metadado “tags” de acordo com a política de indexação social assistida.

As informações acerca da aplicação do Modelo de Indexação Social, bem como a de manuseio do Tainacan será realizada através da explicação do Manual de Indexação Social, que consta como Apêndice B.

**Etapa 5: Analisar o resultado do processo de indexação colaborativa realizada pelos grupos de usuários, mediante comparação daquilo que consta disponibilizado no campo de metadado “assuntos” a partir das métricas de avaliação da qualidade da indexação.:**

Pretendendo condensar os aspectos coletados a partir da utilização por parte dos usuários selecionados, num primeiro instante se faz necessário regressar aos fundamentos teóricos dos modelos colaborativos de indexação evidenciados ao longo da pesquisa. Com a realização da verificação e discussão dos dados coletados a partir dos conceitos e métodos dos modelos colaborativos de indexação social existentes na literatura da Ciência da Informação e aqui discutido.

Para tanto, é válido reforçar que foram identificados três trabalhos sobre modelos colaborativos de indexação social na literatura da Ciência da Informação e que suas propostas servem para embasar a análise do que fora coletado.

Será realizado o confronto entre o que foi produzido pelo autor-indexador e pelo usuário-indexador, levando em conta as discussões sobre os modelos colaborativos de indexação sociais já existentes na literatura da Ciência da Informação de Santarem Segundo (2010), Silva (2013) e Santos (2015).

---

<sup>2</sup> Nome da Coleção criada no Tainacan para desenvolvimento da pesquisa.

Em termos objetivos, será realizada uma comparação das palavras-chaves dos autores que constam no corpus de Souza (2005) com as palavras-chave atribuídas pelos usuários-indexadores a partir da utilização do vocabulário controlado oriundo do TBCI, verificando também a ocorrência das mesmas ou expressões equivalentes no texto dos documentos.

Para isso os resultados da indexação feitos pelos usuários serão analisados considerando o cenário do experimento, avaliando os índices de consistência, precisão, revocação e medida F, formulas essas apresentadas na **seção 2.7.1** que discorre sobre avaliação de Sistemas de Indexação Automática. Isso permitirá reconhecer informações importantes para o uso da indexação automática, principalmente no que se refere à avaliação de Sistemas de Recuperação da Informação para bases de dados científicas.

Outro ponto a ser discutido, é o papel da moderação no que diz respeito à aprovação dos termos escolhidos pelos usuários-indexadores. Isso por que existe a possibilidade de muitos usuários relatarem a não existência de termos que correspondam ao conteúdo a ser indexado, bem como a escolha inadequada, com erros possíveis erros de grafia ou que não constam na base do TBCI.

#### **Etapa 6: Definir e aplicar regras para compilação do corpus de referência na área de Ciência da Informação para avaliação de sistemas de indexação automática:**

Por fim, a proposta da pesquisa é ser capaz de determinar se a folksonomia assistida pode ser aplicada na construção de corpus de referência para avaliação de sistemas de indexação automática. Como resultado para a presente pesquisa é ressaltada a importância de diretrizes que visem estabelecer parâmetros que possam validar a qualidade da indexação.

É importante frisar nesse ponto, a necessidade de ser capaz de determinar critérios que visem identificar, por exemplo, quais os descritores podem ou não ser atribuídos a um conteúdo informacional. Nesse ponto, é válido destacar que a possibilidade de alinhar o grau de entendimento do que o autor tem sobre a própria obra e o que o leitor, que irá indexar, tem sobre essa mesma obra passa a ser um bom indício a ser levado em consideração no que se refere à qualidade da indexação.

Dentre os modelos de indexação social pensadas, uma condição interessante de execução é a de permitir a correção da indexação feita por um usuário. Isso garante a construção coletiva na perspectiva de um consenso entre o que os indexadores-usuários pensam a respeito do conteúdo.

Sendo prático, a pesquisa aqui desenvolvida apresentará em seus resultados a Folksonomia Assistida gerada pelos Grupos de usuários-indexadores fazendo relações com a indexação do próprio autor, evidenciando o que fora dito anteriormente acerca do nível de entendimento sobre o mesmo documento dos dois grupos (autor e leitor).

Além disso, a proposta nessa etapa perpassa pela busca de um dos principais intuitos no processo de indexação, à correção na indexação ou garantia de ausência de erros. Dentre todo o processo de contextualização da pesquisa, foi relatada a preocupação com a quantidade de informação gerada em ambientes colaborativos, principalmente em decorrência dos problemas que são gerados pela liberdade no processo de interação.

A Folksonomia Assistida, busca garantir, dentro de seu contexto de aplicação, a resolução das questões que acompanham, por exemplo, a folksonomia. Soergel (1994) argumenta que esta qualidade tem uma importância primordial para o desempenho da recuperação, pois a indexação é suscetível a dois tipos de falhas:

- por omissão (quando um termo que deve ser atribuído é omitido);
- por inclusão (adição de um termo que não deve ser atribuído).

Para tanto, a indexação é avaliada como bem-sucedida quando permite a quem realiza as buscas localizar itens de que precisa sem ter de examinar muitos de que não precisa (LANCASTER, 2004).

Por fim, o intuito é permitir, através de análises, que a aplicação de um modelo que tem como resultado a Folksonomia Assistida seja uma alternativa que garanta, dentre as possibilidades, uma forma de recuperação da informação. Além disso, para esta pesquisa, o importante nesta etapa vem de encontro ao estabelecimento de regras que permitam definir quais termos do TBCI serão usados como descritores para cada artigo do corpus. Como outro aspecto importante, se tem a ideia de como o corpus será disponibilizado na forma de arquivos. Tais pontos darão condições de no ato de compilação do corpus de referência se tenha condições de criar um mecanismo de avaliação de sistemas de indexação automática.

## 4 ANÁLISE DOS RESULTADOS

Neste capítulo será apresentada a análise dos resultados envolvendo o experimento realizado. Para isso, foram definidos subcapítulos, que abaixo serão descritos conforme seu enquadramento nas etapas descritas na metodologia, no intuito de expor detalhadamente os relatos e inferências dos dados coletados.

Na seção 4.1 são manifestados aspectos inerentes à análise, bem como a discussão acerca do modelo colaborativo de indexação social assistida escolhido para a aplicação da metodologia, principalmente levando-se em consideração o que fora descrito no subtópico 2.3 desta dissertação.

No que se refere à seção 4.2 será relatado o desenvolvimento da configuração da coleção, apontando às vantagens e desvantagens da adoção do modelo colaborativo de indexação social assistida com o intuito de melhorar a qualidade da RI e, conseqüentemente, da recuperação dos documentos nessa base de dados.

A seção 4.3 discorre sobre a Política de Indexação Social Assistida, serão discutidos os pontos observados no desenvolvimento da Política juntamente aos grupos escolhidos para aplicação da pesquisa.

Posteriormente, na seção 4.4 será abordado a Folksonomia Assistida gerada a partir dos objetivos traçados para a dissertação, tendo como propósito apresentar o resultado obtido.

Nas seções 4.5 e 4.6 tem como foco apresentar a compilação do corpus de referência a partir da Folksonomia Assistida revisada de forma a envolver a avaliação da qualidade da indexação por meio dos índices de consistência, revocação, precisão e medida F.

É importante relatar que todos os aspectos abordados nesse capítulo são embasados e descritos a partir do que se entendeu do referencial teórico desenvolvido nesta pesquisa.

### 4.1 MODELO COLABORATIVO DE INDEXAÇÃO SOCIAL ASSISTIDA

Os resultados a serem aqui explorados giram em torno do papel desempenhado por cada ator do sistema, e que foram apresentados anteriormente através do Quadro 4. Vale lembrar que, do modelo desenvolvido por Silva (2013)

em relação ao que fora aplicado nesta pesquisa, os atores e funções passaram por uma adequação de acordo com as condições encontradas para aplicação do modelo.

Dessa forma, inicialmente as questões que podem ser abordadas na construção de um Modelo Colaborativo de Indexação Social Assistida, ou no caso dessa pesquisa, a aplicação de um modelo adaptado, se refere ao papel do super administrador. Nesse ponto, 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 trás 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.

Uma dificuldade identificada na aplicação do Modelo pode ser visto 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 explicação de como proceder na pesquisa, à escolha de um termo alternativo, capaz de se aproximar da representação que se desejava inicialmente.

Vale ressaltar que a finalidade da ferramenta autocomplete disponibilizada através do site <http://www.vocabularyserver.com/tbci/>, principalmente para casos como esses descritos, possibilita o 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. Como observação a ser relatada, dar condições de os usuários realizarem ações de moderação, ainda que com a possibilidade de volta da ação, pode acarretar certas dificuldades, como a exclusão de termos equivocadamente.

A proposta, levando as questões levantadas, é que o Modelo de colaboração defina adequadamente quem desempenhará o papel de administrador, até para que se consiga manter uma coordenação adaptável e extensível.

É importante atentar para o fato de que a moderação a ser realizada estará sempre intimamente ligada ao fluxo de trabalho. O administrador, a partir desse fluxo terá condições de avaliar de forma a configurar como o comportamento colaborativo para cada perfil de usuário deverá ocorrer, garantindo a sua participação da melhor forma possível conforme seu propósito e a comunidade de usuários indexadores.

No que se refere ao papel do membro (usuário-indexador), que seguindo a proposta é o usuário cadastrado e que deve estar logado no sistema, o aspecto mais importante que deve ser levado em consideração diz respeito até que ponto ele tem liberdade para desempenhar a função a qual o Modelo propõe. Essa função seria a de indexar conteúdos em espaços que comumente não contariam com essa possibilidade, garantindo uma nova possibilidade de RI e conseqüentemente a recuperação.

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 conseqüência de forma mais lenta. Como efeito desse aspecto, o que pode vir a ocorrer é 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, bem como menos 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 que com a indexação sendo feita livremente, precisam 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, 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. Para por fim, como uma das principais questões atreladas a aplicação de um modelo colaborativo, está o fato de a pesquisa baseada em *tags* proporcionarem uma busca semântica de documentos.

## 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 facilitada e voltada para usuários independente de 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, ao recurso de auto completar 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, onde

dentre eles, está à dificuldade em ajustar as configurações tendo como base a definição das atribuições de cada grupo de atores, bem como em alguns momentos a instabilidade em momentos de grande fluxo de usuários logados. Para tanto, inicialmente será relatado como ocorreu a configuração da coleção em toda suas etapas.

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 que a nortearão, e que podem ser vista na figura a seguir:

**Figura 6 – Permissões para Coleção Corpus de Referência em CI no Tainacan**

Permissões - Escolha as permissões para cada uma das ações abaixo			
Entidade	Criar	Editar	Deletar
Categoria	Membros	Aprovação	Aprovação
Classificação	Membros		Aprovação
Item	Não permitido		Aprovação
Comentários	Membros	Aprovação	Aprovação
Tags	Membros	Aprovação	Aprovação
Metadado de dado	Não permitido	Aprovação	Aprovação
Valor do Metadado de Dado		Aprovação	
Metadado de objeto	Não permitido	Aprovação	Aprovação
Valor do Metadado de Objeto		Aprovação	
Metadado de termo	Não permitido	Aprovação	Aprovação

Fonte: Dados da Pesquisa (2018)

Como permissão, foi definida 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 permitiu que diante da aplicação aos 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, e que funcionaria como o 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 todo termo da hierarquia, uma por uma, e a segunda através de importação da taxonomia. Em ambos os casos é possível visualizar na figura abaixo:

**Figura 7 – Inserção de categorias para Coleção Corpus de Referência em CI no Tainacan**

The screenshot displays a web interface for managing categories in Tainacan. At the top, there is a text input field containing 'Assunto de Ciência da Informação'. To its right is a blue button labeled '+Adicionar sub-categoria', followed by four small icons: an upward arrow, a downward arrow, a leftward arrow, and a rightward arrow. Further right is a button labeled 'Importar Taxonomia'. Below these elements is a large, empty rectangular area with the text 'Clique aqui para criar uma subcategoria'. At the bottom left of the interface is a button labeled 'Voltar para a coleção', and at the bottom right is a green button labeled 'Salvar e Continuar'.

Fonte: Dados da Pesquisa (2018)

Definido que o TBCI seria a base hierárquica a ser inserida nas categorias, a forma de importação descrita pelo Tainacan seria por meio de arquivo no formato TXT. 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 mais uma opção.

Diante da problemática apresentada, à saída foi à busca de uma alternativa para o andamento das etapas da pesquisa, de forma a não comprometer os objetivos. Tal ação resultou no uso site <http://www.vocabularyserver.com/tbci/> já descrito na seção anterior.

Na sequência, a próxima etapa da configuração da Coleção passa pela definição de quais são os metadados que irão constar nos itens. Aproveitando-se disso, a pesquisa definiu como metadados os campos Dublin Core, acrescentando apenas o metadado *tags*. Assim sendo, os metadados e regras de entrada de dados foram:

- **Autor** – (DC.creator) preencha com o nome de cada autor, informe do primeiro ao último nome, com nomes do meio abreviados pela inicial seguida de ponto.
- **Título** – (DC.title) preencha com o título em português
- **Descrição** – (DC.description) informe o resumo em português
- **Linguagem** – (DC.language) preencha com pt-BR

- **Fonte** – (DC.source) informe a referência bibliográfica (use a BRAPCI para obter isto mais facilmente, desprezando a informação de “Disponível em...”. Normalize o nome dos autores)
- **Relação** – (DC. Relation) deixar vazio
- **Cobertura** – (DC. Coverage) deixar vazio
- **Contribuição** – (DC. Contributor) deixar vazio
- **Licença** – (DC.Rights)deixar vazio
- **Publicador** – (DC.Publisher) informe o nome da revista (Ex.: Ciência da Informação, Datagramazero)
- **Data** – (DC.data) informe o ano de publicação do artigo
- **Formato** – (DC.format) preencha com pdf ou htm
- **Identificação** – (DC.identifier) preencha com o número do artigo
- **Tipo** – (DC.type) preencha com: Artigo
- **Tags** – (DC.subject) entre com cinco termos autorizados do TBCI, adicionando novo campo para cada assunto, priorize termos mais específicos que apareçam no título, resumo ou palavras-chave

Como forma de garantir um nível de usabilidade por parte do administrador, o Tainacan permite a definição de 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.

**Figura 8 – Itens da Coleção Corpus de Referência em CI no Tainacan**

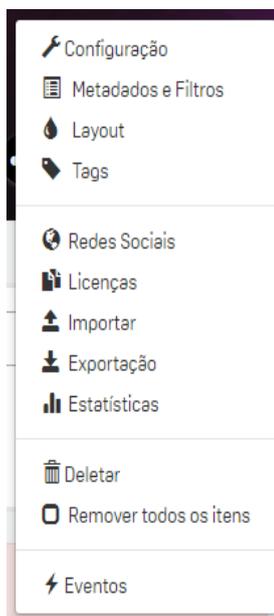
	<b>Artigo 01 – Transferência da Informação: análise para valoração de unidades de conhecimento</b>	Criado por: Bruno Silva Criado em: 20/10/2017	 
	<b>Artigo 02 – Popularização do Conhecimento Científico</b>	Criado por: Bruno Silva Criado em: 20/10/2017	 
	<b>Artigo 03 – O Valor da Informação: um desafio permanente</b>	Criado por: Bruno Silva Criado em: 20/10/2017	 
	<b>Artigo 04 – Auto-arquivamento: uma opção inovadora para a produção científica</b>	Criado por: Bruno Silva Criado em: 20/10/2017	 
	<b>Artigo 05 – Análise Contrastiva: memória da construção de uma metodologia para investigar a tradução do conhecimento científico em</b>	Criado por: Bruno Silva Criado em: 20/10/2017	 
	<b>Artigo 06 – O Tesouro Eletrônico do Mundo do Trabalho: produto de um esforço interdisciplinar</b>	Criado por: Bruno Silva Criado em: 20/10/2017	 
	<b>Artigo 07 – Inteligência Competitiva em Organizações: dado, informação e conhecimento</b>	Criado por: Bruno Silva Criado em: 20/10/2017	 
	<b>Artigo 08 – A conceituação de massa documental e o ciclo de interação entre tecnologia e o registro do conhecimento</b>	Criado por: Bruno Silva Criado em: 20/10/2017	 
	<b>Artigo 09 – A conceituação de massa documental e o ciclo de interação entre tecnologia e o registro do conhecimento</b>	Criado por: Bruno Silva Criado em: 20/10/2017	 
	<b>Artigo 10 – Implicações da “nova economia” para a mensuração estatística: desajustes conceituais e metodológicos</b>	Criado por: Bruno Silva Criado em: 20/10/2017	 
	<b>Artigo 11 – Por uma nova Ciência da Informação: ensino, pesquisa e formação</b>	Criado por: Bruno Silva Criado em: 20/10/2017	 
	<b>Artigo 12 – Ensino e pesquisa em ciência da informação</b>	Criado por: Bruno Silva Criado em: 20/10/2017	 

« | »    Página 1 de 5    « | »    Exibindo itens: 1 - 12 de 60    Itens por página: 12 ▾

Fonte: Dados da Pesquisa (2018)

Com todos os artigos inseridos, o usuário tem a sua disposição a possibilidade de interagir como toda ferramenta Web 2.0, incluindo fazer buscas personalizadas, dentre elas as *Tags* que foram definidas pelos usuários. Finalizado todas as etapas de configuração da Coleção, o Tainacan permite que todas elas sejam revistas caso necessárias (FIGURA 8), incluindo o espaço para aprovação dos itens inseridos (FIGURA 9).

**Figura 9 – Configuração do Corpus de Referência em CI no Tainacan**



Fonte: Dados da Pesquisa (2018)

**Figura 10 – Página de aprovação das ações do Corpus de Referência em CI no Tainacan**

Eventos não verificados(21)		Eventos verificados(33)		
Mostrar <input type="text" value="10"/> itens por página		Pesquisar: <input type="text"/>		
Data	Usuário	Tipo do Evento	Descrição do Evento	Estado
13/11/2017		Deletar Categoria	Deletar a Categoria[ Assunto de Corpus de referência ] na coleção Corpus de referência	Não verificado
13/11/2017		Remover o metadado de dado	Remover o metadado de dado [ Tags ] na coleção Corpus de referência	Não verificado
13/11/2017		Remover o metadado de dado	Remover o metadado de dado [ Tags ] na coleção Corpus de referência	Não verificado
13/11/2017		Deletar Classificação	Remover a tag : Ciência da informação no item Artigo 06	Não verificado
13/11/2017		Deletar Classificação	Remover a tag : Ciência da informação no item Artigo 06	Não verificado
13/11/2017		Deletar Classificação	Remover a tag : Ciência da informação no item Artigo 06	Não verificado
13/11/2017		Deletar Classificação	Remover a tag : Ciência da informação no item Artigo 06	Não verificado
13/11/2017		Deletar Classificação	Remover a tag : Ciência da informação no item Artigo 06	Não verificado
13/11/2017		Deletar Classificação	Remover a tag : Ciência da informação no item Artigo 06	Não verificado
13/11/2017		Deletar Classificação	Remover a tag : Ciência da informação no item Artigo 06	Não verificado

Exibindo de 1 até 10 de 21 itens

Anterior **1** 2 3 Próxima

Fonte: Dados da Pesquisa (2018)

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 as atividades de forma homogênea, foi criada como instrumento a política de indexação social assistida. Por se tratar de um Modelo Colaborativo, onde se tem características híbridas (junção de taxonomia e folksonomia), observaram-se diversos aspectos que reiteram a necessidade de a aplicação de uma política se fazer necessária 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 comum. Isso é reforçado quando observado que algumas das recomendações feitas na política não são seguidas, mostrando assim a necessidade de direcionamento das ações a serem executadas.

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

Assim, para que de fato a Indexação caminhasse como previsto, 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 dos conceitos válidos para a recuperação.

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 o artigo, o usuário executasse a simples ação de copiar do site <http://www.vocabularyserver.com/tbci/> o termo 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 constante. 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 co-gestão estabelecida na participação dos grupos de atores (administrador e membros) em torno do modelo de indexação colaborativa social.

#### 4.4 ANÁLISE DA FOLKSONOMIA ASSISTIDA GERADA

No decorrer do capítulo 2, especificamente na seção 2.2, foi abordada a construção da Folksonomia Assistida, de forma a contribuir para um maior entendimento das possibilidades e implicações existentes em torno do termo.

Dentro do processo de desenvolvimento da Folksonomia Assistida, a funcionalidade que caracteriza o seu surgimento gira em torno do apoio ao usuário no exato momento de inserção de uma *tag* que vise descrever o 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”.

A Folksonomia Assistida resultante da aplicação junto aos grupos de usuários pode ser vista no **APÊNDICE D** onde constam as *tags* utilizadas, da exata forma que foi gerada, e sua recorrência, incluindo os problemas no uso de caracteres e letras maiúsculas e minúsculas indevidamente.

Diante da Folksonomia Assistida gerada, foram contabilizadas 261 *tags* diferentes, totalizando 583 registros totais. As *tags* mais utilizadas pelos usuários foram:

**Tabela 1 – 6 (seis) *tags* mais recorrentes na Folksonomia Assistida gerada**

<b><i>Tags</i></b>	<b>Recorrência</b>
ciência da informação	29
Sociedade da Informação	21
Gestão da informação.	14
tecnologias da informação e comunicação	13
acesso a informação	12
gestão do conhecimento	12

Fonte: da Pesquisa (2018)

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 ou minúsculas. Em um único caso, foi identificada uma *tag* que fora utilizada por um usuário-indexador para representar o Artigo 45 e que não constava como termo autorizado no TBCI, mesmo com toda orientação de como proceder em sua inserção.

Como Folksonomia Assistida gerada, o termo “Métodos quantitativos de avaliação” foi mantido a título de indicação de sua existência, contudo, no processo de revisão, que cabe ao Administrador, ele foi retirado.

De imediato, tendo como base a **Tabela 6**, é possível visualizar já dentre as *tags* mais utilizadas, problema de inserção indevida. O termo “Gestão da informação.”, por exemplo, apresenta uma letra maiúscula e “.”, que em ambas as situações são formas inválidas levando-se em consideração o que consta no TBCI, pois a forma correta a ser inserida deveria ter sido “gestão da informação”.

Os **Quadros 7 e 8**, apresentam separadamente os problemas identificados na Folksonomia Assistida gerada, sendo a primeira com o erro de uso de caracteres indevidamente e a segunda com os erros no uso de letras maiúsculas e minúsculas.

**Quadro 7** – 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: Dados da Pesquisa (2018)

**Quadro 8** – 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 à 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 científica	produtividade científica
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

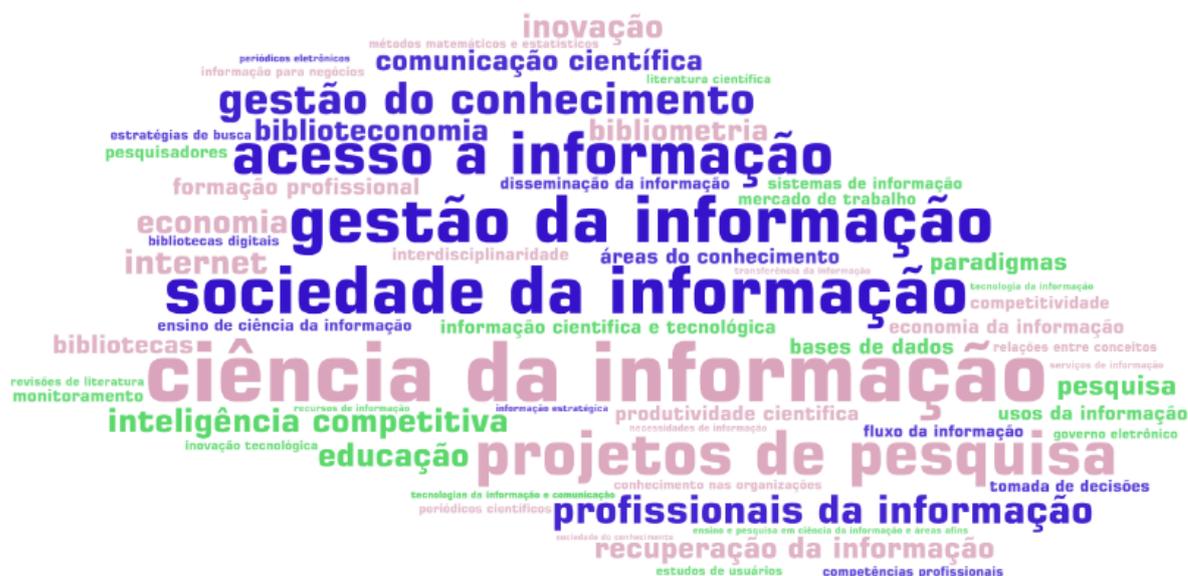
Fonte: Dados da Pesquisa (2018)

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 o 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 quando se levado em consideração que diante de suas características, que são oriundas da Folksonomia, os pontos negativos foram reduzidos. Isso garantiu que o resultado possa ser utilizado para Avaliar Sistemas de Indexação Automática, proposta da presente pesquisa.

Na **figura 11** é possível visualizar a nuvem de *tags* gerada como resultado, onde os termos mais usados se encontram em maior evidência. Vale salientar que a nuvem de *tag* carrega consigo a possibilidade de evidenciar um cenário, e que por essa razão, para essa pesquisa, palavras como “Ciência da Informação”, “Sociedade da Informação”, entre outras, são tão recorrentes.

Figura 11 – Nuvem de tags do Corpus



Fonte: Dados da Pesquisa (2018)

Dado o cenário encontrado, foi constatado que o resultado da Folksonomia Assistida gerada apresenta-se como alternativa ao uso das palavras-chave dos autores na avaliação de sistemas de indexação automática, agregando novas possibilidades no processo de recuperação dos artigos, contudo tendo uma devida atenção aplicada aos problemas de inserção relatados anteriormente.

A primeira vista, não existe uma determinação por parte das revistas analisadas quanto à quantidade “obrigatória” de palavras-chave a serem utilizadas nos documentos podendo ser causado por uma falha na elaboração das normas editoriais do periódico, como consequência surge uma grande variação de um artigo para outro, como pode ser observada na **Tabela 2**.

**Tabela 2** – Quantitativo de palavras-chave dos documentos do *Corpus*

Identificação	Nº. de palavras-chave	Identificação	Nº. de palavras-chave
Artigo 01	3	Artigo 31	5
Artigo 02	2	Artigo 32	6
Artigo 03	5	Artigo 33	5
Artigo 04	5	Artigo 34	3
Artigo 05	5	Artigo 35	5
Artigo 06	7	Artigo 36	4
Artigo 07	5	Artigo 37	2
Artigo 08	5	Artigo 38	3
Artigo 09	5	Artigo 39	5
Artigo 10	5	Artigo 40	2
Artigo 11	4	Artigo 41	4
Artigo 12	6	Artigo 42	5
Artigo 13	3	Artigo 43	4
Artigo 14	6	Artigo 44	3
Artigo 15	2	Artigo 45	5
Artigo 16	5	Artigo 46	6
Artigo 17	3	Artigo 47	3
Artigo 18	4	Artigo 48	3
Artigo 19	5	Artigo 49	3
Artigo 20	4	Artigo 50	2
Artigo 21	5	Artigo 51	5
Artigo 22	9	Artigo 52	5
Artigo 23	3	Artigo 53	5
Artigo 24	6	Artigo 54	7
Artigo 25	2	Artigo 55	4
Artigo 26	4	Artigo 56	5
Artigo 27	3	Artigo 57	7
Artigo 28	7	Artigo 58	7
Artigo 29	3	Artigo 59	5
Artigo 30	3	Artigo 60	4

Fonte: Dados da Pesquisa (2018)

Referente à Folksonomia Assistida gerada como resultado, por ter sido aplicado a três grupos de usuários poderiam conter até 15 *tags* por documento, podendo variar para menos quando houvesse coincidência dentre as *tags* escolhidas. Por esse 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 3** a seguir.

**Tabela 3** - Quantitativo de *tags* dos documentos do Corpus

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

Fonte: Dados da Pesquisa (2018)

É importante frisar que cabe ao Administrador o papel de revisar todas as questões e problemas levantados até então, gerando assim o que aqui será chamado de Folksonomia Assistida Revisada. A proposta é dar condições de estabelecer os devidos parâmetros para avaliação da qualidade de indexação a partir da compilação do *Corpus de Referência* sejam executados da melhor forma.

#### 4.5 ANÁLISE DA FOLKSONOMIA ASSISTIDA REVISADA

Partindo do ponto que foi gerada uma Folksonomia Assistida, a sua revisão se faz necessária. Nesse ponto, passado pelo processo de revisão partiu-se para estabelecer o que se chamou de “Termos Comuns”.

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 termos alternou de um mínimo de 0 (zero), em apenas um caso onde nenhum termo coincidiu, a um máximo de 6 (seis) termos comuns entre eles, para dois casos, os **Artigo 12** e **58**. A média de termos comum variou entre 2 e 3, tendendo ao maior.

**Tabela 4** - Quantitativo de termos em comuns das palavras-chave dos autores e *tags* dos usuários-indexadores

Identificação	Nº. de palavras-chave	Identificação	Nº. De palavras-chave
Artigo 01	2	Artigo 31	3
Artigo 02	2	Artigo 32	4
Artigo 03	1	Artigo 33	3
Artigo 04	2	Artigo 34	2
Artigo 05	2	Artigo 35	5
Artigo 06	5	Artigo 36	3
Artigo 07	5	Artigo 37	2
Artigo 08	2	Artigo 38	2
Artigo 09	3	Artigo 39	4
Artigo 10	2	Artigo 40	2
Artigo 11	3	Artigo 41	3
Artigo 12	6	Artigo 42	3
Artigo 13	2	Artigo 43	3
Artigo 14	3	Artigo 44	3
Artigo 15	3	Artigo 45	4
Artigo 16	4	Artigo 46	4
Artigo 17	4	Artigo 47	2
Artigo 18	2	Artigo 48	2
Artigo 19	4	Artigo 49	3
Artigo 20	4	Artigo 50	2
Artigo 21	1	Artigo 51	2
Artigo 22	5	Artigo 52	3
Artigo 23	0	Artigo 53	4
Artigo 24	1	Artigo 54	3

Artigo 25	2	Artigo 55	3
Artigo 26	2	Artigo 56	3
Artigo 27	1	Artigo 57	2
Artigo 28	3	Artigo 58	6
Artigo 29	3	Artigo 59	3
Artigo 30	2	Artigo 60	4

Fonte: Dados da Pesquisa (2018)

Através dos dados coletados, foi utilizado as formulas que geram índices de consistência, revocação, precisão e medida F, bem como descritas na seção 2.7.1 e que podem ser vistas no APÊNDICE E. O intuito disso foi a de realizar uma análise da Folksonomia Assistida, quando aplicada, para determinar o grau de qualidade dos sistemas indexação.

Partindo desse ponto, a avaliação foi realizado levando em consideração os termos presentes no campo palavras-chave de cada documento que compõe o Corpus de Referência, e que teve identificado um mínimo de 2 e um máximo de 9 termos atribuídos para indexação, gerando uma média de 4 a 5 descritores atribuídos por documento.

Ao realizar de imediato uma avaliação generalista dos resultados alcançados por meio da coleta, é possível afirmar, quando confrontado as palavras-chave com as *tags*, que a grande quantidade de termos do segundo item influenciou diretamente nos cálculos dos índices de coeficiente de consistência e, os índices de precisão, revocação e medida F. Os valores médios de cada índice podem ser vistos conforme ilustrado na **Tabela 5**:

**Tabela 5** - Índices médios de coeficiente de consistência, precisão, revocação e medida F entre as palavras-chave dos autores e tags dos usuários-indexadores

<b>Coeficiente de consistência</b>	28%
<b>Precisão</b>	32%
<b>Revocação</b>	69%
<b>Medida F</b>	37%

Fonte: Dados da Pesquisa (2018)

Bandim e Corrêa (2018, p. 72) buscam categorizar os valores alcançados para o índice de consistência no intuito de definir o que é satisfatório para sistemas de indexação automática. Assim, de acordo com a fórmula do índice de consistência os autores definem as seguintes faixas de valores para o índice de consistência médio correspondentes aos níveis de desempenho:

- 0 a 11% - desempenho insatisfatório (corresponde a uma média de 0 a 1 termo em comum);
- 12 a 25% - desempenho satisfatório (corresponde a uma média de 1 a 2 termos em comum);
- 26 a 43% - desempenho bom (corresponde a uma média de 2 a 3 termos em comum);
- 44 a 67% - desempenho ótimo (corresponde a uma média de 3 a 4 termos em comum);
- 68 a 100% - desempenho excelente (corresponde a uma média de 4 a 5 termos em comum). Para cálculo dos limites dos intervalos 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 é bom. Na sequência, a aplicação das formulas geraram um índice de precisão de apenas 32%, número muito abaixo do que foi descrito por Hlava (2002) que seria por volta de 60% para ser considerado bom, logo, proporcionalmente um número de revocação maior do que o esperado, 45%.

Como a medida F é a média harmônica entre precisão e revocação, entende-se que quanto mais perto de 100%, maior foi a qualidade da indexação, e quando tende a 0% menos qualificado. Seguindo essa lógica, os 36% se apresentam como um percentual conflituoso de harmonia, podendo demonstrar que o entendimento do autor se apresenta falho ao ser comparado ao nível de entendimento do usuário acerca do mesmo documento quando usado os termos do TBCI como descritores. Como descrito por Bandim (2017, p. 69) para se conseguir valores médios altos para F é necessário que se tenha valores altos para revocação e precisão.

Diante do cenário apresentado, alguns pontos precisam ser destacados individualmente no que se refere à relação do número de palavras-chave pelo número de termos da folksonomia assistida. Quando confrontado, esses dois números apresentam um resultado, aqui denominado como número de termos em comum. Assim, nas tabelas a seguir é possível visualizar aqueles que apresentam o melhor (Tabela 5) e o pior (Tabela 6) caso dentre os 60 artigos.

**Tabela 6 – Os 6 (seis) melhores casos de termos em comum entre as Palavras-chave e Folksonomia Assistida**

ID	Nº. DE PALAVRAS-CHAVE	Nº. DE TERMOS DA FOLKSONOMIA ASSISTIDA	Nº. DE TERMOS EM COMUM	COEFICIENTE DE CONSISTÊNCIA	P	R	MEDIDA F
Artigo 06	7	10	5	42%	50%	71%	59%
Artigo 07	5	10	5	50%	50%	100%	67%
Artigo 12	6	10	6	60%	60%	100%	75%
Artigo 22	9	8	5	42%	63%	56%	59%
Artigo 35	5	6	5	83%	83%	100%	91%
Artigo 58	7	7	6	75%	86%	86%	86%

Fonte: Dados da Pesquisa (2018)

**Tabela 7 – Os 5 (seis) piores casos de termos em comum entre as Palavras-chave e Folksonomia Assistida**

ID	Nº. DE PALAVRAS-CHAVE	Nº. DE TERMOS DA FOLKSONOMIA ASSISTIDA	Nº. DE TERMOS EM COMUM	COEFICIENTE DE CONSISTÊNCIA	P	R	MEDIDA F
Artigo 03	5	11	1	7%	9%	20%	13%
Artigo 21	5	6	1	10%	17%	20%	18%
Artigo 23	3	9	0	0%	0%	0%	0%
Artigo 24	6	8	1	8%	13%	17%	14%
Artigo 27	3	12	1	7%	8%	33%	13%

Fonte: Dados da Pesquisa (2018)

Diante dos dados expostos, é interessante observar que o fato de possuir maior número de termos da folksonomia assistida, não necessariamente garante que vá existir mais termos em comum, sendo os casos mais emblemáticos nesse ponto os **Artigo 58** e o **Artigo 23**. Esse cenário pode demonstrar que quanto maior o número de indivíduos indexando, maior são as possibilidades, principalmente quando levado em consideração todas as questões cognitivas existentes no desenvolvimento da atividade de indexar.

No **Artigo 58**, é possível notar que a quantidade de palavras-chave é idêntica, e ainda sim o resultado de termos em comum entre eles é de quase 100%, já no caso do **Artigo 23**, mesmo tendo um número de 9 *tags*, o número de termos em comum é 0. Nesse caso em específico, isso pode ocorrer por falta de consenso do que o autor compreende sobre a publicação, e o que de fato os leitores compreenderam, causando várias interpretações.

Nesse ponto é importante acrescentar a discussão o fato de que muitas das palavras-chave utilizadas pelos autores não constam no TBCI, resultando numa falta de controle terminológico e conseqüentemente, 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 dado, dando maior respaldo para melhores resultados.

O fato de o índice de precisão estar diretamente relacionada à quantidade de palavras-chave encontradas dentre as *tags* apresentam um cenário um pouco diferente do apresentado acima, principalmente dentre os melhores casos.

**Tabela 8 – Os 6 (seis) melhores casos de índice de precisão (P)**

ID	Nº. DE PALAVRAS-CHAVE	Nº. DE TERMOS DA FOLKSONOMIA ASSISTIDA	Nº. DE TERMOS EM COMUM	COEFICIENTE DE CONSISTÊNCIA	P	R	MEDIDA F
Artigo 12	6	10	6	60%	60%	100%	75%
Artigo 22	9	8	5	42%	63%	56%	59%
Artigo 35	5	6	5	83%	83%	100%	91%
Artigo 46	6	6	4	50%	67%	67%	67%
Artigo 53	5	6	4	57%	67%	80%	73%
Artigo 58	7	7	6	75%	86%	86%	86%

Fonte: Dados da Pesquisa (2018)

Dessa forma, ao fazer uma breve comparação das **Tabelas 6 e 7**, é possível ver que de fato a Precisão é determinada pela quantidade de termos que coincidem entre as indexações. Assim, os **Artigos 58 e 35**, passam a listar como tendo os melhores índices de precisão, tendo os dois 86% e 83% respectivamente.

Por fim, como proposta de alcançar uma média harmônica, a medida F encontrada apresentou o cenário a ser visto na **Tabela 9**.

**Tabela 9** – Os 6 (seis) melhores casos do índice de Medida F

ID	Nº. DE PALAVRAS-CHAVE	Nº. DE TERMOS DA FOLKSONOMIA ASSISTIDA	Nº. DE TERMOS EM COMUM	COEFICIENTE DE CONSISTÊNCIA	P	R	MEDIDA F
<b>Artigo 07</b>	5	10	5	50%	50%	100%	67%
<b>Artigo 12</b>	6	10	6	60%	60%	100%	75%
<b>Artigo 35</b>	5	6	5	83%	83%	100%	91%
<b>Artigo 46</b>	6	6	4	50%	67%	67%	67%
<b>Artigo 53</b>	5	6	4	57%	67%	80%	73%
<b>Artigo 58</b>	7	7	6	75%	86%	86%	86%

Fonte: Dados da Pesquisa (2018)

De acordo com a tabela acima, o destaque fica por conta do **Artigo 07** que até então não aparecia, mas que sua média entre precisão e revocação resultou num índice satisfatório quando se partindo da premissa que quanto mais perto de 100% maior foi à qualidade da indexação. No caso do artigo, o valor foi de 67% e no que se refere a um cenário mais que satisfatório, o **Artigo 35** apresentou um índice de Medida F de 91%, representando uma qualidade e equilíbrio na indexação realizada tanto pelo autor, quanto pelos usuários.

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

Num outro momento, 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 a conclusão de que a utilização da indexação automática com base no Tesouro 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 boa quando se leva em consideração as palavras-chaves e *tags* com base no TBCI.

Bandim (2017) ainda afirma que sempre será importante o olhar do profissional indexador no processo de entender e decidir qual o melhor termo ou termos que representarão determinado objeto informacional, fazendo com que seja necessário considerar tanto a indexação automática quanto a manual.

Na busca de entender o índice de revocação na indexação automática por sintagmas nominais de artigos de Periódicos em ciência da informação, Celerino e Corrêa (2017) utilizam os títulos e resumos do mesmo *corpus*. Os autores concluem 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.

Celerino (2018) relata que através da aplicação das regras desenvolvidas no método aplicado para alcance dos seus resultados, juntamente ao fato de considerar as palavras-chave do autor, se torna conveniente sua utilização para um sistema automático de normalização dos SNs. A forma positiva coincide com o fato das regras estabelecidas conseguiram desenvolver novos SNs descritores para os documentos, ampliando assim as possibilidades de recuperação através da Indexação automática.

Celerino e Corrêa (2017) descrevem que a maior adoção pela comunidade científica das normas existentes para elaboração de resumo e atribuição de termos advindos de tesouros especializados, possam minimizar tal limitação no futuro quanto a uniformidade quanto a expressividade e coerência da descrição e representação dos trabalhos.

Por fim, é importante observar que em ambos os trabalhos, de alguma forma a necessidade de normalização trazem resultados positivos, precisando assim ser investido tempo e estudo.

#### 4.6 COMPILAÇÃO DO CORPUS DE REFERÊNCIA

Tendo em posse as palavras-chave geradas pelos autores e a Folksonomia Assistida resultante da inserção de *tags* nos documentos da *Coleção Corpus de Referência em CI* é possível apresentar a compilação do *Corpus* por meio da descrição das duas listas de palavras. Para isso algumas regras de correção e exclusão foram levadas em consideração para que de fato se chegasse ao resultado final. Foram elas:

1. Uso indevido de caracteres;
2. Uso indevido de letras maiúsculas e minúsculas;
3. Não ser um termo autorizado no TBCI.

No **APÊNDICE F** é possível visualizar a Folksonomia Assistida revisada de grupo de usuários no qual foi o método aplicado. Nesse ponto é importante destacar que os termos que de alguma forma não seguiram as regras acima apresentadas, por essa razão foram desconsiderados no processo final de compilação. Os **Quadros 7 e 8** desta pesquisa, contemplam os itens 1 e 2 listados acima, na sequência, o **Quadro 9**, apresenta o resultado do item 3, os termos descartados por não serem autorizados no TBCI.

**Quadro 9 – Termos descartados da Folksonomia Assistida**

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

Fonte: Dados da Pesquisa (2018)

A partir do processo de exclusão dos termos não verificados no TBCI, a etapa subsequente contemplou a junção de todos termos em comum gerados pelos grupos, formando um quadro geral de Folksonomia Assistida e que pode ser vista no quadro a seguir:

### Quadro 10 – Folksonomia Assistida resultante da junção dos grupos

ARTIGO FOLKSONOMIA ASSISTIDA REVISADA	
<b>Artigo 01</b>	acesso a informação    ciência da informação    competitividade    conhecimento nas organizações    gestão da informação    gestão de documentos    gestão do conhecimento    métodos de pesquisa e análise    repositórios digitais    transferência da informação
<b>Artigo 02</b>	acesso a informação    áreas do conhecimento    ciência da informação    comunicação científica    divulgação científica    educação    gestão da informação    notícias    popularização da ciência    sociedade da informação
<b>Artigo 03</b>	acesso a informação    análise de informação na indústria    atributos da informação    biblioteconomia    ciência da informação    direito à informação    ética na informação    história    liberdade de pensamento    recuperação da informação    sociedade da informação
<b>Artigo 04</b>	acesso livre    autoarquivamento de documentos    comunicação científica    direito autoral    literatura científica    OAI-PMH    OAIS    publicações científicas: periódicos    repositórios digitais
<b>Artigo 05</b>	acesso a informação    áreas do conhecimento    comunicação científica    comunidades científicas    dados científicos    divulgação científica    estudantes    gestão da informação    globalização da informação    informação científica e tecnológica    sociedade da informação
<b>Artigo 06</b>	ciência da computação    ciência da informação    elaboração de linguagens documentárias    gerenciamento    interação homem-computador    interdisciplinaridade    manutenção de tesouros    mercado de trabalho    recuperação da informação    sistemas de informação
<b>Artigo 07</b>	ciência da informação    economia da informação    fluxo da informação    gerenciamento    gestão da informação    gestão do conhecimento    informação para negócios    inteligência competitiva    organização do conhecimento    transferência da informação
<b>Artigo 08</b>	áreas do conhecimento    ciência da informação    conceitos de informação    conteúdos da informação    disseminação da informação    explosão da informação    gestão de documentos    gestão do conhecimento    modelagem do conhecimento    registros bibliográficos    sociedade da informação    tecnologias da informação e comunicação
<b>Artigo 09</b>	ciência da informação    comunicação nas organizações    cultura organizacional    fluxo da informação    gestão da informação    informação estratégica    inteligência competitiva    obsolescência da literatura    tomada de decisões    universidades
<b>Artigo 10</b>	atributos da informação    comunicação    conceitos    economia    inteligência competitiva    métodos matemáticos e estatísticos    probabilidade e estatística    sociedade da informação    sociedade do conhecimento    tecnologias da informação e comunicação
<b>Artigo 11</b>	áreas do conhecimento    armazenamento de dados    ciência da informação    educação    ensino de ciência da informação    gestão da informação    interdisciplinaridade    propriedade intelectual    recuperação da informação    unidades de informação
<b>Artigo 12</b>	acesso a informação    arquivologia    biblioteconomia    ciência da informação    educação    ensino de ciência da informação    pesquisa    profissionais de informação    sistemas de informação    teoria da informação
<b>Artigo 13</b>	ciência da informação    competências profissionais    conhecimento nas organizações    disseminação da informação    economia    formação profissional    gestão do conhecimento nas organizações    habilidades profissionais    mediadores da informação    profissionais de informação    sociedade da informação

<b>Artigo 14</b>	competências profissionais    fluxo da informação    formação profissional    gestão da informação    gestão do conhecimento    gestores    mediadores da informação    mercado de trabalho    profissionais de informação    sociedade da informação
<b>Artigo 15</b>	ciência da informação    competências profissionais    educação superior    ensino de ciência da informação    ensino e pesquisa em ciência da informação e áreas afins    formação profissional    instituições de ensino e pesquisa    paradigmas    profissionais de informação    sociedade do conhecimento
<b>Artigo 16</b>	biblioteconomia    ciência da informação    educação em ciência da informação    educação superior    ensino de biblioteconomia    ensino e pesquisa em ciência da informação e áreas afins    formação profissional    profissionais de informação    sociedade da informação
<b>Artigo 17</b>	armazenamento de dados    comunicação científica    controle bibliográfico    dados    gestão do conhecimento    inovação tecnológica    internet    produtividade científica    propriedade intelectual    segurança da informação    sociedade da informação    tecnologias da informação e comunicação
<b>Artigo 18</b>	categorias    ciência da informação    construtivismo    desenvolvimento de acervos    economia da informação    escolas e correntes filosóficas    planejamento    produtividade científica    relações entre conceitos    sociedade da informação    sociedade do conhecimento    sociedades científicas    taxonomias    tecnologias da informação
<b>Artigo 19</b>	ambiente organizacional    centros de informação (bibliotecas especializadas)    competitividade    conhecimento organizacional    gestão da informação    gestão do conhecimento    inteligência competitiva    inteligência empresarial    monitoramento ambiental    relevância    tecnologias da informação e comunicação
<b>Artigo 20</b>	avaliação    avaliação de publicações científicas    ciência da informação    dissertações e teses    ensino de ciência da informação    pesquisa    pesquisa em ciência da informação    profissionais de informação    programas de pós-graduação    programas de pós-graduação em ciência da informação    projetos de pesquisa
<b>Artigo 21</b>	ciência cognitiva    inteligência artificial    linguística    multidisciplinaridade    promoção da leitura    sociedade da informação
<b>Artigo 22</b>	ciência da informação    complexidade    hipertextos    organização da informação    paradigmas    processamento de informações    tecnologia da informação    tecnologias da informação e comunicação
<b>Artigo 23</b>	ciência da informação    disseminação da informação    documentação    interdisciplinaridade    recuperação da informação    relações entre conceitos    sociedade da informação    teoria da informação    tipos de documento
<b>Artigo 24</b>	ciência da informação    comunicação científica    informação científica e tecnológica    institutos de pesquisa    paradigmas    pesquisadores    políticas públicas    transdisciplinaridade
<b>Artigo 25</b>	agências de fomento    agências de fomento à pesquisa    bolsas de pesquisa    ciência da informação    frente de pesquisa    gestão de recursos humanos    levantamentos    pesquisa e desenvolvimento    pesquisadores    programas de pós-graduação em ciência da informação    projetos de pesquisa
<b>Artigo 26</b>	compressão de dados    consistência na indexação    fluxo da informação    informação estratégica    medidas de avaliação de sistemas de recuperação da informação    monitoramento    palavras-chaves    precisão    processamento de textos    relevância    revocação    semântica
<b>Artigo 27</b>	agências de fomento    análise comparativa    análise de dados    bases de dados    bases de dados cadastrais    bolsas de pesquisa    documentos primários    extração de informação em bases de dados    indicadores de C&T    pesquisadores    produtividade científica    projetos de pesquisa

<b>Artigo 28</b>	bibliotecas digitais    comércio eletrônico    fontes primárias    gestão da informação    governo eletrônico    inovação    marketing da informação    relações entre conceitos    serviços de informação    sistemas de informação    tecnologias da informação e comunicação
<b>Artigo 29</b>	acesso a informação    gestão da informação    gestão de conteúdos na web    governo eletrônico    informação governamental    políticas de informação    serviços de informação
<b>Artigo 30</b>	análise de dados    coleta de dados    comunicação e acesso à informação    estudos de usuários    perfil do usuário    periódicos científicos    periódicos eletrônicos    processos de gestão    publicações de acesso livre    registros de uso    ruído na recuperação da informação    webmetria    world wide web
<b>Artigo 31</b>	economia    economia da informação    governo eletrônico    infraestrutura de informação    internet    políticas de informação    regimes de informação    relações entre conceitos    sociedade da informação
<b>Artigo 32</b>	base de dados    controle de vocabulário    estratégias de busca    linguagens documentárias    processamento da linguagem natural    recuperação da informação    revisões de literatura    vocabulários controlados
<b>Artigo 33</b>	artigos de periódico    bibliotecas digitais    bibliotecas virtuais    citações bibliográficas    informação científica e tecnológica    inovação    produtividade científica    produtividade de periódicos    revisões de literatura    sociedade da informação    tecnologias da informação e comunicação
<b>Artigo 34</b>	acesso a informação    ciência e tecnologia de alimentos    comunicação científica    fluxo da informação    informação científica e tecnológica    informação tecnológica    inovação    inovação tecnológica    interação universidade-empresa    pesquisa e desenvolvimento    tecnologias da informação e comunicação    transferência da informação
<b>Artigo 35</b>	HTML    internet    intranetes    linguagens de marcação    SGML (uf standard generalized markup language)    XML
<b>Artigo 36</b>	bibliometria    ciência da informação    disseminação da informação    economia    formação profissional    lei de Lotka    literatura científica    métodos matemáticos e estatísticos    produtividade de autor    produtividade de periódicos    profissionais de informação
<b>Artigo 37</b>	acesso a informação    bases de dados    bases de dados factuais    informação financeira    informação jurídica    informação para negócios    recursos de informação na internet    sociedade da informação    tecnologias da informação    tomada de decisões
<b>Artigo 38</b>	acesso a informação    bibliotecas digitais    bibliotecas híbridas    educação a distância    ensino a distância    necessidades de informação    recursos de informação    serviços de biblioteca    serviços de informação    suportes de informação    tecnologias da informação e comunicação    usos da informação
<b>Artigo 39</b>	artigos de revisão    bases de dados    estratégias de busca    interdisciplinaridade    mecanismos de busca    recuperação da informação    revisões de literatura    sistemas de recuperação da informação    sociedade da informação
<b>Artigo 40</b>	ciência da informação    explosão da informação    formação profissional    gestão da informação    gestão do conhecimento    mercado de trabalho    necessidades de informação    profissionais de informação    recursos de informação    sociedade da informação    tecnologias da informação e comunicação
<b>Artigo 41</b>	análise estatística da literatura científica    ciência da informação    comunidades científicas    história da ciência da informação    literatura científica    literatura cinzenta    produtividade científica    programas de pós-graduação    representação do conhecimento    sociedade da informação    tipos de documento    usos da informação

<b>Artigo 42</b>	bibliotecas    consultores de informação    gestão do conhecimento    gestão financeira    informação para negócios    inteligência competitiva    profissionais da informação    serviços bibliotecários    sistemas de informação    unidades de informação
<b>Artigo 43</b>	análise de informação na indústria    conhecimento nas organizações    informação científica e tecnológica    informação tecnológica    inovação    inovação tecnológica    modelos de bases de dados    necessidades de informação    produtividade científica    tomada de decisões
<b>Artigo 44</b>	ambiente organizacional    competitividade    conhecimento nas organizações    ensino e pesquisa em ciência da informação e áreas afins    gestão do conhecimento    informação estratégica    inovação    métodos de análise na inteligência competitiva    processos de gestão    tecnologias da informação e comunicação    tomada de decisões
<b>Artigo 45</b>	análise quantitativa    bibliometria    cibermetria    cienciometria    cientometria    informetria    métodos matemáticos e estatísticos    webmetria
<b>Artigo 46</b>	bibliometria    ciência da informação    lei de Bradford    periódicos    teoria do caos
<b>Artigo 47</b>	buscas em linha    ciência da informação    economia da informação    estudos de usuários    indexação automática    periódicos científicos    periódicos eletrônicos    usabilidade    usos da informação    World Wide Web
<b>Artigo 48</b>	custos    economia da informação    empresas    gestão da informação    informação para negócios    inovação    inteligência competitiva    tecnologias da informação e comunicação    usos da informação
<b>Artigo 49</b>	automação de bibliotecas    bibliometria    biblioteconomia    ciência da informação    descritores    gestão do conhecimento    inteligência competitiva    modelos cognitivos    monitoramento    periódicos científicos    periódicos eletrônicos    representação do conhecimento
<b>Artigo 50</b>	bibliotecários    biblioteconomia    educação    educação em biblioteconomia    ensino de biblioteconomia    formação profissional    gestão da informação    gestão do conhecimento    profissionais de informação    sociedade da informação
<b>Artigo 51</b>	acessibilidade    acesso a informação    bibliotecas    ciberespaço    comunicação mediada por computador    perfil do usuário    sistemas de informação    tecnologias da informação e comunicação
<b>Artigo 52</b>	ciência da informação    disseminação da informação    estudos de usuários    hipertextos    internet    usuários
<b>Artigo 53</b>	comportamento do usuário    estratégias de busca    linguagens documentárias    mecanismos de busca    recuperação da informação    serviços de referência
<b>Artigo 54</b>	bibliotecários    bibliotecas    competência em informação    educação    gestão do conhecimento    recursos de informação    sociedade da informação    usos da informação
<b>Artigo 55</b>	competências profissionais    ensino de ciência da informação    formação profissional    inteligência competitiva    mercado de trabalho    profissionais de informação
<b>Artigo 56</b>	biblioteconomia    ciência da informação    cientistas da informação    comunicação científica    história da ciência da informação    informação científica e tecnológica    interdisciplinaridade    sociologia    teorias na ciência da informação

<b>Artigo 57</b>	áreas do conhecimento    ciência da informação    ensino e pesquisa em ciência da informação e áreas afins    epistemologia da ciência da informação    paradigmas    pesquisa    recuperação da informação    regimes de informação    sociedade da informação
<b>Artigo 58</b>	categorização automática de textos    ciência cognitiva    ciência da informação    indexação    interação homem-computador    processamento de informações    recuperação da informação
<b>Artigo 59</b>	análise de citação    bibliometria    ciência da informação    cientometria    comunicação científica    educação    epistemologia    epistemologia arqueológica    produtividade científica
<b>Artigo 60</b>	acesso a informação    agentes inteligentes    automação    gestão da informação    inovação    inteligência competitiva    internet    monitoramento    usos da informação

Fonte: Dados da Pesquisa (2018)

A compilação de todos os termos gerados a partir do quadro apresentado anteriormente resultou na representação do que seria o *Corpus de Referência de CI* para os usuários. O **Quadro 11**, disponibilizado seguindo a ordem de recorrência dos termos, se torna representativo partindo da premissa de que no ato da busca na base de dados, a possibilidade de uso de um desses termos seria maior do que os termos utilizados pelos autores para descrever os próprios artigos.

**Quadro 11 – Compilação da Folksonomia Assistida Revisada**

Coluna 1	Coluna 2	Coluna 3
ciência da informação	processos de gestão	fontes primárias
sociedade da informação	produtividade de periódicos	frente de pesquisa
gestão da informação	programas de pós-graduação	gestão de conteúdos na web
tecnologias da informação e comunicação	programas de pós-graduação em ciência da informação	gestão de recursos humanos
acesso a informação	propriedade intelectual	gestão do conhecimento nas organizações
gestão do conhecimento	regimes de informação	gestão financeira
profissionais da informação	relevância	gestores
inteligência competitiva	repositórios digitais	globalização da informação
recuperação da informação	representação do conhecimento	habilidades profissionais
comunicação científica	teoria da informação	história
formação profissional	tipos de documento	HTML
inovação	unidades de informação	indexação
produtividade científica	webmetria	indexação automática
biblioteconomia	World Wide Web	indicadores de C&T
educação	acessibilidade	informação financeira
informação científica e tecnológica	agências de fomento à pesquisa	informação governamental
usos da informação	agentes inteligentes	informação jurídica
áreas do conhecimento	análise comparativa	informetria

bibliometria	análise de citação	infraestrutura de informação
disseminação da informação	análise estatística da literatura científica	instituições de ensino e pesquisa
economia da informação	análise quantitativa	institutos de pesquisa
ensino de ciência da informação	arquivologia	inteligência artificial
fluxo da informação	artigos de periódico	inteligência empresarial
interdisciplinaridade	artigos de revisão	interação universidade-empresa
internet	autoarquivamento de documentos	intranetes
sistemas de informação	automação	lei de bradford
bases de dados	automação de bibliotecas	lei de lotka
competências profissionais	avaliação	levantamentos
conhecimento nas organizações	avaliação de publicações científicas	liberdade de pensamento
economia	bases de dados cadastrais	linguagens de marcação
ensino e pesquisa em ciência da informação e áreas afins	bases de dados factuais	linguística
informação para negócios	bibliotecas híbridas	literatura cinzenta
mercado de trabalho	bibliotecas virtuais	manutenção de tesouros
métodos matemáticos e estatísticos	buscas em linha	marketing da informação
paradigmas	categorias	medidas de avaliação de sistemas de recuperação da informação
relações entre conceitos	categorização automática de textos	métodos de análise na inteligência competitiva
tomada de decisões	centros de informação (bibliotecas especializadas)	métodos de pesquisa e análise
bibliotecas	ciberespaço	métodos quantitativos de avaliação
bibliotecas digitais	cibernetria	modelagem do conhecimento
competitividade	ciência da computação	modelos cognitivos
estratégias de busca	ciência e tecnologia de alimentos	modelos de bases de dados
estudos de usuários	cienciometria	monitoramento ambiental
governo eletrônico	cientistas da informação	multidisciplinaridade
informação estratégica	citações bibliográficas	notícias
inovação tecnológica	coleta de dados	OAI-PMH
literatura científica	comércio eletrônico	OAIS
monitoramento	competência em informação	obsolescência da literatura
necessidades de informação	complexidade	organização da informação
periódicos científicos	comportamento do usuário	organização do conhecimento
periódicos eletrônicos	compressão de dados	palavras-chaves
pesquisa	comunicação	periódicos
pesquisadores	comunicação e acesso à informação	pesquisa em ciência da informação.
projetos de pesquisa	comunicação mediada por computador	planejamento

recursos de informação	comunicação nas organizações	políticas públicas
revisões de literatura	conceitos	popularização da ciência
serviços de informação	conceitos de informação	precisão
sociedade do conhecimento	conhecimento organizacional	probabilidade e estatística
tecnologia da informação	consistência na indexação	processamento da linguagem natural
transferência da informação	construtivismo	processamento de textos
agências de fomento	consultores de informação	produtividade de autor
ambiente organizacional	conteúdos da informação	promoção da leitura
análise de dados	controle bibliográfico	publicações científicas: periódicos
análise de informação na indústria	controle de vocabulário	publicações de acesso livre
armazenamento de dados	cultura organizacional	recursos de informação na internet
atributos da informação	custos	registros bibliográficos
bibliotecários	dados	registros de uso
bolsas de pesquisa	dados científicos	revocação
ciência cognitiva	descritores	ruído na recuperação da informação
cientometria	desenvolvimento de acervos	segurança da informação
comunidades científicas	direito à informação	semântica
divulgação científica	direito autoral	serviços bibliotecários
educação superior	dissertações e teses	serviços de biblioteca
ensino de biblioteconomia	documentação	serviços de referência
explosão da informação	documentos primários	sgml (uf standard generalized markup language)
gerenciamento	educação á distância	sistemas de recuperação da informação
gestão de documentos	educação em biblioteconomia	sociedades científicas
hipertextos	educação em ciência da informação	sociologia
história da ciência da informação	elaboração de linguagens documentárias	suportes de informação
informação tecnológica	empresas	taxonomias
interação homem-computador	ensino a distância	teoria do caos
linguagens documentárias	epistemologia	teorias na ciência da informação
mecanismos de busca	epistemologia arqueológica	transdisciplinaridade.
mediadores da informação	epistemologia da ciência da informação	universidades
perfil do usuário	escolas e correntes filosóficas	usabilidade
pesquisa e desenvolvimento	estudantes	usuários
políticas de informação	ética na informação	vocabulários controlados
processamento de informações	extração de informação em bases de dados	XML

Fonte: Dados da Pesquisa (2018)

Para chegar a reunião dos termos mais representativos do *Corpus de Referência em CI* foi confrontado os termos gerados pelos autores por meio das palavras-chave com os termos oriundos da Folksonomia Assistida apresentada no **Quadro 10**. O resultado dessa etapa consta no quadro a seguir:

**Quadro 12 –**Termos representativos da Folksonomia Assistida Revisada

ARTIGO	FOLKSONOMIA ASSISTIDA REVISADA
<b>Artigo 01</b>	transferência da informação    gestão do conhecimento
<b>Artigo 02</b>	popularização da ciência    comunicação científica
<b>Artigo 03</b>	direito à informação
<b>Artigo 04</b>	acesso livre    autoarquivamento de documentos
<b>Artigo 05</b>	comunicação científica    divulgação científica
<b>Artigo 06</b>	recuperação da informação    interdisciplinaridade    interação homem-computador    manutenção de tesouros    sistemas de informação
<b>Artigo 07</b>	inteligência competitiva    gestão do conhecimento    gestão da informação    fluxo da informação    transferência da informação
<b>Artigo 08</b>	conceitos de informação    tecnologias da informação e comunicação
<b>Artigo 09</b>	fluxo da informação    universidades    inteligência competitiva
<b>Artigo 10</b>	economia    métodos matemáticos e estatísticos
<b>Artigo 11</b>	ciência da informação    ensino de ciência da informação    recuperação da informação
<b>Artigo 12</b>	ciência da informação    biblioteconomia    ensino de ciência da informação    sistemas de informação    arquivologia    pesquisa
<b>Artigo 13</b>	formação profissional    profissionais da informação
<b>Artigo 14</b>	profissionais da informação    gestão da informação    gestão do conhecimento
<b>Artigo 15</b>	ensino de ciência da informação    ensino e pesquisa em ciência da informação e áreas afins    formação profissional
<b>Artigo 16</b>	biblioteconomia    ciência da informação    educação superior    formação profissional
<b>Artigo 17</b>	comunicação científica    internet    produtividade científica    tecnologias da informação e comunicação
<b>Artigo 18</b>	sociedade da informação    tecnologias da informação
<b>Artigo 19</b>	inteligência empresarial    gestão da informação    gestão do conhecimento    monitoramento ambiental
<b>Artigo 20</b>	ciência da informação    pesquisa    pesquisa em ciência da informação    programas de pós-graduação em ciência da informação
<b>Artigo 21</b>	sociedade da informação
<b>Artigo 22</b>	paradigmas    ciência da informação    tecnologia da informação    hipertextos    complexidade
<b>Artigo 23</b>	-

<b>Artigo 24</b>	Pesquisadores
<b>Artigo 25</b>	agência de fomento à pesquisa    ciência da informação
<b>Artigo 26</b>	monitoramento    palavras-chave
<b>Artigo 27</b>	indicadores de C&T
<b>Artigo 28</b>	bibliotecas digitais    governo eletrônico    sistemas de informação
<b>Artigo 29</b>	governo eletrônico    políticas de informação    informação governamental
<b>Artigo 30</b>	análise de dados    periódicos eletrônicos
<b>Artigo 31</b>	políticas de informação    sociedade da informação    internet
<b>Artigo 32</b>	base de dados    estratégias de busca    linguagens documentárias    recuperação da informação
<b>Artigo 33</b>	artigo de periódico    bibliotecas digitais    bibliotecas virtuais
<b>Artigo 34</b>	informação tecnológica    transferência da informação
<b>Artigo 35</b>	HTML    internet    intranetes    linguagens de marcação    XML
<b>Artigo 36</b>	bibliometria    lei de Lokta    produtividade de autor
<b>Artigo 37</b>	informação para negócios    bases de dados
<b>Artigo 38</b>	bibliotecas híbridas    serviços de informação
<b>Artigo 39</b>	estratégias de buscas    recuperação da informação    bases de dados    artigos de revisão
<b>Artigo 40</b>	ciência da informação    gestão da informação
<b>Artigo 41</b>	literatura científica    literatura cinzenta    ciência da informação
<b>Artigo 42</b>	gestão do conhecimento    informação para negócios    sistemas de informação
<b>Artigo 43</b>	informação tecnológica    inovação    necessidades de informação
<b>Artigo 44</b>	competitividade    conhecimento nas organizações    gestão do conhecimento
<b>Artigo 45</b>	bibliometria    cienciometria    infometria    webmetria
<b>Artigo 46</b>	bibliometria    lei de Bradford    ciência da informação    teoria do caos
<b>Artigo 47</b>	periódicos eletrônicos    usabilidade
<b>Artigo 48</b>	usos da informação    economia da informação
<b>Artigo 49</b>	biblioteconomia    ciência da informação    monitoramento
<b>Artigo 50</b>	ensino de biblioteconomia    profissionais de informação
<b>Artigo 51</b>	acessibilidade    bibliotecas
<b>Artigo 52</b>	estudos de usuários    internet    hipertextos
<b>Artigo 53</b>	comportamento dos usuários    estratégias de busca    mecanismos de busca    recuperação da informação
<b>Artigo 54</b>	bibliotecários    bibliotecas    competência em informação
<b>Artigo 55</b>	competências profissionais    profissionais da informação    mercado de trabalho

<b>Artigo 56</b>	teorias na ciência da informação    história da ciência da informação    comunicação científica
<b>Artigo 57</b>	recuperação da informação    sociedade da informação
<b>Artigo 58</b>	ciência da informação    ciência cognitiva    processamento de informações    indexação    recuperação da informação    interação homem-computador
<b>Artigo 59</b>	comunicação científica    bibliometria    cientometria
<b>Artigo 60</b>	agentes inteligentes    inteligência competitiva    internet    monitoramento

Fonte: Dados da Pesquisa (2018)

É válido destacar que tendo sido gerado um total de 582 termos únicos de 900 possíveis na Folksonomia Assistida e 266 termos indexados pelos próprios autores, ao confrontado gerou um total de 171 termos em comum, representando 29% de coincidência entre a Folksonomia Assistida e as palavras-chave dos autores. Esse dado reforça que boa parte dos termos utilizados pelos autores como palavras-chave, além de não coincidir com os termos descritos como *tags* pelos usuários-indexadores, também não são termos no TBCI, o que fez com que boa parte das palavras-chave não fossem categorizadas como mais representativas.

Dentre questões importantes que valem menção, uma delas é a de que o fato de alguns termos terem um baixo número de ocorrências em um corpus não os tornam menos importantes enquanto representações de conceitos-chave de uma determinada área de conhecimento.

## 5 CONSIDERAÇÕES FINAIS

Sempre tendo como norte a base teórico-metodológica escolhida e desenvolvida ao longo do presente trabalho, observou-se que os objetivos específicos oriundos do objetivo geral traçado foram alcançados, assim como o problema de pesquisa devidamente respondido. De forma a ficar mais claro, nos próximos parágrafos será possível ver tais pontos discutidos.

No que se refere ao primeiro objetivo específico definido, onde a proposta perpassava pela construção da Folksonomia assistida por meio de uma adaptação de modelo colaborativo de indexação social assistida, deu condições de reforçar o que vem sendo estudado pelos pesquisadores da Ciência da Informação. 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 alcance do objetivo supracitado foi possível utilizando-se da configuração da plataforma Tainacan que possui como características os aspectos principais da web 2.0, como interação, participação e colaboração de todos os atores envolvidos. Isso permitiu que houvesse condições de garantir o suporte no processo de indexação social assistida, no qual seria ressaltado os papéis, atividades e regras de interação entre usuários, assim como previstas para o modelo colaborativo.

Nesse ponto, a aplicação do modelo possibilitou tomar conhecimento acerca da compreensão e dimensão dos desdobramentos da participação ativa coordenada dos usuários frente à construção de uma linguagem para a organização e recuperação da informação em ambientes colaborativos.

Das muitas possibilidades de entendimento acerca da Folksonomia Assistida foi possível verificar que estudos como esses trazem consigo a chance de investigação dos perfis, características culturais e cognitivas dos usuários. Para tanto, questões referentes ao papel de interventor, sendo desempenhado pelo usuário nas atividades de organização de conteúdos informacionais, ampliam as noções dos principais fatores que podem vir a motivar na criação de *tags* para recursos em ambientes digitais.

Portanto, no que se trata dos ambientes digitais e suas características frente à produção de informação de forma indiscriminada, como o próprio caso da Folksonomia, é preciso ter um olhar de forma a não permitir que seja negligenciada

pela comunidade científica, e que seja verificado o potencial de aplicação dessas características para outro contexto. Nesse ponto, se faz necessário realizar estudos sobre as tendências de representação e organização do conhecimento que visam à construção da memória coletiva em rede.

A aplicação de um modelo colaborativo de indexação social assistida vem de encontro ao papel que os profissionais da informação precisam desempenhar na criação de metodologias que possam vir a utilizar metadados oriundos da participação dos usuários para adicionar ou aperfeiçoar instrumentos de controle terminológico. Isso permitiria o surgimento de uma forma de auxiliar estruturas que nem sempre são flexíveis para acondicionar novos assuntos em sua estrutura.

O surgimento de uma metodologia que se utiliza da construção informacional dos usuários pode ser uma maneira de aumentar a autonomia no que se refere a organização da informação, sendo tudo isso de acordo com características próprias desses usuários. Além disso, podem colaborar na garantia de aperfeiçoamento dos objetos, processos e produtos construídos no âmbito da Ciência da Informação como os tesouros, as ontologias e as taxonomias.

Por essa razão, foi desenvolvido para esta pesquisa uma política de indexação social assistida, servindo de apoio para entendimento do modelo que haveria de ser aplicado, e principalmente para apresentar a adoção de um tesouro da área de Ciência da Informação para a realização de um experimento de indexação social assistida de artigos científicos do corpus de Souza (2005).

O uso do TBCI se apresentou coerente ao *Corpus* escolhido, mas que em alguns pontos não abrangeu de forma satisfatória o que os usuários entendiam como propício para representar o conteúdo, reforçando a necessidade de atualização (aperfeiçoamento) do Tesouro em questão.

Assim, para o modelo aqui aplicado, a coexistência dos vocabulários controlados e da Folksonomia resultante da indexação dos recursos em ambientes digitais se fez necessário, e ainda pôde garantir a potencialização da inteligência coletiva dos usuários.

Tendo em vista a aplicação com grupos de usuários que não foram os autores dos artigos do *corpus*, a possibilidade do autor, diante das *tags* já sugeridas, realizar novamente a representação do conteúdo do seu trabalho permite que ele enriqueça de forma oportuna a indexação visando torná-la mais recuperável.

Se bem construído e atualizado de forma a garantir as bases literárias, estrutural e de uso, o instrumento de controle terminológico pode aumentar a confiança do usuário-indexador na identificação de conceitos relacionados para a indexação e recuperação dos documentos.

O fato de qualquer usuário cadastrado e logado no Tainacan ter permissão para definir quais *tags* irá representar as produções científicas de seu interesse, torna possível que ele realize uma espécie de "lembrete" atribuindo pontos de acesso para futura leitura ou recuperação das produções que ele julga relevantes no sistema.

Bandim (2017) que em sua pesquisa utiliza o mesmo *corpus*, alerta sobre o fato de alguns dos 60 artigos que o compuseram possuíam apenas duas palavras-chave e, outros, com mais de 8 palavras-chave, causando uma discrepância nos resultados alcançados.

No caso do presente trabalho esses dados tornaram o uso da Folksonomia Assistida ainda mais relevante. Isso porque levando em conta que o experimento foi aplicado a 3 (três) grupos de usuários, gerando em média um total de 9 a 10 *tags* por artigo, houve uma ampliação na possibilidade de representatividade e consequentemente recuperação dos artigos.

Como resultado, os índices verificados para determinar o grau de qualidade da indexação apresentaram em sua maioria um número abaixo do esperado, como as médias de 32% de precisão, 69% de revocação, 28% do coeficiente de consistência e 41% de medida F. Quando avaliados conjuntamente, esses índices médios nos leva a entender que o nível de harmonização entre as palavras-chave produzidas pelos autores e as produzidas, em forma de *tags*, pelos usuários alcançam um resultado bom. Como consequência, é possível vislumbrar o surgimento de problemas na recuperação dos artigos, tudo isso quando se leva em consideração que os termos que representariam os documentos para os usuários-indexadores, em sua grande maioria não são os mesmos usados pelos autores.

Com a aplicação das regras definidas para o modelo colaborativo de indexação social assistida nesta pesquisa chegou-se a algumas vantagens, que são:

1. a produção significativa de metadados semânticos;
2. a correção dos descritores com erros ortográficos;
3. a capacidade de associar as necessidades dos usuários e a linguagem utilizada pelo sistema.

Tal resultado permite que o modelo, aplicado aqui em escala micro, possa vir a ser utilizado ou aperfeiçoado para bases de dados de artigos de periódicos eletrônicos ou repositórios digitais.

Levando-se em consideração os resultados alcançados, fica como sugestão a utilização dos metadados de *tags*, e da Folksonomia Assistida em bases de dados científicas como alternativas de indexação pelos próprios leitores para alcançar melhores resultado na recuperação.

Para Ciência da Informação, como contribuição, o presente trabalho deixa reflexões em torno da validação da produção de metadados semânticos ou estruturados, onde espaços, ao exemplo das bases de dados científicas, se valeriam amplamente no processo de busca e recuperação. Nesse ponto surge uma possibilidade de descrever e representar trabalhos por meio das *tags*, a serem fornecidas pelos usuários, com valor semântico.

Tendo a indexação a devida importância para o desenvolvimento das bases de dados, a utilização da Folksonomia Assistida agregaria valor a espaços como esses que fazem uso em sua grande maioria, de buscas por palavras-chave dos autores ou da indexação intelectual realizada por equipe de indexadores.

O papel de um administrador no que se refere à correção e exclusão dos termos com erros de grafia, bem como falta de autorização junto a base terminológica escolhida, traz possibilidades de aplicação para esse universo de aumento significativo da produção de informação desordenada. Nesse ponto, é importante mencionar que a proposta não é descaracterizar a Folksonomia, mas ampliar estudos e refletir em torno dos seus principais problemas para aplicação em ambientes específicos.

A possibilidade de colaboração do usuário pode garantir uma maior familiaridade dos termos a serem utilizados nos espaços que apliquem o modelo de Folksonomia Assistida. Os termos aplicados para representar uma produção científica são os mesmos que serão utilizados para recuperá-los posteriormente por determinado usuário em um sistema, tudo isso reforçado pela capacidade de associar as necessidades dos usuários com a linguagem utilizada pelo próprio sistema.

Dentre as possibilidades em torno da Folksonomia Assistida aplicada na construção de corpus de referência, propõe-se a utilização desse resultado para avaliação de sistemas de indexação automática. Assim, se teria mais uma forma de

avaliação da qualidade da indexação levando em conta agora, as palavras-chaves dos autores e as *tags* definidas pelos usuários-indexadores com base no TBCI.

Por outro lado, é necessário levar em consideração alguns desafios encontrados no decorrer da pesquisa, dentre eles estão:

1. construção um modelo colaborativo de indexação assistida de acordo com os recursos disponíveis.
2. garantia de que de fato a interação e participação dos usuários ocorra;
3. entendimento de como o instrumento de controle terminológico que objetiva melhorar o processo de representação e recuperação da informação deve ser aplicado e avaliado de forma sistemática.

É importante a partir de este ponto reforçar que os estudos acerca de modelos de colaboração voltados para a atividade de indexação dos recursos em um sistema são poucos dentro da literatura da área de Ciência da Informação, necessitando assim de mais intervenções.

A maior limitação verificada no decorrer desta pesquisa foi conseguir estabelecer um parâmetro claro do papel do administrador frente ao que fora produzido pelos usuários-indexadores. Isso se reforça quando a plataforma escolhida para aplicação do método, no caso o Tainacan, não dá as devidas condições para a realização do que se tinha estabelecido inicialmente para a pesquisa.

A consequência disso foi à limitação da forma com que o usuário interagiu com o sistema, tendo que haver uma adaptação para que a pesquisa continuasse sem maiores problemas. Assim, a principal questão foi operacional, e que por ocorrer, fez com que alguns problemas de inserção de termos fossem identificados.

Vale frisar que o principal ponto da Folksonomia Assistida é servir de apoio ao usuário no ato de escolha de um termo representativo, trazendo consigo a possibilidade de resolver problemas recorrentes na Folksonomia, que seria a polissemia, erros de grafia, entre outros.

Sabendo da importância de se ter continuidade aos estudos acerca da temática, fica como sugestão a aplicação em escala macro do modelo aqui apresentado com o intuito de dar condições de aperfeiçoamento dos objetos, processos e produtos construídos no âmbito da Ciência da Informação, como tesauros, ontologias, entre outros. Além disso, o uso da Folksonomia Assistida permite vislumbrar maiores reflexões em torno mediação do processo de indexação

e de recuperação da informação em determinados ambientes, principalmente de caráter digital.

Por fim, como trabalho futuro, outra possibilidade seria 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 entender o grau de comprometimento do usuário no ato da indexação.

## REFERÊNCIAS

ALVARENGA, Lídia. Organização da informação nas bibliotecas digitais. In: NAVES, Madalena M. L.; KURAMOTO, Hélio (Org.). **Organização da informação: princípios e tendências**. Brasília: Briquet de Lemos Livros, 2006. p.79-98.

BANDIM, Márcio Aécio Silva. **Indexação automática por atribuição de artigos científicos da área de Ciência da Informação**. 2017. 144 f. Dissertação (Mestrado) - Curso de Ciência da Informação, Departamento de Ciência da Informação, Universidade Federal de Pernambuco, Recife, 2017.

BANDIM, Marcio Aercio Silva; CORRÊA, Renato Fernandes. 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. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p64/37377>>. Acesso em: 10 nov. 2018.

BLATTMANN, Ursula; SILVA, Fabiano Couto Corrêa da. Colaboração e interação na Web 2.0 e biblioteca 2.0. **Revista ACB**, v. 12, n. 2, p. 191-215, jul./dez. 2007. Disponível em: <<https://revista.acbsc.org.br/racb/article/view/530>>. Acesso em: 18 abr. 2018.

BORKO, H. Information Science: what is it? **American Documentation**, v.19, n.1, p.3-5, jan. 1968.

BROWN, P. et al. The democratic indexing of images. **New Review of Hypermedia and Multimedia**, v. 2, 1996, p. 107-120.

CAMPOS, Maria Luiza de Almeida; GOMES, Hagar Espanha. Metodologia de elaboração de tesouro conceitual: a categorização como princípio norteador. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 3, p.348-359, 2006. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/273>>. Acesso em: 10 nov. 2017.

CARNEIRO, M. V. Diretrizes para uma política de indexação. **R. Esc. Biblioteconomia**. UFMG, Belo Horizonte, v.14, n.2, p.221-241, set. 1985. Disponível em: <<http://www.brapci.inf.br/index.php/article/download/13788>>. Acesso em: 17 jun. 2018.

CASTELLS, Manuel. **A sociedade em rede**. 4.ed. São Paulo: Paz e Terra, 2000. 618 p.

CATARINO, Maria Elisabete; BAPTISTA, Ana Alice. Folksonomia: um novo conceito para a organização dos recursos digitais na Web. **Revista Data Grama Zero**, v.8, n.3, jun. 2007.

CESARINO, M. A. N; PINTO, M. C. M. Cabeçalho de assunto como linguagem de indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 7, n. 2, p. 268-288, set. 1978.

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

CELERINO, Victor Galvão; CORRÊA, Renato Fernandes. 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.

DODEBEI, Vera Lucia. **Tesouro**: linguagem de representação da memória documentária. Niterói: Intertexto, 2002. 119 p.

FARMER, Linda. Automatic categorization: what's it all about?. **Serials Librarian**, v. 51, n. 2, p. 91-101, 2006.

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. Disponível em: <<http://www.brapci.inf.br/index.php/article/download/7561>>. Acesso em: 14 mai. 2018.

FUNK, M. E.; REID, C. A. **Indexing consistency**. In MEDLINE. *Bulletin of the Medical Library Association*, v. 71, p. 176-183, 1983.

GEROSA, M. A. **Desenvolvimento de groupware componentizado com base no modelo 3c de colaboração**. 2006. 275f. Tese (Doutorado) - Curso de Informática, Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2006. Disponível em: <<http://www.seconlab.inf.puc-rio.br/public/papers/teseGerosa/TeseGerosa.pdf>>. Acesso em: 16 jan. 2018.

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

GIL LEIVA, Isidoro. **La automatización de la indización, propuesta teórico-metodológica**: aplicación al área de Biblioteconomía y Documentación. 1997. 268f. Tese (Doutorado) – Universidad de Murcia, Murcia, España, 1997. Disponível em: <<http://webs.um.es/isgil/resources/PhDissertation%20Gil-Leiva.pdf>>. Acesso em: 10 jul. 2018

GIL LEIVA, Isidoro. **Manual de indización**: teoría y práctica. Gijón: Trea, 2008, 432 p.

GIL LEIVA, Isadoro; RUBI, Milena Polsinelli; FUJITA, Mariângela Spotti Lopes. Consistência na indexação em bibliotecas universitárias brasileiras. **Transformação**, v. 20, n. 3, p. 233-253, set./dez. 2008. Disponível em: <<http://www.scielo.br/pdf/tinf/v20n3/03.pdf>>. Acesso em: 12 jun. 2018

GUEDES, V. L. S. Estudos de um critério para indexação automática derivativa de textos científicos e tecnológicos. **Ciência da Informação**, v. 23, n. 3, p. 318-326, set./dez. 1994. Disponível em: <<http://revista.ibict.br/ciinf/article/view/529>>. Acesso em: 12 jan. 2018.

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. Disponível em: <<https://www.iberid.eu/ojs/index.php/scire/article/download/1133/1115>>. Acesso em: 14 mai. 2018.

GUINCHAT, C.; MENO, M. **Introdução geral às ciências técnicas da informação e documentação**. Brasília: IBICT, 1994. 540 p.

IIVONEN, M.; KIVIMAKI, K. Common entities and missing properties: similarities and differences in the indexing of concepts. *Knowledge Organization*, v. 25, n. 3, p. 90-102, 1998.

HASAN, Kazi Saidul; NG, Vincent. 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. Disponível em: <<http://acl2014.org/acl2014/P14-1/pdf/P14-1119.pdf>>. Acesso em: 04 jul. 2018.

HASSAN-MONTERO, Yusef. Indización Social y Recuperación de Información. **No Solo Usabilidad Journal**, Granada, n. 5, 2006. Disponível em: <[http://www.nosolousabilidad.com/articulos/indizacion\\_social.htm](http://www.nosolousabilidad.com/articulos/indizacion_social.htm)>. Acesso em: 10 fev. 2018.

HLAVA, M. M. Automatic indexing: a matter of degree. **Bulletin of the American Society for Information Science and Technology**, v. 29, n. 1, p. 12-15, out./nov. 2002. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1002/bult.261>>. Acesso em: 18 abr. 2018.

HJORLAND, Birger. Toward a theory of aboutness, subject, topicality, theme, domain, field, content... and relevance. **Journal of the American Society for Information Science and Technology**, v. 52, n. 9, p. 249-298, 2001. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.1131>>. Acesso em: 14 abr. 2018

KIM, H. et al. Integrating Tagging into the Web of Data: Overview and Combination of Existing Tag Ontologies. **Journal Of Internet Technology**, 2011, v. 12, n. 4, p. 561-572, 2011. Disponível em: <[http://johnbreslin.org/files/publications/20110700\\_jit\\_2011.pdf](http://johnbreslin.org/files/publications/20110700_jit_2011.pdf)>. Acesso em: 04 nov. 2017.

KIPP, M. E. I. Exploring Measures of Inter-Tagger Consistency. *SIG-CR Workshop Poster, Annual Meeting of the American Society for Information Science and Technology*, 2009, Vancouver, British Columbia, Canada.

KOBASHI, Nair Yumiko; LARA, Marilda Lopes Ginez de; TÁLAMO, Maria de Fátima G. Moreira. Contribuição da terminologia para a elaboração de tesouros. **Ciência da Informação**, Brasília, v. 21, n. 3, p.197-200, 1992. Disponível em: <<http://revista.ibict.br/ciinf/article/view/432/432>>. Acesso em: 10 nov. 2017.

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

KURAMOTO, H. Sintagmas nominais: uma nova abordagem no processo de indexação. In: NAVES, M. M. L.; KURAMOTO, H. (Orgs.). **Organização da informação: princípios e tendências**. Brasília: Briquet de Lemos, 2006.

KUHN, Thomas S. **A estrutura das revoluções científicas**. 5.ed. São Paulo: Editora Perspectiva S.A, 1998. 257 p.

LANCASTER, F.W. **Evaluation of the MEDLARS demand search service**. Washington, D.C. National Library of Medicine, 1968.

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

LAPA, Remi Correia. **Indexação automática no Brasil no âmbito da ciência da informação (1973-2012)**. 2014. 287 f. Dissertação (Mestrado) - Curso de Ciência da Informação, Departamento de Ciência da Informação, Universidade Federal de Pernambuco, Recife, 2014. Disponível em: <<https://repositorio.ufpe.br/handle/123456789/12947>>. Acesso em: 04 maio 2018.

LE COADIC, Yves François. **A ciência da informação**. 2.ed. Brasília: Briquet de Lemos Livros, 2004. 124 p.

LEININGER, K. Interindexer consistency in PsycINFO. **Journal of Librarianship and Information Science**, v. 32, n. 1, p. 4-8, 2000.

LEONARD, L. E. **Inter-indexer consistency and retrieval effectiveness: measurement of relations**. Champaign: University of Illinois, 1975 (PhD Thesis).

LIMA, G. A. B. Interfaces entre a ciência da informação e a ciência cognitiva. **Ciência de Informação**, 2003, v.32, n.1, p.77-87. Disponível em: <<http://revista.ibict.br/ciinf/article/view/1021>>. Acesso em: 04 nov. 2017.

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

MARKEY, K. Interindexer consistency tests: A literature review and report of a test of consistency in indexing visual materials. **Library and Information Science Research**, v. 6, n. 2, p. 155-177, 1984.

MAXIMIANO, A. C. A. **Introdução à administração**. 5.ed. São Paulo: Atlas, 2000.

MIDDLETON, M.R. A comparison of indexing consistency and coverage in the AEI, ERIC, and APAIS databases. **Behavioral & Social Sciences Librarian**, v. 3, n. 4, p. 33-43, 1984.

MORIN, E. **Introdução ao pensamento complexo**. 4.ed. Porto Alegre: Sulina, 2011. 120 p.

MOURA, M. Folksonomias, Redes Sociais e a formação para o tagging literacy. **Informação & Informação**, Londrina, 14, p.2-5, 2009. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/2196>>. Acesso em: 13 nov. 2017.

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

NARUKAWA, C. M. **Estudo de Vocabulário Controlado na Indexação Automática**: Aplicação no Processo de Indexação do Sistema de Indización Semiautomática (SISA). 2011. 222 f. Dissertação (Mestrado) – Curso de Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2011.

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: Est.**, João Pessoa, v. 19, n. 2, p. 99-118, maio/ago. 2009. Disponível em: <<http://www.periodicos.ufpb.br/ojs/index.php/ies/article/view/2925>>. Acesso em: 15 jun. 2018.

NESHAT, N.; HORRI, A. A study of subject indexing consistency between the National Library of Iran and Humanities Libraries in the Area of Iranian Studies. **Cataloging & Classification Quarterly**, v. 43, n. 1, p. 67-76, 2006.

PINHEIRO, Lena Vânia Ribeiro. Campo interdisciplinar da Ciência da Informação: fronteiras remotas e recentes. In: \_\_\_\_\_. **Ciência da Informação, ciências sociais e interdisciplinaridade**. Brasília: IBICT, 1999. p. 155-182.

PINHEIRO, Lena Vania Ribeiro; FERREZ, Helena Dodd. Rede metodológica integrando epistemologia, organização do conhecimento, bibliometria e tesouros: concepção e construção do Tesouro brasileiro de ciência da informação. In: ROBREDO, Jaime; BRÄSCHER, Marisa (Orgs.). **Passeios pelo bosque da informação**: estudos sobre a representação e organização da informação e do conhecimento. Brasília: Ibiict, 2010. p. 259-276.

PRIBERAN. **Significado de modelo**. 2015. Disponível em:

<<https://www.priberam.pt/DLPO/modelo>>. Acesso em: 15 out. 2017.

PRIMO, Alex. O aspecto relacional das interações na Web 2.0. **E- Compôs**, Brasília, v. 9, p.1-21, 2007. Disponível em: <<http://www6.ufrgs.br/limc/PDFs/web2.pdf>>.

Acesso em: 20 mar. 2018.

RAFFERTY, Pauline; HIDDENLEY, Rob. Flickr and democratic indexing: dialogic approaches to indexing. **Aslib Proceedings**, v. 59, n. 4, 2007. p. 397-410.

RECTOR, Monica; YUNES, Eliana. **Manual de semântica**. Rio de Janeiro: Ao Livro Técnico, 1980. 171 p.

ROBREDO, J. A indexação automática de textos: o presente já entrou no futuro,

**Estudos Avançados em Biblioteconomia e Ciência da Informação**, Brasília, v. 1, n. 1, p. 235-274, 1982. Disponível em:

<<http://www.brapci.inf.br/index.php/article/download/17365>>. Acesso em: 14 jan. 2018.

ROSCHELLE, Jeremy; TEASLEY, Stephanie D.. The Construction of Shared Knowledge in Collaborative Problem Solving. **Computer Supported Collaborative Learning**, Berlim, v. 25, p.69-97, 1995. Disponível em:

<<http://umdperg.pbworks.com/f/RoschelleTeasley1995OCR.pdf>>. Acesso em: 10 mar. 2018.

ROWLEY, Jennifer. **A biblioteca eletrônica**. Brasília: Briquet de Lemos Livros, 2002. 399 p.

SAARTI, J. Consistency of subject indexing of novels by public library professionals and patrons. **Journal of Documentation**, v. 58, n. 1, p. 49-65, 2002.

SALTON, G.; MCGILL, M. J. **Introduction to modern information retrieval**. 2.ed. New York: McGraw-Hill, 2003. 474 p.

SANTARÉM SEGUNDO, José Eduardo. **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. Disponível em: <<https://repositorio.unesp.br/handle/11449/103346>>. Acesso em: 01 set. 2017.

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, José Eduardo.; 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. Disponível em:

<<http://revista.ibict.br/liinc/article/view/3300>>. Acesso em: 15 jan. 2018

SANTOS, Raimunda Fernanda 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) - Curso de Ciência da Informação, Departamento de Ciência da Informação, Universidade Federal de Pernambuco, Recife, 2016. Disponível em: <<https://repositorio.ufpe.br/handle/123456789/17218>>. Acesso em: 01 nov. 2017.

SANTOS, Raimunda Fernanda dos; CORRÊA, Renato Fernandes. A Folksonomia e a representação colaborativa da informação em ambientes digitais. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 8, p. 69-84, 2015. Disponível em: <<http://ojs.lat/tpbci/index.php/tpbci/article/view/225>>. Acesso em: 11 set. 2017.

SARACEVIC, Tefko. Interdisciplinary nature of information science. **Ciência da Informação**, v. 24, n. 1, 1995.

SAYÃO, Luís Fernando. Bases de dados: a metáfora da memória científica. **Ciência da Informação**, v. 25, n. 3, p. 314-318, set./dez. 1996. Disponível em: <<http://revista.ibict.br/ciinf/article/view/629/633>>. Acesso em: 11 abr. 2018.

SAYÃO, Luís Fernando. Modelos teóricos em Ciência da Informação: abstração e método científico. **Ciência da Informação**, Brasília, v.30, n.1, p.82-91, jan./abr. 2001. Disponível em: <<http://revista.ibict.br/ciinf/article/view/941>>. Acesso em: 11 abr. 2018.

SIEVERT, M. E.; ANDREWS, M. J. Indexing consistency in Information Science Abstracts. **Journal of the American Society for Information Science**, v. 42, n. 1, p. 1-6, 1991.

SILVA, C. K. P. e. **Web 2.0: A migração para a Web social**. 2007. 68f. Trabalho de Conclusão de Curso (Dissertação) – Curso de Ciência da Computação, Universidade Federal de Juiz de Fora, Minas Gerais, 2007.

SILVA, Marcel Ferrante. **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. Disponível em: <<http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/BUOS-9JLJF>>. Acesso em: 01 set. 2017.

SOERGEL, D. Indexing and retrieval performance: the logiciel evidence. *Journal of the American Society for Information Science*, v. 45, n.8, p.589-599, 1994. Disponível em: <<https://pdfs.semanticscholar.org/546b/5f8efce62029d44342927ee115d12a667370.pdf>>. Acesso em: 10 out. 2018.

SOUZA, Renato Rocha. **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. Disponível em: <<http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/RRSA-6GGGUF>>. Acesso em: 12 set. 2017.

TARTAROTTI, R. D. E.; EVEDOVE, P. R.; FUJITA, M. S. L. Elaboração de tesauro conceitual de política de indexação para bibliotecas universitárias. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 16., 2015. João Pessoa, **Anais...** João Pessoa: Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação, 2015.

TENNIS, Joseph. Social Tagging and the next steps for indexing. In: SIG/CR CLASSIFICATION RESEARCH WORKSHOP, 17., Austin (USA), 2006. **Papers...** American Society for Information Science and Technology, 2006.

TESAURO BRASILEIRO DE CIÊNCIA DA INFORMAÇÃO – TBCI. **Labcom - Laboratório de Pesquisas em Comunicação Científica - IBICT**, Rio de Janeiro: 2014. 384 p. Disponível em: <<http://www.ibict.br/publicacoes-e-institucionais/tesauro-brasileiro-de-Ciência-da-informacao-1>>. Acesso em: 17 set. 2017.

TONTA, Y. A Study of indexing consistency between Library of Congress and BritishLibrary Catalogers. **Library Resources and Technical Services**, v. 35, n. 2, p. 177-185, 1991.

TRASK, Robert Lawrence. **Dicionário de Linguagem e Lingüística**. São Paulo: Contexto. 2004, 364 p.

VAN AMSTEL, F. Folcsonomia: Vocabulário Descontrolado, Anarquitectura da Informação ou Samba do Crioulo Doido? In: ENCONTRO BRASILEIRO DE ARQUITETURA DA INFORMAÇÃO, 1., 2007, **Anais...** 2007. Disponível em: <[http://usabilidoido.com.br/arquivos/folcsonomia\\_anarquitectura.pdf](http://usabilidoido.com.br/arquivos/folcsonomia_anarquitectura.pdf)>. Acesso em: 20. jun. 2017.

VOSS, Jakob. Tagging, Folksonomy & Co: Renaissance of Manual Indexing?. In: INTERNATIONAL SYMPOSIUM FOR INFORMATION SCIENCE, 10., 2007, Colônia. **Anais...** . Colônia: Kmedu Hub, 2007. p. 1 - 12. Disponível em: <<https://arxiv.org/pdf/cs/0701072.pdf>>. Acesso em: 04 mar. 2018.

WAL, Thomas Vander. **Folksonomy definition and wikipedia**. 2005. Disponível em: <<http://www.vanderwal.net/random/entrysel.php?blog=1750>>. Acesso em: 22 fev. 2018.

WEINBERGER, D. **A nova desordem digital**: os novos princípios que estão reiventando os negócios, a educação, a política, a ciência e a cultura. Rio de Janeiro: Elsevier. 2007, 275p.

WIKI DO TAINACAN. [online]. Disponível em: <<http://wiki.tainacan.medialab.ufg.br/doku.php>>. Acesso em: 11 set. 2017.

YIN, Robert K. **Estudo de caso**: planejamento e métodos. 2 ed. Porto Alegre: Bookman, 2001, 200 p.

## APÊNDICE A - POLÍTICA DE INDEXAÇÃO SOCIAL

### APRESENTAÇÃO

A elaboração dessa Política de Indexação consiste numa tentativa de padronizar o processo de indexação dos artigos disponibilizado na Coleção “Corpus de Referência em CI” através da plataforma Tainacan.

Para o desenvolvimento da Indexação no Tainacan, o primeiro ponto que se tomou conhecimento foi acerca da realidade da indexação realizada em ambientes digitais, que se caracterizam pela construção colaborativa, tendo assim características próprias. Com isso foi possível identificar a comunidade usuária no qual seriam destinados os assuntos da seleção de documentos e dos tipos de materiais, bem como na escolha da linguagem e dos critérios de indexação mais adequados.

O objetivo então é que esta política seja norteadora nos processos de indexação não somente por estar pautada numa fundamentação teórica e metodológica, mas principalmente pela necessidade avaliar a relação existente da indexação realizada pelo usuário-indexador.

Como forma de embasar a elaboração desta política, é de fundamental importância se valer de publicações que abordam a temática da indexação na Ciência da Informação. Alguns dos principais conceitos de indexação, que se remonta de décadas atrás, tem se aperfeiçoado com o passar do tempo, tendo em vista a evolução dos acervos físicos e digitais.

Para Chaumier (1988, p. 63) a indexação é:

[...] a parte mais importante da análise documentária. Consequentemente, é ela que condiciona o valor de um sistema documentário. Uma indexação inadequada ou uma indexação insuficiente representam 90% das causas essenciais para a aparição de “ruídos” ou de “silêncios” em uma pesquisa. Os 10% restantes serão devidos a causas mecânicas [...]

Com base nas palavras do autor é nítida a necessidade do indexador conhecer as características de cada documento, bem como a demanda por informação de seus usuários, para que, assim, possa atribuir termos descritores que sejam mais apropriados à representação do conteúdo informacional.

Para a indexação considere os seguintes aspectos: 1. Compreensão do conteúdo do documento como um todo; 2. Identificação dos conceitos que representam este conteúdo e 3. Seleção dos conceitos válidos para a recuperação.

## **ELEMENTOS DA POLÍTICA DE INDEXAÇÃO**

Para essa política foram determinados os objetivos da indexação para auxiliar o processo de recuperação da informação. Alguns elementos foram traçados para isso, são eles: cobertura temática, composição do acervo e público alvo/indexador.

### **COBERTURA TEMÁTICA**

A coleção “Corpus de Referência em CI” é composto por publicações das revistas **DataGramZero**, sendo um total de 29 documentos e da **Ciência da Informação**, com 31 documentos. Dessa forma, a cobertura temática abrange a área Ciência da Informação, bem como suas respectivas disciplinas.

### **COMPOSIÇÃO DO ACERVO**

Os documentos que constam na coleção foram retirados de Revistas Científicas da área da Ciência da Informação, sendo artigos científicos e que constam em plataformas online. A seleção de artigos foi definida a partir da tese de Souza (2005)<sup>3</sup> e foi escolhido pelo prestígio das revistas escolhidas junto ao QUALIS da CAPES, bem como a sua utilização em pesquisas acerca da indexação automática.

### **PÚBLICO-ALVO/INDEXADOR**

A indexação dos artigos disponibilizados através da Coleção Corpus de Referência em CI tem como público-alvo o mesmo público-indexador, sendo eles ligado a Ciência da Informação e suas disciplinas (Biblioteconomia, Gestão da Informação, entre outras),

---

<sup>3</sup> <http://www.gercinalima.com/mhtx/pages/prototipo-btdeci/teses/souza-rr/anexo-a.php> ou [http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/RRSA-6GGUF/doutorado\\_renato\\_rocha\\_souza.pdf?sequence=1](http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/RRSA-6GGUF/doutorado_renato_rocha_souza.pdf?sequence=1)

Para o usuário-indexador ele deverá fazer uso do “Manual de Indexação”, como orientação para desenvolver as ações na plataforma escolhida, o Tainacan.

## **TESAURO PARA INDEXAÇÃO**

Sabendo da existência de diversos Tesouros online, por exemplo, recomenda-se que os cadastros das *tags* (indexação) sejam realizados a partir do Tesouro Brasileiro de Ciência da Informação.

Segundo o próprio Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), responsável pelo TBCI, o tesouro conta com cerca de 1.800 termos, sendo a maioria contemplada também com versão inglesa e espanhola, e por definições. O IBICT reforça que a expectativa é a de que o Tesouro Brasileiro de Ciência da Informação tenha um papel central na recuperação da informação no Brasil e para países lusófonos.

Diante desse cenário, a consulta a tesouros de diferentes décadas se tornou importante por deixar mais evidente a evolução terminológica ao longo dos anos. Por outro lado, o fato de ser em diferentes idiomas e culturas, além de enriquecedor do ponto de vista do trabalho em si, motivou a expansão do tesouro brasileiro, razão pelo qual é importante estar atento e flexível com as mudanças que podem vir a ocorrer.

## APÊNDICE B – MANUAL DE INDEXAÇÃO

Para a execução das ações de indexação, o site a ser utilizado é o Acervo de Cultura Afro Brasileiro (<http://afro.culturadigital.br/>), desenvolvido numa parceria do Ministério da Cultura (MinC), Universidade Federal de Goiás (UFG) e Universidade Federal de Pernambuco (UFPE), desenvolvido a partir da plataforma de gerenciamento de coleções, Tainacan.

Antes de iniciar de fato a indicação de como as etapas de execução vão transcorrer, é importante compreender o funcionamento da Plataforma e quais as suas possibilidades operacionais.

O Tainacan foi criado em 2009, através do próprio Ministério da Cultura - Minc, a partir de discussões em torno da Cultura Digital no Fórum da Cultura Digital Brasileira, reforçando a ideia de que é preciso encontrar formas de entender e atender o cenário que se desenha.

A plataforma tem como proposta, soluções técnicas de integração que anseiam pela simplicidade e 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 científica.

Seu desenvolvimento teve como intuito efetivar ao máximo a inteligência coletiva por meio da colaboração, se apresentando de fácil configuração e personalização, principalmente no processo de gestão das coleções digitais.

A ideia de coletividade no qual se baseia a web 2.0 pode ser visualizada de forma efetiva na maneira amigável e responsiva dada pelo usuário final quando utiliza a ferramenta. Assim, se tem uma importante ferramenta para análises de comportamento do que pode ser desenvolvido em repositórios digitais.

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, por exemplo, como administração, armazenamento, preservação e disponibilização de conteúdos informacionais nos diversos formatos eletrônicos.

## DESENVOLVIMENTO DAS AÇÕES NO TAINACAN

A partir desse ponto as atividades serão divididas em passo para uma melhor visualização do que deve ser feito, e elas são:

Passo 1 – Cadastro no Acervo da Cultura Afro Brasileira;

Passo 2 – Busca da coleção “Corpus de Referência em CI”;

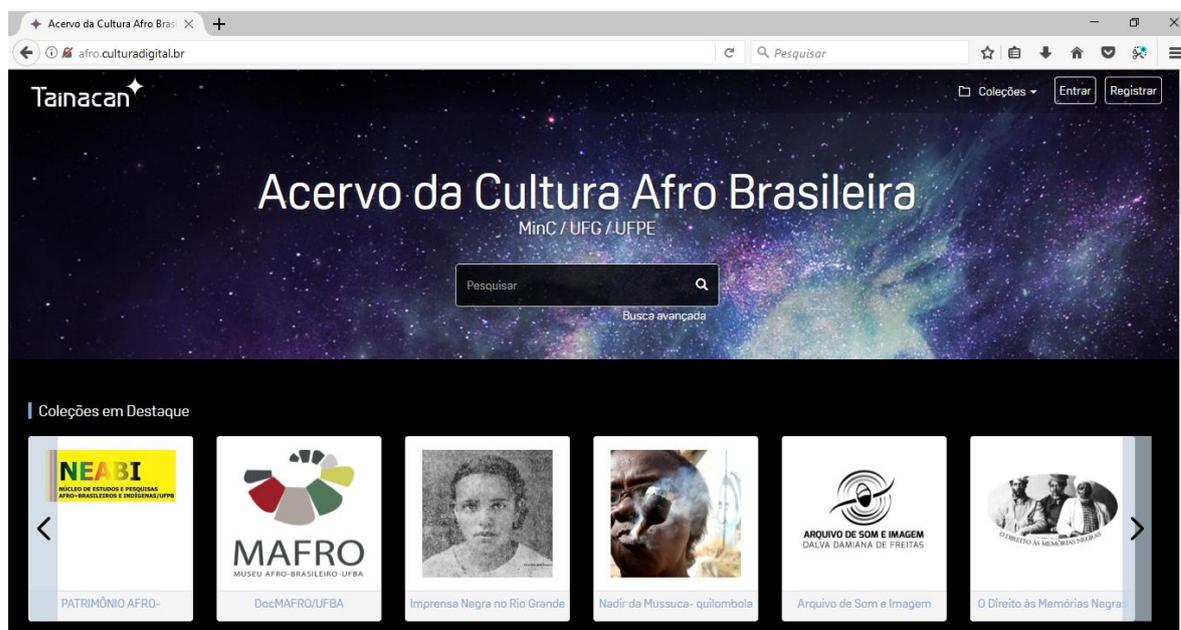
Passo 3 – Acesso a coleção “Corpus de Referência em CI”;

Passo 4 – Acesso aos itens da coleção;

### Passo 1 - Cadastro no Acervo da Cultura Afro Brasileira

Na primeira etapa deverá ser realizado o cadastro enquanto usuário no endereço [www.afro.culturadigital.br](http://www.afro.culturadigital.br) para desenvolver todas as ações de interação na plataforma. Essa ação permitirá que aquilo que for realizado fique registrado e devidamente salvo.

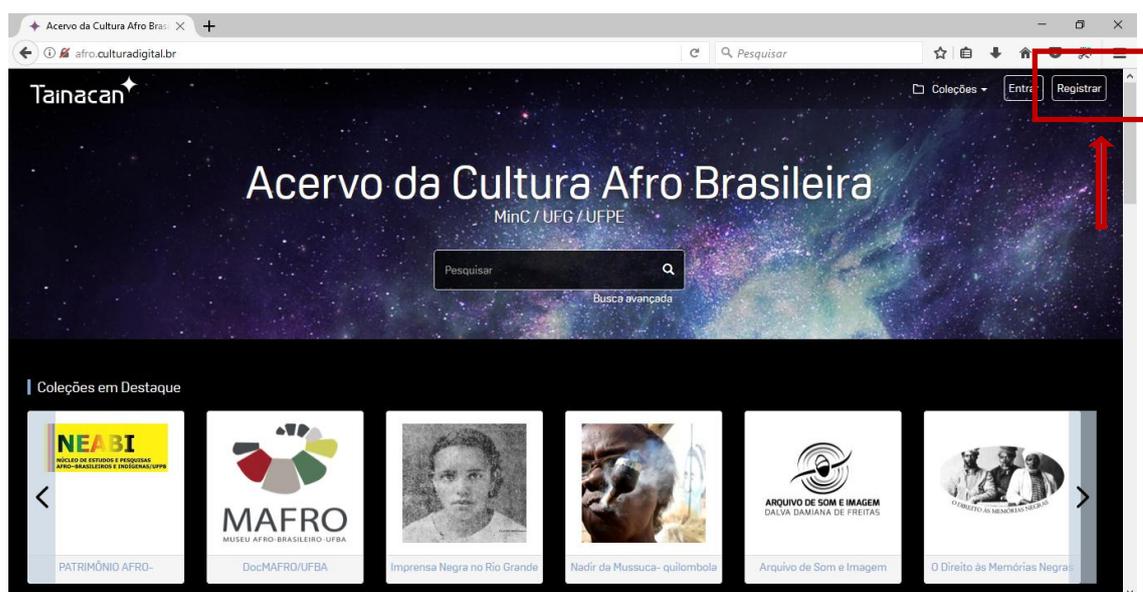
**Figura 1** – Página Inicial do Acervo de Cultura Afro Brasileira (Tainacan)



Fonte: Acervo de Cultura Afro Brasileira (2018, *printscreen*)

O cadastro é feito a partir da opção **Registrar** que fica no canto superior direito do site e que pode ser visto na Figura 2 destacado.

**Figura 2 – Cadastro (Tainacan)**



Fonte: Acervo de Cultura Afro Brasileira (2018, *printscren*)

Ao clicar na opção registrar será direcionada a uma página que dará duas opções de registro, a primeira utilizando o login do Facebook e a segunda através de um cadastro no qual será necessário inserir algumas informações.

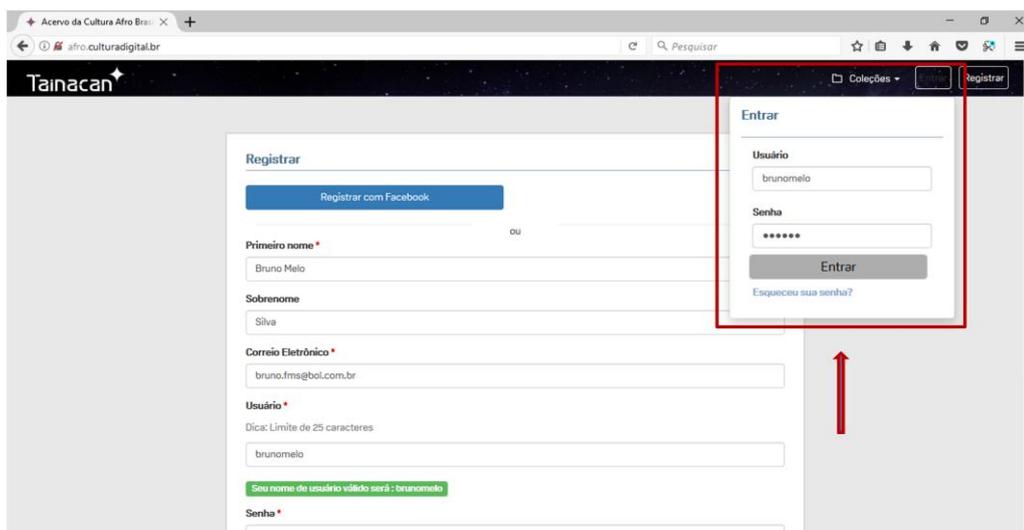
**Obs.:** Nesse ponto fica recomendado a criação de um login a partir do preenchimento do formulário, por conta de instabilidade no processo de convergência entre o Facebook e o Tainacan.

**Figura 3 – Página de Cadastro (Tainacan)**

Fonte: Acervo de Cultura Afro Brasileira (2018, *printscren*)

Confirmado a criação do cadastro, é de suma importância que seja feito em seguida o login na plataforma como indica a Figura 4.

**Figura 4 – Busca na página principal**

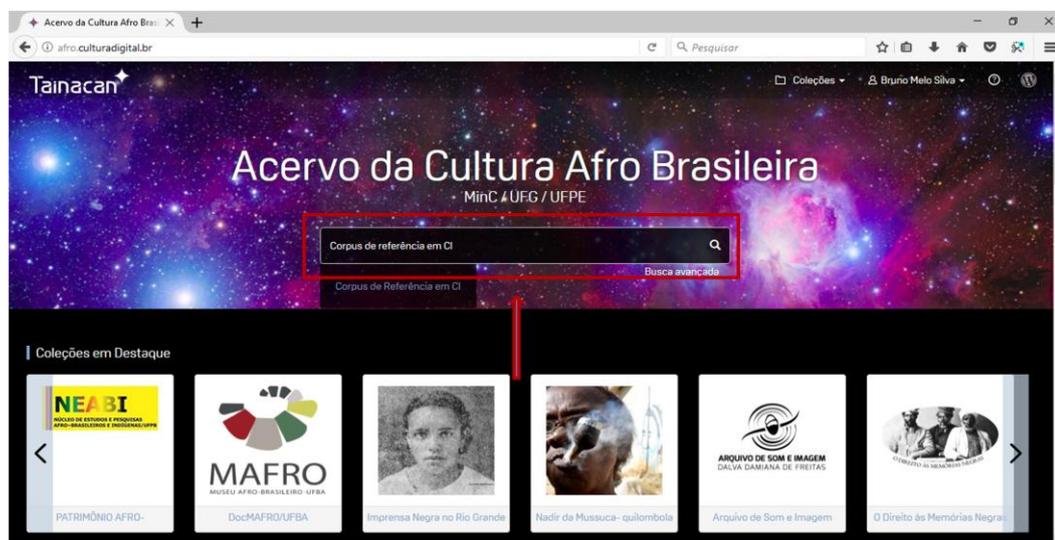


Fonte: Acervo de Cultura Afro Brasileira (2018, *printscreen*)

## **Passo 2 – Busca da coleção “Corpus de Referência em CI”**

O segundo momento a ser realizado diz respeito à localização da Coleção que deverá ser trabalhada. Com o cadastro e login efetuado na plataforma, é preciso voltar à página inicial para efetuar a busca da coleção “Corpus de Referência em CI” bem como está indicado na **Figura 5**.

**Figura 5 – Busca na página principal**



Fonte: Acervo de Cultura Afro Brasileira (2018, *printscreen*)

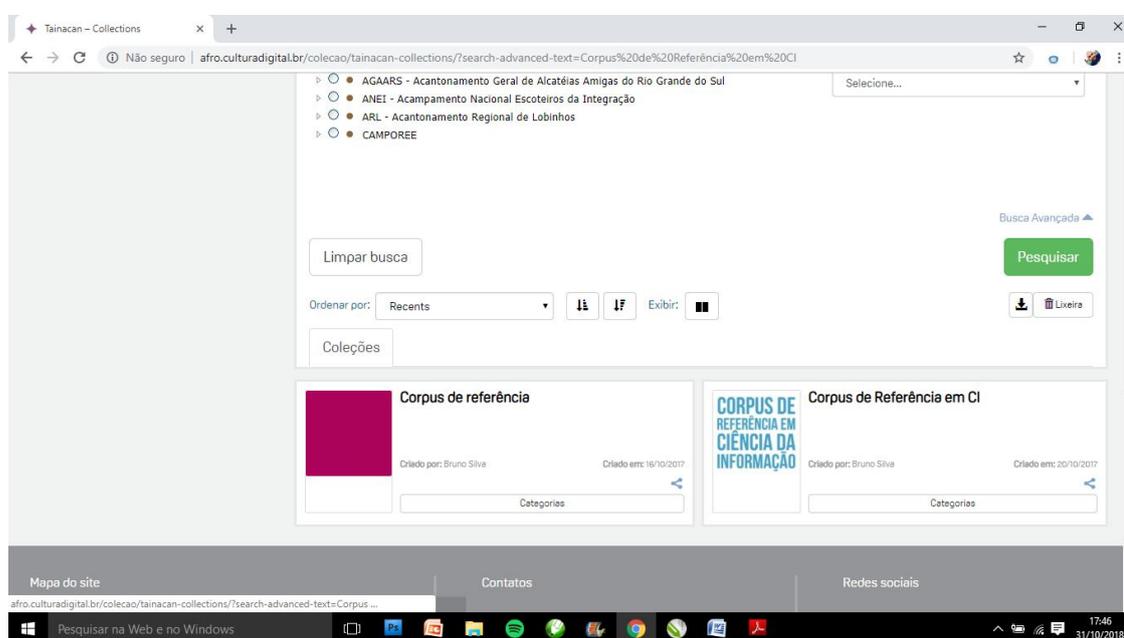
Quando efetuar a busca, a plataforma irá direcionar automaticamente para outra página com os “Resultados da Busca”. Dentre as opções deverá ser identificada aquela que consta como título “Corpus de Referência em CI” ou que tem como imagem de perfil a **Figura 6**.

**Figura 6 – Imagem de perfil da Coleção Corpus de Referência em CI**



Fonte: Elaborada pelo autor (2018).

**Figura 7 – Resultado da busca**

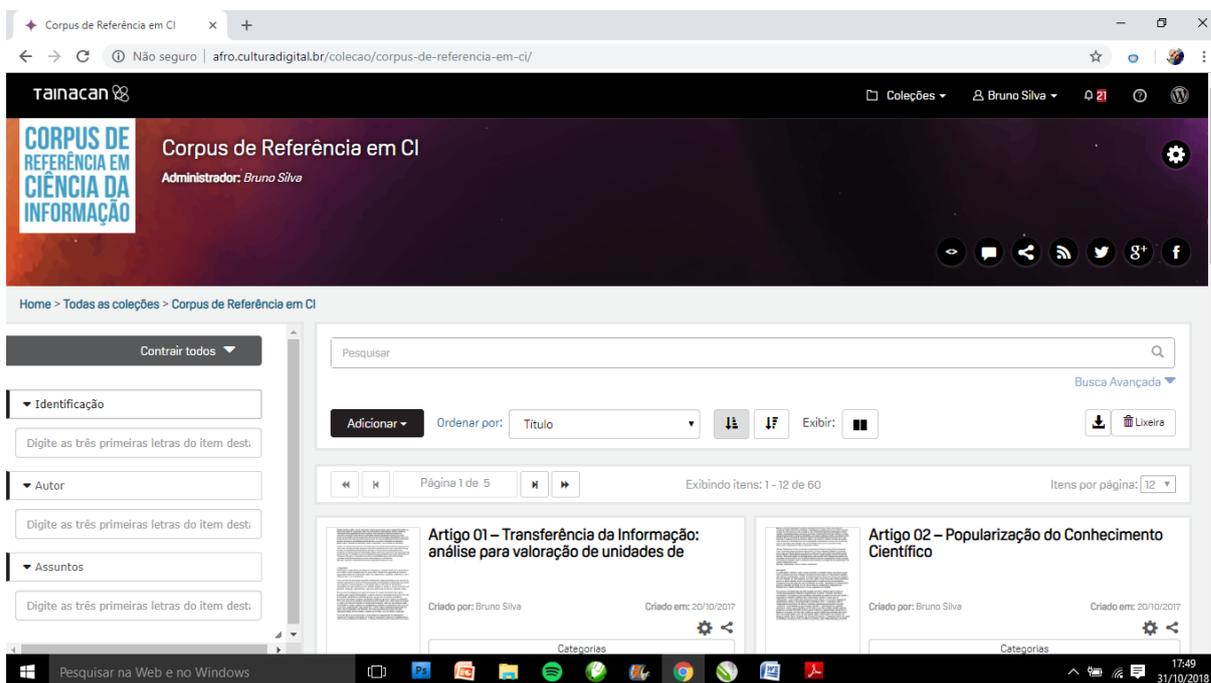


Fonte: Acervo de Cultura Afro Brasileira (2018, *printscren*)

### **Passo 3 – Acesso a coleção “Corpus de Referência em CI”**

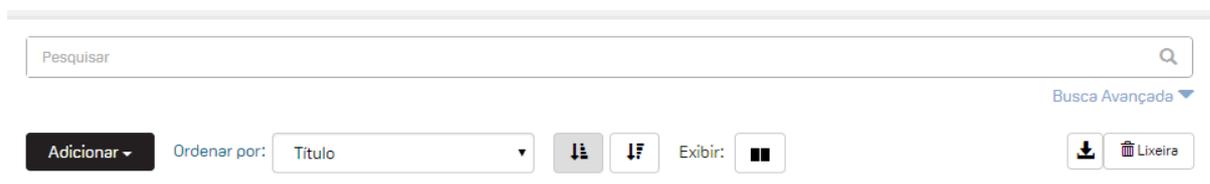
Ao identificar e acessar a coleção constará todos os artigos (itens) que devem ser etiquetados/indexados seguindo as próximas recomendações. A coleção também pode ser acessada diretamente pelo endereço: <http://afro.culturadigital.br/colecao/corpus-de-referencia-em-ci/> (**Figura 8**) e tem como figuração em sua página inicial, além dos itens, uma ferramenta de busca exclusiva para coleção (**Figura 9**), dando assim mais autonomia ao seu controle.

**Figura 8 – Página da Coleção “Corpus de Referência em CI”**



Fonte: Acervo de Cultura Afro Brasileira (2018, *printscreen*)

**Figura 9 – Ferramenta de busca da Coleção “Corpus de Referência em CI”**

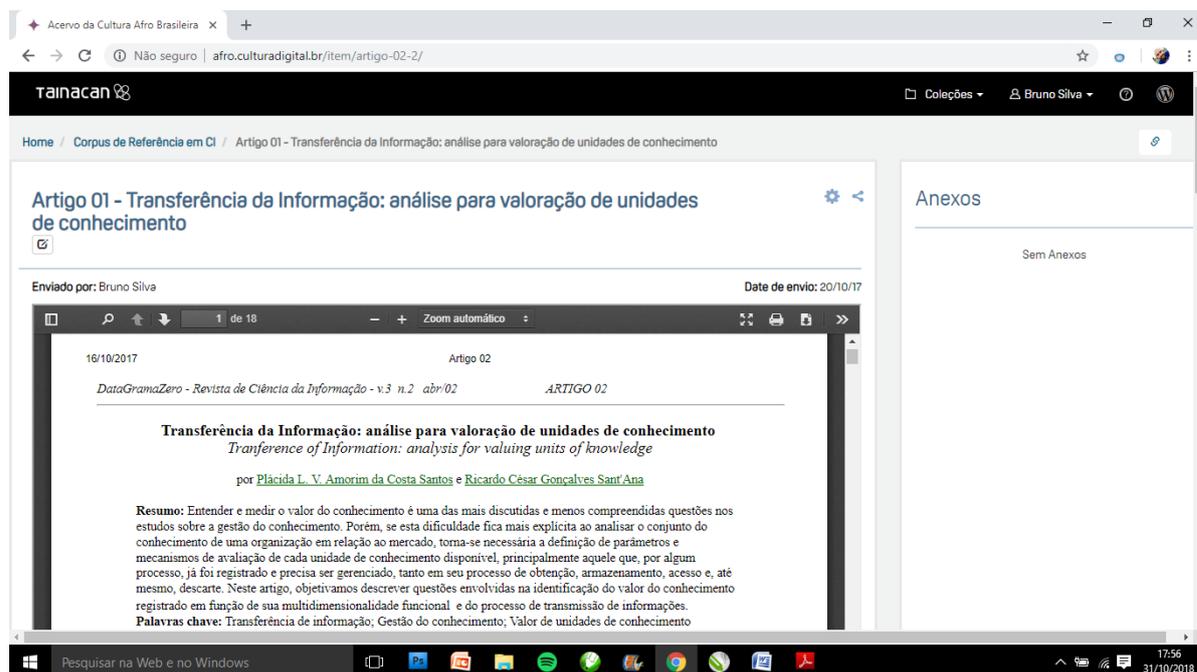


Fonte: Acervo de Cultura Afro Brasileira (2018, *printscreen*)

#### **Passo 4 – Acesso/Inserção das tags aos itens da coleção**

Na última etapa, ao clicar em cada item será possível visualizar o arquivo em formato PDF (**Figura 10**), bem como os metadados (**Figura 11**) referentes ao artigo disponibilizados. Nesse ponto é importante destacar que a etiquetagem/indexação deverá seguir algumas recomendações:

**Figura 10** – Página do item da coleção “Corpus de Referência em CI”



Fonte: Acervo de Cultura Afro Brasileira (2018, *printscreen*)

**Figura 11** – Metadados dos itens da coleção

Metadados	
<b>Identificação</b> Artigo 01	<b>Título</b> Transferência da Informação: análise para valoração de unidades de conhecimento
<b>Autor</b> Plácida L. V. Amorim da Costa Santos Ricardo César Gonçalves Sant'Ana	<b>Descrição</b> Entender e medir o valor do conhecimento é uma das mais discutidas e menos compreendidas questões nos estudos sobre a gestão do conhecimento. Porém, se esta dificuldade fica mais explícita ao analisar o conjunto do conhecimento de uma organização em relação ao mercado, torna-se necessária a definição de parâmetros e mecanismos de avaliação de cada unidade de conhecimento disponível, principalmente aquele que, por algum processo, já foi registrado e precisa ser gerenciado, tanto em seu processo de obtenção, armazenamento, acesso e, até mesmo, descarte. Neste artigo, objetivamos descrever questões envolvidas na identificação do valor do conhecimento registrado em função de sua multidimensionalidade funcional e do processo de transmissão de informações.
<b>Palavras-chave</b>	<b>Publicador</b> Datagramazero

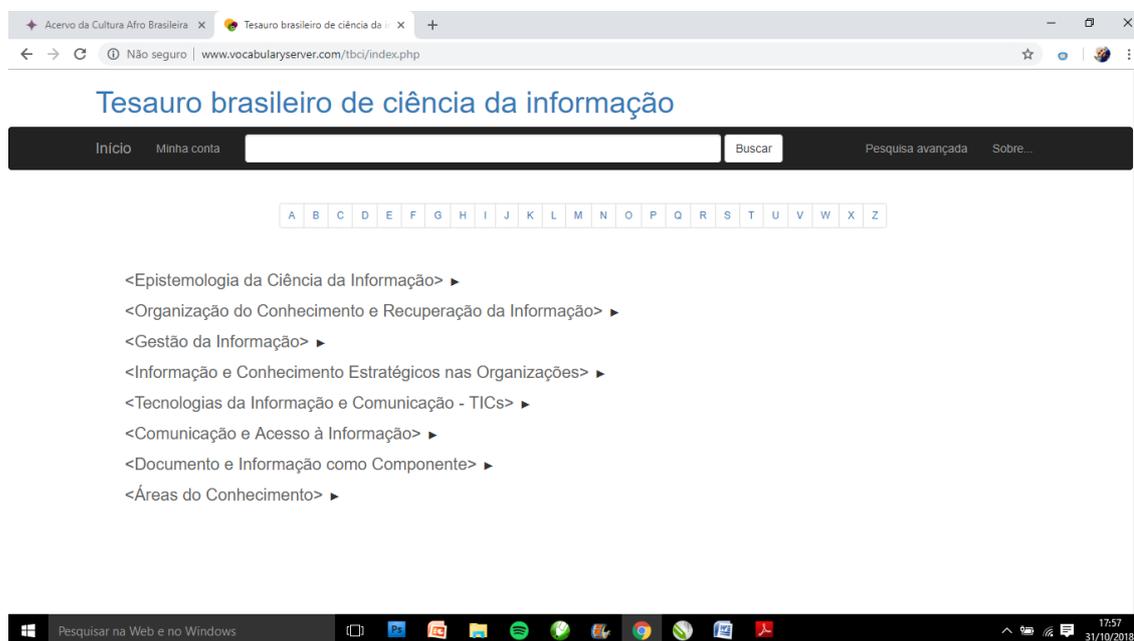
Fonte: Acervo de Cultura Afro Brasileira (2018, *printscreen*)

A seleção das *tags* deverá ser feita a partir da utilização base <http://www.vocabularyserver.com/tbci/index.php> (**Figura 12**).

Obs: A escolha da plataforma se deu em decorrência da viabilidade de se ter, a medida que for digitando uma palavra na busca, a ferramenta de autocomplete listando as opções existentes (**Figura 13**). Além disso, é de suma importância que

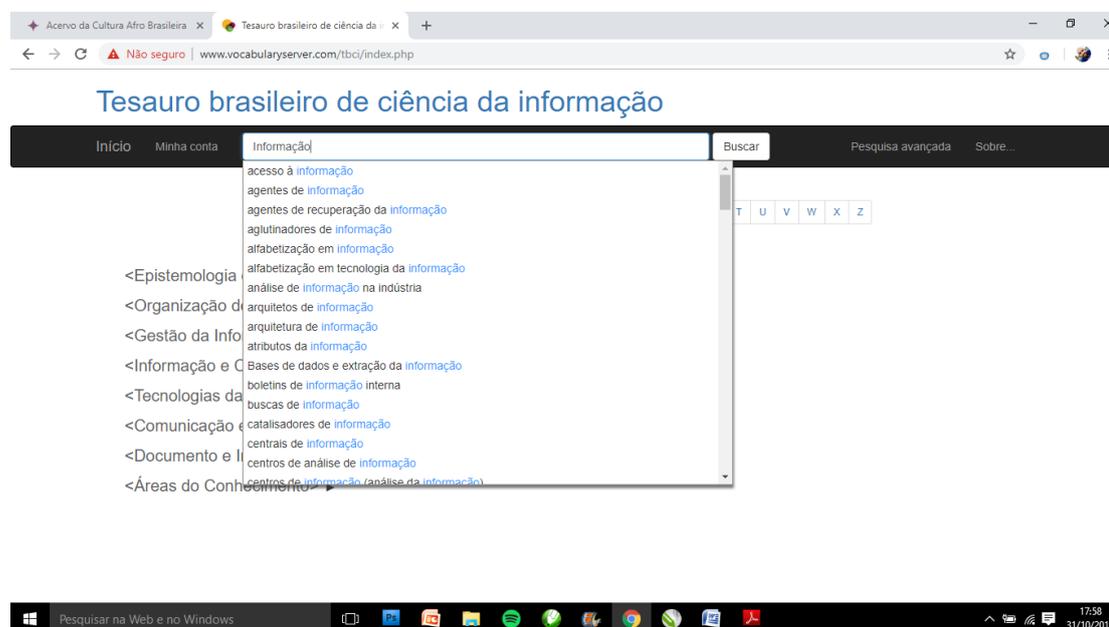
se entenda que todos os termos que constam no endereço em questão fazem parte do Tesouro Brasileiro de Ciência da Informação.

**Figura 12** – Página principal do Tesouro Brasileiro de Ciência da Informação



Fonte: Tesouro Brasileiro de Ciência da Informação (2018, *printscreen*)

**Figura 13** – Resultado da Busca no Tesouro Brasileiro de Ciência da Informação



Fonte: Tesouro Brasileiro de Ciência da Informação (2018, *printscreen*)

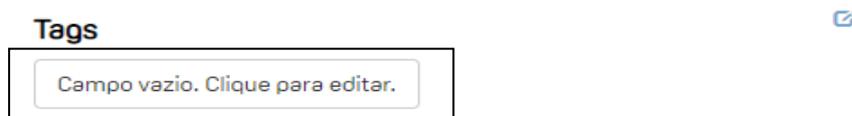
Para cada item da coleção deverá ser escolhidos **5 termos** para representá-los com base nos termos dispostos no endereço apresentado anteriormente. Escolhidos

os 5 termos que irão representar os itens, a próxima ação será colocá-los no campo de metadado *Tags*.

Obs: No metadado “*tags*” é preciso que a inserção seja feita um termo de cada vez, onde ao digitar, seja salvo na sequência, repetindo a ação 5 vezes. Além disso, **COLOCAR TODAS AS LETRAS EM MINÚSCULO.**

A sequência de ações que deverá ser seguida para salvar todas as *tags* pode ser visualizada abaixo:

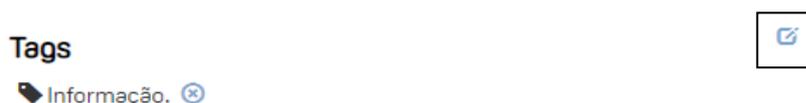
**1º Passo:** Clicar sob espaço “Campo vazio. Clique para editar”



**2º Passo:** Digitar o termo escolhido e em seguida clicar no ícone que se assemelha a um disquete:



**3º Passo:** Após salvar a *tag*, para adicionar uma nova é preciso clicar no ícone “editar”:



**4º Passo:** Aparecerá um campo para digitar uma nova *tag* e nele deverá ser inserido o outro termo escolhido. Essa ação deverá ser repetida até completar a inserção dos 5 termos.

Obs.: Lembrando que para cada termo inserido é preciso salvar clicando no ícone “disquete”.



**Passo 5:** Ao inserir os 5 termos, a visualização deverá ficar assim, só assim as *tags* estarão salvas.

### Tags

 arquivologia  biblioteconomia  ciência da informação  Documentação  Informação.

## CONCEITOS IMPORTANTES NO TAINACAN

Tendo em vista o uso do Tainacan para a indexação de artigos, é importante conhecer os conceitos utilizados pela plataforma:

- **Coleção:** Conjunto de itens agrupados por uma determinada temática;
- **Etiqueta (tags):** É um termo digitado livremente pelo usuário para a descrição de um item sem um controle ou normalização;
- **Assunto:** É um termo que faz parte de um vocabulário controlado utilizado na descrição de um item. Pode estar ordenada em uma lista de termos alfabeticamente ou pode ser estruturado hierarquicamente;
- **Item:** Um item é formado pelo objeto de conteúdo original;
- **Metadado:** É um atributo ou relacionamento de um item ou objeto;  
Ex.: Autor, Descrição, Data, Licença, etc.

Como política básica de catalogação, os campos agrupados como Metadados (Dublin core) foram preenchidos da seguinte forma:

**Autor** – (DC.creator) preencha com o nome de cada autor, informe do primeiro ao último nome, com nomes do meio abreviados pela inicial seguida de ponto.

**Título** – (DC.title) preencha com o título em português

**Descrição** – (DC.description) informe o resumo em português

**Linguagem** – (DC.language) preencha com pt-BR

**Fonte** – (DC.source) informe a referência bibliográfica (use a BRAPCI para obter isto mais facilmente, desprezando a informação de “Disponível em...”. Normalize o nome dos autores)

**Relação** – (DC. Relation) deixar vazio

**Cobertura** – (DC. Coverage) deixar vazio

**Contribuição** – (DC. Contributor) deixar vazio

**Licença** – (DC.Rights) deixar vazio

**Publicador** – (DC.Publisher) informe o nome da revista (Ex.: Ciência da Informação, Datagramazero)

**Data** – (DC.data) informe o ano de publicação do artigo

**Formato** – (DC.format) preencha com pdf ou htm

**Identificação** – (DC.identifier) preencha com o número do artigo

**Tipo** – (DC.type) preencha com: Artigo

**Tags** – (DC.subject) entre com cinco termos autorizados do TBCI, adicionando novo campo para cada assunto, priorize termos mais específicos que apareçam no título, resumo ou palavras-chaves, pesquise no site:

<http://www.vocabularyserver.com/tbci/index.php>

## APÊNDICE C – TERMOS DE INDEXAÇÃO DOS AUTORES E USUÁRIOS- INDEXADORES<sup>4</sup>

ARTIGO	ASSUNTOS	FOLKSONOMIA ASSISTIDA REVISADA	TERMOS COMUNS <sup>5</sup>
<b>Artigo 01</b>	Transferência de informação   Gestão do conhecimento   Valor de unidades de conhecimento	acesso a informação   ciência da informação   competitividade   conhecimento nas organizações   gestão da informação   gestão de documentos   gestão do conhecimento   métodos de pesquisa e análise   repositórios digitais   transferência da informação	transferência da informação   gestão do conhecimento
<b>Artigo 02</b>	Popularização da Ciência   Comunicação Científica	acesso a informação   áreas do conhecimento   ciência da informação   comunicação científica   divulgação científica   educação   gestão da informação   notícias   popularização da ciência   sociedade da informação	popularização da ciência   comunicação científica
<b>Artigo 03</b>	Informação   Valor Informacional   Direito à Informação   Memória Social   Estoque Informacional	acesso a informação   análise de informação na indústria   atributos da informação   biblioteconomia   ciência da informação   direito à informação   ética na informação   história   liberdade de pensamento   recuperação da informação   sociedade da informação	direito à informação
<b>Artigo 04</b>	Arquivos-abertos   Sistema de Publicação   Budapest Open Access Initiative   Acesso Livre   Auto-arquivamento	acesso livre   autoarquivamento de documentos   comunicação científica   direito autoral   literatura científica   OAI-PMH   OAI   publicações científicas: periódicos   repositórios digitais	acesso livre   autoarquivamento de documentos
<b>Artigo 05</b>	Conhecimento Científico   Conhecimento Privado   Conhecimento	acesso a informação   áreas do conhecimento   comunicação científica   comunidades científicas   dados científicos	comunicação científica   divulgação científica

<sup>4</sup> Os termos dos campos **Folksonomia assistida revisada** e **Termos comuns** foram colocados como constam no TBCI. Os erros de mal uso de alguns caracteres, bem como o uso indevido de letras maiúsculas e minúsculas podem ser consultados nos quadros 7 e 8 desta pesquisa.

<sup>5</sup> Os termos comuns correspondem também àqueles termos autorizados segundo o TBCI, bem como os conceitos comuns no contexto de uma indexação semântica.

	Escolar   Democratização da Ciência   Comunicação Científica	divulgação científica   estudantes   gestão da informação   globalização da informação   informação científica e tecnológica   sociedade da informação	
<b>Artigo 06</b>	Tesouro Eletrônico   Mundo do Trabalho   Recuperação da Informação   Interface de Consulta   Sistema de Informação   Interdisciplinaridad e   Interação Humano- Computador (IHC)	ciência da computação   ciência da informação   elaboração de linguagens documentárias   gerenciamento   interação homem-computador   interdisciplinaridade   manutenção de tesouros   mercado de trabalho   recuperação da informação   sistemas de informação	recuperação da informação   interdisciplinaridade   interação homem- computador   manutenção de tesouros   sistemas de informação
<b>Artigo 07</b>	Inteligência Competitiva   Gestão do Conhecimento   Gestão da Informação   Fluxos Informacionais   Transferência da Informação	ciência da informação   economia da informação   fluxo da informação   gerenciamento   gestão da informação.   gestão do conhecimento   informação para negócios   inteligência competitiva   organização do conhecimento   transferência da informação	inteligência competitiva   gestão do conhecimento   gestão da informação   fluxo da informação   transferência da informação
<b>Artigo 08</b>	Informação   Massa Documental   Conceito de Informação   Tecnologia   Registro do Conhecimento	áreas do conhecimento   ciência da informação   conceitos de informação   conteúdos da informação   disseminação da informação   explosão da informação   gestão de documentos   gestão do conhecimento   modelagem do conhecimento   registros bibliográficos   sociedade da informação   tecnologias da informação e comunicação	conceitos de informação   tecnologias da informação e comunicação
<b>Artigo 09</b>	Universidade   Gestão do fluxo de Informação na Universidade   Inteligência Competitiva   Barreiras na Comunicação da Informação   Pecados Informacionais	ciência da informação   comunicação nas organizações   cultura organizacional   fluxo da informação   gestão da informação   informação estratégica   inteligência competitiva   obsolescência da literatura   tomada de decisões   universidades	fluxo da informação   universidades   inteligência competitiva
<b>Artigo 10</b>	Informação Estatística   Nova Economia   Mensuração Estatística   Desajuste	atributos da informação   comunicação   conceitos   economia   inteligência competitiva   métodos matemáticos e	economia   métodos matemáticos e estatísticos

	Conceitual   Metodologia Estatística	estatísticos   probabilidade e estatística   sociedade da informação   sociedade do conhecimento   tecnologias da informação e comunicação	
<b>Artigo 11</b>	Ciência da Informação   Armazenamento e recuperação   Curso em informação   Unidade e especificidade da informação	áreas do conhecimento   armazenamento de dados   ciência da informação   educação   ensino de ciência da informação   gestão da informação   interdisciplinaridade   propriedade intelectual   recuperação da informação   unidades de informação acesso a informação	ciência da informação   ensino de ciência da informação   recuperação da informação
<b>Artigo 12</b>	Ciência da Informação   Biblioteconomia   Sistema de Informação   Arquivologia   Ensino   Pesquisa	arquivologia   biblioteconomia   ciência da informação   educação   ensino de ciência da informação   pesquisa   profissionais de informação   sistemas de informação   teoria da informação	ciência da informação   biblioteconomia   ensino de ciência da informação   sistemas de informação   arquivologia   pesquisa
<b>Artigo 13</b>	Profissional da informação   Informação organizacional   Formação e profissional da informação	ciência da informação   competências profissionais   conhecimento nas organizações   disseminação da informação   economia   formação profissional   gestão do conhecimento nas organizações   habilidades profissionais   mediadores da informação   profissionais de informação   sociedade da informação	formação profissional   profissionais da informação
<b>Artigo 14</b>	Profissionais da informação   Funções sociais   Perfis de profissionais da informação   Inclusão digital   Gestão da informação   Gestão do conhecimento	competências profissionais   fluxo da informação   formação profissional   gestão da informação   gestão do conhecimento   gestores   mediadores da informação   mercado de trabalho   profissionais de informação   sociedade da informação	profissionais da informação   gestão da informação   gestão do conhecimento
<b>Artigo 15</b>	Formação profissional   Ensino e pesquisa	ciência da informação   competências profissionais   educação superior   ensino de ciência da informação   ensino e pesquisa em ciência da informação e áreas afins   formação profissional   instituições de ensino e pesquisa   paradigmas   profissionais de informação   sociedade do conhecimento	ensino de ciência da informação   ensino e pesquisa em ciência da informação e áreas afins   formação profissional

<b>Artigo 16</b>	Ciência da Informação - Formação profissional   Educação Superior no Brasil   Sociedade da Informação - educação   Ciência da Informação e Biblioteconomia   Ciência da Informação - curso de graduação.	biblioteconomia   ciência da informação   educação em ciência da informação   educação superior   ensino de biblioteconomia   ensino e pesquisa em ciência da informação e áreas afins   formação profissional   profissionais de informação   sociedade da informação	biblioteconomia   ciência da informação   educação superior   formação profissional
<b>Artigo 17</b>	Internet e Produção Científica   Novas Tecnologias de Informação e de Comunicação   Produção Científica e Novas Tecnologias	armazenamento de dados   comunicação científica   controle bibliográfico   dados   gestão do conhecimento   inovação tecnológica   internet   produtividade científica   propriedade intelectual   segurança da informação   sociedade da informação   tecnologias da informação e comunicação	comunicação científica   internet   produtividade científica   tecnologias da informação e comunicação
<b>Artigo 18</b>	Estudos Sociais da Ciência   Sociologia da Ciência   Ciência e Sociedade   Tecnologia e Sociedade	categorias   ciência da informação   construtivismo   desenvolvimento de acervos   economia da informação   escolas e correntes filosóficas   planejamento   produtividade científica   relações entre conceitos   sociedade da informação   sociedade do conhecimento   sociedades científicas   taxonomias   tecnologias da informação	sociedade da informação   tecnologias da informação e comunicação
<b>Artigo 19</b>	Inteligência Empresarial   Monitoração Ambiental   Fontes de Informação   Gestão do Conhecimento   Gestão da Informação	ambiente organizacional   centros de informação (bibliotecas especializadas)   competitividade   conhecimento organizacional   gestão da informação   gestão do conhecimento   inteligência competitiva   inteligência empresarial.   monitoramento ambiental   relevância   tecnologias da informação e comunicação	inteligência empresarial   gestão da informação   gestão do conhecimento   monitoramento ambiental

<b>Artigo 20</b>	Ciência da Informação no Brasil   Avaliação 2001 CAPES   Pós-graduação em Ciência da Informação   Pesquisa em Ciência da Informação no Brasil	avaliação   avaliação de publicações científicas   ciência da informação   dissertações e teses   ensino de ciência da informação   pesquisa   pesquisa em ciência da informação   profissionais de informação   programas de pós-graduação   programas de pós-graduação em ciência da informação   projetos de pesquisa.	ciência da informação   pesquisa   ensino e pesquisa em ciência da informação e áreas afins   programas de pós-graduação em ciência da informação
<b>Artigo 21</b>	Leitura-teoria   Cognóscio   Conhecimento-introjeção   Leitura e Sociedade   Informação e Sociedade	ciência cognitiva   inteligência artificial   linguística   multidisciplinaridade   promoção da leitura   sociedade da informação	sociedade da informação
<b>Artigo 22</b>	Paradigma   Holografia   Ciência da Informação   Tecnologia da Informação   Hipertexto   Complexidade   Interatividade   Virtual   Totalidade	ciência da informação   complexidade   hipertextos   organização da informação   paradigmas   processamento de informações   tecnologia da informação   tecnologias da informação e comunicação	paradigmas   ciência da informação   tecnologia da informação   hipertextos   complexidade
<b>Artigo 23</b>	Informação   Memória Social   Espaço Prisional	ciência da informação   disseminação da informação   documentação   interdisciplinaridade   recuperação da informação   relações entre conceitos   sociedade da informação   teoria da informação.   tipos de documento	-
<b>Artigo 24</b>	Contrato Social   Ciência   Pesquisa   Pesquisadores   Autonomia   Ecologia dos Conhecimentos	ciência da informação   comunicação científica   informação científica e tecnológica   institutos de pesquisa   paradigmas   pesquisadores   políticas públicas   transdisciplinaridade. agências de fomento   agências de fomento à pesquisa   bolsas de pesquisa   ciência da informação   frente de pesquisa   gestão de recursos humanos   levantamentos   pesquisa e desenvolvimento   pesquisadores   programas de pós-graduação em	pesquisadores
<b>Artigo 25</b>	Fomento à pesquisa - Ciência da Informação   CNPq - fomento a pesquisa em Ciência da Informação	agências de fomento à pesquisa   bolsas de pesquisa   ciência da informação   frente de pesquisa   gestão de recursos humanos   levantamentos   pesquisa e desenvolvimento   pesquisadores   programas de pós-graduação em	agências de fomento à pesquisa   ciência da informação

		ciência da informação   projetos de pesquisa	
<b>Artigo 26</b>	Compressão Semântica   Monitoramento da Informação   Estoques de Informação   Palavras-chave	compressão de dados   consistência na indexação   fluxo da informação   informação estratégica   medidas de avaliação de sistemas de recuperação da informação   monitoramento   palavras-chaves   precisão   processamento de textos   relevância   revoação   semântica	monitoramento   palavras-chave
<b>Artigo 27</b>	Indicadores Científicos   Política Científica e Tecnológica   Gestão de Ciência e Tecnologia	agências de fomento   análise comparativa   análise de dados   bases de dados   bases de dados cadastrais   bolsas de pesquisa   documentos primários   extração de informação em bases de dados   indicadores de C&T   pesquisadores   produtividade científica   projetos de pesquisa	indicadores de C&T
<b>Artigo 28</b>	Governo Eletrônico   Arquitetura de Sistemas de Informação   Integração de Informações   Gestão de C&T   Bibliotecas Digitais   Plataforma Lattes   Rede ScienTI	bibliotecas digitais   comércio eletrônico   fontes primárias   gestão da informação   governo eletrônico   inovação   marketing da informação   relações entre conceitos   serviços de informação   sistemas de informação   tecnologias da informação e comunicação	bibliotecas digitais   governo eletrônico   sistemas de informação
<b>Artigo 29</b>	Governo Eletrônico   Políticas de Informação   Informação Governamental	acesso a informação   gestão da informação   gestão de conteúdos na web   governo eletrônico   informação governamental   políticas de informação   serviços de informação	governo eletrônico   políticas de informação   informação governamental
<b>Artigo 30</b>	Periódicos eletrônicos   Avaliação de acesso   Arquivo de log de acesso	análise de dados   coleta de dados   comunicação e acesso à informação   estudos de usuários   perfil do usuário   periódicos científicos   periódicos eletrônicos   processos de gestão   publicações de acesso livre   registros de uso	análise de dados   periódicos eletrônicos

		ruído na recuperação da informação   webmetria   world wide web	
<b>Artigo 31</b>	Política de informação   Sociedade da informação   Internet   Institucionalização da informação   Estado	economia   economia da informação   governo eletrônico   infraestrutura de informação   internet   políticas de informação   regimes de informação   relações entre conceitos   sociedade da informação	políticas de informação   sociedade da informação   internet
<b>Artigo 32</b>	Bases de dados   Estratégia de busca   Linguagem controlada   Linguagem natural   Recuperação da informação   Artigo de revisão	base de dados   controle de vocabulário   estratégias de busca   linguagens documentárias   processamento da linguagem natural   recuperação da informação   revisões de literatura   vocabulários controlados	base de dados   estratégias de busca   linguagens documentárias   recuperação da informação
<b>Artigo 33</b>	Biblioteca digital   Biblioteca virtual   Produção científica   Produção bibliográfica   Periódicos	artigos de periódico   bibliotecas digitais   bibliotecas virtuais   citações bibliográficas   informação científica e tecnológica   inovação   produtividade científica   produtividade de periódicos   revisões de literatura   sociedade da informação   tecnologias da informação e comunicação	artigos de periódico   bibliotecas digitais   bibliotecas virtuais
<b>Artigo 34</b>	Informação tecnológica   Transferência de informação   Transferência tecnológica	acesso a informação   ciência e tecnologia de alimentos   comunicação científica   fluxo da informação   informação científica e tecnológica   informação tecnológica   inovação   inovação tecnológica   interação universidade-empresa   pesquisa e desenvolvimento   tecnologias da informação e comunicação   transferência da informação	informação tecnológica   transferência da informação
<b>Artigo 35</b>	XML   HTML   Linguagens de marcação   Internet   Intranet	HTML   internet   intranetes   linguagens de marcação   SGML (uf standard generalized markup language)   XML	html   internet   intranetes   linguagens de marcação   xml
<b>Artigo 36</b>	Bibliometria   Lei de Lotka   Produtividade de autores   Brasil	bibliometria   ciência da informação   disseminação da informação   economia   formação profissional   lei de lotka	bibliometria   lei de lotka   produtividade de autor

		literatura científica   métodos matemáticos e estatísticos   produtividade de autor   produtividade de periódicos   profissionais de informação   acesso a informação   bases de dados   bases de dados factuais   informação financeira   informação jurídica   informação para negócios   recursos de informação na internet   sociedade da informação   tecnologias da informação   tomada de decisões	
<b>Artigo 37</b>	Informação para negócios   Bases de dados		informação para negócios   bases de dados
		acesso a informação   bibliotecas digitais   bibliotecas híbridas   educação á distância   ensino a distância   necessidades de informação   recursos de informação   serviços de biblioteca   serviços de informação   suportes de informação   tecnologias da informação e comunicação   usos da informação	
<b>Artigo 38</b>	Biblioteca híbrida   Tipos de usuários   Bens e serviços		bibliotecas híbridas   serviços de informação
		artigos de revisão   bases de dados   estratégias de busca   interdisciplinaridade   mecanismos de busca   recuperação da informação   revisões de literatura   sistemas de recuperação da informação   sociedade da informação	
<b>Artigo 39</b>	Estratégia de busca   Recuperação da informação   Técnicas de estratégia de busca   Bases de dados   Artigo de revisão		estratégias de buscas   recuperação da informação   bases de dados   artigos de revisão
		ciência da informação   explosão da informação   formação profissional   gestão da informação   gestão do conhecimento   mercado de trabalho   necessidades de informação   profissionais de informação   recursos de informação   sociedade da informação   tecnologias da informação e comunicação	
<b>Artigo 40</b>	Ciência da informação   Gestão da informação		ciência da informação   gestão da informação

<b>Artigo 41</b>	Produção científica   Literatura branca   Literatura cinzenta   Ciência da informação	análise estatística da literatura científica   ciência da informação   comunidades científicas   história da ciência da informação   literatura científica   literatura cinzenta   produtividade científica   programas de pós-graduação   representação do conhecimento   sociedade da informação   tipos de documento   usos da informação	literatura científica   literatura cinzenta   ciência da informação
<b>Artigo 42</b>	Gestão do conhecimento   Capital intelectual   Informação para negócios   Sistemas de informação para negócios   Agentes do conhecimento	bibliotecas   consultores de informação   gestão do conhecimento   gestão financeira   informação para negócios   inteligência competitiva   profissionais da informação   serviços bibliotecários   sistemas de informação   unidades de informação	gestão do conhecimento   informação para negócios   sistemas de informação
<b>Artigo 43</b>	Necessidade de informação tecnológica   Informação tecnológica   Setor industrial   Inovação	análise de informação na indústria   conhecimento nas organizações   informação científica e tecnológica   informação tecnológica   inovação   inovação tecnológica   modelos de bases de dados   necessidades de informação   produtividade científica   tomada de decisões	informação tecnológica   inovação   necessidades de informação
<b>Artigo 44</b>	Gestão do conhecimento   Informação e competitividade   Processos organizacionais	ambiente organizacional   competitividade   conhecimento nas organizações   ensino e pesquisa em ciência da informação e áreas afins   gestão do conhecimento   informação estratégica   inovação   métodos de análise na inteligência competitiva   processos de gestão   tecnologias da informação e comunicação   tomada de decisões	competitividade   conhecimento nas organizações   gestão do conhecimento
<b>Artigo 45</b>	Bibliometria   Cienciometria   Informetria   Webometria   Métodos quantitativos de avaliação	análise quantitativa   bibliometria   cibermetria   cienciometria   cientometria   informetria   métodos matemáticos e estatísticos   webmetria	bibliometria   cienciometria   informetria   webmetria
<b>Artigo 46</b>	bibliometria   Lei de Bradford   Pesquisa operacional	bibliometria   ciência da informação   lei de bradford   métodos matemáticos e	bibliometria   lei de bradford   ciência da informação

	Caos   Ciência da informação   Inferência bayesiana	estatísticos   periódicos   teoria do caos	teoria do caos
<b>Artigo 47</b>	Periódicos eletrônicos   Usabilidade   Novas tecnologias	buscas em linha   ciência da informação   economia da informação   estudos de usuários   indexação automática   periódicos científicos   periódicos eletrônicos   usabilidade   usos da informação   World Wide Web	periódicos eletrônicos   usabilidade
<b>Artigo 48</b>	Uso da informação   Economia da Informação   Modelo genérico	custos   economia da informação   empresas   gestão da informação   informação para negócios   inovação   inteligência competitiva   tecnologias da informação e comunicação   usos da informação	usos da informação   economia da informação
<b>Artigo 49</b>	Monitoramento da informação   Biblioteconomia   Ciência da informação	automação de bibliotecas   bibliometria   biblioteconomia   ciência da informação   descritores   gestão do conhecimento   inteligência competitiva   modelos cognitivos   monitoramento   periódicos científicos   periódicos eletrônicos   representação do conhecimento	biblioteconomia   ciência da informação   monitoramento
<b>Artigo 50</b>	Educação dos bibliotecários   Profissional da informação	representação do conhecimento   bibliotecários   biblioteconomia   educação   educação em biblioteconomia   ensino de biblioteconomia   formação profissional   gestão da informação   gestão do conhecimento   profissionais de informação   sociedade da informação	ensino de biblioteconomia   profissionais de informação
<b>Artigo 51</b>	Acessibilidade   Espaço digital   Bibliotecas   Pessoas portadoras de deficiência   Ajudas técnicas	acessibilidade   acesso a informação   bibliotecas   ciberespaço   comunicação mediada por computador   perfil do usuário   sistemas de informação   tecnologias da informação e comunicação	acessibilidade   bibliotecas
<b>Artigo 52</b>	Estudos de usuários   Educação	ciência da informação   disseminação da informação   estudos de usuários	estudos de usuários   internet   hipertextos

	ambiental   Internet   Hipertexto   Pesquisa participante	hipertextos   internet   usuários	
<b>Artigo 53</b>	Análise de logs   Máquinas de busca   Recuperação de informação   Comportamento de usuários   Estratégias de busca	comportamento do usuário   estratégias de busca   linguagens documentárias   mecanismos de busca   recuperação da informação   serviços de referência	comportamento do usuário   estratégias de busca   mecanismos de busca   recuperação da informação
<b>Artigo 54</b>	Information literacy   Competência em informação   Alfabetização informacional   Biblioteca aprendente   Bibliotecário educador   Sociedade de aprendizagem   Habilidades informacionais.	bibliotecários   bibliotecas   competência em informação   educação   gestão do conhecimento   recursos de informação   sociedade da informação   usos da informação	bibliotecários   bibliotecas   competência em informação
<b>Artigo 55</b>	Profissional da informação   Profissional da informação – habilidades   Perfil e atuação profissional   Mercado de trabalho	competências profissionais   ensino de ciência da informação   formação profissional   inteligência competitiva   mercado de trabalho   profissionais de informação	competências profissionais   profissionais da informação   mercado de trabalho
<b>Artigo 56</b>	Teoria da ciência da informação   Sociologia da informação   História da ciência da informação   Comunicação científica   Responsabilidade social	biblioteconomia   ciência da informação   cientistas da informação   comunicação científica   história da ciência da informação   informação científica e tecnológica   interdisciplinaridade   sociologia   teorias na ciência da informação	teorias na ciência da informação   história da ciência da informação   comunicação científica
<b>Artigo 57</b>	Recuperação da informação   Inteligência científica   Integração dos conhecimentos   Estado   Ciência   Sociedade   Informação	áreas do conhecimento   ciência da informação   ensino e pesquisa em ciência da informação e áreas afins   epistemologia da ciência da informação   paradigmas   pesquisa   recuperação da informação   regimes de informação   sociedade da informação	recuperação da informação   sociedade da informação

<b>Artigo 58</b>	Ciência da informação   Ciência cognitiva   Processamento da informação   Categorização   Indexação   Recuperação da informação   Interação homem-computador	categorização automática de textos   ciência cognitiva   ciência da informação   indexação   interação homem-computador   processamento de informações   recuperação da informação.	ciência da informação   ciência cognitiva   processamento de informações   indexação   recuperação da informação.   interação homem-computador
<b>Artigo 59</b>	Comunicação científica   bibliometria   comunicação e educação   estudos de citações   cienciometria	análise de citação   bibliometria   ciência da informação   cientometria   comunicação científica   educação   epistemologia   epistemologia arqueológica   produtividade científica	comunicação científica   bibliometria   cientometria
<b>Artigo 60</b>	Inteligência competitiva   Internet   Monitoramento de fontes de informação   Agentes inteligentes	acesso a informação   agentes inteligentes   automação   gestão da informação   inovação   inteligência competitiva   internet   monitoramento   usos da informação	agentes inteligentes   inteligência competitiva   internet   monitoramento

Fonte: Dados da Pesquisa (2018)

**APÊNDICE D – TAGS POR FREQUÊNCIA DE UTILIZAÇÃO PARA DESCRIÇÃO  
DOS DOCUMENTOS NA FOLKSONOMIA GERADA**

<b>Tags</b>	<b>Qtd.</b>
Acessibilidade	1
acesso a informação	12
agências de fomento	2
agências de fomento à pesquisa	1
Agentes Inteligentes	1
ambiente organizacional	2
análise comparativa	1
análise de citação	1
Análise de dados	2
análise de informação na indústria	2
análise estatística da literatura científica	1
Análise quantitativa	1
Áreas do conhecimento	5
armazenamento de dados	2
Arquivologia	1
Artigos de periódico	1
artigos de revisão	1
atributos da informação	2
autoarquivamento de documentos	1
Automação	1
automação de bibliotecas	1
Avaliação	1
avaliação de publicações científicas	1
bases de dados	4
Bases de Dados Cadastrais	1
bases de dados factuais	1
Bibliometria	5
Bibliotecários	2
Bibliotecas	3
Bibliotecas digitais	3
bibliotecas híbridas	1
Bibliotecas virtuais	1
biblioteconomia	6
Bolsas de Pesquisa	2
buscas em linha	1
Categorias	1
categorização automática de textos	1
centros de informação (bibliotecas especializadas)	1
Ciberespaço	1
Cibermetria	1
ciência cognitiva	2
ciência da computação	1

ciência da informação	29
ciência e tecnologia de alimentos	1
Cienciometria	1
cientistas da informação	1
cientometria	2
citações bibliográficas	1
coleta de dados	1
Comércio eletrônico	1
competência em informação	1
competências profissionais	4
competitividade	3
Complexidade	1
comportamento do usuário	1
Compressão de Dados	1
comunicação	1
Comunicação científica	8
comunicação e acesso à informação	1
comunicação mediada por computador	1
comunicação nas organizações	1
comunidades científicas	2
Conceitos	1
conceitos de informação	1
conhecimento nas organizações	4
Conhecimento organizacional	1
consistência na indexação	1
construtivismo	1
consultores de informação	1
conteúdos da informação	1
controle bibliográfico	1
controle de vocabulário	1
cultura organizacional	1
Custos	1
Dados	1
dados científicos	1
Descritores	1
desenvolvimento de acervos	1
Direito à informação	1
direito autoral	1
Disseminação da informação.	5
dissertações e teses	1
divulgação científica	2
Documentação	1
documentos primários	1
Economia	4
economia da informação	5
Educação	6

Educação á distância	1
educação em biblioteconomia	1
Educação em ciência da informação	1
educação superior	2
elaboração de linguagens documentárias	1
Empresas	1
ensino a distância	1
Ensino de biblioteconomia	2
ensino de ciência da informação	5
Ensino e Pesquisa em Ciência da Informação e Áreas Afins	4
Epistemologia	1
epistemologia arqueológica	1
Epistemologia da Ciência da Informação	1
escolas e correntes filosóficas	1
estratégias de busca	3
Estudantes	1
Estudos de Usuários	3
ética na informação	1
explosão da informação	2
extração de informação em bases de dados	1
fluxo da informação	5
fontes primárias	1
formação profissional	8
frente de pesquisa	1
gerenciamento	2
Gestão da informação.	14
gestão de conteúdos na web	1
gestão de documentos	2
gestão de recursos humanos	1
gestão do conhecimento	12
gestão do conhecimento nas organizações	1
gestão financeira	1
Gestores	1
globalização da informação	1
Governo eletrônico	3
habilidades profissionais	1
Hipertextos	2
História	1
história da ciência da informação	2
HTML	1
Indexação	1
indexação automática	1
Indicadores de C&T	1
informação científica e tecnológica	6
informação estratégica	3

informação financeira	1
informação governamental	1
informação jurídica	1
informação para negócios	4
Informação tecnológica.	2
Informetria	1
infraestrutura de informação	1
Inovação	7
Inovação Tecnológica	3
instituições de ensino e pesquisa	1
institutos de pesquisa	1
inteligência artificial	1
Inteligência Competitiva;	9
Inteligência empresarial.	1
interação homem-computador	2
interação universidade-empresa	1
interdisciplinaridade	5
Internet	5
Intranetes	1
lei de Bradford	1
Lei de lotka	1
Levantamentos	1
liberdade de pensamento	1
linguagens de marcação	1
linguagens documentárias	2
Linguística	1
literatura científica	3
Literatura cinzenta	1
manutenção de tesouros	1
marketing da informação	1
mecanismos de busca	2
Mediadores da informação	2
medidas de avaliação de sistemas de recuperação da informação	1
mercado de trabalho	4
métodos de análise na inteligência competitiva	1
métodos de pesquisa e análise	1
métodos matemáticos e estatísticos	4
Métodos quantitativos de avaliação	1
modelagem do conhecimento	1
modelos cognitivos	1
modelos de bases de dados	1
monitoramento	3
monitoramento ambiental	1
multidisciplinaridade	1
necessidades de informação	3

Notícias	1
OAI-PMH	1
OAIS	1
obsolescência da literatura	1
Organização da informação	1
Organização do conhecimento	1
palavras-chaves	1
paradigmas	4
perfil do usuário	2
Periódicos	1
periódicos científicos	3
periódicos eletrônicos	3
Pesquisa	3
pesquisa e desenvolvimento	2
Pesquisa em ciência da informação.	1
pesquisadores	3
planejamento	1
políticas de informação	2
Políticas públicas	1
popularização da ciência	1
Precisão	1
probabilidade e estatística	1
processamento da linguagem natural	1
processamento de informações	2
processamento de textos	1
processos de gestão	2
Produtividade científica	7
produtividade de autor	1
produtividade de periódicos	2
profissionais da informação	11
Programas de pós-graduação	2
programas de pós-graduação em ciência da informação	2
Projetos de pesquisa.	3
promoção da leitura	1
propriedade intelectual	2
Publicações científicas: periódicos	1
publicações de acesso livre	1
Recuperação da informação.	9
recursos de informação	3
recursos de informação na Internet	1
regimes de informação	2
registros bibliográficos	1
registros de uso	1
relações entre conceitos	4
Relevância	2
repositórios digitais	2

representação do conhecimento	2
revisões de literatura	3
Revocação	1
ruído na recuperação da informação	1
segurança da informação	1
Semântica	1
serviços bibliotecários	1
serviços de biblioteca	1
Serviços de informação.	3
serviços de referência	1
SGML (UF standard generalized markup language)	1
sistemas de informação	5
sistemas de recuperação da informação	1
Sociedade da Informação	21
sociedade do conhecimento	3
sociedades científicas	1
Sociologia	1
suportes de informação	1
taxonomias	1
Tecnologia da informação	3
tecnologias da informação e comunicação	13
Teoria da informação.	2
teoria do caos	1
teorias na ciência da informação	1
tipos de documento	2
tomada de decisões	4
Transdisciplinaridade.	1
TRANSFERÊNCIA DA INFORMAÇÃO	3
Unidades de informação	2
universidades	1
Usabilidade	1
usos da informação	6
Usuários	1
vocabulários controlados	1
Webmetria	2
World Wide Web	2
XML	1

Fonte: Dados da Pesquisa (2018)

**APÊNDICE E – ÍNDICES DE AVALIAÇÃO, PRECISÃO, REVOCAÇÃO E  
MEDIDA F**

ID	Nº. DE PALAVRAS-CHAVE	Nº. DE TERMOS DA FOLKSONOMIA ASSISTIDA	Nº. DE TERMOS EM COMUM	COEFICIENTE DE 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%	150%	50%
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%

<b>Artigo 34</b>	3	12	2	15%	17%	67%	27%
<b>Artigo 35</b>	5	6	5	83%	83%	100%	91%
<b>Artigo 36</b>	4	11	3	25%	27%	75%	40%
<b>Artigo 37</b>	2	10	2	20%	20%	100%	33%
<b>Artigo 38</b>	3	12	2	15%	17%	67%	27%
<b>Artigo 39</b>	5	9	4	40%	44%	80%	57%
<b>Artigo 40</b>	2	11	2	18%	18%	100%	31%
<b>Artigo 41</b>	4	12	3	23%	25%	75%	38%
<b>Artigo 42</b>	5	10	3	25%	30%	60%	40%
<b>Artigo 43</b>	4	10	2	17%	20%	50%	29%
<b>Artigo 44</b>	3	11	3	27%	27%	100%	43%
<b>Artigo 45</b>	5	8	4	44%	50%	80%	62%
<b>Artigo 46</b>	6	6	4	50%	67%	67%	67%
<b>Artigo 47</b>	3	10	2	18%	20%	67%	31%
<b>Artigo 48</b>	3	9	2	20%	22%	67%	33%
<b>Artigo 49</b>	3	12	3	25%	25%	100%	40%
<b>Artigo 50</b>	2	10	2	20%	20%	100%	33%
<b>Artigo 51</b>	5	8	2	18%	25%	40%	31%
<b>Artigo 52</b>	5	6	3	38%	50%	60%	55%
<b>Artigo 53</b>	5	6	4	57%	67%	80%	73%
<b>Artigo 54</b>	7	8	3	25%	38%	43%	40%
<b>Artigo 55</b>	4	6	3	43%	50%	75%	60%
<b>Artigo 56</b>	5	9	3	27%	33%	60%	43%
<b>Artigo 57</b>	7	9	2	14%	22%	29%	25%
<b>Artigo 58</b>	7	7	6	75%	86%	86%	86%
<b>Artigo 59</b>	5	9	3	27%	33%	60%	43%
<b>Artigo 60</b>	4	9	4	44%	44%	100%	62%
<b>Média</b>	4,43	9,70	2,85	28%	32%	68%	41%
<b>Desvio</b>	1,52	1,89	1,23	16%	18%	29%	18%
<b>Mínimo</b>	2	6	0	0%	0%	0%	0%
<b>Máximo</b>	9	14	6	83%	86%	150%	91%

Fonte: Dados da Pesquisa (2018)

## APÊNCIDE F – FOLKSONOMIA ASSISTIDA REVISADA POR GRUPO<sup>6</sup>

ID	GRUPO 1	GRUPO 2	GRUPO 3
<b>Artigo 01</b>	acesso a informação   gestão da informação   gestão do conhecimento   métodos de pesquisa e análise   transferência da informação	transferência da informação   gestão do conhecimento   conhecimento nas organizações   competitividade   repositórios digitais	ciência da informação   gestão da informação   gestão de documentos   gestão do conhecimento   transferência da informação
<b>Artigo 02</b>	áreas do conhecimento   comunicação científica   educação   notícias   popularização da ciência	divulgação científica   comunicação científica   sociedade da informação   ciência da informação   acesso à informação	ciência da informação   comunicação científica   divulgação científica   gestão da informação   popularização da ciência
<b>Artigo 03</b>	acesso a informação   análise de informação na indústria   atributos da informação   direito à informação   recuperação da informação	liberdade de pensamento   direito à informação   sociedade da informação   história   acesso à informação	acesso à informação   biblioteconomia   ciência da informação   direito à informação   ética na informação
<b>Artigo 04</b>	acesso livre   autoarquivamento de documentos   comunicação científica   direito autoral   publicações científicas: periódicos	comunicação científica    autoarquivamento de documentos   acesso livre   literatura científica   repositórios digitais	acesso livre   autoarquivamento de documentos   comunicação científica   OAI-PMH   OAIS
<b>Artigo 05</b>	acesso a informação   comunicação científica   comunidades científicas   dados científicos   estudantes	comunicação científica   acesso à informação   divulgação científica   sociedade da informação   informação científica e tecnológica	acesso a informação   áreas do conhecimento   gestão da informação   globalização da informação   sociedade da informação
<b>Artigo 06</b>	interação homem-computador   interdisciplinaridade   mercado de trabalho   recuperação da informação   sistemas de informação	elaboração de linguagens documentárias   recuperação da informação   manutenção de tesouros   interdisciplinaridade   interação homem-computador	ciência da computação   ciência da informação   gerenciamento   interdisciplinaridade   recuperação da informação
<b>Artigo 07</b>	business information   economia da informação   gestão do conhecimento   inteligência competitiva   transferência da informação	inteligência competitiva   gestão do conhecimento   gestão da informação   fluxo da informação   transferência da informação	ciência da informação   gerenciamento   gestão da informação   gestão do conhecimento   organização do conhecimento

<sup>6</sup> Os termos que constam em vermelho não são autorizados pelo TBCI e foram removidos

<b>Artigo 08</b>	ciência da informação   conteúdos da informação   disseminação da informação   explosão da informação   sociedade da informação	conceitos de informação   registros bibliográficos   modelagem do conhecimento   tecnologias da informação e comunicação   ciência da informação	áreas do conhecimento   ciência da informação   conceitos de informação   gestão de documentos   gestão do conhecimento
<b>Artigo 09</b>	ciência da informação   fluxo da informação   informação estratégica   inteligência competitiva   tomada de decisões	universidades   gestão da informação   inteligência competitiva   comunicação nas organizações   tomada de decisões	cultura organizacional   fluxo da informação   gestão da informação   inteligência competitiva   obsolescência da literatura
<b>Artigo 10</b>	atributos da informação   métodos matemáticos e estatísticos   probabilidade e estatística   sociedade da informação   sociedade do conhecimento	métodos matemáticos e estatísticos   economia   conceitos   sociedade da informação   tecnologias da informação e comunicação	comunicação   inteligência competitiva   probabilidade e estatística   sociedade da informação   tecnologias da informação e comunicação
<b>Artigo 11</b>	áreas do conhecimento   armazenamento de dados   ciência da informação   propriedade intelectual   recuperação da informação.	ciência da informação   recuperação da informação   ensino de ciência da informação   unidades de informação   interdisciplinaridade	armazenamento de dados   ciência da informação   educação   gestão da informação   recuperação da informação
<b>Artigo 12</b>	arquivologia   biblioteconomia   ciência da informação   educação   pesquisa	ciência da informação   biblioteconomia   sistemas de informação   arquivologia   ensino de ciência da informação	acesso à informação   biblioteconomia   educação   profissionais de informação   teoria da informação
<b>Artigo 13</b>	ciência da informação   disseminação da informação   economia   formação profissional   profissional da informação	profissionais de informação   habilidades profissionais   formação profissional   gestão do conhecimento nas organizações   sociedade da informação	competências profissionais   conhecimento nas organizações   formação profissional   mediadores da informação   profissionais de informação
<b>Artigo 14</b>	fluxo da informação   gestão da informação   gestão do conhecimento   profissionais da informação   sociedade da informação	profissionais de informação   competências profissionais   gestores   sociedade da informação   mercado de trabalho	formação profissional   gestão da informação   gestão do conhecimento   mediadores da informação   profissionais de informação

<b>Artigo 15</b>	educação superior   formação profissional   paradigmas   profissional da informação   sociedade do conhecimento	formação profissional   ciência da informação   ensino de ciência da informação   instituições de ensino e pesquisa   paradigmas	competências profissionais   educação superior   ensino e pesquisa em ciência da informação e áreas afins   formação profissional   profissionais de informação
<b>Artigo 16</b>	biblioteconomia   ciência da informação   formação profissional   profissionais de informação   sociedade da informação	ciência da informação   formação profissional   biblioteconomia   sociedade da informação   ensino de ciência da informação	educação em ciência da informação   educação superior   ensino de biblioteconomia   ensino e pesquisa em ciência da informação e áreas afins   formação profissional
<b>Artigo 17</b>	armazenamento de dados   gestão do conhecimento   inovação tecnológica   internet   tecnologia da informação	internet   tecnologias da informação e comunicação   produtividade científica   segurança da informação   sociedade da informação	comunicação científica   controle bibliográfico   dados   propriedade intelectual   tecnologias da informação e comunicação
<b>Artigo 18</b>	categorias   ciência da informação   desenvolvimento de acervos   sociedade da informação   tecnologias da informação	sociologia do conhecimento   sociedades científicas   relações entre conceitos   economia da informação   produtividade científica	construtivismo   escolas e correntes filosóficas   planejamento   sociedade do conhecimento   taxonomias
<b>Artigo 19</b>	conhecimento organizacional   gestão da informação   gestão do conhecimento   inteligência competitiva   inteligência empresarial	inteligência competitiva   monitoramento ambiental   tecnologias da informação e comunicação   gestão do conhecimento   gestão da informação	ambiente organizacional   centros de informação (bibliotecas especializadas)   competitividade   inteligência competitiva   relevância
<b>Artigo 20</b>	ciência da informação   pesquisa em ciência da informação   pós-graduação   programas de pós- graduação   projetos de pesquisa	ensino de ciência da informação   ciência da informação   pesquisa   avaliação de publicações científicas   programas de pós- graduação em ciência da informação	avaliação   dissertações e teses   profissionais de informação   programas de pós- graduação em ciência da informação   projetos de pesquisa

<b>Artigo 21</b>	código   conhecimento   contexto   informação e sociedade   promoção da leitura	multidisciplinaridade   inteligência artificial   linguística   ciência cognitiva   sociedade da informação	estudo da leitura   estudos mnemônicos   sociedade da informação   sociologia da leitura   teorias da leitura
<b>Artigo 22</b>	ciência da informação   hipertexto   organização da informação   tecnologia da informação   virtual	paradigmas   ciência da informação   tecnologias da informação e comunicação   hipertextos   complexidade	hipertextos   paradigmas   processamento de informações   tecnologias da informação e comunicação   vannevar bush
<b>Artigo 23</b>	ciência da informação   disseminação da informação   documento   recuperação da informação   teoria da informação	relações entre conceitos   interdisciplinaridade   disseminação da informação   tipos de documento   sociedade da informação	documentação   espaço prisional   informação histórica   memória social   Rio de Janeiro
<b>Artigo 24</b>	ciência da informação   ciência e tecnologia   ciências sociais   pesquisadores   transdisciplinaridade	comunicação científica   institutos de pesquisa   pesquisadores   políticas públicas   paradigmas	autonomia   conhecimento científico   contrato social   informação científica e tecnológica
<b>Artigo 25</b>	agências de fomento   bolsas de pesquisa   frente de pesquisa   pesquisa e desenvolvimento   projetos de pesquisa	agências de fomento à pesquisa   ciência da informação   bolsas de pesquisa   programas de pós- graduação em ciência da informação   pesquisadores	agências de fomento   bolsas de pesquisa   ciência da informação   gestão de recursos humanos   levantamentos
<b>Artigo 26</b>	consistência na indexação   medidas de avaliação de sistema de recuperação da informação   monitoramento   precisão   revocação	informação estratégica   semântica   monitoramento   palavras-chave   fluxo da informação	compressão de dados   monitoramento   processamento de textos   relevância   semântica
<b>Artigo 27</b>	análise de dados   bolsas de pesquisa   indicadores C&T   produtividade em pesquisa   projetos de pesquisa	bolsas de pesquisa   agências de fomento   extração de informação em bases de dados   indicadores de C&T   produtividade científica	análise comparativa   bases de dados   bases de dados cadastrais   documentos primários   pesquisadores

<b>Artigo 28</b>	bibliotecas digitais   comércio eletrônico   governo eletrônico   marketing da informação   sistemas de informação	governo eletrônico   sistemas de informação   bibliotecas digitais   gestão da informação   relações entre conceitos	fontes primárias   inovação   serviços de informação   sistemas de informação   tecnologias da informação e comunicação
<b>Artigo 29</b>	gestão de conteúdos na web   governo eletrônico   informação governamental   políticas de informação   serviços de informação	governo eletrônico   políticas de informação   informação governamental   acesso à informação   gestão da informação	acesso a informação   gestão da informação   governo eletrônico   informação governamental   políticas de informação
<b>Artigo 30</b>	coleta de dados   periódicos científicos   periódicos eletrônicos   processos de gestão   registros de uso	periódicos eletrônicos   publicações de acesso livre   análise de dados   estudos de usuários   World Wide Web	comunicação e acesso à informação   perfil do usuário   periódicos eletrônicos   ruído na recuperação da informação   webmetria
<b>Artigo 31</b>	economia   infraestrutura de informação   internet   regimes de informação   sociedade da informação	políticas de informação   sociedade da informação   internet   governo eletrônico   relações entre conceitos	economia da informação   infraestrutura de informação   internet   políticas de informação   sociedade da informação
<b>Artigo 32</b>	bases de dados   estratégias de busca   recuperação da informação   revisões de literatura   vocabulários controlados	bases de dados   estratégias de busca   linguagens documentárias   recuperação da informação   revisões de literatura	base de dados   controle de vocabulário   linguagens documentárias   processamento da linguagem natural   recuperação da informação
<b>Artigo 33</b>	informação científica e tecnológica   informação   inovação   sociedade da informação   tecnologias da informação e comunicação	bibliotecas digitais   bibliotecas virtuais   produtividade científica   revisões de literatura   artigos de periódico	artigos de periódico   bibliotecas digitais   bibliotecas virtuais   citações bibliográficas   produtividade de periódicos
<b>Artigo 34</b>	acesso à informação   fluxo da informação   inovação   informação tecnológica   inovação tecnológica	informação científica e tecnológica   transferência da informação   interação universidade- empresa   comunicação científica   tecnologias da informação e comunicação	ciência e tecnologia de alimentos   informação científica e tecnológica   inovação tecnológica   pesquisa e desenvolvimento   transferência da informação

<b>Artigo 35</b>	HTML   internet   intranetes   linguagens de marcação   XML	XML   HTML   linguagens de marcação   internet   intranetes	HTML   internet   linguagens de marcação   SGML (UF standard generalized markup language)   XML
<b>Artigo 36</b>	ciência da informação   disseminação da informação   economia   formação profissional   profissionais de informação	bibliometria   lei de Lotka   produtividade de autor   métodos matemáticos e estatísticos   literatura científica	bibliometria   lei de Lotka   literatura científica   produtividade de autor   produtividade de periódicos
<b>Artigo 37</b>	acesso à informação   bases de dados   informação para negócios   sociedade da informação   tecnologias da informação	bases de dados   informação para negócios   tomada de decisões   informação financeira   recursos de informação na internet	bases de dados   bases de dados factuais   informação financeira   informação jurídica   informação para negócios
<b>Artigo 38</b>	acesso à informação   bibliotecas híbridas   serviços de informação   tipo de usuário   usos da informação	bibliotecas híbridas   ensino a distância   tecnologias da informação e comunicação   necessidades de informação   serviços de biblioteca	bibliotecas digitais   bibliotecas híbridas   educação à distância   recursos de informação   suportes de informação
<b>Artigo 39</b>	bases de dados   estratégias de busca   recuperação da informação   sistemas de recuperação da informação   sociedade da informação	estratégias de busca   recuperação da informação   bases de dados   revisões de literatura   mecanismos de busca	artigos de revisão   bases de dados   estratégias de busca   interdisciplinaridade   recuperação da informação
<b>Artigo 40</b>	formação profissional   gestão da informação   gestão do conhecimento   profissionais da informação   recursos de informação	gestão da informação   ciência da informação   profissionais de informação   mercado de trabalho   sociedade da informação	ciência da informação   explosão da informação   gestão da informação   necessidades de informação   tecnologias da informação e comunicação
<b>Artigo 41</b>	ciência da informação   literatura cinzenta   sociedade da informação   tipos de documento   usos da informação	literatura científica   literatura cinzenta   programas de pós- graduação   ciência da informação   produtividade científica	análise estatística da literatura científica   comunidades científicas   história da ciência da informação   literatura cinzenta   representação do conhecimento

<b>Artigo 42</b>	bibliotecas   gestão do conhecimento   profissionais da informação   sistemas de informação   unidades de informação	informação para negócios   consultores de informação   gestão do conhecimento   sistemas de informação   inteligência competitiva	consultores de informação   gestão financeira   informação para negócios   inteligência competitiva   serviços bibliotecários
<b>Artigo 43</b>	conhecimento nas organizações   informação científica e tecnológica   inovação   necessidades de informação   tomada de decisões	informação científica e tecnológica   necessidades de informação   produtividade científica   inovação   tomada de decisões	análise de informação na indústria   informação tecnológica   inovação tecnológica   modelos de bases de dados   necessidades de informação
<b>Artigo 44</b>	competitividade   gestão do conhecimento   informação estratégica   inovação   processos de gestão	gestão do conhecimento   informação estratégica   ambiente organizacional   tecnologias da informação e comunicação   conhecimento nas organizações	ensino e pesquisa em ciência da informação e áreas afins   gestão do conhecimento   métodos de análise na inteligência competitiva   processos de gestão   tomada de decisões
<b>Artigo 45</b>	análise quantitativa   bibliometria   cientometria   informetria   webmetria	métodos matemáticos e estatísticos   bibliometria   cientometria   informetria   webmetria	bibliometria   cibermetria   cientometria   informetria   métodos quantitativos de avaliação
<b>Artigo 46</b>	bibliometria   ciência da informação   lei de Bradford   periódicos   teoria do caos	bibliometria   lei de Bradford   teoria do caos   ciência da informação   métodos matemáticos e estatísticos	bibliometria   ciência da informação   lei de Bradford   métodos matemáticos e estatísticos   teoria do caos
<b>Artigo 47</b>	ciência da informação   economia da informação   periódicos eletrônicos   usabilidade (programas de computador)   World Wide Web	periódicos eletrônicos   periódicos científicos   buscas em linha   indexação automática   usabilidade	ciência da informação   estudos de usuários   periódicos científicos   periódicos eletrônicos   usos da informação
<b>Artigo 48</b>	custos   economia da informação   empresas   inovação   usos da informação	usos da informação   economia da informação   inteligência competitiva   tecnologias da informação e comunicação   gestão da informação	economia da informação   informação para negócios   inovação   inteligência competitiva   usos da informação

<b>Artigo 49</b>	automação de bibliotecas   biblioteconomia   ciência da informação   gestão do conhecimento   representação do conhecimento	monitoramento   modelos cognitivos   bibliometria   periódicos científicos   inteligência competitiva	bibliometria   ciência da informação   descritores   monitoramento   periódicos eletrônicos
<b>Artigo 50</b>	educação   formação profissional   gestão do conhecimento   profissionais de informação   sociedade da informação	profissionais de informação   ensino de biblioteconomia   educação   sociedade da informação   gestão da informação	bibliotecários   biblioteconomia   educação em biblioteconomia   ensino de biblioteconomia   profissionais de informação
<b>Artigo 51</b>	acessibilidade   acesso à informação   bibliotecas   ciberespaço   sistemas de informação	acessibilidade   bibliotecas   comunicação mediada por computador   perfil do usuário   tecnologias da informação e comunicação	-
<b>Artigo 52</b>	ciência da informação   estoques informacionais   estrutura informacional   hipertexto   usuários	estudos de usuários   internet   hipertextos   ciência da informação   disseminação da informação	-
<b>Artigo 53</b>	comportamento do usuário   estratégias de busca   mecanismos de busca   recuperação da informação   serviços de referência	mecanismos de busca   recuperação da informação   comportamento do usuário   estratégias de busca   linguagens documentárias	-
<b>Artigo 54</b>	bibliotecários   bibliotecas   gestão do conhecimento   recursos de informação   sociedade da informação	competência em informação   sociedade da informação   bibliotecas   usos da informação   educação	-
<b>Artigo 55</b>	competências profissionais   formação profissional   mercado de trabalho   profissão e mercado de trabalho   profissionais de informação	profissionais de informação   competências profissionais   mercado de trabalho   inteligência competitiva   ensino de ciência da informação	-

<b>Artigo 56</b>	ciência da informação   cientistas da informação   comunicação científica   história da ciência da informação   teorias na ciência da informação	história da ciência da informação   ciência da informação   comunicação científica   informação científica e tecnológica   sociologia	biblioteconomia   cientistas da informação   comunicação científica   história da ciência da informação   interdisciplinaridade
<b>Artigo 57</b>	áreas do conhecimento   ciência da informação   pesquisa   recuperação da informação   sociedade da informação	ciência da informação   pesquisa   recuperação da informação   paradigmas   regimes de informação	ciência da informação   ensino e pesquisa em ciência da informação e áreas afins   epistemologia da ciência da informação   recuperação da informação   regimes de informação
<b>Artigo 58</b>	categorização automática de textos   ciência cognitiva   ciência da informação   interação homem- computador   recuperação da informação	ciência cognitiva   ciência da informação   recuperação da informação   interação homem- computador   processamento de informações	ciência cognitiva   ciência da informação   indexação   processamento de informações   recuperação da informação
<b>Artigo 59</b>	análise de citação   bibliometria   ciência da informação   comunicação científica   epistemologia arqueológica	produtividade científica   bibliometria   análise de citação   educação   comunicação científica	bibliometria   cientometria   comunicação científica   educação   epistemologia
<b>Artigo 60</b>	gestão da informação   inovação   inteligência competitiva   internet   acesso à informação	inteligência competitiva   agentes inteligentes   internet   monitoramento   gestão da informação	agentes inteligentes   automação   inteligência competitiva   internet   usos da informação

Fonte: Dados da Pesquisa (2018)

## ANEXO A – CORPUS DE REFERÊNCIA DE SOUZA (2005)

O presente anexo apresenta os artigos que compõem o *corpus de referência*, nele consta: o título, os autores, o resumo, as palavras-chave, a edição da revista que os contêm e o endereço eletrônico aonde pôde ser encontrado na Internet.

### I) Revista DataGramaZero (29 artigos)

la) *DataGramaZero - Revista de Ciência da Informação - v.3 n.2 abr/02*  
Disponível no endereço: <http://www.dgz.org.br/abr02/index.htm>

#### **Artigo 1: Transferência da Informação: análise para valoração de unidades de conhecimento**

**Autores**

**Autoria:** Plácida L. V. Amorim da Costa Santos e Ricardo César Gonçalves Sant'Ana

**Resumo:** Entender e medir o valor do conhecimento é uma das mais discutidas e menos compreendidas questões nos estudos sobre a gestão do conhecimento. Porém, se esta dificuldade fica mais explícita ao analisar o conjunto do conhecimento de uma organização em relação ao mercado, torna-se necessária a definição de parâmetros e mecanismos de avaliação de cada unidade de conhecimento disponível, principalmente aquele que, por algum processo, já foi registrado e precisa ser gerenciado, tanto em seu processo de obtenção, armazenamento, acesso e, até mesmo, descarte. Neste artigo, objetivamos descrever questões envolvidas na identificação do valor do conhecimento registrado em função de sua multidimensionalidade funcional e do processo de transmissão de informações.

**Palavras chave:** Transferência de informação; Gestão do conhecimento; Valor de unidades de conhecimento.

#### **Artigo 2: Popularização do Conhecimento Científico Autora**

**Autoria:** Suzana P. M. Mueller

**Resumo:** A questão da popularização da ciência é apresentada como um tema de interesse para estudos da comunicação científica pela ciência da informação. A participação da sociedade no processo de criação e divulgação da ciência por meio de pressões sociais e econômicas vêm se tornando fator importante na formulação de políticas científicas, especialmente em países com alto grau de educação. A opinião pública sobre fatos científicos, no entanto, depende em grande medida da atuação de intermediários, que traduzam a linguagem científica, especialmente a mídia imprensa e televisiva. As questões da distorção do sentido da notícia científica inerente ao processo, mas às vezes intencional, e o tratamento de notícias sobre fatos científicos que contém risco potencial à sociedade são exemplos usados para mostrar as dificuldades do processo de popularização e argumentar que o assunto, pouco estudado pela ciência da informação, é parte integrante e cada vez mais influente no processo de comunicação científica.

**Palavras chave:** Popularização da Ciência; Comunicação Científica.

lb) *DataGramaZero - Revista de Ciência da Informação - v.3 n.3 jun/02*  
Disponível no endereço: <http://www.dgz.org.br/jun02/index.htm>

#### **Artigo 3: O Valor da Informação: um desafio permanente Autora**

**Autoria:** Ana Lúcia Siaines de Castro

**Resumo:** O trabalho discute a questão da informação como uma alternativa de sobrevivência, de garantia jurídica frente a dispositivos de exceção, como ocorrido no período da ditadura militar no Brasil. A análise da informação como um valor estratégico de harmonização do indivíduo à sua capacidade de escolha, de avaliação e de reflexão. Na medida em que relatos e memórias alcançam o espaço público, passíveis de institucionalização, a vinculação entre memória e informação revela-se confluyente, interliga-se à confiabilidade dos mecanismos de produção, organização e recuperação da informação. A memória passa a representar um estoque informacional de valor social que possibilita a sociedade e os indivíduos disporem de garantias para exercitarem da melhor forma seu direito à informação.

**Palavras chave:** Informação; Valor Informacional; Direito à Informação; Memória Social; Estoque Informacional.

**Artigo 4: Auto-arquivamento: uma opção inovadora para a produção científica** **Autoria:** Ligia Café e Márcia Basílio Lage

**Resumo:** Trata do conceito inovador de auto-arquivamento e suas implicações no sistema de publicações científicas. Esta nova filosofia procura minimizar as conseqüências provocadas pelo controle editorial, pela revisão severa entre os pares e pela reserva dos direitos autorais. A experiência da Budapest Open Access Initiative (BOAI) é relatada com o objetivo de mostrar uma ação efetiva que viabiliza o auto-arquivamento. Fundamentada no acesso livre (open access), a BOAI pretende uma reorganização dos mecanismos de produção do meio científico, baseada em conceitos mais democráticos de acesso ao conteúdo.

**Palavras chave:** Arquivos-abertos, Sistema de Publicação, Budapest Open Access Initiative, Acesso Livre, Auto-arquivamento.

**Artigo 5: Análise Contrastiva: memória da construção de uma metodologia para investigar a tradução de conhecimento científico em conhecimento público**

**Autoria:** Teresinha Fróes Burnham

**Resumo:** Este artigo é uma reconstrução sumária e parcial da dinâmica de construção de uma metodologia de análise de processos de tradução do conhecimento científico - conhecimento privado a uma comunidade específica - já submetido a uma primeira tradução como conhecimento escolar, para acesso a um público de não-cientistas: estudantes de nível médio. Toma-se a escola como locus de investigação, levando em conta que esta é a instituição socialmente responsabilizada pela democratização da informação científica, na perspectiva de que esta se transforme em conhecimento pessoal de indivíduos sociais, compreendidos como sujeitos do conhecimento, pela agregação de significados relevantes à formação da cidadania. O texto caracteriza-se como uma memória teórica e experiencialmente referenciada, de uma investigação realizada há mais de duas décadas, através da qual foi produzida a primeira formulação desta metodologia que, depois de várias reconstruções (que continuam a se processar contemporaneamente), vem sendo a base dos trabalhos realizados pela Rede Cooperativa de Pesquisa em (In)formação, Currículo e Trabalho - REDPECT / UFBA, dedicada a participar na construção do novo campo interdisciplinar e multirreferencial da Info-Educação. O texto limita-se a apenas um dos "componentes" do processo de tradução: a dupla dimensão de (des)construção e (re)construção de estruturas conceituais formais de um ou mais corpos teóricos da área de Biologia.

**Palavras chave:** Conhecimento Científico, Conhecimento Privado, Conhecimento Escolar, Democratização da Ciência, Comunicação Científica.

lc) *DataGramaZero - Revista de Ciência da Informação* - v.3 n.4 ago/02

Disponível no endereço: <http://www.dgz.org.br/ago02/index.htm>

**Artigo 6: O Tesouro Eletrônico do Mundo do Trabalho: produto de um esforço interdisciplinar**

**Autoria:** Marília Levacov, Nadia Vanti, Júlio César Zancan e Maria Lizete Gomes Mendes

**Resumo:** O presente artigo relata a implementação, de uma ferramenta para o gerenciamento do Tesouro Eletrônico do Mundo do Trabalho, criado para a Unitrabalho, uma fundação voltada a pesquisas acadêmicas sobre o trabalho, agregando 84 universidades brasileiras. A ferramenta é constituída de duas interfaces: uma para consulta e navegação e outra para gerenciamento. Esta atividade foi realizada por uma equipe interdisciplinar, do ponto de vista da Interação Humano-Computador, buscando alternativas para o diálogo entre dois universos: o dos profissionais da Ciência da Informação e o dos profissionais da Ciência da Computação.

**Palavras-chave:** Tesouro Eletrônico; Mundo do Trabalho; Recuperação da Informação; Interface de Consulta; Sistema de Informação; Interdisciplinaridade; Interação Humano-Computador (IHC).

**Artigo 7: Inteligência Competitiva em Organizações: dado, informação e conhecimento**

**Autoria:** Marta Lígia Pomim Valentim

**Resumo:** O conjunto 'dados, informações e conhecimento' tem sido importante fator de competitividade em diferentes tipos de organizações. Prospeccionar, filtrar e transferir esse conjunto é essencial para a consolidação do processo de inteligência competitiva organizacional. Através do gerenciamento desses recursos informacionais pode-se subsidiar várias atividades para a melhoria contínua do negócio da organização. O papel do conjunto 'dados, informações e conhecimento' no processo de inteligência competitiva é fundamental para o aumento da produtividade e da qualidade

da organização. Estabelecer fluxos formais e informais, bem como mapear e reconhecer os dados, informações e conhecimento estruturados, estruturáveis e não-estruturados para o negócio também são ações que contribuem para o desenvolvimento da inteligência competitiva organizacional.

**Palavras chave:** Inteligência Competitiva; Gestão do Conhecimento; Gestão da Informação; Fluxos Informacionais; Transferência da Informação.

#### **Artigo 8: A conceituação de massa documental e o ciclo de interação entre tecnologia e o registro do conhecimento**

**Autoria:** Antonio Miranda e Elmira Simeão

**Resumo:** A polissemia do conceito de "informação" parece ser uma decorrência natural da apropriação do termo por diferentes áreas do conhecimento e está ligada ao fenômeno conhecido como "definição consuetudinária" em que diferentes especialistas se expressam conforme o estado da arte dos conhecimentos sobre determinado fenômeno. Tais definições estariam, conseqüentemente, sujeitas a reformulações e reconceitualizações pari passu com a evolução da pesquisa. A questão que se levanta constantemente é se a Ciência da Informação deveria ou não ter uma concepção única para o termo, o que parece não só impraticável, quanto inócuo.

**Palavras chave:** Informação; Massa Documental; Conceito de Informação; Tecnologia; Registro do Conhecimento.

#### **Artigo 9: Informação e Universidade: os pecados informacionais e barreiras na comunicação da informação para a tomada de decisão na universidade**

**Autoria:** Claudio Starec

**Resumo:** O trabalho analisa o fluxo de informação nos Campi Rebouças e Nova América da Universidade Estácio de Sá. O objetivo deste trabalho é discutir as dificuldades, ruídos, os problemas e barreiras da comunicação da informação e seus efeitos no fluxo informacional numa organização voltada para o aprendizado. Duas questões levantadas pelo poeta americano T.S.Elliot retratam o viés deste trabalho: "Quanta informação perdemos devido à comunicação? e quanto conhecimento perdemos por causa da informação?" A base teórica é da Ciência da Informação, mais especificamente os conceitos de informação de Barreto, de Relevância de Saracevic, os Sistemas de Recuperação e Disseminação Seletiva de Informação de Araújo, as Barreiras de Freire. O foco da Inteligência Competitiva está em destacar a questão da informação como um dos maiores ativos de estratégias no setor e, possivelmente, ferramenta mais importante para ajudar os gestores da universidade a tomar decisões acadêmicas e administrativas à tempo e em tempo real. O modelo escolhido foi a Mandala Tibetana de Paul Carro adaptada na Mandala da Informação Universitária.

**Palavras chave:** Universidade; Gestão do fluxo de Informação na Universidade; Inteligência Competitiva; Barreiras na Comunicação da Informação; Pecados Informacionais.

#### **Artigo 10: Implicações da "nova economia" para a mensuração estatística: desajustes conceituais e metodológicos**

**Autoria:** Rosa Maria Porcaro

**Resumo:** Este artigo discute como importantes transformações que marcam a sociedade atual se refletem na pertinência das informações estatísticas oficiais, construídas a partir de representações da realidade social. Questiona-se se tais transformações estão sendo apreendidas com o arcabouço conceitual-metodológico dos atuais levantamentos estatísticos construído e consolidado para "retratar" a sociedade capitalista industrial moderna de escopo nacional, hoje completamente modificada.

**Palavras chave:** Informação Estatística; Nova Economia; Mensuração Estatística; Desajuste Conceitual; Metodologia Estatística.

Id) *DataGramaZero - Revista de Ciência da Informação - v.3 n.5 out/02*

Disponível no endereço: <http://www.dgz.org.br/out02/index.htm>

#### **Artigo 11: Por uma nova Ciência da Informação: ensino, pesquisa e formação** **Autoria:** Luiz Carlos Brito Paternostro

**Resumo:** O armazenamento e a recuperação de informações incluem sua organização, classificação, proteção, difusão e transferência. O armazenamento e a recuperação dividem o mundo da experiência entre os movimentos de guardar e de tomar, inclusive sob um ponto de vista histórico e cultural. Sob a regência destes movimentos, podemos estudar qualquer coisa ligada à *informação*.

Um conjunto de disciplinas interdependentes tratando especificamente do armazenamento e da recuperação de dados pode vir a compor um curso de Ciência da Informação capaz de tratar, de

forma unificada, questões que variam da *propriedade intelectual* até os *fundamentos da modelagem de dados*.

**Palavras-chave:** Ciência da Informação, Armazenamento e recuperação, Curso em informação, Unidade e especificidade da informação.

#### **Artigo 12: Ensino e pesquisa em ciência da informação**

**Autoria:** Eduardo Wense Dias

**Resumo:** Considerando-se o acesso à informação como a questão básica da ciência da informação, constata-se que é possível segmentar esse campo pelo tipo de informação a que se procura facilitar o acesso: informação publicada especializada, informação publicada não-especializada e informação não-publicada. As características peculiares desses segmentos vão determinar a forma que os nomes dos profissionais neles atuantes podem tomar, as disciplinas importantes, a pesquisa, além de outros aspectos relacionados com a formação na área do conhecimento.

**Palavras chave:** Ciência da Informação, Biblioteconomia, Sistema de Informação, Arquivologia, Ensino, Pesquisa.

#### **Artigo 13: O Profissional da Informação: O Humano Multifacetado**

**Autoria:** Kátia de Carvalho

**Resumo:** O profissional que na sua origem se forma no seio da biblioteca com a função de zelar pelo acervo acompanha o desenvolvimento da sociedade e se transforma em um ser humano multifacetado que além de desta primeira função citada passa a ser o responsável pela preservação da memória humana sem perder de vista o objetivo primordial que é a disseminação do conhecimento e da informação. O profissional nessa sociedade amplia as suas competências para dar conta do seu papel nos sistemas de informação. Ele, no contexto atual, deve ser um indivíduo que faz experiências e é sensível a aprendizagem sendo a sua presença insubstituível nas organizações, além de ser um mediador, entre usuário e acervos. Esse profissional representa o elemento humano nas relações com o meio em um mundo em transformação, com um modelo de economia global baseada no conhecimento.

**Palavras chave:** Profissional da informação, Informação organizacional, Formação e profissional da informação.

#### **Artigo 14: Funções Sociais e Oportunidades para Profissionais da Informação** **Autoria:** Kira Tarapanoff, Emir Suaiden, Cecília Leite Oliveira

**Resumo:** No contexto da sociedade em rede são discutidas funções sociais e delineados alguns perfis de atuação para profissionais da informação. Dentre as funções sociais delineadas estão as educativa e a de mediação. A educativa relaciona-se à alfabetização em informação e a segunda à animação da inteligência coletiva. Dentre os papéis profissionais emergentes são enumerados e brevemente discutidos os seguintes perfis: gestores da informação; trabalhadores do conhecimento; gestores e engenheiros do conhecimento; especialistas de informação. Conclui-se que não há um perfil único para o profissional da informação, que como um "soldado universal" atenderia a todas as demandas de informação nas organizações e na sociedade. Há papéis a serem preenchidos e demandas específicas a serem atendidas por profissionais com os mais diversos perfis, consagrados e emergentes, mas que têm como único objetivo o trabalho com a informação e o conhecimento, agregando valor à primeira e facilitando o acesso e transferindo informação e o conhecimento para todos.

**Palavras chave:** Profissionais da informação, Funções sociais, Perfis de profissionais da informação, Inclusão digital, Gestão da informação, Gestão do conhecimento.

#### **Artigo 15: Relação Ensino-Pesquisa: em discussão a formação do Profissional da Informação**

**Autoria:** Mara Eliane Fonseca Rodrigues

**Resumo:** Tendo como referência as mudanças paradigmáticas que se avizinham para a educação, em geral, e para a universidade, em particular, discute a formação do profissional da informação no Brasil. Após, tomando por pressuposto que a formação, a prática profissional e a pesquisa, compõem a base de uma profissão e que estes três componentes devem interagir constantemente, enfoca a pesquisa como elemento capaz de permitir o repensar da formação e da prática do profissional da informação, considerando-a como um princípio também educativo.

**Palavras chave:** Formação profissional, Ensino e pesquisa.

#### **Artigo 16: Educação para a Informação: desafios contemporâneos para a Ciência da Informação**

**Autoria:** Ana Maria Pereira Cardoso

**Resumo:** O artigo situa os desafios para a formação de profissionais de informação no contexto das mudanças no ensino superior no Brasil. Aborda a consolidação do campo da Ciência da Informação e as influências recebidas por via das literaturas americana e francesa. Discute as especificidades da Ciência da Informação em contraponto com a Biblioteconomia. Partindo destas referências apresenta o projeto de formação de analistas de informação conforme implementado na PUC Minas; destacando o perfil do profissional visado, os eixos temáticos do curso, as estratégias de ensino/aprendizagem.

**Palavras chave:** Ciência da Informação - Formação profissional, Educação Superior no Brasil, Sociedade da Informação - educação, Ciência da Informação e Biblioteconomia, Ciência da Informação - curso de graduação.

le) *DataGramaZero - Revista de Ciência da Informação - v.3 n.6*

dez/02

Disponível no endereço: <http://www.dgz.org.br/dez02/index.htm>

### **Artigo 17: Novas Tecnologias e Produção Científica: uma relação de causa e efeito ou uma relação de muitos efeitos?**

**Autoria:** Maria das Graças Targino

**Resumo:** Discute a relação entre novas tecnologias e o desenvolvimento da produção científica e da publicação eletrônica, enfatizando a Internet. Sem negar sua relevância como elemento interveniente da realidade contemporânea, prioriza as desvantagens trazidas pelas facilidades de produção no espaço cibernético, no caso particular da produção científica, tais como: a inconsistência, instantaneidade e efemeridade das informações; a complexidade de armazenamento; a dificuldade do controle bibliográfico; a banalização da autoria e o desrespeito à propriedade intelectual; o uso aético da informação; a invasão da privacidade x relações impessoais.

**Palavras-chave:** Internet e Produção Científica, Novas Tecnologias de Informação e de Comunicação, Produção Científica e Novas Tecnologias.

### **Artigo 18: Enfoques sobre a relação Ciência, Tecnologia e Sociedade: Neutralidade e Determinismo**

**Autoria:** Renato Dagnino

**Resumo:** De uma forma bastante genérica e mesmo ingênua, mas adequada à finalidade deste trabalho, é possível classificar as formas de abordar o campo dos Estudos Sociais da Ciência e Tecnologia ou, mais especificamente, a relação Ciência, Tecnologia e Sociedade, em duas grandes categorias. A primeira possui como foco privilegiado de análise, ou como elemento determinante da dinâmica da relação, o seu primeiro pólo, a C&T; enquanto que, a segunda, a Sociedade.

**Palavras-chave:** Estudos Sociais da Ciência, Sociologia da Ciência, Ciência e Sociedade, Tecnologia e Sociedade.

### **Artigo 19: Inteligência Empresarial: uma avaliação de fontes de informação sobre o ambiente organizacional externo**

**Autoria:** Ricardo Rodrigues Barbosa

**Resumo:** O artigo relata um estudo sobre o processo de monitoração do ambiente organizacional externo. Os 91 participantes da pesquisa registraram, dentre outros fatores, a frequência com que utilizam diversos tipos de fontes de informação. Essas fontes foram também analisadas de acordo com o seu grau de relevância e confiabilidade. Os resultados indicam uma elevada taxa de utilização de fontes eletrônicas de informação, porém as mesmas são vistas como pouco confiáveis e relevantes. As pessoas (colegas, subordinados e superiores hierárquicos) são vistas como as fontes mais confiáveis. As bibliotecas e centros de informação internos, embora considerados as fontes mais confiáveis, encontram-se entre as menos utilizadas e menos relevantes.

**Palavras-chave:** Inteligência Empresarial, Monitoração Ambiental, Fontes de Informação, Gestão do Conhecimento, Gestão da Informação

### **Artigo 20: Contribuição da Pós-graduação para a Ciência da Informação no Brasil: uma visão**

**Autoria:** Johanna W. Smit, Eduardo Wense Dias, Rosali Fernandez de Souza

**Resumo:** Síntese da avaliação continuada dos programas de pós-graduação em Ciência da Informação reconhecidos pela CAPES (PUC/CAMP, UFBA, UFMG, UFRJ/IBICT, UnB e UNESP/Marília), relativa ao ano de 2001. A partir da constituição dos corpos docente e discente, números de dissertações e teses defendidas e publicações do corpo docente, propõe-se um diagnóstico da pós-graduação na área, finalizando por uma discussão das características da

pesquisa em Ciência da Informação realizada nos programas e a fragilidade da área em relação ao Sistema Nacional de Pós-Graduação. Em anexo uma tabela transcreve as áreas de concentração e linhas de pesquisa, com respectivas ementas, dos programas da área em 2001.

**Palavras-chave:** Ciência da Informação no Brasil, Avaliação 2001 CAPES, Pós-graduação em Ciência da Informação, Pesquisa em Ciência da Informação no Brasil.

### **Artigo 21: Os múltiplos aspectos e interfaces da leitura**

**Autoria:** Lígia Maria Moreira Dumont

**Resumo:** Este trabalho apresenta uma visão panorâmica referente às áreas do conhecimento que se entrelaçam e propiciam um melhor entendimento do ato de ler. Os estudos sobre leitura caracterizam-se pela multidisciplinaridade, portanto, estão sempre abertos à interferência de outras áreas do conhecimento, dependendo certamente de determinado recorte, dentre os múltiplos e diversos ângulos de análise possíveis na temática da leitura. Por suposto, está-se diante de um processo complexo; torna-se tarefa difícil estabelecer os limites de cada olhar, pois o ato da leitura não se efetiva em ações isoladas, lineares, mas sim em decorrência de complexa reação em cadeia de ações, sentimentos, motivações, especulações no cognóscio do leitor, suas análises e críticas. No artigo, são abordadas diversas teses e teorias sobre a temática da leitura, centradas na premissa da leitura como ação social. Primeiramente, são delineados os estudos de Mme. de Staël, Taine e Marx, pioneiros a destacarem o componente social na leitura. A seguir, são analisados os estudos sistemáticos desenvolvidos nos Estados Unidos e na França, nas décadas de 1930 e 1950, respectivamente, que se constituem nas teorias cunhadas de "sociologia da leitura": a Teoria dos fatores subjacentes de Holmes e os modelos de Carrigan e de Gray, de fundamentação organística e funcionalista. Por fim, são delineadas algumas abordagens de autores contemporâneos estrangeiros, como Escarpit, Barthes, Compagnon, Chartier, Allen e Spiro, bem como dos brasileiros Silva, Maria, Sodré e Kato. As teorias que se baseiam na área da psicolinguística e na teoria da computação (inteligência artificial) são destacadas por Kato e Spiro. As abordagens culminam com a tese de Paulo Freire, que imbrica definitivamente a vivência dos sujeitos ao aprendizado e ao desenvolvimento do ato da leitura.

**Palavras-chave:** Leitura-teoria, Cognóscio, Conhecimento-introjeção, Leitura e Sociedade, Informação e Sociedade.

### **Artigo 22: A Informação e o Paradigma Holográfico: a Utopia de Vannevar Bush** **Autoria:** Nilton Bahlis dos Santos

**Resumo:** A Ciência da Informação tem dois elementos constituintes: por um lado ela nasce como acúmulo teórico e de experiências de processamento de informações, em particular da biblioteconomia e da documentação, com suas tecnologias capazes de processar volumes finitos de informação. Por outro como utopia, resultado da ampliação e alargamento do horizonte da ciência, nos esforços aliados na segunda guerra mundial e o desejo de Bush de um novo ordenamento para a Informação. O aspecto mais importante não é a "explosão informacional" como aumento quantitativo, mas a interconexão de experiências e pesquisas, que gera a necessidade de processamentos para a circulação de grandes massas de informação; utopia alimentada pela possibilidade vislumbrada de processar um volume infinito com o surgimento da tecnologia informática. Nossa reflexão é que se o primeiro aspecto está estruturado no paradigma do moderno, com sua visão determinista e racional, resumindo-se a estudar o processo de informação em sistemas fechados, homogêneos e passíveis de serem organizados à priori, o segundo, a utopia, não consegue encontrar uma resposta no interior deste paradigma. Este segundo aspecto constituinte, isto é a busca da capacidade de processar informações em um número infinito e independente de linguagens controladas e de disciplinas, tem como marco o texto "Como nós pensamos" de Vannevar Bush. Ele aponta para a necessidade e possibilidade da Ciência da Informação enfrentar de uma maneira nova o problema da complexidade e interatividade, características cada vez mais presentes em nosso mundo, colocando em questão o próprio paradigma vigente. Esta utopia, no entanto, foi posta em segundo plano devido aos objetivos produtivistas colocados pelas opções práticas que a marcaram. Para recolocá-la na ordem do dia é necessário rever a própria definição de Ciência da Informação, seus limites como campo de conhecimento, seus métodos, suas técnicas e tecnologias. O Paradigma Holográfico apresenta determinados caminhos e opções para uma nova discussão e o hipertexto o evidencia em termos práticos.

**Palavras-chave:** Paradigma, Holografia, Ciência da Informação, Tecnologia da Informação, Hipertexto, Complexidade, Interatividade, Virtual, Totalidade.

Disponível no endereço: <http://www.dgz.org.br/fev03/index.htm>

### **Artigo 23: Informação, Memória e Espaço Prisional no Rio de Janeiro**

**Autoria:** Icléia Thiesen Magalhães Costa

**Resumo:** As relações entre informação, memória e espaço prisional são discutidas nessa proposta de estudo que tem por objetivo principal analisar as formações institucionais e jurídicas, direcionadas à constituição, implantação, reprodução e permanência do chamado *Panoptismo*, em especial na definição e configuração do espaço prisional, no Rio de Janeiro, no período de 1830 a 1930. A Ciência da Informação, de caráter interdisciplinar, propicia a ampliação das fronteiras da Ciência, aproximando saberes de diferentes naturezas e, por essa razão, contribuindo não apenas para a recuperação e disseminação da informação histórica contida nos escaninhos da memória, mas também para a formação de novas relações conceituais, tais como informação e história, espaço e poder, memória e documento, em suas diferentes combinatórias.

**Palavras-chave:** Informação, Memória Social, Espaço Prisional.

### **Artigo 24: O Contrato Social da Pesquisa: em busca de uma nova equação entre a autonomia epistêmica e autonomia política**

**Autoria:** Maria Nélide González de Gómez

**Resumo:** Consideramos próprio das modernas formações ocidentais o desenvolvimento dos conhecimentos científicos por procedimentos complementares a) de diferenciação e autonomização da atividade de pesquisa e b) de conversão da validade científica em valores econômicos ou sociais. Pergunta-se, nesse contexto, pela possibilidade de reformulação do contrato social da ciência, revisando as definições dos sujeitos e dos princípios que organizam os programas de pesquisa, em seu escopo e abrangência, tal que essa nova versão do contrato seja capaz de orientar uma ecologia política dos conhecimentos.

**Palavras-chave:** Contrato Social, Ciência, Pesquisa, Pesquisadores, Autonomia, Ecologia dos Conhecimentos.

### **Artigo 25: A Ciência da Informação no CNPq - fomento à formação de recursos humanos e à pesquisa entre 1994-2002**

**Autoras:** Suzana Pinheiro Machado Mueller e Maria Gorette Santana

**Resumo:** Levantamento dos dados referentes às ações de fomento de CNPq para a área de Ciência da Informação, para o período de 1994 e 2002. Após breve introdução sobre as origens do CNPq em que é enfatizada sua vocação inicial como agência de fomento para as áreas de ciências exatas e naturais, o artigo mostra dados sobre quantidade e dispêndio do órgão com a área de Ciência da Informação. As ações do CNPq relatadas são as que se destinam à formação de recursos humanos no exterior e no país e a pesquisas no país. Os dados mostram que considerando todas as áreas financiadas pelo CNPq, uma parte muito reduzida do orçamento tem sido destinada à Ciência da Informação. Por outro lado, pode-se argumentar que dado o número de cursos pós-graduação, especialmente doutorado, existentes no período considerado e especialmente o número de pesquisa em andamento cujos relatórios foram relatados na reunião de 2000 da ANCIB, a sociedade que congrega os pesquisadores da área, e ainda, a demanda bruta registrada no CNPq, os auxílios recebidos e vigentes, embora ainda insuficientes, parecem menos inadequados. No entanto, a estagnação no volume de bolsas concedidas entre 1994 e 2002 levanta preocupações a respeito da evolução da área.

**Palavras-chave:** Fomento à pesquisa - Ciência da Informação; CNPq - fomento à pesquisa em Ciência da Informação.

Ig) *DataGramaZero - Revista de Ciência da Informação - v.4 n.2 abr/03*

Disponível no endereço: <http://www.dgz.org.br/abr03/index.htm>

### **Artigo 26: Políticas de Monitoramento da Informação por Compressão Semântica dos seus Estoques**

**Autoria:** Aldo de Albuquerque Barreto

**Resumo:** Este artigo se orienta para o estudo da estrutura do texto escrito e sua análise morfológica com a finalidade de extrair informações para uso na gestão estratégica da informação, localizada em estoques específicos. Visa, ainda, fornecer subsídios para um processo de monitoração de conteúdos informacionais em língua portuguesa e a realização de outros estudos de administração da informação. Procura indicar subsídios técnicos e teóricos para construção de softwares para o estudo de contextos de informação utilizando o instrumental da ciência da informação e do processamento

computacional do português em linguagem natural. Ambiciona ser um instrumento estratégico para localizar e caracterizar através de palavras-chave conteúdos de famílias de textos visando a gestão e o controle de um estoque específico de informação.

**Palavras-chave:** Compressão Semântica, Monitoramento da Informação, Estoques de Informação, Palavras-chave.

**Artigo 27: Bolsas de Pesquisador do CNPq: informações sobre política de C&T a partir da base que contém os dados cadastrais dos bolsistas**

**Autoria:** Gilda Olinto

**Resumo:** As bolsas de pesquisador concedidas pelo CNPq são aqui analisadas a partir das bases de dados da agência de fomento que contém o cadastro dos pesquisadores bolsistas. Destaca-se inicialmente a relevância deste objeto de estudo em função das características destas bolsas e, também, em função da riqueza de informações e possibilidades de análises que se apresentam através da transformação dessa base de dados gerada com fins administrativos para uma base com a finalidade de gerar indicadores científicos e tecnológicos. As análises aqui apresentadas focalizando apenas algumas das informações contidas nestas bases de dados – área acadêmica, estado e instituição do trabalho do bolsista – mostram que muitas informações podem ser geradas e revelar algumas características e desequilíbrios que podem ser úteis para subsidiar políticas de governo e o monitoramento da C&T no país.

**Palavras-chave:** Indicadores Científicos, Política Científica e Tecnológica, Gestão de Ciência e Tecnologia

**Artigo 28: Arquitetura conceitual e resultados da integração de sistemas de informação e gestão da ciência e tecnologia**

**Autoria:** Roberto Pacheco e Vinícius Kern

**Resumo:** Iniciativas governamentais na área de gestão da informação esbarram freqüentemente na falta de integração e baixa qualidade da informação, incluindo iniciativas de governo eletrônico. Este artigo apresenta a concepção de sistemas de informação governamentais a partir da consideração dos interesses de todos os atores, configurando uma arquitetura conceitual para projetos de governo eletrônico. A Plataforma Lattes é apresentada como exemplo de implementação desta arquitetura. O papel das bibliotecas digitais de teses e dissertações é destacado, ressaltando seu papel em relação a outros provedores de informação do sistema nacional de ciência, tecnologia e inovação. A internacionalização da Plataforma Lattes é comentada à luz da oferta e demanda de informação que vem provocando.

**Palavras-chave:** Governo Eletrônico, Arquitetura de Sistemas de Informação, Integração de Informações, Gestão de C&T, Bibliotecas Digitais, Plataforma Lattes, Rede ScienTI.

**Artigo 29: Políticas de Informação Governamental: a construção de Governo Eletrônico na Administração Federal do Brasil**

**Autoria:** Carlos Henrique Marcondes e José Maria Jardim

**Resumo:** Políticas de informação governamental têm sido implementadas em diversos países sob a noção de governo eletrônico, ainda pouco estruturada do ponto de vista teórico. No Brasil, a Administração Federal tem desenvolvido diversas ações desde 2000. Limitações de ordem sócio-econômica dificultam o acesso da maioria da população a sistemas de telefonia e a equipamentos de informática. Outro obstáculo ao Governo Eletrônico é a deficiência na gestão das informações governamentais. Após dois anos de implantação, o impacto do Governo Eletrônico revela-se maior na gestão interna da Administração Federal do que no atendimento ao cidadão.

**Palavras-chave:** Governo Eletrônico, Políticas de Informação, Informação Governamental.

**II) Revista Ciência da Informação (31 artigos)**

Ila) Ciência da Informação, v. 31, n. 1, jan./abr. 2002

Disponível no endereço: <http://www.ibict.br/cienciadainformacao/viewissue.php?id=14>

**Artigo 30: Avaliação do acesso a periódicos eletrônicos na web pela análise do arquivo de log de acesso**

**Autoria:** Guilherme Ataíde Dias

**Resumo:** Este artigo apresenta uma abordagem sobre a avaliação do acesso a periódicos eletrônicos disponibilizados na World Wide Web por meio da análise do arquivo de log de acesso. O arquivo de log de acesso da revista **Informação & Sociedade: Estudos** é processado e apresentado

como um exemplo de aplicação do uso de uma ferramenta automatizada de análise para arquivo de **log** de acesso. As características inerentes à análise do arquivo de **log** de acesso são apresentadas e discutidas.

**Palavras-chave:** Periódicos eletrônicos; Avaliação de acesso; Arquivo de log de acesso.

### **Artigo 31: Novos cenários políticos para a informação**

**Autoria:** Maria Nélide González de Gómez

**Resumo:** Poderíamos dizer que hoje, nos cenários mundiais, a economia do conhecimento é proposta, sem mais nem menos, como o novo conteúdo e referência da política da informação ou, em certa forma, da totalidade do político. Consideramos que contribui, para essa subversão de sentido, um terceiro termo, que para uns seria “infra-estrutura”, e para outros, “sociedade da informação”. Se o modus operandi dessa virada estratégica seria a transubstanciação do informacional e semiótico no econômico, através da mediação tecnológica e dos mercados, optamos por considerar as mudanças do papel do Estado – como modus cognoscendi dessas transformações, que afetam profundamente o que, até agora, denominara-se – em sentido restrito – “Política de informação”. Nossa análise remeter-se-á à revisão do conceito “governança”, adotando como apoio argumentativo o conceito de “regime de informação”. A partir da consideração de alguns dos pressupostos da governança, indagaremos quais estruturas de informação poderiam sustentar os processos de formação, circulação e institucionalização do poder, em um horizonte democrático.

**Palavras-chave:** Política de informação; Sociedade da informação; Internet; Institucionalização da informação; Estado.

### **Artigo 32: Uso das linguagens controlada e natural em bases de dados: revisão da literatura**

**Autoria:** Ilza Leite Lopes

**Resumo:** O trabalho tem como objetivo examinar o uso da linguagem controlada ou da linguagem natural, no planejamento da estratégia de busca em um ambiente de bases de dados em CD-ROM ou em linha. São revisados os estudos que abordam o uso das linguagens controlada e natural nas estratégias de busca, suas vantagens e desvantagens, proporcionando uma perspectiva sobre a complexidade para a busca da informação bibliográfica e referencial, incluindo a seleção de termos para as estratégias e a função do vocabulário controlado ou da linguagem natural nesse contexto.

**Palavras-chave:** Bases de dados; Estratégia de busca; Linguagem controlada; Linguagem natural. Recuperação da informação; Artigo de revisão.

### **Artigo 33: Bibliotecas virtuais e digitais: análise de artigos de periódicos brasileiros (1995/2000)**

**Autoria:** Maria Lourdes Blatt Ohira e Noêmia Schoffen Prado

**Resumo:** A evolução da temática biblioteca virtual e biblioteca digital como assunto de artigos de periódicos brasileiros publicados de 1995 a 2000 é o objetivo deste trabalho. Analisa 33 artigos apresentando os aspectos metodológicos adotados para o planejamento e criação de bibliotecas virtuais e digitais, o desenvolvimento de coleções diante dessa nova realidade, o impacto causado nas unidades de informação e nos profissionais da informação, as estatísticas das bibliotecas na Internet e programas institucionais, além da produção bibliográfica sobre bibliotecas virtuais e digitais. Avalia a produção no período, a produtividade dos autores e tipo de autoria, número de referências bibliográficas por artigo, tipo de documentos e o idioma dos documentos citados. Aponta, como resultados, que não há convergência sobre o conceito de biblioteca eletrônica, polimídia, digital e virtual e, para a preocupação dos autores, centrados nos aspectos metodológicos visando à implantação de bibliotecas digitais e virtuais.

**Palavras-chave:** Biblioteca digital; Biblioteca virtual; Produção científica; Produção bibliográfica; Periódicos.

### **Artigo 34: Experiência do Leaal/UFPE na produção e transferência de tecnologia** **Autoria:** Cecília Prysthon e Susana Schmidt

**Resumo:** O Laboratório de Experimentação e Análise de Alimentos (Leaal) do Departamento de Nutrição do Centro de Ciências da Saúde da Universidade Federal de Pernambuco cria/desenvolve e transfere tecnologia para o setor produtivo. Na sociedade industrial, transferência de tecnologia implica comunicação de informação tecnológica relevante para a produção de bens e serviços. Além das barreiras que interferem na comunicação final da informação, os mecanismos de visibilidade muitas vezes não são adequados ao acesso nos sistemas tradicionais e/ou automatizados de comunicação da informação tecnológica. Este trabalho trata de informação tecnológica, sua criação, desenvolvimento de bens e serviços até sua transferência e incorporação pela sociedade. Apresenta

ações e mudanças necessárias ao laboratório para tornar visíveis e mais adequados os mecanismos de transferência tecnológica no âmbito da universidade e comunidade externa.

**Palavras-chave:** Informação tecnológica; Transferência de informação; Transferência tecnológica..

IIb) Ciência da Informação, v. 31, n. 2, maio/ago. 2002

Disponível no endereço: <http://www.ibict.br/cienciadainformacao/viewissue.php?id=13>

### **Artigo 35: Uma introdução ao XML, sua utilização na Internet e alguns conceitos complementares**

**Autoria:** Maurício Barcellos Almeida

**Resumo:** O HTML – Hypertext Markup Language – é uma linguagem de marcação, inicialmente concebida como uma solução para a publicação de documentos científicos em meios eletrônicos, que ganhou popularidade e se tornou padrão para a Internet. Diversos tipos de aplicações, como navegadores, editores, programas de e-mail, bancos de dados etc., tornam possível atualmente o uso intensivo do HTML. Ao longo dos anos, recursos têm sido adicionados ao HTML para que ele possa atender às expectativas de usuários e sistemas computadorizados, aumentando sua complexidade. Estima-se que a versão 4.0 do HTML possua aproximadamente cem diferentes marcações fixas (conhecidas como tags), sem contar aquelas específicas para cada tipo de navegador da Internet. É comum se encontrarem páginas HTML que possuem mais marcações do que conteúdo. Uma possível solução para novas demandas nessa área é a utilização do Extended Markup Language (XML), uma linguagem de marcação que pode introduzir novas possibilidades e trazer melhor integração entre dados e usuários. Este artigo se propõe a abordar, de forma introdutória, o XML, sua utilização na Internet, alguns conceitos complementares necessários ao entendimento do assunto em apresentar vantagens no uso do XML, em relação ao HTML. Além disso, pretende apresentar o assunto como um campo fértil para discussões, proposições e estudo por profissionais da ciência da informação.

**Palavras-chave:** XML; HTML; Linguagens de marcação; Internet; Intranet.

### **Artigo 36: A Lei de Lotka na bibliometria brasileira**

**Autoria:** Rubén Urbizagástegui Alvarado

**Resumo:** Usando os dados reportados em artigos publicados em revistas brasileiras e trabalhos apresentados em congressos nacionais, replicaram-se as aplicações da Lei de Lotka à literatura brasileira em 10 campos diferentes. Utilizou-se o modelo do poder inverso pelos métodos do mínimo quadrado e probabilidade máxima. Das 10 literaturas nacionais analisadas, somente a literatura de medicina, siderurgia, jaca e biblioteconomia ajustaram-se ao modelo do poder inverso generalizado pelo método dos mínimos quadrados. No entanto, só duas literaturas (veterinária e cartas do Arquivo Privado de Getúlio Vargas) não se ajustaram ao modelo quando se usou o método da máxima probabilidade. Para ambas literaturas, tentaram-se modelos diferentes. A literatura de veterinária ajustou-se à distribuição binomial negativa, e as cartas do Arquivo Privado de Getúlio Vargas ajustaram-se melhor à distribuição Gauss-Poisson Inversa Generalizada.

**Palavras-chave:** Bibliometria; Lei de Lotka; Produtividade de autores; Brasil.

### **Artigo 37: Bases de dados de informação para negócios**

**Autoria:** Beatriz Valadares Cendón

**Resumo:** O conjunto de informações usadas por administradores para a tomada de decisão tem sido chamado de "informação para negócios" e inclui informações mercadológicas, financeiras, estatísticas, jurídicas, sobre empresas e produtos e outras informações fatuais e analíticas sobre tendências nos cenários político-social, econômico e financeiro nos quais operam organizações empresariais. Este artigo categoriza e descreve algumas das principais bases de dados estrangeiras sobre informação para negócios, mostrando o universo de informações que elas disponibilizam em forma eletrônica. Para fins de discussão, as bases foram agrupadas em 10 categorias: (1) notícias em geral; (2) informações sobre empresas e setores industriais; (3) diretórios de empresas; (4) informações sobre produtos; (5) informações biográficas; (6) informações financeiras; (7) informações para investimento; (8) pesquisas de mercado; (9) informações jurídicas e (10) informações estatísticas. Agrupadas dessa forma para fins didáticos, na prática essas categorias se sobrepõem, e muitas bases de dados se enquadram em mais de uma categoria desta classificação. O artigo revê também algumas das principais empresas produtoras e distribuidoras de bases de dados sobre informação para negócios e as tendências da indústria de informação eletrônica.

**Palavras-chave:** Informação para negócios; Bases de dados

**Artigo 38: Biblioteca híbrida: um novo enfoque no suporte à educação a distância****Autoria:** Eliane Maria Stuart Garcez e Gregório J. Varvakis Rados**Resumo:** Discute-se o papel das bibliotecas híbridas no contexto atual, em face das transições pelas quais passam as bibliotecas convencionais, principalmente pelo surgimento da Internet e pela intensificação dos cursos no ensino a distância nas universidades, resultado do incremento da utilização da tecnologia da informação e comunicação no ensino. Enfoca-se a importância da flexibilização dos bens e serviços que devem ser oferecidos pelas bibliotecas híbridas para atender às necessidades de uma diversidade de tipos de usuários existentes na educação a distância.**Palavras-chave:** Biblioteca híbrida; Tipos de usuários; Bens e serviços.**Artigo 39: Estratégia de busca na recuperação da informação: revisão da literatura****Autoria:** Ilza Leite Lopes**Resumo:** Os sistemas de recuperação de informação, também denominados de bancos de dados, apresentam uma complexidade indiscutível no processo de armazenamento e busca da informação, envolvendo uma série de aspectos que são interdependentes. Dentre estes, podem ser destacados os seguintes fatores: a tecnologia eletrônica conduz os usuários ao acesso democrático à informação ampliando a busca de informação em bases de dados geograficamente distantes; o alcance da qualidade na informação recuperada requer o planejamento de estratégias de busca específicas para cada base de dados. Esse artigo revisa a literatura publicada sobre estratégias de busca abordando os tópicos relativos a seus conceitos, suas principais técnicas e etapas para sua operacionalização.**Palavras-chave:** Estratégia de busca; Recuperação da informação; Técnicas de estratégia de busca; Bases de dados; Artigo de revisão.**Artigo 40: A ciência e a gestão da informação: compatibilidades no espaço profissional****Autoria:** Patricia Zeni Marchiori**Resumo:** Apresenta o contexto contemporâneo que embasa as atividades de profissionais da informação, tendo em vista mercados de trabalhos com crescentes níveis de exigência e a necessidade de se solucionarem problemas de informação cada vez mais complexos e dinâmicos. Define-se gestão da informação, assim como sua abrangência acadêmico operacional tendo como base os pressupostos teóricos da área de ciência da informação, em especial o núcleo de conteúdos relacionados à gestão integral dos recursos de informação de indivíduos, grupos e organizações. Ressaltase que a gestão da informação compartilha com demais profissões afins, os processos de criação, seleção e avaliação, gerenciamento, divulgação, utilização, preservação e políticas de direitos (privacidade, direitos autorais e outros) relacionados ao trinômio dado, informação e conhecimento. São descritas habilidades e conhecimentos necessários ao desempenho profissional do gestor, assim como as dificuldades inerentes à atuação no campo de atividades de informação.**Palavras-chave:** Ciência da informação; Gestão da informação.**Artigo 41: Produção das literaturas “branca” e “cinzenta” pelos docentes/doutores dos programas de pós-graduação em ciência da informação no Brasil****Autoria:** Dinah Aguiar Población e Daisy Pires Noronha**Resumo:** Estudo cienciométrico da produção científica de docentes/doutores de programas de pós-graduação do Brasil. Objetivo: identificar o perfil dos docentes/doutores e as tendências das literaturas “branca” e “cinzenta” produzidas segundo as linhas de pesquisa dos programas. Método: dados coletados por meio de comunicação contínua e interativa com os docentes através da técnica da “Conferência de Delfos” para identificar o perfil dos docente/doutor e caracterizar a respectiva produção científica. Resultados: na análise dos 5 Programas em Ciência da Informação e na Área de Concentração do Programa de Comunicação da ECA/USP foram identificadas 22 linhas de pesquisa às quais estavam vinculados 66 docentes/doutores, sendo 54,5% titulados na área da ciência da informação. Dos 1.108 documentos produzidos no período de 1990 a 1999, 59,8% referem-se a publicações de literatura branca, na qual os artigos de periódicos detêm o maior índice, com 37,8% da produção total. Da literatura cinzenta produzida (40,2%), destacam-se as comunicações em eventos que detêm 29,8% do total da produção. Verificou-se o predomínio da autoria única (73,2%), confirmando as características dos trabalhos individuais da área de humanidades. A produção científica vinculada à linha de pesquisa influencia a formação de grupos de trabalhos e núcleos de pesquisa.**Palavras-chave:** Produção científica; Literatura branca; Literatura cinzenta; Ciência da informação.**Artigo 42: Informação para negócios: os novos agentes do conhecimento e a gestão do capital intelectual**

**Autoria:** Yara Rezende

**Resumo:** A evolução das características e necessidades dos diferentes usuários de informação em empresas vem determinando, ao longo do tempo, não apenas a criação de diversos tipos de sistemas de informação para atendê-los, como também uma constante adaptação do perfil de formação acadêmica e de atuação dos profissionais da informação. Ao primeiro e tradicional modelo de biblioteca técnica de empresa seguiram-se os centros de documentação, os centros de informação, as bibliotecas virtuais, os sistemas de inteligência competitiva e, atualmente, os programas de gestão do conhecimento. O reconhecimento da importância estratégica da administração do conhecimento e do capital intelectual das empresas configura-se como a mais recente fase de evolução na gestão da informação. Os diferentes modelos de sistemas de informação para empresas já surgidos, apesar de distintos, não se excluem e convivem, ainda que parcela significativa dos profissionais da informação não venha acompanhando e se adaptando a essa evolução e esteja perdendo espaço de atuação para profissionais de outras áreas.

**Palavras-chave:** Gestão do conhecimento; Capital intelectual; Informação para negócios; Sistemas de informação para negócios; Agentes do conhecimento.

#### **Artigo 43: Análise metodológica dos estudos de necessidades de informação sobre setores industriais brasileiros: proposições**

**Autoria:** Janete Fernandes Silva, Marta Araújo Tavares Ferreira e Mônica Erichsen Nassif Borges

**Resumo:** Este trabalho apresenta uma investigação sobre os diagnósticos de necessidade de informação tecnológica detectados em empresas brasileiras do setor industrial. Propõe procedimentos metodológicos que permitam orientar as futuras pesquisas sobre necessidades informacionais ditadas pelos processos de aprendizagem e da inovação tecnológica. Discutiu-se o nível de detalhamento, bem como o grau de abrangência e profundidade destes estudos. Os escolhidos para compor a pesquisa destacaram os setores com potencial de crescimento nos mercados internos e externos como possíveis participantes na geração do desenvolvimento tecnológico e econômico. No entanto, existe uma insatisfação com estes estudos, especialmente pela sua incapacidade em reconhecer as reais necessidades de informação e tecnologia.

**Palavras-chave:** Necessidade de informação tecnológica; Informação tecnológica; Setor industrial; Inovação.

#### **Artigo 44: Informação e competitividade: a contextualização da gestão do conhecimento nos processos organizacionais**

**Autoria:** Sergio Luis da Silva

**Resumo:** No âmbito das organizações empresariais, este artigo se propõe a discutir a visualização da gestão do conhecimento na organização em três níveis diferentes, mas fortemente inter-relacionados: o estratégico, o tático e o operacional. O primeiro nível trata da ligação entre competitividade da empresa e o trabalho com os conhecimentos para a criação de competências organizacionais. O segundo nível destaca a importância de se considerar a gestão de conhecimentos na organização como sendo parte relevante de seus processos de negócio e não somente de suas áreas departamentais. E finalmente, em um terceiro nível, está o lado operacional da gestão do conhecimento ligado à aprendizagem, aos formatos que o conhecimento assume e ao papel desempenhado pela tecnologia da informação. Este artigo, baseado principalmente em um estudo diversificado de várias referências bibliográficas, procura trazer algumas contribuições iniciais para esta discussão.

**Palavras-chave:** Gestão do conhecimento; Informação e competitividade; Processos organizacionais.

#### **Artigo 45: Da bibliometria à webometria: uma exploração conceitual dos utilizados para medir o registro da informação e a difusão do conhecimento**

**Autora:** Nadia Aurora Peres Vanti

**Resumo:** Este é um estudo comparativo de quatro subdisciplinas que permitem medir os fluxos da informação, a comunicação acadêmica e a difusão do conhecimento científico: a bibliometria, a cienciometria, a informetria e a webometria. Mediante a leitura de renomados autores que têm abordado estes temas, é realizada uma discussão teórico-conceitual e uma análise das semelhanças e diferenças que unem e separam os quatro métodos quantitativos no que diz respeito ao seu histórico, objeto de estudo, variáveis, técnicas, objetivos e campos de aplicação. Uma ênfase maior é dada à caracterização da webometria, por se tratar de uma área emergente dentro da ciência da informação, ainda pouco explorada no Brasil e com grandes potencialidades derivadas da expansão mundial da Internet.

**Palavras-chave:** Bibliometria; Cienciometria; Informetria; Webometria; Métodos quantitativos de avaliação

Ilc) Ciência da Informação, v. 31, n. 3, set./dez. 2002

Disponível no endereço: <http://www.ibict.br/cienciadainformacao/viewissue.php?id=12>

**Artigo 46: Métodos quantitativos de apoio à bibliometria: a pesquisa operacional pode ser uma alternativa?**

**Autoria:** Paulo César Rodrigues Borges

**Resumo:** O objetivo deste trabalho é apresentar uma forma alternativa para aplicar os métodos da Pesquisa Operacional aos fenômenos bibliométricos que surgiram no início do século XX, até hoje muito polêmicos. Dentre as várias formulações no campo da bibliometria, a chamada “lei de Bradford” foi o foco da investigação. Tentativas deste gênero podem ser uma saída para sistematizar conceitos na bibliometria, confirmando ou descartando descrições e princípios oriundos de suas formulações empíricas. Tendo por base uma linha de analogia entre fenômenos físicos da Teoria do Caos – resolvidos pela Pesquisa Operacional (PO) – e casos de oferta e procura de periódicos, é possível encontrar uma explicação para o comportamento anômalo da curva de Bradford em certas condições críticas. Para aduzir alguma evidência empírica para este ensaio, dois casos práticos na área da PO foram adaptados para a resolução de problemas bibliométricos típicos. Além disso, ao longo de todo o texto, foram assinalados alguns pontos que parecem comuns entre a bibliometria e a Teoria do Caos. Este ensaio, portanto, enseja uma nova questão: a PO poderá contribuir com a ciência da informação, suprimindo-a com modelos determinísticos e bayesianos para explicar os fenômenos bibliométricos?

**Palavras-chave:** bibliometria; Lei de Bradford; Pesquisa operacional; Caos; Ciência da informação; Inferência bayesiana.

**Artigo 47: Periódicos eletrônicos: considerações relativas à aceitação deste recurso pelos usuários**

**Autoria:** Guilherme Ataíde Dias

**Resumo:** Este artigo apresenta algumas reflexões sobre a aceitação de periódicos eletrônicos disponibilizados na World Wide Web. Assuntos que freqüentemente são ignorados durante a elaboração dos mesmos são discutidos. Citam-se como exemplo alguns periódicos científicos eletrônicos brasileiros na área da ciência da informação. Analisam-se também algumas barreiras tecnológicas que impedem o uso mais amplo e irrestrito deste recurso.

**Palavras-chave:** Periódicos eletrônicos; Usabilidade; Novas tecnologias.

**Artigo 48: Alguns aspectos do uso da informação na economia da informação** **Autoria:** Max F. Cohen

**Resumo:** Se a sociedade encontra-se em uma economia da informação, como as empresas estão usando a informação para competir no mercado? Este artigo busca a estruturação do referencial teórico para a construção do modelo que permita medir o uso da informação por parte das organizações. Com base nos levantamentos realizados, entende-se que as empresas usam a informação em busca de seis estratégias genéricas: redução de custos, criação de valor, inovação, redução do risco, virtualização e diferenciação de produto. Destacam-se, na economia da informação, as firmas que conseguem criar a interação entre os atores econômicos, tirar proveito da interconectividade e sincronizar as suas operações.

**Palavras-chave:** Uso da informação; Economia da Informação; Modelo genérico.

**Artigo 49: Ferramentas alternativas para monitoramento e mapeamento automatizado do conhecimento**

**Autoria:** Lúcia Cunha Ortiz, Wilson Aires Ortiz e Sergio Luis da Silva

**Resumo:** A análise da informação é uma excelente estratégia para monitoramento, pesquisa e desenvolvimento em todos os ramos do conhecimento. O objetivo primordial deste trabalho foi consolidar um método alternativo empregando ferramentas eletrônicas na realização do monitoramento automatizado da informação e em sua análise bibliométrica. O trabalho foi desenvolvido tendo como suporte a base Web of Science, do Institute for Scientific Information (ISI), e o uso de softwar como Word, Excel, Reference Manager e Origin. A título de exemplo, aplicamos o método à área de desenvolvimento de produtos, obtendo como resultados uma lista de descritores, a relação dos periódicos mais importantes da área, os autores mais produtivos e uma indicação das parcerias mais freqüentes entre eles.

**Palavras-chave:** Monitoramento da informação; Biblioteconomia; Ciência da informação.

**Artigo 50: A formação profissional no século XXI: desafios e dilemas**

**Autoria:** Edna Lúcia da Silva e Miriam Vieira da Cunha

**Resumo:** Reflexão sobre a educação no século XXI com enfoque especial à educação dos bibliotecários. Destaca os quatros pilares básicos e essenciais, preconizados pela Unesco, a um novo conceito de educação: aprender a conhecer, aprender a viver juntos, aprender a fazer e aprender a ser. Apresenta as ponderações elaboradas por Morin , a pedido da Unesco, que poderão melhorar a educação do futuro. Com base em tais fundamentos, discute o papel e a formação do bibliotecário no século XXI. Declara que os dilemas dos educadores, nesses novos tempos, estão centrados em três questionamentos: O que ensinar? Como ensinar? Para que ensinar? Pondera que a formação do bibliotecário deverá enfatizar sua função educativa e que a base deve ser polivalente alicerçada em um conjunto de valores que possibilite alterar percepções, maneiras de pensar e instaure a cooperação e a sabedoria em detrimento do tecnicismo hoje privilegiado. Conclui que o papel mais importante do bibliotecário no século XXI parece ainda ser o de gerenciador da informação.

**Palavras-chave:** Educação dos bibliotecários; Profissional da informação.

**Artigo 51: A acessibilidade à informação no espaço digital**

**Autoria:** Elisabeth Fátima Torres, Alberto Angel Mazzoni e João Bosco da Mota Alves

**Resumo:** O trabalho aborda aspectos referentes à acessibilidade no espaço digital. Uma ênfase especial é dada às situações relacionadas à interação das pessoas portadoras de deficiência com a informação, em ambientes de bibliotecas. O texto propõe algumas adequações para a acessibilidade ao espaço digital, conforme categorias de usuários, com o intuito de contribuir para um maior nível de acessibilidade à informação, nesse espaço.

**Palavras-chave:** Acessibilidade; Espaço digital; Bibliotecas; Pessoas portadoras de deficiência; Ajudas técnicas.

**Artigo 52: Estudos de usuários: o padrão que une três abordagens**

**Autoria:** Isa Maria Freire, Bruno Macedo Nathanhson, Carla Tavares e Carmelita do Espírito Santo

**Resumo:** Trata-se de três projetos de pesquisa em andamento no Programa de Pósgraduação em Ciência da Informação – PPGCI/IBICT/UFRJ. O primeiro visa a um estudo de usuários com base em uma experiência de interatividade na rede Internet, tendo como objeto de estudo o informativo [www.clippirata.com.br](http://www.clippirata.com.br). O segundo aposta no papel da informação para a educação ambiental. Para tanto, objetiva demonstrar como oficinas de reciclagem artesanal de papel podem funcionar como agregados de informação para a produção do conhecimento. O último projeto tem como objetivo a construção de um instrumento digital sobre informação cultural com base na estrutura do hipertexto. A responsabilidade social da ciência da informação é a base conceitual que une as três abordagens. O fator comum aos três projetos é a participação dos usuários de informação no desenvolvimento de cada um deles, um pressuposto básico da metodologia participante adotada nas pesquisas.

**Palavras-chave:** Estudos de usuários; Educação ambiental; Internet; Hipertexto; Pesquisa participante.

lId) Ciência da Informação, v. 32, n. 1, jan./abr. 2003

Disponível no endereço: <http://www.ibict.br/cienciadainformacao/viewissue.php?id=11>

**Artigo 53: Como incrementar a qualidade dos resultados das máquinas de busca: da análise de logs à interação em português**

**Autoria:** Rachel Virgínia Xavier Aires e Sandra Maria Aluísio

**Resumo:** Com o intuito de avaliar a submissão de consultas em língua natural, especificamente em português, a máquinas de busca na Web, e contrastar com as consultas por palavras-chave, realizou-se um experimento com alunos, professores e funcionários de uma universidade brasileira. Particularmente, analisaram-se as consultas para verificar se os usuários expressavam bem seus objetivos em palavras-chave; como expressariam seus objetivos em língua natural, caso esta possibilidade fosse oferecida; se as consultas em língua natural forneciam informações que pudessem facilitar a recuperação de informação. O pedido de colaboração foi enviado a 440 pessoas de um instituto de computação da universidade. Foram obtidas 63 consultas, correspondentes a 42 objetivos. Observou-se que, para o item **a**, na maioria dos casos (71,43%), as consultas por meio de palavras-chave não trazem todas as informações declaradas importantes no objetivo; para o item **b** as consultas foram feitas por meio de perguntas (71,87%), afirmações (18,75%) e ordens (9,37%); e,

para o item **c** todas as perguntas diretas deixavam claro o objetivo da consulta já com a primeira palavra da frase, ou com as duas ou três primeiras, com exceção das iniciadas pela palavra "qual".

**Palavras-chave:** Análise de **logs**; Máquinas de busca; Recuperação de informação; Comportamento de usuários; Estratégias de busca.

#### **Artigo 54: Information literacy: princípios, filosofia e prática**

**Autoria:** Elisabeth Adriana Dudziak

**Resumo:** Surgida na literatura em 1974, a information literacy liga-se à necessidade de se exercer o domínio sobre o sempre crescente universo informacional. Incorporando habilidades, conhecimentos e valores relacionados à busca, acesso, avaliação, organização e difusão da informação e do conhecimento. A information literacy é a própria essência da competência em informação. O objetivo deste trabalho é definir a information literacy a partir do entendimento do conceito, objetivos e práticas relacionadas, com ênfase no papel educacional das bibliotecas e do bibliotecário. Inicialmente, apresenta-se a evolução do conceito segundo um referencial histórico. Examina-se a information literacy enquanto processo de interiorização de conhecimentos, habilidades e valores ligados à informação e ao aprendizado. Define-se a expressão, suas características e objetivos. Discutem-se diferentes concepções de information literacy, segundo três referenciais: informação, conhecimento e aprendizado. Em seguida, são elencados pontos relevantes de atuação de bibliotecas e bibliotecários na implementação de uma educação voltada para a information literacy. Explorando a information literacy education, evidencia-se a necessidade de construção de um novo paradigma educacional ante a sociedade atual que incorpore a competência em informação.

**Palavras-chave:** Information literacy; Competência em informação; Alfabetização informacional; Biblioteca aprendente; Bibliotecário educador; Sociedade de aprendizagem; Habilidades informacionais.

#### **Artigo 55: Profissional da informação: perfil de habilidades demandadas pelo mercado de trabalho**

**Autoria:** Danielle Thiago Ferreira

**Resumo:** Doze empresas de consultoria em recrutamento e seleção de recursos humanos foram estudadas para obter informações acerca da demanda atual do mercado de trabalho. Foram levantadas e analisadas as literaturas sobre o mercado de trabalho, as qualificações profissionais requeridas pelo mercado e as informações obtidas em depoimentos de empregadores. O estudo trouxe quatro conclusões principais: (1) os profissionais devem desenvolver continuamente suas habilidades técnicas típicas de ciência da informação, bem como suas atitudes comportamentais; (2) as potencialidades desses profissionais nem sempre são reconhecidas pelo mercado de trabalho; (3) como conseqüência, não é comum encontrar profissionais da informação ocupando posições superiores como analistas ou gerentes; (4) as causas principais das deficiências são tanto a falta de desenvolvimento dessas habilidades durante o período de formação, quanto a falta de reconhecimento do perfil dos profissionais da informação pelo mercado e da auto-imagem por eles mesmos.

**Palavras-chave:** Profissional da informação; Profissional da informação – habilidades; Perfil e atuação profissional; Mercado de trabalho.

#### **Artigo 56: O olhar da consciência possível sobre o campo científico**

**Autoria:** Isa Maria Freire

**Resumo:** O artigo descreve o exercício de tecer, no tear da ciência da informação, uma rede para apreender e explicar um evento de comunicação da informação no campo científico. Como objeto de estudo, foi selecionado o artigo em que G. Wersig e U. Neveling propõem, em 1975, um fundamento social para a ciência da informação. A pesquisa encontrou os indícios de que os autores compartilhavam com outros cientistas uma visão socialista da ciência da informação, fundada na importância da organização da informação científica e tecnológica e de sua comunicação no campo científico. Contudo, os autores foram além da consciência real do seu grupo, ao antever a relevância da informação para todos os grupos sociais na sociedade contemporânea. Nesse contexto, a proposição de uma "responsabilidade social" é retomada como fundamento à práxis dos cientistas da informação e como 'padrão que une' ciência e ética, no campo da ciência da informação.

**Palavras-chave:** Teoria da ciência da informação; Sociologia da informação; História da ciência da informação; Comunicação científica; Responsabilidade social.

#### **Artigo 57: As relações entre ciência, Estado e sociedade: um domínio de visibilidade para as questões da informação**

**Autoria:** Maria Nélide González de Gómez

**Resumo:** Se a origem da ciência da informação está marcada pelas alianças de pós-guerra entre ciência, Estado, sociedade, a pesquisa em questões da informação recebe hoje as demandas de articulação dos três principais eixos de integração e avaliação dos conhecimentos, no Brasil e na América latina: o eixo paradigmático, o eixo *corporativo* e o eixo territorial.

**Palavras-chave:** Recuperação da informação; Inteligência científica; Integração dos conhecimentos; Estado; Ciência; Sociedade; Informação.

#### **Artigo 58: Interfaces entre a ciência da informação e a ciência cognitiva**

**Autoria:** Gercina Ângela Borém Lima

**Resumo:** Estudo panorâmico sobre aspectos da ciência da informação (CI) e da ciência cognitiva (CC), apontando recentes contribuições em quatro de suas possíveis interseções: categorização, indexação, recuperação da informação (RI) e interação homem-computador.

**Palavras-chave:** Ciência da informação; Ciência cognitiva; Processamento da informação; Categorização; Indexação; Recuperação da informação; Interação homem-computador.

#### **Artigo 59: A produção científica da Anped e da Intercom no GT da Educação e Comunicação**

**Autoria:** Solange Puntel Mostafa e Luis Fernando Máximo

**Resumo:** Analisa as literaturas publicadas no período 1994-2001 nos grupos de trabalho da Sociedade Interdisciplinar para os Estudos da Comunicação (Intercom) e da Associação Nacional de Pesquisa em Educação (Anped) no tema da comunicação educativa, em que foram analisadas respectivamente 1.023 e 1.049 citações bibliográficas presentes nos trabalhos apresentados. O objetivo da pesquisa foi perguntar quais autores nacionais e internacionais constituem a frente de pesquisa (autores mais influentes) nas duas literaturas e, se possível, visualizar tendências epistemológicas na produção científica. Os resultados apontam o humanismo e as teorias críticas da recepção na Intercom, enquanto na Anped o pós-estruturalismo parece ser a tendência dominante.

**Palavras-chave:** Comunicação científica; bibliometria; comunicação e educação; estudos de citações; cienciometria

#### **Artigo 60: Inteligência competitiva na Internet: um processo otimizado por agentes inteligentes**

**Autoria:** Helena Pereira da Silva

**Resumo:** Apresenta a proposta de um processo de inteligência competitiva (IC) na Internet, utilizando agentes inteligentes na tarefa de monitoramento de fontes de informação disponíveis na rede. O processo foi aplicado como projeto-piloto no Núcleo de Estudos em Inovação, Gestão e Tecnologia de Informação (IGTI) da Universidade Federal de Santa Catarina. Em seguida, foi verificada a aplicabilidade em mais três estudos de caso. Pode-se afirmar que, pelos casos estudados, foi possível vislumbrar a possibilidade efetiva de utilização do processo proposto em diferentes tipos de organizações. Os resultados ainda confirmam, como proposta, a necessidade de formalização do uso da informação e do processo de gestão da informação nas organizações, bem como a automação do processo por meio de agentes inteligentes.

**Palavras-chave:** Inteligência competitiva; Internet; Monitoramento de fontes de informação; Agentes inteligentes.