



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

SÂMELA ROUSE DE BRITO SILVA

**SISTEMAS DE INDEXAÇÃO AUTOMÁTICA POR ATRIBUIÇÃO: uma
análise comparativa**

RECIFE

2020

SÂMELA ROUSE DE BRITO SILVA

SISTEMAS DE INDEXAÇÃO AUTOMÁTICA POR ATRIBUIÇÃO: uma
análise comparativa

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

Área de concentração: Informação, Memória e Tecnologia.

Linha de Pesquisa: Comunicação e Visualização da Memória.

Orientador: Prof. Dr. Renato Fernandes Corrêa

RECIFE

2020

Catálogo na fonte
Bibliotecária Andréa Carla Melo Marinho, CRB-4/1667

S586s Silva, Sâmela Rouse de Brito
Sistemas de indexação automática por atribuição: uma análise comparativa / Sâmela Rouse de Brito Silva. – Recife, 2020.
190f.: il.

Orientador: Renato Fernandes Corrêa.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. Centro de Artes e Comunicação. Programa de Pós-Graduação em Ciência da Informação, 2020.

Inclui referências, anexo e apêndices.

1. Indexação automática. 2. Indexação automática por atribuição. 3. Ciência da Informação. 4. Sistemas de Indexação automática. 5. Avaliação da Indexação automática.
I. Corrêa, Renato Fernandes (Orientador). II. Título.

020 CDD (22. ed.)

UFPE (CAC 2020-51)

SÂMELA ROUSE DE BRITO SILVA

SISTEMAS DE INDEXAÇÃO AUTOMÁTICA POR ATRIBUIÇÃO: uma
análise comparativa

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

Aprovada em: 28/02/2020

BANCA EXAMINADORA

Prof. Dr. Renato Fernandes Corrêa (Orientador)

Universidade Federal de Pernambuco

Prof. Dr. Fabio Assis Pinho (Examinador Interno)

Universidade Federal de Pernambuco

Prof. Dr. Isidoro Gil Leiva (Examinador Externo)

Universidade de Murcia

À minha família pelo apoio incondicional e
à crença de que sonhos são sempre possíveis.

AGRADECIMENTOS

A Deus, pela oportunidade de realização de mais um sonho.

Aos meus pais Arnon e Meire, por mostrarem que o aprendizado é a melhor herança que os pais podem oferecer a um filho.

Às minhas irmãs, pelo incentivo e apoio durante esta jornada.

Ao meu filho Neto, pela motivação constante. Que este trabalho contemple exemplos de superação e determinação.

Ao meu esposo Márcio, pela confiança, carinho e paciência dedicados durante a nossa caminhada.

Ao orientador, Renato Fernandes Corrêa, pela sapiência, paciência e dedicação prestadas durante a construção deste trabalho.

Ao professor, Fábio Assis Pinho, pela grande contribuição durante o exame de qualificação.

Aos professores, Fábio Assis Pinho e Isidoro Gil Liva, pela sabedoria e generosidade dispensadas durante a defesa da dissertação.

Ao Programa de Pós de Graduação em Ciência da Informação - PPGCI da Universidade Federal de Pernambuco - UFPE, pelo acolhimento e pela descoberta de que a Pesquisa vai além de cartas marcadas e que a responsabilidade da aprovação e dos bons resultados reside na competência e vontade de ser pesquisador.

À Universidade Federal de Alagoas - UFAL, pela possibilidade de realizar meu sonho.

Aos amigos, Tati e Rafa pelas incansáveis conversas e pelo carinho de sempre. Renata, pelas responsabilidades assumidas durante meu afastamento...estou chegando!

À Fernanda, pela paciência e compreensão de que é necessário ter calma, pois não fomos feitos para sermos fortes o tempo todo.

Aos amigos do PPGCI, pela paciência, amizade, acolhimento e companheirismo durante esta jornada.

*“É impossível para um homem aprender aquilo
que ele acha que já sabe.”*

(Epicteto)

RESUMO

Avalia e compara a qualidade de indexação dos softwares SISA (Sistema de Indización Semi-Automático) e MAUI (Multi-Purpose Automatic Topic Indexing) na perspectiva da indexação automática por atribuição de artigos científicos escritos em português da área da Ciência da Informação. O SISA é bastante utilizado na área da Ciência da Informação e os experimentos envolvendo o software apontam resultados satisfatórios. O MAUI é muito utilizado na área na Ciência da Computação, porém, o software possui a característica de utilizar um modelo de indexação intelectual para realizar a indexação automática. A pesquisa visa responder às seguintes indagações: Qual a qualidade da indexação dos sistemas SISA e MAUI? Como se dá o processo de indexação em ambos os sistemas? Qual o sistema que apresenta melhor desempenho? Quais os tipos de erros cometidos pelos sistemas na indexação? Para tanto, estabelece objetivos específicos que são alcançados em etapas. A pesquisa classifica-se como exploratória e empírica, cujo método utilizado para construção deste trabalho é um estudo de caso e a forma de coleta de dados se dá por meio de um experimento de indexação. O experimento constitui-se da indexação automática dos artigos científicos do *corpus de referência* composto por 60 artigos da tese de doutorado de Souza (2005). A avaliação do SISA apresentou a média de 28% no índice de consistência, 63% de precisão, 34% de revocação e 42% de medida F. A avaliação do MAUI apresentou a média de 36% no índice de consistência, 54% de precisão, 51% de revocação e 52% de medida F. Apesar de alguns erros cometidos pelos sistemas como a atribuição de termos no plural e singular; dificuldade em atribuir termos compostos, diferenças entre as estruturas nos termos de indexação e atribuição automática de termos gerais e específicos, pode-se concluir que os softwares apresentaram bons resultados, porém o MAUI se destacou por apresentar melhores índices e desempenho.

Palavras-Chave: Indexação Automática. Indexação automática por atribuição. Ciência da Informação. Sistemas de indexação automática. Avaliação da indexação automática.

ABSTRACT

This work evaluates and compares the quality in automatic indexing of the SISA (System Indexing Semi-Automatic) and MAUI (Multi-Purpose Automatic Topic Indexing) software as they are applied to scientific articles written in Portuguese in Information Science. The SISA is widely used in the field of Information Science and experiments with the software point towards satisfactory results. The MAUI is widely used in the area in the Computer Science, however the software can use a model of intellectual property indexing in order to execute automatic indexing. The research aims to answer the following questions: What is the quality of SISA and MAUI indexing? How does the indexing process in both systems happen? Which system has the best performance? What kinds of errors do those systems make in indexing? To this end, this work establishes specific objectives that are accomplished in stages. The research is classified as exploratory and empirical, using case study as method and an indexing experiment as data gathering. The experiment consists of automatic indexing of scientific articles in the reference corpus made from 60 articles cited by Souza (2005). The evaluation of the SISA shows an average of 28% in consistency, 63% in precision, 34% in recalls and 42% of F measure. Evaluation of MAUI shows an average of 36% in consistency, 54% in precision, 51% in recalls and 52% of F measure. Although the systems have made some mistakes, such as the allocation of plural and singular terms; difficulty in assigning compound terms; differences between structures in terms of indexing; assignment of general and specific terms, it can be concluded that the software has shown good results, however MAUI stands out with better indexes and performance.

Keywords: Automatic Indexing. Automatic Indexing by assignment. Information Science. Automatic indexing systems. Evaluation of automatic indexing.

LISTA DE FIGURAS

Figura 1 - Fatores que influem na busca em base de dados.....	38
Figura 2 - Informações da planilha do SISA.....	74
Figura 3 - Informações da planilha do MAUI	74

LISTA DE QUADROS

Quadro 1 - Descrição de Organização da Informação e Organização do Conhecimento.....	21
Quadro 2 – Tipos de Indexação.....	25
Quadro 3 – Etapas da indexação.....	27
Quadro 4 – Elementos considerados numa comparação de indexação.....	40
Quadro 5 – Software de indexação automática utilizados no Brasil.....	43
Quadro 6 – Momentos históricos da automatização da indexação.....	45
Quadro 7 – Principais métodos utilizados na indexação automática.....	46
Quadro 8 – Argumentos acerca da Indexação automática.....	49
Quadro 9 – Tipos de Avaliação da Indexação.....	57
Quadro 10 – Objetivos e etapas da pesquisa.....	66
Quadro 11 – Descrição dos softwares SISA e MAUI.....	77
Quadro 12 – Análise comparativa dos termos dos artigos 1 ao 5.....	94
Quadro 13 – Análise comparativa dos termos dos artigos 6 ao 10.....	96
Quadro 14 – Análise comparativa dos termos dos artigos 11 ao 15.....	99
Quadro 15 – Análise comparativa dos termos dos artigos 16 ao 20.....	101
Quadro 16 – Análise comparativa dos termos dos artigos 21 ao 25.....	104
Quadro 17 – Análise comparativa dos termos dos artigos 26 ao 30.....	107
Quadro 18 – Análise comparativa dos termos dos artigos 31 ao 35.....	109
Quadro 19 – Análise comparativa dos termos dos artigos 36 ao 40.....	112
Quadro 20 – Análise comparativa dos termos dos artigos 41 ao 45.....	116
Quadro 21 – Análise comparativa dos termos dos artigos 46 ao 50.....	118
Quadro 22 – Análise comparativa dos termos dos artigos 51 ao 55.....	121
Quadro 23 – Análise comparativa dos termos dos artigos 55 ao 60.....	125

LISTA DE TABELAS

Tabela 1 - Resultado do SISA.....	79
Tabela 2 - Resultado do MAUI.....	82
Tabela 3 - Resultado das métricas do SISA.....	86
Tabela 4 - Resultado das métricas do MAUI.....	86
Tabela 5 - Os 7 melhores casos de indexação realizado pelo SISA.....	89
Tabela 6 - Os 7 melhores casos de indexação realizado pelo MAUI.....	89
Tabela 7 - Os 7 piores casos de indexação realizado pelo SISA.....	90
Tabela 8 - Os 7 piores casos de indexação realizado pelo MAUI.....	91

LISTA DE SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
CAPES	Coordenação De Aperfeiçoamento de Pessoal De Nível Superior
CDD	Classificação Decimal de Dewey
CDU	Classificação Decimal Universal
CI	Ciência Da Informação
DeCS	Descritores em Ciências da Saúde
FINEP	Financiadora de Estudos e Projetos
IBBD	Instituto de Bibliografia e Documentação
IBICT	Instituto Brasileiro de Informação em Ciência Tecnologia
KWIC	Keyword in Context
MAUI	Multi-Purpose Automatic Topic Indexing
PLN	Processamento de Linguagem Natural
RI's	Repositórios Institucionais
SISA	Sistema de Indización Automático
SRI's	Sistemas de Recuperação da Informação
TBCI	Tesouro Brasileiro de Ciência da Informação
TIC'S	Tecnologias da Informação e Comunicação

SUMÁRIO

1 INTRODUÇÃO	15
2 FUNDAMENTAÇÃO TEÓRICA.....	20
2.1 Organizar para recuperar informações.....	20
2.2 Indexação: conceitos, tipologias e processo	23
2.2.1 Políticas de indexação.....	27
2.2.2 Linguagens de indexação	29
2.2.3 Vocabulário controlado	33
2.2.4 Tesauro Brasileiro em Ciência da Informação (TBCI)	36
2.3 Avaliação da Indexação	37
2.4 Indexação automática.....	41
2.4.1 Indexação automática por atribuição.....	50
2.4.2 SISA (Sistema de Indización Automático)	52
2.4.3 MAUI (Multi-purpose Automatic Topic Indexing)	54
2.5 Avaliação da indexação automática.....	56
2.5.1 Avaliação intrínseca	57
2.5.2 Avaliação extrínseca	59
2.5.3 Corpus de referência na avaliação de sistemas de indexação automática.....	62
3 METODOLOGIA.....	64
3.1 Classificação da pesquisa.....	64
3.2 Parâmetros de qualidade da indexação	65
3.3 Etapas da pesquisa	66
4 ANÁLISE E DISCUSSÃO DE RESULTADOS	77
4.1 Análise comparativa das características dos sistemas	77
4.2. Análise comparativa da qualidade na indexação automática.....	79
4.2.1 Indexação automática do SISA	79
4.2.2 Indexação automática do MAUI	82
4.2.3 Análise comparativa da qualidade da indexação automática entre os sistemas SISA e MAUI	85
4.2.4 Análise conjunta da indexação automática dos sistemas	91
5 CONCLUSÃO.....	127
REFERÊNCIAS	130
APÊNDICE A - Lista de termos atribuídos pelo SISA	140

APÊNDICE B - Lista de termos atribuídos pelo MAUI.....	147
APÊNDICE C - Folksonomia assistida utilizada na pesquisa.....	155
APÊNDICE D - Planilha de resultados do SISA.....	165
APÊNDICE E - Planilha de resultados do MAUI.....	169
APÊNDICE F - Análise dos termos de indexação do SISA e MAUI.....	172
ANEXO A - Corpus de Souza (2005) que será utilizado no experimento.	186

1 INTRODUÇÃO

Com o avanço da tecnologia, bem como seu constante desenvolvimento, um volume crescente de publicações é gerado por meio da comunidade acadêmica em diversos domínios, estruturas, formatos e suportes. Este volume é resultado de pesquisas e estudos realizados no ambiente acadêmico pela comunidade científica.

A partir do momento em que o pesquisador publica os resultados de sua pesquisa, é gerado um produto da ciência que requer disseminação e disponibilização aos pares com a mesma velocidade da qual foi produzido.

À medida que o aumento das publicações científicas enaltece aos pesquisadores, gera novas descobertas e desenvolve a ciência, eis que surge a preocupação referente aos processos de organização dessas publicações, destacando-se a disseminação e recuperação dessas informações nas bases de dados.

De acordo com Fujita (2003), a Organização da Informação compreende as atividades e operações do tratamento da informação que abrange conhecimento teórico e metodológico tanto para o tratamento descritivo, quanto para o tratamento temático de conteúdo das informações.

A recuperação consiste na localização de documentos inseridos em bases de dados e requer técnicas e métodos da organização da informação, a fim de satisfazer às necessidades de busca. Dessa forma, a organização e recuperação da informação se concretizam por meio da indexação, que por sua vez é realizada com a finalidade de determinar, por meio do conteúdo dos documentos, um conjunto de palavras-chave ou assuntos, facilitando sua armazenagem em bases de dados e atendendo deste modo, as necessidades de recuperação da informação (FUJITA; GIL-LEIVA, 2010).

Ao passo em que a evolução da tecnologia desenvolve a produção de informações na comunidade científica, ela também possibilita o uso de instrumentos e ferramentas capazes de colaborar nos processos de representação e posterior recuperação das informações.

Os profissionais da informação como bibliotecários, arquivistas, museólogos, entre outros, viabilizam meios para que a informação armazenada nas unidades informacionais possa também ser acessada de forma rápida e eficaz. Para tanto, utilizam técnicas e métodos da Organização da Informação, com a preocupação de promover e disponibilizar acesso à informação por meio das novas tecnologias. A indexação faz parte da Organização da Informação, e pode ser realizada de duas formas: manualmente ou automática por meio de tecnologia.

O cenário atual dos profissionais da informação requer a utilização das Tecnologias de Informação e Comunicação (TICs), advindas do avanço tecnológico empregado aos instrumentos e técnicas capazes de armazenar, recuperar e disponibilizar informações mais rapidamente. A tecnologia é essencial tanto para os cientistas desenvolverem relatos de suas pesquisas, aumentando assim a produção científica, como também para os profissionais da informação que buscam permitir a recuperação de itens do acervo das unidades informacionais.

Na indexação, o uso da tecnologia é definido por Lancaster (2004) como indexação automática. O autor a define como “um processo que ocorre quando o computador é utilizado para substituir a indexação manual realizada por um indexador”. Neste tipo de indexação a tecnologia é empregada e o processo acontece automaticamente em diversos ambientes informacionais.

Com relação à indexação automática, sua pesquisa e aplicações no Brasil se deram no final dos anos 1960, por meio do programa *Keyword In Context* (KWIC) com o objetivo de organizar os índices das bibliografias especializadas publicadas pelo Instituto de Bibliografia e Documentação (IBBD), que atualmente corresponde ao Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) (VIEIRA, 1988).

Em se tratando de indexação automática, Lancaster (2004) afirma que o processo se dá por: *extração* – quando palavras ou expressões presentes num documento são selecionadas para representar seu conteúdo; e por *atribuição* – quando envolve a atribuição de termos a um documento de uma fonte que pode não ser o próprio documento, a exemplo podemos citar o tesouro.

Autores sabem mais sobre o conteúdo de suas obras, do que o indexador, mas eles podem não estar familiarizados com o vocabulário controlado que eles deveriam utilizar para expressar o conteúdo de suas obras. O uso de vocabulário controlado durante o processo de indexação é imprescindível para o controle da terminologia, portanto o uso de tesouros da área científica da qual pertença os documentos é recomendado. Além de controlar o uso de sinônimos, ambiguidades e homógrafos, o tesouro possui a capacidade de ligação semântica entre termos que mantém relação entre si. (LANCASTER, 2004).

Portanto, o controle terminológico por meio de vocabulário controlado possibilita uma maior qualidade de indexação, bem como o sucesso na disseminação e recuperação das informações nas bases de dados.

Neste trabalho aborda-se a indexação automática por atribuição pelo fato de se fazer uso de vocabulário controlado e, além disso, devido ao escasso conhecimento registrado tocante à

literatura da Ciência da Informação no Brasil em razão do tema, bem como, a sua importância para se construir bases de dados científicas.

Considerando que na literatura brasileira de Ciência da Informação há uma escassez de trabalhos que comparem sistemas de indexação automática por atribuição e que o software SISA é bastante mencionado na literatura pertinente, faz-se necessário o estudo para comparar tais sistemas e sua eficácia. Na literatura internacional de Ciência da Computação pode ser encontrado um sistema de indexação automática por atribuição denominado MAUI que até então não foi abordado na literatura nacional de Ciência da Informação.

De tal modo, esta pesquisa consiste na investigação do processo e da qualidade da representação do conteúdo de artigos científicos da área da Ciência da Informação por meio da indexação automática por atribuição, mediante a utilização do Tesouro Brasileiro de Ciência da Informação (TBCI) nos sistemas: Sistema de Indexação Semi-Automática (SISA) e no *Multi-Purpose Automatic Topic Indexing* (MAUI).

Neste contexto indaga-se: Qual a qualidade da indexação dos sistemas SISA e MAUI? Como se dá o processo de indexação em ambos os sistemas? Qual o sistema que apresenta melhor desempenho? Quais os tipos de erros cometidos pelos sistemas na indexação?

A presente pesquisa busca respostas às indagações apresentadas justificando-se pelo fato de a indexação manual ser um processo de alto custo que necessita tempo e disponibilidade do profissional da informação, além de capacidade para exercer tal função de forma homogênea. Em contrapartida, a indexação automática possui baixo custo, facilita a aplicação em grande quantidade de documentos de forma rápida e homogênea.

Este trabalho é fruto de pesquisas do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco (PPGCI-UFPE), tendo como área de concentração *Informação, Memória e Tecnologia*, discutido na linha de pesquisa *Comunicação e visualização da memória*.

Segundo Lancaster (2004), a qualidade da indexação está relacionada à experiência, concentração e capacidade de leitura, entre outros fatores, como o tempo e a quantidade de documentos. Para Borges (2008, p. 182), “o tempo restrito do indexador e a quantidade cada vez maior de documentos passíveis de tratamento são fatores preocupantes”. Sendo assim, o estudo da aplicação das TIC’s tem se mostrado eficaz aos Sistemas de Recuperação da Informação (SRI’s).

Portanto, este trabalho tem como objetivo geral analisar comparativamente os softwares SISA e MAUI por meio da avaliação da qualidade na indexação automática por atribuição dos softwares na indexação de artigos de periódicos científicos da área da Ciência da Informação.

Desta forma, os objetivos específicos são:

- a) Especificar e descrever o processo de indexação automática por atribuição envolvendo os softwares SISA e MAUI;
- b) Analisar os resultados obtidos por ambos os sistemas, avaliando a qualidade da representação dos assuntos dos documentos do corpus pelos sistemas SISA e MAUI;
- c) Analisar comparativamente a indexação automática por atribuição dos sistemas SISA e MAUI.

Nesta pesquisa será formatado um conjunto de arquivos de texto para ser utilizado durante o experimento pelos softwares SISA e MAUI, que consiste do corpus composto de 60 artigos de periódicos em Ciência da Informação, selecionado por Souza (2005) em sua tese de doutorado.

Adicionalmente, o Tesouro Brasileiro de Ciência da Informação (TBCI) será formatado para entrada nos softwares. O TBCI compreende cerca de 1.800 termos os quais, em sua maioria, possui versão em língua inglesa e espanhola e são complementados por definições. Possui a finalidade de assumir um papel essencial na recuperação da informação no Brasil e em outros países. O TBCI será aplicado como linguagem de indexação no processo de indexação automática por atribuição do corpus desta pesquisa. Cabe ressaltar que esse instrumento é direcionado aos indexadores, pesquisadores e demais profissionais da informação em geral.

Considerando que a pesquisa tem como foco principal analisar comparativamente softwares de indexação automática, esperamos que seus resultados contribuam para a área da CI, no sentido de elucidar dúvidas acerca da aplicação e uso dos softwares estudados, tornando o trabalho do indexador mais célere e permitindo, destarte, que existam maiores discussões sobre essa temática entre os diversos profissionais da informação que lidam com indexação. Para o PPGI-UFPE, este trabalho representa a continuidade dos estudos acerca da Indexação automática e contribui para o desenvolvendo da temática no âmbito da Ciência da Informação brasileira.

Esta dissertação é constituída de 5 capítulos: o primeiro capítulo é dedicado à introdução da pesquisa em que as justificativas, objetivos e a metodologia adotados são apresentados. No segundo capítulo encontra-se a fundamentação teórica da presente pesquisa que discorre acerca da Organização da Informação, Indexação, Indexação Automática, Softwares de Indexação Automática e Vocabulário Controlado. Ainda no segundo capítulo reforça-se a avaliação da Indexação Automática por meio das métricas de qualidade da indexação (GIL LEIVA, 2018). Por fim, o capítulo ainda aborda acerca de Políticas de Indexação Automática, Tesouro Brasileiro em Ciência da Informação (TBCI) e Corpus de referência na avaliação de sistemas

de indexação automática, além de descrever os softwares SISA e MAUI. Este capítulo é considerado relevante por fornecer os aportes teóricos necessários para a compreensão dos resultados da pesquisa.

O terceiro capítulo apresenta a metodologia utilizada para a obtenção dos objetivos propostos pela pesquisa. Este capítulo encontra-se subdividido em Classificação da pesquisa, Parâmetros de qualidade da indexação e Etapas da pesquisa, sendo esta última apresentada detalhadamente por meio de procedimentos metodológicos.

O quarto capítulo foi dedicado a análise e discussão dos resultados, a fim de avaliar e qualificar a indexação automática dos softwares SISA e MAUI, por meio do experimento de indexação proposto pela pesquisa. Por fim, será possível visualizar os resultados da pesquisa a partir da metodologia aplicada, assim como as conclusões de acordo com o que foi discutido no referencial teórico.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo trata-se do referencial teórico acerca da indexação automática por atribuição com o uso do tesauro como linguagem de indexação. De acordo com o exposto acima, na subseção 2.1 discute-se a organização da informação no contexto de recuperação da informação. Em seguida, na subseção 2.2, discute-se os conceitos e definições referentes ao processo de indexação. Na sequência, na subseção 2.3, discute-se acerca da importância e do papel da avaliação da indexação nos ambientes informacionais. Na subseção 2.4, discute-se a indexação automática, destacando a indexação automática por atribuição, bem como a descrição dos sistemas de indexação automática SISA e MAUI. Na subseção 2.5, enfatiza-se a avaliação da indexação automática e as formas de avaliação intrínseca e extrínseca da qual a indexação pode ser submetida.

2.1 Organizar para recuperar informações

Com o desenvolvimento da sociedade, o mundo passou a produzir mais conhecimento e a registrá-lo por meio de publicações, gerando um volume crescente de documentos que necessitam de tratamento técnico adequado para serem recuperados eficazmente.

A invenção da imprensa por Gutemberg no século XV e o surgimento das universidades possibilitaram o compartilhamento do conhecimento mais democrático, por meio dos livros e pela produção de periódicos científicos. As publicações científicas passaram a ser um veículo de comunicação, o que à época ocasionou uma forma de disseminação da informação científica e em avanços no conhecimento como um todo, porém gerou um crescimento exponencial na quantidade de publicações.

Durante toda a história da humanidade, a informação sempre foi considerada instrumento de poder e de desenvolvimento. Conforme Le Coadic (2004), a informação é fundamental para o homem, sendo comparada a uma necessidade vital para o desenvolvimento de toda a atividade humana. Por compreender esta importância para a sociedade, pensou-se na utilização de tecnologia para solucionar os problemas referentes à organização e disponibilização da informação.

Atualmente uma das grandes preocupações para os profissionais da informação é a grande quantidade de informação e conhecimento produzidos pela sociedade. Este fato deve-se ao desenvolvimento da tecnologia e do conhecimento científico.

Para Pagani, Kovaleski e Resende (2017), dois fatores contribuem para o aumento das publicações:

[...] primeiro, as novas tecnologias da informação e comunicação (TICs) que possibilitam novas formas de pesquisa científica, favorecendo o descobrimento de novos estudos relacionados à pesquisa em andamento; segundo, a necessidade de especialização e construção de novos conhecimentos, imposta pelos mercados e pela sociedade do conhecimento (PAGANI; KOVALESKI; RESENDE, 2017 p. 163).

O aumento no número de publicações se deve ao desenvolvimento das TICs, bem como ao desenvolvimento da ciência. Estes fatos modificaram as formas de comunicação e informação até então dispostas na sociedade. Estas modificações são fruto de pesquisas científicas que utilizaram o conhecimento adquirido para produção de meios de comunicação e de suportes da informação capazes de armazenar, organizar e disponibilizar informação.

As TICs se desenvolveram em duas vertentes. A primeira foi devido ao desenvolvimento da ciência. Novas descobertas, ocasionaram o crescimento exponencial no número de publicações científicas, gerando a explosão informacional. A segunda foi em decorrência da primeira: necessidade em desenvolver técnicas, métodos e instrumentos capazes de tratar os documentos a fim de organizar, recuperar, disponibilizar e disseminar informações à medida em que eram produzidas.

Nessa perspectiva de tratamento documental, a área da Organização do Conhecimento compreende processos, produtos e instrumentos a fim de representar e recuperar informações. Um dos processos da organização do conhecimento é a organização da informação, definida por Brasher e Café (2008, p. 5) como “um processo que envolve a descrição física e de conteúdo dos objetos informacionais”. Para compreendermos a discussão acerca de Organização do Informação e Organização do Conhecimento, as autoras as descrevem no quadro 1:

Quadro 1: Descrição de Organização da Informação e Organização do Conhecimento

Organização da Informação	Organização do Conhecimento
---------------------------	-----------------------------

Processo que envolve a descrição física e de conteúdo dos objetos informacionais.	Processo que se aplica às unidades do pensamento (conceitos) e visa a construção de modelos de mundo que se constituem em abstrações da realidade.
Seu produto é a representação da informação , conjunto de elementos descritivos que representam os atributos de um objeto informacional específico.	Seu resultado é a representação do conhecimento que reflete, assim, um modelo de abstração do mundo real, construído para determinada finalidade.
<i>Mundo dos objetos físicos</i>	<i>Mundo da cognição ou ideias</i>

Fonte: Brasher e Café (2008)

Na tentativa de solucionar os problemas relacionados ao acesso e ao uso da informação, a CI consolidou a organização da informação como área de estudo, no que diz respeito ao tratamento informacional. Para Café e Sales (2010):

A organização da informação é abordada na ciência da informação sob dois aspectos, enquanto espaço investigativo que fornece os pressupostos teóricos e metodológicos ao tratamento da informação, e enquanto atividade operacional inerente ao fazer profissional relativo ao tratamento da informação. Os saberes da organização da informação no espaço investigativo alicerçam os fazeres dessa organização como atividade aplicada. (CAFÉ; SALES, 2010, p.116)

Assim, os métodos e técnicas da organização da informação, aliados à tecnologia, promovem a recuperação da informação possibilitando o acesso e uso em variados suportes.

Por ser caracterizada de natureza efêmera, a informação científica necessita de um tratamento documental bastante eficiente e eficaz, pois sua produção requer imediata disponibilização para não se tornar obsoleta diante do montante de informações produzidas devido ao constante desenvolvimento científico.

Ortega (2008, p. 8) atenta para o fato de que é necessário considerar que “as informações documentárias são entendidas como aquelas apreendidas, registradas e armazenadas em sistemas de informação documentária a fim de que sejam passíveis de recuperação e uso”. Neste sentido, a compreensão de recuperar para usar é imprescindível e justifica as ações voltadas para o tratamento informacional.

A disponibilização e o acesso às informações nas bases de dados, sejam de qual natureza for, se tornou uma condição *sine qua non* para a comunidade científica. Esta constatação se

deve ao fato de vivermos em uma sociedade em rede (CASTELLS, 2002), sociedade da informação (MATTELART, 2006) ou mesmo sociedade do conhecimento (DOWBOR, 1994).

2.2 Indexação: conceitos, tipologias e processo

De acordo com Saracevic (1992, 1995, 1999), uma nova área de conhecimento denominada Recuperação da Informação resultou da grande multiplicidade de conceitos teóricos, empíricos e práticos. Conforme Calvin Mooers (1951, p. 25), “a recuperação da informação engloba os aspectos intelectuais da descrição de informações e suas especificidades para a busca, além de quaisquer sistemas, técnicas ou máquinas empregadas para o desempenho da operação”.

Representar conteúdo dentro de um Sistema de Recuperação da Informação (SRI) é considerado uma das atividades mais relevantes de uma unidade organizacional, uma vez que este processo reflete a finalidade ou missão junto ao usuário de permitir a recuperação da informação de forma rápida e prática.

A construção de índices, considerados os primeiros instrumentos para armazenagem e recuperação de informações, marca a época em que o homem passa a preocupar-se em registrar e organizar tudo o que era produzido por meio da representação de conteúdo. Nessa perspectiva, a indexação é um processo de representação sintetizada da informação e requer uma análise documental acerca do conteúdo do documento, auxiliando na identificação, tratamento e disponibilização da informação.

De acordo com o Sistema Mundial de Informação Científica, a indexação é: “a atividade que descreve e identifica o conteúdo de um documento, através de termos. Os conceitos dos documentos podem ser representados por termos selecionados através da linguagem natural ou por símbolos” (UNISIST, 1981, p. 83).

Fujita e Gil Leiva (2010) definem indexação como:

“A indexação é um processo de análise documental realizado sobre os documentos com a finalidade de determinar-lhes um conjunto de palavras-chave ou assuntos para facilitar sua armazenagem em bases de dados e sua posterior recuperação para atender necessidades de informação” (FUJITA, GIL LEIVA, 2010, p. 2).

Os profissionais da informação utilizam diversos métodos, técnicas e instrumentos para organizar informações com o objetivo de representá-las e recuperá-las posteriormente, e pode ser realizada a partir das características físicas ou do conteúdo que compõem os documentos.

Conforme Silva e Fujita (2004) o conceito de indexação se deu a partir da construção de índices, apesar de atualmente estar mais diretamente vinculada à conceituação e análise de assuntos. Para as autoras, a necessidade dos usuários de uma recuperação mais precisa e especializada, tornou a indexação um aparato metodológico indispensável para instituições informacionais.

Para Lapa e Corrêa (2014, p. 15):

A indexação é um processo de tratamento temático essencial, pois consiste no ato de identificar e descrever um documento de acordo com o seu assunto, e cujo principal objetivo é orientar o usuário sobre esse conteúdo intelectual, permitindo, dessa forma, a sua recuperação de forma ágil e eficiente.

Para muitos autores, a indexação é discutida como um mecanismo de representação documentária cujo propósito é unicamente voltado para a recuperação da informação contida em determinado documento, como por exemplo, Borko e Bernier (1978) ao afirmarem que a indexação é o processo de analisar o conteúdo informacional contido nos registros informacionais.

Lancaster (2004) afirma que o ato de indexar transcende a ideia de apenas construir índices, uma vez que o avanço tecnológico global exige uma necessidade de condensação mais abrangente para uma recuperação mais aproximada daquilo que necessita o usuário.

Diante dessas definições pode-se inferir a importância da indexação e do seu valor dentro de uma unidade informacional, bem como o papel imprescindível atribuído ao profissional da informação no que tange às atividades de indexação e responsabilidade diante da representação e recuperação da informação.

Assim, é nítida a importância que a indexação tem para o funcionamento ideal dos Sistemas de Recuperação da Informação, já que ela é responsável pela descrição dos documentos e, portanto, facilita a sua recuperação. (KURAMOTO, 1995).

Durante o processo de indexação é necessário que o documento seja minuciosamente analisado pelo indexador e que sejam estudadas as estratégias de busca, as necessidades de informação e o usuário efetivo e/ou potencial daquela determinada unidade informacional para qual a indexação está direcionada.

A partir do momento em que o indexador atribui descritores a um documento, ele afirma que estes descritores possuem alto grau de relevância para o conteúdo do documento. A indexação é simplesmente a “operação pela qual se escolhe os termos mais apropriados para descrever o conteúdo de um documento” (GUINCHAT; MENO, 1994, p. 175).

Segundo Narukawa (2011), existem três tipos de indexação, como podemos verificar no Quadro 2:

Quadro 2: Tipos de Indexação

Indexação Intelectual	A indexação intelectual ou manual como é usualmente chamada na literatura, é caracterizada pela atividade humana durante todo o processo, não utilizando assim de qualquer dispositivo tecnológico. Portanto, o homem é o responsável pela indexação em todas as etapas do processo.
Indexação Semiautomática	A indexação semiautomática é aquela em que, após a seleção automática dos termos por software, esses são avaliados por um profissional indexador. Relaciona-se ao processo em que um sistema computacional realiza a atividade de análise de texto do documento e, posteriormente, um indexador humano avalia os termos para indexação indicados pelo sistema, escolhendo assim, quais termos serão efetivamente empregados para representar o documento.
Indexação Automática	A seleção automática dos termos é realizada por softwares, levando em conta critérios estatísticos da ocorrência dos termos no texto dos documentos.

Fonte: Dados baseados em Narukawa (2011)

O tipo de indexação dependerá da instituição e dos recursos disponíveis para tal, mas independente dela ocorrer ou não por meio da aplicação de tecnologia, o indexador sempre direcionará seu trabalho de representar e recuperar informações, levando em consideração, principalmente, o seu usuário.

Na literatura pertinente, o processo de indexação se dá em algumas etapas. A quantidade de etapas é bastante divergente entre os autores da área, variando entre duas a cinco, dependendo do ponto de vista do autor.

De acordo com UNISIST (1981, p. 85), “essencialmente a indexação consiste em dois estágios: estabelecimento dos conceitos tratados num documento, isto é, o assunto e tradução dos conceitos nos termos da linguagem de indexação”.

Chaumier (1988, p. 23, tradução nossa) aponta que o processo de indexação se dá em duas etapas: “o reconhecimento e extração de conceitos informativos e tradução dos conceitos em linguagem documental”.

A Norma ABNT 12676 (1992) apresenta três estágios da indexação: a análise do documento e estabelecimento do assunto de seu conteúdo; a identificação dos conceitos presentes no assunto e a tradução desses conceitos nos termos de uma linguagem de indexação.

Gil Leiva (2008, p. 245, tradução nossa) afirma que “as etapas da indexação são: a análise dos documentos e as questões para a seleção dos conceitos explícitos ou implícitos e armazenamento dessas palavras-chave como elas são, ou sua normalização por meio de um vocabulário controlado”.

Para Lancaster (2004, p. 8), “a indexação de assuntos envolve duas etapas: análise conceitual e tradução. Intelectualmente são etapas totalmente distintas, embora nem sempre sejam diferenciadas com clareza e possam, de fato, ocorrer de modo simultâneo”. Fujita (2013) concorda com Lancaster (2004), ao afirmar que o processo de indexação é composto pelas mesmas etapas: análise conceitual e tradução.

Ainda com relação a quantidade de etapas no processo de indexação, Robredo (2005, p. 165) afirma que são quatro a saber:

A análise conceitual do conteúdo significativo do documento, ou seja, a identificação do assunto; a expressão desta análise, através de um conjunto de palavras, frases ou códigos que representem o assunto; a tradução das descrições dos assuntos relevantes para a linguagem de indexação; a organização das descrições padronizadas dos assuntos de acordo com a sintaxe da linguagem de indexação.

Para Van Slype (1991, p. 116, tradução nossa) a indexação também se dá em quatro etapas: “contato com o documento; identificação dos conceitos explícitos e implícitos do documento; tradução dos conceitos expressos em linguagem natural para descritores e estabelecimento de ligações sintática entre os descritores”.

Segundo Rubi (2008, p. 27):

Os autores divergem quanto ao número de etapas da indexação, porém, elas tratam basicamente das mesmas operações:

- análise: leitura e segmentação do texto para identificação e seleção de conceitos;
- síntese: construção do texto documentário com os conceitos selecionados. Está relacionada especificamente à elaboração de resumos;
- representação: por meio de linguagens documentárias.

Dessa maneira, podemos inferir que o processo de indexação acontece em várias etapas e de diversas formas, contudo, todas convergem para o objetivo que é representar a informação contida nos documentos a fim de torná-la recuperável em ambientes informacionais.

Para melhor visualização, o quadro abaixo apresenta os autores e seus pontos de vista acerca das etapas do processo de indexação:

Quadro 3: Etapas da indexação

AUTORES	ETAPAS
UNISIST (1981)	Estabelecimento dos conceitos tratados num documento e Tradução dos conceitos nos termos da linguagem de indexação.
CHAUMIER (1988)	Reconhecimento e extração de conceitos informativos e Tradução dos conceitos em linguagem documental.
NORMA ABNT 12676 (1992)	Análise do documento e estabelecimento do assunto de seu conteúdo; Identificação dos conceitos presentes no assunto e Tradução desses conceitos nos termos de uma linguagem de indexação.
LANCASTER (2004)	Análise conceitual e Tradução.
GIL LEIVA (2008)	Análise dos documentos e as questões para a seleção dos conceitos explícitos ou implícitos; Armazenamento dessas palavras-chave como elas são, ou sua normalização por meio de um vocabulário controlado.
VAN SLYPE (1991)	Contato com o documento; Identificação dos conceitos explícitos e implícitos do documento; Tradução dos conceitos expressos em linguagem natural para descritores; Estabelecimento de ligações sintática entre os descritores.
ROBREDO (2005)	Análise conceitual do conteúdo significativo do documento, ou seja, a identificação do assunto; Expressão desta análise, através de um conjunto de palavras, frases ou códigos que representem o assunto; Tradução das descrições dos assuntos relevantes para a linguagem de indexação e Organização das descrições padronizadas dos assuntos de acordo com a sintaxe da linguagem de indexação.
FUJITA (2013)	Análise conceitual e Tradução.

Fonte: Dados da pesquisa.

Como podemos observar, a partir da análise de que trata o documento, ou seja, da análise do assunto, o indexador identifica e seleciona os termos que representarão seu conteúdo, portanto estudos acerca da leitura documentária são muito relevantes (FUJITA, 2003).

2.2.1 Políticas de indexação

As unidades informacionais têm a preocupação de garantir que o conteúdo de seus documentos seja representado com eficiência e eficácia, contemplando assim, a necessidade informacional de seus usuários. Para tanto, é necessário que a indexação seja realizada de forma

imparcial, considerando a satisfação do usuário e respeitando os valores e recursos cabíveis à unidade por meio das políticas de indexação.

Carneiro (1985, p. 221), numa perspectiva de gestão, afirma que a política de indexação:

[...] deve servir como um guia para tomada de decisões, deve levar em conta os seguintes fatores: a) características e objetivos da organização, determinantes do tipo de serviço a ser oferecido; b) identificação dos usuários, para atendimento de suas necessidades de informação; c) recursos humanos, materiais e financeiros, que delimitam o funcionamento de um sistema de recuperação de informações.

Considerando as afirmações da autora, a política de indexação busca direcionar a indexação aos objetivos dos usuários e deve ser considerada em todas as atividades de indexação, desde sua implementação, durante o desenvolvimento dos sistemas de recuperação da informação, às atividades inerentes à indexação, para assim, garantir sua eficácia.

Nesse sentido, Oliveira (2017) afirma que:

[...] para tanto, deve-se constituir um documento formal, estruturado e alinhado com o contexto informacional, que seja capaz de regular e orientar o trabalho do indexador e a maneira segundo a qual a representação temática deve ser conduzida (OLIVEIRA, p. 42, 2017).

Para Fujita e Rubi (2006) a política de indexação deve ser compreendida como uma filosofia pertinente aos objetivos da unidade de informação para a posterior recuperação da informação, e não apenas uma lista de procedimentos a serem seguidos. As autoras afirmam ainda que, a falta de reconhecimento acerca da importância da política de indexação compromete os estudos acerca da temática. Gil Leiva (2016) concorda com as autoras ao afirmar que o conceito de política de indexação não é recente, mas continua pouco conhecido na realidade e prática das unidades e centros de informação, estendo aos planos de estudo da Academia.

Segundo Fujita (2012, p. 22):

[...] a política de indexação é um conjunto de procedimentos, materiais, normas e técnicas orientadas por decisões que refletem a prática e princípios teóricos da cultura organizacional de um sistema de informação.

A inexistência e o desconhecimento quanto à política de indexação são fatores que interferem diretamente na recuperação da informação. Outros fatores que interferem nos resultados de busca são ocasionados por questões da linguagem adotada na indexação.

Na política de indexação deverá conter a decisão de utilizar ou não uma linguagem documentária. De acordo com Cintra et al. (1994), as linguagens documentárias são construídas

para indexação, armazenamento e recuperação da informação e correspondem a símbolos, que permitem a tradução dos conteúdos dos documentos, de uma linguagem natural para uma linguagem controlada, fornecendo um padrão de linguagem que favorecerá a comunicação entre o usuário e a unidade.

Conforme Fujita (2003, p. 176):

Saber usar a linguagem documentária e dominar seu vocabulário é importante, mas não é só isso. É preciso usar estratégias de exploração textual e ter domínio da estrutura textual complementado pelo conhecimento linguístico. É preciso conhecer os objetivos do sistema de informação e a demanda da comunidade usuária para a seleção dos conceitos.

Dessa forma, é preciso que as unidades informacionais adotem estratégias no processo de indexação, que estas sejam coerentes às buscas dos usuários, para tanto, conhecer os usuários e suas necessidades se torna uma condição extremamente necessária para que a recuperação da informação seja obtida de maneira rápida e eficaz. A escolha da linguagem de indexação é considerada de extrema importância, pois segundo Carneiro (1985, p. 222), a linguagem de indexação deverá estar de acordo com os objetivos, usuários e recursos da instituição, caso contrário, poderá haver o comprometimento de desempenho do sistema.

Uma vez estabelecida a política de indexação num centro informacional, a recuperação da informação se torna mais eficaz, pois a partir de suas diretrizes o profissional indexador será direcionado a selecionar os melhores descritores para representar o conteúdo de documentos, acarretando numa otimização do sistema.

2.2.2 Linguagens de indexação

Na literatura pertinente à indexação, encontramos duas formas de linguagem: linguagem natural e linguagem artificial. Ambas são utilizadas para representar e recuperar informações nos sistemas de recuperação da informação.

De acordo com Lancaster (1993), a linguagem natural é a linguagem encontrada nos documentos com a terminologia usada pelo autor do texto. Está presente nas diversas formas da comunicação humana, como na fala e na escrita, e que é o contrário de vocabulário controlado.

Vogel (2009) considera linguagem artificial é uma linguagem que apresenta estrutura hierárquica, com o intuito de recuperar informações por meio do controle de vocabulário, que é realizado a partir da padronização de termos ao realizar a representação e busca num sistema.

As unidades informacionais adotam linguagens de indexação que podem ou não utilizar o controle de vocabulário. Quando a unidade informacional não adota controle de vocabulário, sua linguagem de indexação é a linguagem natural, quando a unidade informacional adota algum tipo de controle de vocabulário, sua linguagem de indexação é a linguagem artificial.

A padronização dos termos se dá por vários tipos de linguagem de indexação: sistemas de classificação, linguagens de estruturas alfabéticas ou combinatórias, tesouros e índices. Para Almeida (2011), as estruturas das linguagens documentárias ou de indexação podem ser:

[...] notacionais ou verbais e pré ou pós-coordenadas. As notacionais utilizam símbolos (notações) que auxiliam na organização e na localização dos documentos na estante. Alguns exemplos mais conhecidos de linguagens documentárias notacionais são: a Classificação Decimal de Dewey (CDD) e a Classificação Decimal Universal (CDU). Já as linguagens documentárias verbais visam o tratamento e a recuperação da informação pela representação do assunto. São exemplos de linguagens documentárias verbais: vocabulários controlados e tesouros (ALMEIDA, 2011, p. 21).

Geralmente a linguagem de indexação é escolhida pela instituição para resolver problemas indesejáveis encontrados na linguagem natural como a polissemia e a ambiguidade. Estes problemas acarretam erros que são prejudiciais ao processo de indexação, bem como na recuperação da informação.

De acordo com Fujita, Cruz e Patrício (2017, n.p), além de erros na representação e recuperação da informação, a polissemia e a ambiguidade “diminuem as possibilidades de disseminação com precisão e especificidade”. Sendo assim, o uso da linguagem de indexação torna o sistema mais preciso por meio do uso de descritores que são termos padronizados, definidos por especialistas ou indexadores, que servem para definir assuntos e recuperar a informação (BANDIM, 2018, p. 20).

As linguagens de indexação ou linguagens documentárias, como alguns autores se referem, atuam como instrumentos de representação da informação e têm fundamental importância no processo de comunicação entre os usuários e a unidade informacional. O principal objetivo das linguagens de indexação é facilitar o processo de comunicação e garantir o controle de vocabulário diante das dificuldades encontradas durante este processo.

A linguagem de indexação é analisada por Fujita, Cruz e Patrício, (2017, n.p) considerando os seguintes aspectos:

A linguagem de indexação é responsável por garantir que os termos escolhidos pelos indexadores sejam precisos, de tal sorte que levem em si não somente as ideias que o documento apresenta, mas também a filosofia e a padronização da instituição à qual a linguagem está associada, de modo a permitir que os usuários se satisfaçam e que seja possível o intercâmbio de informações entre as bases de dados.

Para Blair (1990), a representação de documentos é fundamentalmente um processo linguístico que requer o uso correto de uma linguagem que contemple a unidade de informação e seus usuários. Blair destaca ainda que existem muitos problemas linguísticos que incidem na recuperação da informação. Deste modo, o uso de uma linguagem de indexação é fundamental para melhorar aspectos que incidem na recuperabilidade da informação referentes à linguagem.

Dias e Naves (2007) ressaltam os estudos de Cintra (1983) acerca de aspectos auxiliares à indexação, apresentando a semântica, sintaxe e a morfologia, como elementos que intervêm na conversão entre a linguagem natural para a linguagem documentária.

De acordo com Cintra (1983), a linguagem corresponde a um sistema, a uma organização onde cada elemento está relacionado e faz parte do mesmo conjunto. A autora considera o uso da linguagem documentária frente às dificuldades que a linguagem natural oferece para operar com a descrição de documentos. A autora destaca “[...] a polissemia, a sinonímia, a homonímia, a antonímia, os modos e expressões de relações complexas[...]” como dificuldades presentes na linguagem natural (CINTRA, 1983, p. 7).

No que tange as dificuldades linguísticas encontradas no processo de indexação, Cintra (1983) afirma que:

Para estudos de indexação importam de modo especial a semântica, disciplina que se ocupa do sentido ou da significação dos elementos, a sintaxe, disciplina que se ocupa das relações que se estabelecem a partir da organização sintagmática dos elementos e, de certo modo, a morfologia que, sintetizando parcialmente aspectos da semântica e da sintaxe, se encarrega da identificação das partes da palavra e de suas condições de ocorrência (CINTRA, 1983, p. 7).

Todos estes fatores deveriam ser compreendidos pelo indexador, a fim de alcançar melhores resultados na representação e recuperação da informação, porém a relação entre as áreas de linguística e Ciência da informação ainda não é explorada como deveria.

À luz do processo de indexação, as unidades informacionais se esforçam para representar conteúdos, uma vez que os objetivos destas instituições se resumem em armazenar e recuperar informação, porém, a linguagem natural nem sempre oferece recursos lógicos durante o indexar, fazendo com que as instituições convertam da linguagem natural para a linguagem de indexação, permitindo assim, o controle de vocabulário (CINTRA, 1983, p. 7).

Cintra et al. (2002, p. 33) entendem linguagens de indexação como linguagens “[...] construídas para indexação, armazenamento e recuperação da informação e correspondem a sistema de símbolos destinados a ‘traduzir’ os conteúdos dos documentos”.

No que se refere a linguagens de indexação no contexto de tratamento temático da informação, Dias e Naves (2007, p. 23) afirmam que há a necessidade de seu uso por se tratar de instrumentos que permitem o fornecimento de termos definidos por um padrão para representar um assunto do documento analisado. Para os autores existem dois tipos de linguagens de indexação:

[...] linguagens alfabéticas e linguagens simbólicas. Os principais tipos de linguagens alfabéticas que existem são as listas de cabeçalhos de assunto e os tesouros. Os cabeçalhos são para uso principalmente de bibliotecas públicas e outros tipos de bibliotecas cuja abrangência de assunto seja mais geral. Os tesouros, por outro lado, são voltados para coleções especializadas. Quanto às linguagens simbólicas, estas se limitam praticamente aos sistemas de classificação bibliográfica, que utilizam símbolos como letras e números para representar os assuntos dos documentos (DIAS; NAVES, 2007, p. 23).

A linguagem natural pode acarretar diversos problemas como a dispersão terminológica e dispersão sintática. O controle de vocabulário ou o uso de linguagem de indexação diminui essa dispersão a um nível aceitável, pois permite o controle de relações semânticas e sintáticas (PINTO, 1985).

Para Guimarães (1988, p. 89), as linguagens artificiais objetivam o “controle do vocabulário e à padronização da linguagem no processo de busca”. Fujita (1988) corrobora com o este pensamento, tendo em vista que, as linguagens de indexação têm o objetivo de “representar o significado e o conteúdo dos documentos a serem recuperados através do índice” (FUJITA, 1988, p. 24-25).

Nessa perspectiva, Fujita e Gil Leiva (2007) atribuem uma grande importância para linguagem de indexação, tendo em vista a função a ela atribuída, de compatibilizar e tornar viável a linguagem entre os usuários e a unidade de informação.

Para Vale (1987, p.14), “a escolha de uma linguagem de indexação é fator essencial para a eficácia de um sistema de recuperação de informação”. Gil Leiva (2008) concorda com esta afirmação ao apontar que, a linguagem de indexação como um instrumento da indexação que visa facilitar a comunicação entre o usuário e a informação, de modo que, sua utilização permite uma recuperação mais eficaz, tornando a representação de assuntos mais consistente e evitando a dispersão de informações relacionadas entre si.

Cabe ressaltar que, ainda que a linguagem natural e a linguagem artificial possuam suas particularidades, durante o processo de indexação, as duas linguagens estão relacionadas, tendo em vista que, “a linguagem natural é a expressão do documento original e da necessidade informacional do usuário” (NARUKAWA, 2011, p. 24-25).

Ao indexador, cabe adotar, escolher, adaptar ou criar a linguagem de indexação que melhor se adapta ao sistema, principalmente aquela que melhor atenda às necessidades do seu usuário, levando em consideração que, uma vez adotada como instrumento auxiliar da indexação, todos os documentos da unidade de informação deverão receber o tratamento temático por meio da linguagem de indexação escolhida, portanto é uma decisão que requer compromisso e responsabilidade.

É importante mencionarmos que a linguagem de indexação pode simplesmente ser a própria linguagem natural, ou seja, a instituição pode não utilizar qualquer forma de controle de vocabulário, deste modo a linguagem de indexação será a linguagem natural.

Posteriormente abordaremos de forma mais detalhada o *tesauro*, tendo em vista que esta linguagem de indexação integra a temática desta pesquisa: indexação automática por atribuição.

2.2.3 Vocabulário controlado

Autores sabem mais sobre o conteúdo de suas obras do que o indexador, mas eles podem não estar familiarizados com o vocabulário controlado que deveriam usar para expressar o conteúdo de suas obras, considerando a posterior indexação das mesmas em uma base de dados. Este fato acarreta intercorrências durante a representação e recuperação da informação. Uma possível solução consiste na utilização de uma linguagem de indexação, cujo objetivo é padronizar a linguagem utilizada tanto pelo sistema informacional, como pelos usuários, servindo assim de tradução da linguagem do autor e realizada por meio de um vocabulário controlado.

Na indexação, representar conteúdo é uma tarefa considerada difícil, dada a sua complexidade que envolve a subjetividade e problemas relacionados à linguagem. Neste caso, o controle do vocabulário é imprescindível para que haja a padronização nos termos escolhidos e utilizados tanto pela unidade informacional, como pelo usuário.

De acordo com Cesarino e Pinto (1978), a utilização do controle de vocabulário foi bastante contestada por alguns estudiosos, porém, várias pesquisas comprovaram a existência de fatores que levaram a um consenso sobre a necessidade de uso.

Cesarino e Pinto (1978) explicam estes: os fatores humanos, linguísticos e hierárquicos

- a) Fatores humanos - diferenças de cultura, de experiência dos autores e domínio de terminologia entre os autores, indexadores e usuários;
- b) Fatores referentes à própria linguagem natural, onde:
 - vários termos podem designar o mesmo conceito (sinônimos);
 - um mesmo termo pode designar conceitos diferentes (homógrafos);
 - o sentido dos termos se altera pela posição que ocupam na frase (sintaxe);
- c) Fatores hierárquicos - um conceito implica em outros mais amplos e/ou mais restritos (CESARINO; PINTO, 1978, p. 272).

O controle de vocabulário é uma solução para estes fatores, atuando por exemplo, na seleção de termos preferidos, diferenciação de homógrafos e estabelecendo relações entre os termos.

Lancaster (2002) concorda com Cesarino e Pinto (1978), ao afirmar que o controle de vocabulário facilita a representação consistente dos assuntos, evita a dispersão de informações relacionadas, além de facilitar a busca sobre determinado assunto. Portanto, o controle terminológico possibilita uma maior qualidade de indexação, bem como o sucesso na disseminação e recuperação das informações nas bases de dados é o reflexo do tratamento documental do qual as mesmas foram submetidas.

Para Barité (2014):

O controle do vocabulário pode ser visto como um conjunto de técnicas e procedimentos aplicados à linguagem para resolver problemas de compreensão, ambiguidade, abrangência e relação entre termos que expressam conceitos, e / ou entre denominações que expressam nomes de pessoas, lugares, produtos ou instituições (BARITÉ, 2014, p. 97, tradução nossa).

O autor afirma que as primeiras noções de controle de vocabulário datam o fim da década de 1940, como consequência do dinamismo do fenômeno da representação do conhecimento, justificando, assim, pesquisas recorrentes acerca de como melhor a representação de conteúdos na área da Organização do Conhecimento e Linguística.

Almeida (2011), em sua dissertação de mestrado intitulada “O vocabulário controlado como instrumento de organização e representação da informação na FINEP”, discorre brevemente sobre as teorias que fundamentam os instrumentos de controle terminológico:

A utilização de instrumentos de controle terminológico para auxiliar a indexação e recuperação da informação é uma prática importante diante dos sistemas de recuperação informatizados. A elaboração destes instrumentos tem sido realizada com base em teorias clássicas da Representação da Informação, tais como: Teoria da Classificação Facetada de Shiyali Ramamrita Ranganathan, Teoria da Terminologia de Eugen Wüster e Teoria do Conceito de Ingetraut Dahlberg (ALMEIDA, 2011, p. 16).

A Teoria do Conceito de Dahlberg foi construída a partir de contribuições da Terminologia e da Categorização de Ranganathan. Estas áreas ofereceram a base teórica para fundamentação do tesouro. Para Campos e Gomes (2006), “Dahlberg agrega princípios metodológicos relacionados ao conteúdo conceitual e sua definição, estabelecendo assim uma teoria consolidada para a determinação de tesouros, ou seja, conceito representado por um termo” domínio” (CAMPOS; GOMES, 2006, p.358).

O tesouro é um tipo de vocabulário controlado muito utilizado na área da CI durante as tarefas de indexação e recuperação de informações. De acordo com Dodebei (2002, p. 66):

A partir de 1940, o termo tesouro começou a ser utilizado na esfera da Ciência da Informação e, em especial, no processo de recuperação da informação, como sendo um instrumento capaz de transpor conceitos e suas relações mútuas, tal como expressos na linguagem dos documentos, em uma língua regular, com controle de sinônimos e estruturas sintáticas simplificadas.

Portanto, o tesouro é uma linguagem documentária que pode ser utilizado na indexação, por permitir o estabelecimento de relações entre os termos de um domínio relacionados entre si, a fim de representar e recuperar informações.

Nesse contexto, Lancaster (1979) discorre acerca de tesouro:

O thesaurus, ou outro instrumento de controle do vocabulário oferece muita ajuda aos intermediários que executam a busca, incluindo, entre outros, o controle de sinônimos e quase-sinônimos; a separação dos homógrafos; o uso da pré-coordenação para evitar falsas coordenações e relações incorretas entre os termos e, ainda, a ligação de termos relacionados, todos com suas respectivas hierarquias. O thesaurus pode ser muito específico, mas nunca tão específico quanto a linguagem natural, que é a linguagem do discurso dos próprios autores (LANCASTER, 1979, p. 281).

O uso de tesouros da área científica da qual pertença os documentos é recomendado. Além de controlar o uso de sinônimos, ambiguidades e homógrafos, o tesouro possui a capacidade de ligação semântica entre termos que mantêm relação entre si. (LANCASTER, 2004).

De acordo com Narukawa (2011, p. 33):

As relações semânticas estabelecidas em um tesouro permitem expressar relações de equivalência, hierárquicas e de associação que, estruturadas de forma alfabética e/ou sistemática e/ou gráfica e, sinalizadas por código especiais (USE, UP, TG, TE, TR), lhe garantem uma organização própria que facilita o processo de uso para representação e recuperação da informação.

Na perspectiva da indexação automática, a autora afirma que:

A aplicação de vocabulário controlado em indexação automática permite definir o significado das palavras, seleciona as candidatas à indexação no contexto de uma área de conhecimento e auxilia na determinação dos termos que, de fato, serão atribuídos para indexação (NARUKAWA, 2011, p.57).

Nesse contexto, cabe ao profissional documentalista decidir quais os termos que comporão a estrutura do tesauro, considerando a garantia de uso na recuperação da informação. A autora aponta ainda que, na elaboração de vocabulários controlados para indexação automática é preciso considerar as particularidades da indexação automática, tendo em vista que antever as variações da linguagem natural é fundamental.

Nesta pesquisa utiliza-se como linguagem de indexação, o Tesauro Brasileiro de Ciência da Informação. O TBCI será aplicado aos sistemas de indexação automática por atribuição SISA e MAUI, portanto, torna-se necessário discorrer acerca deste tipo de controle de vocabulário.

2.2.4 Tesauro Brasileiro em Ciência da Informação (TBCI)

A história do Tesauro Brasileiro em Ciência da Informação (TBCI) teve início há mais de vinte anos sob iniciativa da linha de pesquisa em Ciência da informação, do curso de Doutorado em Comunicação e Cultura da Universidade Federal do Rio de Janeiro (UFRJ). Mesmo recebendo apoio da Financiadora de Estudos e Projetos (FINEP) e do Instituto Brasileiro em Informação, Ciência e Tecnologia (IBICT), o tesauro não foi publicado de imediato, passando a ser utilizado apenas internamente. Somente no ano de 2014 foi divulgado por meio de uma apresentação durante a realização do XV Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB).

À época, o TBCI elaborado pela professora e pesquisadora do IBICT, Lena Vania Ribeiro Pinheiro, que contou com a colaboração de Helena Dodd Ferrez, autora principal do tesauro para acervos museológicos.

O TBCI é resultado de pesquisas acerca da temática que envolve o tesauro, bem como da colaboração de profissionais envolvidos, como técnicos de Ciência da Informação, professores e pesquisadores empenhados em oferecer aos profissionais da informação um instrumento

fundamental para a consistência de terminologia e de vocabulário do domínio da Ciência da Informação.

O TBCI compreende, em média, 1.800 termos os quais, em sua maioria, possui versão em língua inglesa e espanhola e são complementados por definições, possui a finalidade de assumir um papel essencial na recuperação da informação no Brasil e em outros países.

De acordo com Ramalho (2017):

O TBCI representou uma importante contribuição para a área de Ciência da Informação no Brasil, fornecendo um mapeamento temático e categorização da literatura da área, sendo recomendado, inclusive, que os aurores da área utilizem, preferencialmente, os termos extraídos do TBCI como palavras-chaves de suas publicações (RAMALHO, 2017).

Sob a ótica da CI, para tornar o trabalho do indexador mais preciso, o uso do TBCI é recomendado, tendo vista que o mesmo está voltado para os aspectos inerentes à área. Com relação ao uso, o tesouro não se restringe apenas aos indexadores, podendo ser utilizado também por professores, pesquisadores e qualquer profissional da informação, de que dele necessite.

O TBCI será aplicado como linguagem de indexação no processo de indexação automática por atribuição do corpus desta pesquisa. Cabe ressaltar que esse instrumento é direcionado aos indexadores, pesquisadores e demais profissionais da informação em geral.

2.3 Avaliação da Indexação

A avaliação da indexação é compreendida como uma ferramenta para medir a qualidade da indexação nas unidades informacionais. A qualidade da indexação é avaliada por meio de métodos que variam de acordo os aspectos a serem analisados e com os objetivos da unidade.

Para Chaumier (1988), a indexação é considerada adequada quando se atribui um valor satisfatório ao sistema de informação, desta forma, a indexação é apontada 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, p. 135).

Avaliar a qualidade da indexação não é uma tarefa simples, pois além de requerer tempo, baseia-se num julgamento de todo o processo subjetivo pelo qual a indexação foi submetida anteriormente, ou seja, a qualidade da indexação está diretamente relacionada à análise temática realizada pelo profissional indexador.

De acordo com Inácio (2012), a avaliação é um instrumento imprescindível para melhorar produtos e serviços dentro de uma unidade informacional. Para a autora, a avaliação “permite

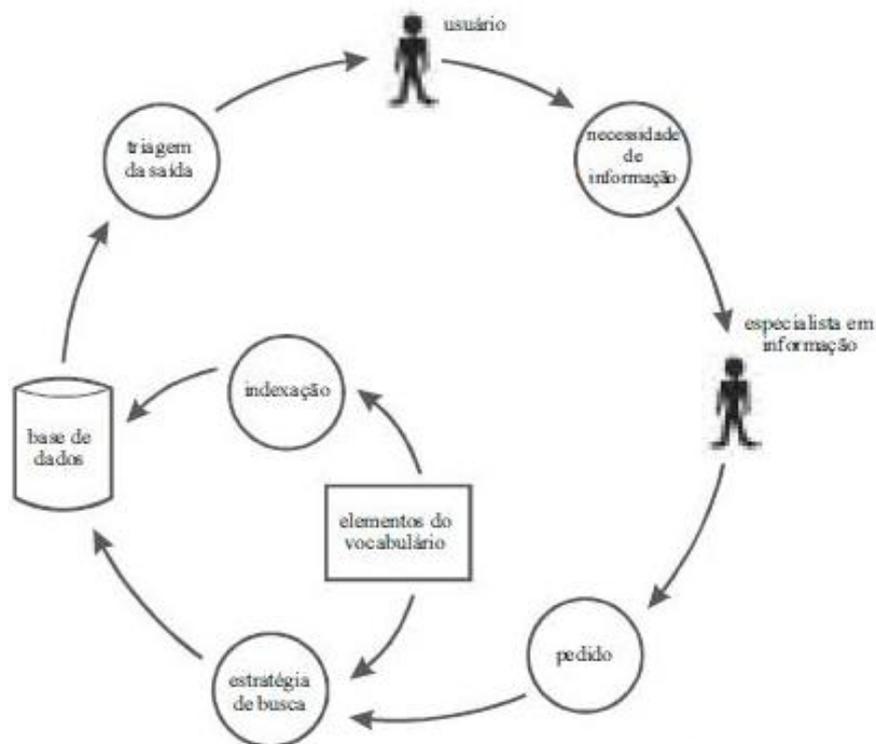
inquirir tanto sobre o fazer da atividade de indexação quanto sobre os sistemas de recuperação da informação, e, ao mesmo tempo, mensurar a satisfação informacional dos usuários” (INÁCIO, 2012, p. 10).

Considerando que o produto da indexação reflete diretamente na recuperação da informação nas bases de dados, bem como na qualidade do atendimento das necessidades informacionais dos usuários, é fundamental que o profissional indexador conheça o público do qual a unidade de informação se destina, para que os objetivos da unidade sejam alcançados, proporcionando assim, a satisfação das necessidades informacionais.

Lancaster (2004) afirma que na recuperação da informação é preciso levar em consideração as estratégias de busca e a base de dados, ambos devem estar em linha para possibilitar o diálogo entre o usuário e o indexador.

Na figura 1 o autor apresenta os fatores que interferem no resultado das buscas dos usuários:

Figura 1: Fatores que influem na busca em base de dados



Fonte: Lancaster, 2004, p. 84.

Na Figura 1, podemos identificar os fatores que influem na qualidade da indexação. Estes fatores estão diretamente relacionados às atividades que envolvem a indexação, de forma que,

um interfere no outro, sendo necessário o consenso entre os profissionais envolvidos quanto à política de indexação, linguagem de indexação, características dos usuários e o sistema de informação adotados durante todo o processo.

Neste sentido Gil Leiva (2008) concorda com Lancaster (2004) ao afirmar que os fatores que mais intervêm na qualidade da indexação são basicamente três: o indexador, o objeto analisado e o contexto. O autor aponta que a formação, a experiência, o domínio do assunto e o profissionalismo do indexador são as principais causas que interferem diretamente na indexação.

Com relação ao fator objeto analisado, o autor aponta que a complexidade, as características, as propriedades e o tamanho, entre outros, agem como pontos de interferência no resultado da indexação. Se tratando do fator contexto, o autor explica que a política de indexação, os tipos de usuários, bem como a carga de trabalho e tempo dedicado à indexação influenciam diretamente em seus resultados.

Em seus estudos, Gil Leiva (2008) ressalta que a qualidade da indexação está relacionada a alguns elementos que caracterizam o processo e o resultado da indexação. Os elementos que implicam diretamente na qualidade da indexação apontados pelo autor são:

- **Exaustividade:** está relacionada com a quantidade de conceitos que caracterizam o conteúdo completo de um documento e nunca com o número de descritores atribuídos a um objeto indexado;
- **Consistência:** é o grau de concordância na representação da informação essencial de um documento por meio de um conjunto de termos de indexação selecionados por cada um dos indexadores de um grupo. A consistência na indexação pode ser estudada tendo como referência a um único indexador ou a vários. Quando um mesmo profissional indexa o mesmo documento em momentos diferentes, falamos de **intraconsistência**. Quando vários profissionais indexam o mesmo documento com o fim de comparar o seu resultado ou enfrentar dois indexadores de diferentes unidades, falamos sobre **Interconsistência**.
- **Especificidade:** é a exatidão com que um termo de indexação representa fielmente um conceito particular presente no documento que está sendo objeto de análise;
- **Correção:** também conhecida como ausência de erros, consiste em falhas na indexação. As falhas podem ser por omissão (quando um termo que representa o conteúdo de um documento é omitido), ou por inclusão (quando há o acréscimo de termos desnecessários) (GIL LEIVA, 2008, p.70-72, tradução nossa).

Todos estes elementos estão diretamente associados à qualidade do atendimento das necessidades dos usuários, porém têm-se observado que o elemento predominante em pesquisas de avaliação da indexação é a consistência.

Acerca dos estudos da consistência da indexação, Saracevic (1987) aponta que no geral, o índice de consistência serve para determinar como indexadores, com certas características em

comum, produzem os mesmos resultados quando realizam a mesma tarefa, ou, caso existam divergências, qual a amplitude da diferença. A avaliação da consistência visa justamente a comparação dos resultados.

Gil Leiva (2008) afirma que ao estudar a consistência, três elementos devem ser levados em consideração: o indexador, o objeto e o contexto, pois interferem em seus resultados. O autor ressalta que, ao avaliar uma unidade informacional, por meio da consistência, deve-se levar em consideração o máximo de elementos que a influenciam conforme podemos verificar no Quadro 4:

Quadro 4: Elementos considerados numa comparação de indexações

INDEXADR	<ul style="list-style-type: none"> - Formação e experiência em indexação: indexador experiente contra principiante - Conhecimento da matéria - Domínio das ferramentas da indexação (linguagens de indexação) - Profissionalismo
CONTEXTO	<ul style="list-style-type: none"> - Políticas de indexação marcadas pela instituição - Objetivo da indexação: temas principais contra especificidade - Tipos e necessidades dos usuários - Carga de trabalho e tempo dedicado
OBJETO	<ul style="list-style-type: none"> - Complexidade do objeto indexado: livro infantil contra patentes - Características e propriedades do objeto indexado: material textual contra material gráfico ou audiovisual - Tamanho: indexação de textos curtos contra textos compridos
MOMENTO	<ul style="list-style-type: none"> - A comparação executa-se com as palavras-chave pinçadas diretamente do texto ou também, uma vez convertidas essas palavras-chave em descritores, após sua filtragem com o vocabulário controlado.
FÓRMULA	<ul style="list-style-type: none"> - São variadas as fórmulas matemáticas utilizadas para conseguir os índices de consistência.

Fonte: LEIVA, 2008, p. 75, tradução nossa.

É fundamental que o indexador tenha o controle sob os fatores que influenciam a indexação, bem como os que norteiam a consistência, para que de fato, a indexação possa representar o conteúdo dos documentos e assim, satisfazer aos usuários.

2.4 Indexação automática

O estudo acerca da automatização da indexação teve início na segunda metade do século XX. Remonta aos últimos anos da década de 50, quando Luhn apresentou o índice KWIC (*Key Word In Context*), em que as palavras do título que servem de entradas no índice são identificadas automaticamente por meio da eliminação das palavras que não tem significado, por comparação com uma lista de palavras vazias de significado, estabelecida previamente (ROBREDO, 1991). Seu desenvolvimento se deu basicamente em momentos históricos e permeiam fatores linguísticos, estatísticos e híbridos.

A indexação automática é compreendida por empregar tecnologia durante a indexação. O uso do computador, por meio de softwares, para indexar documentos, é a maneira mais rápida para o processamento de dados e informações.

Segundo Robredo (2005), para disponibilizar o acesso rápido às bases de dados científicas, o suporte do computador é de suma importância no processamento de dados e informações. Para o autor, atualmente há uma preocupação em oferecer o acesso mais rápido às publicações técnico-científicas por meio do computador durante o processamento de dados e informações.

O uso do computador incide na necessidade de indexar uma grande quantidade de informações em um intervalo de tempo curto, a fim de alimentar às bases de dados e mantê-las atualizadas. Este fato inviabiliza a indexação manual (humana ou intelectual), e nos leva a compreensão de que ela por si só, não deve ser a única forma de análise e representação do conteúdo dos documentos (ROBREDO, 2005).

A indexação automática é amplamente adotada pelos sistemas de recuperação da informação (SRI's), utilizados em bibliotecas, centros de documentação e bases de dados para a construção de índices para busca, mas em poucos casos sua aplicação é feita de forma transparente para o profissional da informação, e os motivos são a falta de conhecimento da existência e uso dos softwares de indexação automática, a resistência quanto à aplicação de tecnologia no processo de indexação e principalmente dúvidas quanto à eficácia na representação da informação (NARUKAWA; GIL LEIVA; FUJITA, 2009).

Lancaster (2004) define que a indexação automática pode ser por extração ou por atribuição. Para o autor, o processo se dá por extração quando palavras ou expressões presentes num documento são selecionadas para representar seu conteúdo; e por atribuição quando envolve a atribuição de termos a um documento de uma fonte que pode não ser o próprio documento (LANCASTER, 2004). Esta fonte é um vocabulário controlado da área do conhecimento de

que se pretende realizar o processo de indexação, que por sua vez, consiste numa lista de termos autorizados que se configura em uma linguagem de indexação, podendo ser um tesouro ou até mesmo uma lista alfabética.

No artigo intitulado “Indexação automatizada de artigos de periódicos científicos: análise da aplicação do software SISA com uso da terminologia DeCS¹ na área de Odontologia” dos autores Narukawa, Gil Leiva e Fujita (2009), há um consenso entre os autores da área ao discorrerem de que o ideal seria que a tecnologia fosse empregada aliada à intervenção humana para avaliar e validar os termos descritores propostos e candidatos à indexação. Este consenso se dá a medida em que os autores apontam que o tempo que se é destinado ao processo de indexação humana pode ser agilizado com o uso de um sistema automatizado, e, portanto, a tecnologia deve ser empregada contribuindo assim no processo de indexação, apontando deste modo, a indexação semiautomática como a melhor alternativa (NARUKAWA; GIL LEIVA; FUJITA, 2009).

Para Hjørland (2008), “a indexação automática é realizada por meio de procedimentos algorítmicos”. Segundo Araújo Júnior (2007) a indexação automática é “qualquer procedimento que permita identificar e selecionar os termos que representam o assunto dos documentos sem a intervenção direta do homem”. Portanto, a indexação automática é realizada por meio de programas de computador que podem contribuir para agilizar o processo de indexação.

De acordo com Lancaster (2004):

O software promete aos profissionais da informação uma contribuição no sentido de diminuir o tempo dedicado ao trabalho atribuído ao processo de indexar, tendo em vista que a indexação é um processo que necessita tempo e técnicas para ser considerada de boa qualidade (LANCASTER, 2004)

Embora os índices de desempenho dos softwares de indexação automática sejam considerados razoáveis, eles não alcançam o mesmo índice da indexação humana. O homem possui capacidades intelectuais superior ao de qualquer máquina por possuir conhecimento inerente ao ser humano. Porém, o uso da indexação automática pode agilizar consideravelmente o processamento de dados e informações, reduzindo a carga de trabalho dos indexadores humanos (LANCASTER, 2004).

¹ Descritores da área da saúde.

Assim, os autores entendem que a automatização da indexação serve como ferramenta de auxílio na prática do profissional da informação, contribuindo para que a indexação possa ser realizada de maneira otimizada por meio da tecnologia empregada em sistemas que permeiam as tarefas do indexador. Agilizar o processo intelectual realizado pelos profissionais da área é principal objetivo da indexação automática.

Os autores Lancaster (2004) e Moreiro González (2004) defendem o pensamento de que os softwares de indexação semiautomáticos são os mais indicados, e que, a utilização destes softwares deve levar em consideração uma análise apurada realizada por um indexador humano, e que somente após a sua aprovação, os descritores resultantes da indexação automática poderão ser permitidos.

Para manter as bases de dados atualizadas, o profissional da informação necessita de tempo para processar e indexar uma grande quantidade de informações produzidas a cada momento. Como alternativa, a indexação automática surgiu com o propósito de contribuir para minimizar o tempo gasto com tratamento da informação, tornando a indexação manual inviável como única forma de analisar a informação (ROBREDO, 2005).

Nesse contexto, o desenvolvimento de pesquisas acerca da indexação automática é fundamental para contribuir com o tratamento documental nas unidades informacionais, disponibilizar alternativas ao profissional indexador no exercício de suas atividades e colaborar para o desenvolvimento da área.

Na literatura, podemos identificar que a CI vem desenvolvendo pesquisas com relação à criação, utilização e avaliação de softwares de indexação automática. No artigo de Corrêa e Lapa (2014), intitulado “A indexação automática no Brasil no âmbito da Ciência da Informação (1973-2012): indicadores bibliométricos”; e nos artigos de Bandim e Corrêa (2018, 2019) intitulados respectivamente: “A consistência na indexação automática por atribuição de artigos científicos na área de Ciência da Informação” e “Indexação automática por atribuição de artigos científicos em português da área de Ciência da Informação”, os autores mencionam os principais softwares utilizados no Brasil.

A seguir no Quadro 5, apresenta-se uma síntese dos softwares de indexação automática citados nos respectivos trabalhos e que possuem relevância para área.

Quadro 5: Softwares de indexação automática utilizados no Brasil

SOFTWARES	ORIGEM	DESCRIÇÃO	TIPO DE INDEXAÇÃO AUTOMÁTICA
OGMA	Universidade Federal de Minas Gerais, Brasil, 2008.	Desenvolvido por Luiz Cláudio Gomes Maia para analisar textos e extrair sintagmas nominais e calcular o peso desses na indexação dos documentos.	Extração
SISA	Universidade de Múrcia, Espanha, 1999-2008.	Desenvolvido por Gil Leiva, extensível a áreas do conhecimento dado um vocabulário controlado.	Atribuição
AUTOMINDEX	Universidade de Brasília, Brasil, 1991.	Desenvolvido por Jaime Robredo, é um software que para propor os termos de indexação, as palavras do texto são confrontadas com as de dois antídicionários concomitantes de palavras vazias: um com palavras invariáveis e outro com radicais de palavras não significativas.	Extração

Fonte: Dados da pesquisa.

Os principais softwares de indexação automática utilizados na literatura científica da área de Ciência da Informação no Brasil são: OGMA, SISA e AUTOMINDEX. A partir do quadro podemos verificar que houve interesse de pesquisadores brasileiros na elaboração de softwares de indexação automática. Verificamos que apenas um dos três softwares realizam a indexação automática por atribuição, o SISA, este será discutido nesta pesquisa mais adiante.

Cabe ressaltar que o surgimento destes softwares é de fundamental importância para a área. Por meio destes sistemas, os profissionais da informação têm a possibilidade de realizar as

atividades de indexação. A partir da utilização dos softwares é possível aprimorá-los e adequá-los para atender às demandas informacionais das unidades de informação.

Para Gil Leiva (1999), o desenvolvimento da indexação automática ocorreu em três momentos históricos: o dos métodos estatísticos, o dos métodos linguísticos e o dos métodos mistos conforme o Quadro 6:

Quadro 6: Momentos históricos da automatização da indexação

MÉTODOS	CARACTERÍSTICAS
Estatísticos	<p>Baseado no princípio do mínimo esforço (Zipf,1949), que se referia ao valor constante que tem a relação entre a frequência das palavras e a posição que ocupam no ordenamento por ordem decrescente de frequência.</p> <p>Contribuições: surgimento da indexação ponderada; valor de discriminação dos termos (Salton e Yang, 1973); função de frequência inversa (Sparck Jones, 1972) e aplicações de métodos de frequência na coleção (Damerau, 1965).</p>
Linguísticos	<p>Na década de 1960 surgiram estudos acerca de processamento automático da Linguagem Natural (PLN), que consiste no estudo e nos aspectos linguísticos de um texto mediante a utilização de programas informáticos e a automatização da indexação.</p> <p>Contribuições: avanços nas áreas de automatização da indexação, recuperação da informação, extração de informação, classificação, indexação e resumos de documentos, e para o reconhecimento da fala. (Haller, 1982), Haller, 1983) e (Andreevsky e Ruas, 1982).</p>
Mistos ou híbridos	<p>Na atualidade, os métodos mistos ou híbridos de indexação automática reúnem aspectos estatísticos, linguísticos e ainda utilizam o tesauro como instrumento de controle de vocabulário.</p> <p>Contribuições: eliminação de problemas como a sinonímia e a identificação de funções sintáticas dos termos, proporcionando benefícios à revocação na recuperação da informação. (Karetnyk, Karlsson e Smart, 1991) e (Godby e Watson, 2001).</p>

Fonte: Baseado em Narukawa, Gil Leiva, Fujita (2009, p. 102-105).

Sendo assim, a evolução da automatização da indexação perpassa momentos históricos distintos, em que diversos estudos acerca dos aspectos que influenciam o processo de indexação contribuíram para o desenvolvimento da indexação automática. Cabe ressaltar que o

computador utiliza métodos como os métodos estatísticos, atribuição de pesos probabilísticos, entre outros, para a identificação dos termos mais significativos do documento.

De acordo com Vieira (1988) *apud* Lapa (2014), os principais métodos utilizados na indexação automática são apresentados no Quadro 7:

Quadro 7: Principais métodos utilizados na indexação automática

MÉTODO	DESCRIÇÃO
Método de Frequência ou Análise estatística (Modelo Booleano)	Este método trata do registro automático do aparecimento da palavra, que pode estar localizada no título, resumo, texto e em diversas combinações entre estes campos, como, por exemplo, em título e resumo.
Método de Atribuição de Peso (Modelo Vetorial)	Este método atribui peso aos descritores como forma de atribuir-lhes valores semânticos para torná-los mais precisos sem, no entanto, diminuir sua capacidade de revocação. É baseado na frequência de cada descritor.
Método Probabilístico (Modelo Probabilístico)	Este método aplica a frequência de co-ocorrência em palavras truncadas automaticamente. As palavras truncadas são extraídas através de um critério estatístico de distribuição binominal, denominado distribuição de Poisson. Aquelas palavras cuja frequência de distribuição for descrita pela função de Poisson serão não significantes.
Análise de “Cluster”	Este método se baseia no reconhecimento automático, por meio de procedimentos estatísticos que podem ser usados para classificar grupos de documentos por observação das semelhanças e dissemelhanças entre eles.
Método de Associação entre Palavras	Este método utiliza a frequência de ocorrência e co-ocorrência de palavras ou pares de palavras para identificar o conteúdo dos documentos. As palavras isoladas e as que se co-associam são identificadas em sentenças. Se as co-associações das mesmas palavras co-ocorrerem com determinada frequência, então, serão consideradas "descritores associados". Este método parte do princípio de que todas as palavras significativas estão relacionadas linearmente.

Fonte: Baseado em Lapa (2014, p. 113-114) *apud* VIEIRA (1988, p. 48-52).

Por meio de programas computacionais, é possível a identificação das palavras ou expressões mais representativas dos documentos. “As palavras significativas são selecionadas automaticamente, através de metodologias específicas, adotadas de acordo com as políticas de indexação e recuperação da informação, desenvolvimento de software e capacidade de hardware dos sistemas de informação”. (VIEIRA, 1988, p. 48)

A indexação automática utiliza estes métodos considerando as políticas de indexação das unidades informacionais. Lembrando que os métodos, bem como as políticas de indexação variam conforme as necessidades dos usuários, tipos de instituição e recursos disponíveis.

Sendo assim, vários estudos apontam a importância do desenvolvimento de ferramentas e métodos para a otimização da indexação automática.

Gil Leiva (2008) afirma que é difícil encontrar um sistema de indexação automática baseado unicamente em apenas um método. Para o autor, geralmente são utilizados vários métodos simultaneamente, além da utilização de algumas ferramentas que representam automaticamente o significado dos textos. Estas ferramentas são também conhecidas por operadores de textos, e se configuram em um conjunto de técnicas computacionais para a análise de textos com o propósito de simular o processamento humano da língua, ou seja, uma área de pesquisa denominada PLN - Processamento da Linguagem Natural (BAEZA-YATES; RIBEIRO-NETO, 2013)

Acerca dessas ferramentas, Gil Leiva (2008) aponta as principais: Listas de palavras vazias; Ponderação dos termos: (Lei de Zipf, frequência do termo e frequência inversa no documento); Valor de discriminação do termo; Analisadores linguísticos nos níveis morfológicos, sintáticos e semânticos; Vocabulários e ontologias; Reconhecedores de nomes próprios e siglas; e Algoritmos e Heurísticas. Vale ressaltar que, algoritmos e heurísticas se aplicam a todo software, porém neste caso específico, o autor os menciona como operadores de texto.

Para Bayeza-Yates e Ribeiro Neto (2013), a lista de termos do documento pode ser reduzida por meio do emprego de operadores de texto, a saber:

- Análise léxica: que elimina dígitos, pontuação entre outros;
- Eliminação de stopwords: artigos, preposições, pronomes, conjunções e alguns adjetivos e advérbios;
- Operação de stemming: que reduz a palavra ao seu radical;

- Seleção de termos de Indexação: determina as palavras que são utilizadas como elementos de indexação (substantivos são mais representativos que adjetivos, verbos, advérbios etc.).

Os operadores são aplicados à indexação automática para obtenção de melhores resultados, podendo ser utilizadas de forma separada ou de forma combinada.

Na literatura existem diversos estudos comparativos entre a indexação manual e a indexação automática. Segundo Vieira (1988, p. 46):

A comparação entre os dois tipos de indexação, automática e manual, é realizada para se verificar as diferenças e semelhanças entre os termos selecionados por programas de um computador e pelo homem. De acordo com os resultados obtidos, avalia-se a aplicabilidade de uma ou outra técnica. Os testes de comparação podem ser divididos em testes de qualidade de indexação e de qualidade de recuperação.

A grande maioria de testes comparativos entre descritores atribuídos manual e automaticamente, segundo Salton (1972), chega a um resultado aproximado de 60% de compatibilidade entre uma linguagem e outra.

Carroll e Roeloffs (1969) *apud* Vieira (1988), realizaram estudos comparativos entre indexação manual e automática na área de Ciência da Informação, e, por meio de análise de correlação estatística, constataram que os termos obtidos pelos indexadores humanos foram semelhantes aos obtidos pela indexação automática, porém, os custos de contratação, treinamento especializado e a inconsistência humana, torna a indexação automática mais viável. (VIEIRA, 1988, p. 46).

Na área médica, Salton (1969), Boyce e Lockard (1975) *apud* Vieira (1988) compararam os mesmos documentos indexados de forma manual, fazendo uso de vocabulário controlado, e indexados automaticamente, pelo programa Automatic Document Analysis and Retrieval System (SMART), para tanto, utilizaram termos livres do resumo. Neste experimento foi constatado que, ao utilizar um controle por meio de tesouros e dicionários, a eficiência da indexação automática é semelhante à da indexação manual. (VIEIRA, 1988, p. 46).

De acordo com Gil Leiva (2008), muitos pesquisadores se posicionam contra e a favor da indexação automática. Essas divergências foram constatadas na década de 1970 e se mantiveram até a década de 1990.

Os primeiros defensores da indexação automática se fundamentaram nos avanços da informática e com o desenvolvimento das ferramentas para PLN, assim como os primeiros sistemas envolvendo a inteligência artificial. Os argumentos contra a indexação automática eram voltados a incapacidade de um programa de computador em indexar os documentos com

qualidade, tendo em vista as tarefas de analisar, detectar, relacionar e seleccionar os conceitos explícitos e implícitos na representação de conteúdo.

No quadro 8, Gil leiva (2008) descreve os argumentos a favor e os argumentos contra a indexação automática, sob a perspectiva de vários autores:

Quadro 8: Argumentos acerca da Indexação automática

Argumentos contra a indexação automática	
Uma indexação automática poderia ser considerada como um caso especial de degeneração da indexação humana, em que o indexador se torna um mero observador dos indicadores estatísticos do programa de computador.	Cooper (1978, p. 108)
Enquanto a indexação automática reconhece os caracteres de primeira sequência que são palavras não vazias, a indexação intelectual distingue conceitos, isto é, representações mentais de objetos do conhecimento.	Slype (1991, p. 116)
A busca da indexação totalmente automática é um esforço inútil porque os termos de indexação propostos necessitam de uma edição humana se si busca uma indexação útil e aceitável	Wellisch (1992)
Na atualidade os sistemas baseados em normas de decisão assim como os baseados em métodos estatísticos, sintáticos e semânticos, tentam imitar os resultados produzidos por indexadores humanos, mas seu sucesso está limitado pelas lacunas existentes na compreensão do processo da indexação.	Milstead (1994, p. 578)
A indexação automática ou assistida, na maioria dos casos, está restrita a áreas específicas do conhecimento. Por outro lado, a indexação automática não poderá substituir a intelectual para a análise de documentos importantes.	Bonura (1994)

Argumentos a favor da indexação automática	
O desenvolvimento contínuo da produção e a publicação de informação em formato eletrônico favorecerá o	

desenvolvimento de investigações e propostas para a obtenção da indexação automática	Sparck Jones (1986, p.12), citada por Pinto Molina (1989, p. 365) e Ward (1996, p. 225)
A indexação humana é subjetiva, lenta e custosa	García Gutiérrez (1984, p.114), Bertrand-Gatald e Pagola (1992, p. 85)
As bases de dados documentais se alimentam da indexação realizada por diferentes indexadores com critérios distintos e variáveis em determinados momentos. Daí a necessidade de unificar procedimentos com a automatização destas operações	García Gutiérrez (1984, p.128)
A indexação automática é mais rápida, econômica, consistente e efetiva que a manual	Anderson (1994, p. 632)
Com a indexação automática se pode ter bons resultados tanto na exaustividade como na precisão da recuperação em comparação com diferentes especialistas	Bordoni e Paziienza (1997, p. 26)
Quando se utilizam palavras como termos de indexação, a indexação automática é tão boa quanto a manual	Hmeidi, Kanaan e Evens (1997, p. 18)
A utilização de alguns parâmetros contrastados na indexação automática poderia reduzir a inconsistência e repercutir positivamente na coerência do sistema de informação	Gil Leiva (1999 e 2003)

Fonte: Baseado em Gil Leiva (2008, p. 320-321).

Os estudos apontam que a indexação automática tem se mostrado uma boa alternativa à demanda das unidades informacionais, tendo em vista a crescente produção científica, os custos referentes à indexação manual e os avanços tecnológicos para tornar viável a recuperação da informação.

2.4.1 Indexação automática por atribuição

O processo de indexação requer a compatibilidade entre a linguagem utilizada pelo indexador e a linguagem de busca do usuário, uma vez que, a não padronização da linguagem poderá acarretar dispersões terminológicas e sintáticas. Como consequências desses problemas, Boccato (2009) afirma que o sistema de recuperação da informação poderá ficar comprometido e que, este fato, gera insatisfação do usuário.

Na tentativa de resolver estes possíveis problemas, Lapa (2014) aponta a linguagem controlada como solução:

Portanto, a linguagem de indexação controlada deve atuar nos sistemas de informação tanto para orientar o indexador sobre quais são os termos mais adequados para representar o assunto de um documento, quanto para guiar os pesquisadores sobre o modo de escolher os termos indexados que representam no sistema o assunto procurado (LAPA, 2014, p. 74).

Nesse contexto, a linguagem de indexação busca uniformizar a linguagem utilizada pelo sistema de indexação e pelo usuário, levando em consideração a comunicação e os objetivos de ambos, além de permitir que seus usuários adotem estratégias de buscas mais eficientes. Na indexação automática os problemas ocorrem de maneira semelhante, porém conhecer bem o sistema é fundamental para obtenção de melhores resultados.

Acerca da indexação automática por extração, Lancaster (2004, p. 289) afirma que “a extração de palavras e/ou expressões dos documentos é tarefa que os computadores executam de modo bastante satisfatório”. O autor complementa evidenciando que há uma nítida vantagem em relação à extração feita por seres humanos: “[...] é totalmente coerente” (LANCASTER, 2004, p. 289).

Porém o autor alerta que, na maioria das vezes, a indexação realizada manualmente não constitui indexação por extração, e sim por atribuição (necessário o uso de um vocabulário controlado), portanto, é considerado um trabalho mais difícil para um computador executar (LANCASTER 2004).

Referente à indexação automática por atribuição, foco desta pesquisa, Lancaster (2004) afirma que este tipo de indexação é mais complexo por envolver o uso de vocabulário controlado contendo termos autorizados para a indexação.

Lancaster (2004, p. 19) define vocabulário controlado como “essencialmente uma lista de termos autorizados”. Porém, o autor ressalta que não se trata meramente de uma lista em si, mas de uma estrutura semântica que se destina a:

[...] controlar sinônimos, optando-se por uma única forma padronizada, com remissivas de todas as outras formas; 2 - diferenciar homógrafos. Por exemplo, PERU (PAÍS) é um termo bastante diferente de PERU (AVE); e 3 - reunir ou ligar termos cujos significados apresentem uma relação mais estreita entre si [...]

Nesse contexto, são identificados dois tipos relações: hierárquicas e não-hierárquicas (ou associativas), onde o termo relaciona-se com outro de maneira hierárquica ou por meio de uma associação. Lancaster exemplifica os tipos de relações utilizando o termo MULHERES

OPERÁRIAS, que segundo o autor, relaciona-se hierarquicamente com MULHERES e DONAS DE CASA (espécies do termo MULHERES), bem como está associado a outros termos como EMPREGO ou FAMÍLIAS MONOPARENTAIS, que aparecem em hierarquias diferentes. (LANCASTER, 2004).

Para Ferneda (2013, p. 64-65), a utilização do uso de vocabulário controlado possui algumas vantagens:

Uma das vantagens subjacentes ao uso de vocabulários controlados é que estes podem ser disponibilizados para os usuários de um sistema de informação, permitindo que tenham acesso à terminologia empregada na indexação dos documentos. Isto possibilita compatibilizar a linguagem dos usuários à linguagem utilizada na representação documentos, resultando em uma recuperação mais eficiente. Além disso, a padronização terminológica proporcionada pelo uso de um vocabulário controlado permite o agrupamento lógico dos documentos de um corpus, criando-se grupos (clusters) ou classes de documentos que são representados por determinados termos de indexação.

Na indexação automática por atribuição, programas de computador desenvolvem, para cada termo a ser indexado, um “perfil” de palavras ou expressões que costumam ocorrer frequentemente nos documentos. Podemos inferir que este processo se resume em duas etapas: a primeira etapa seria extrair palavras ou expressões do texto por meio de técnicas estatísticas; a segunda etapa seria consultar um vocabulário controlado, e a partir dele, atribuir o termo que melhor o represente.

2.4.2 SISA (Sistema de Indización Automático)

No artigo “Indexação automatizada de artigos de periódicos científicos: análise da aplicação do software sisa com uso da terminologia DeCS na área de odontologia” (NARUKAWA, LEIVA e FUJITA, 2009), os autores apontam que este software foi desenvolvido no ano de 1999-2008 pelo professor e pesquisador Isidoro Gil Leiva, da Universidade de Múrcia, na Espanha.

Originalmente o SISA foi proposto como um sistema de indexação semiautomático aplicado à área da Biblioteconomia e Documentação, porém permite que seja aplicado a qualquer área do conhecimento, desde que sua configuração seja adaptada de acordo com algumas exigências do software (NARUKAWA, 2011, p.106).

A seguir são descritos os requisitos de entrada de dados do SISA:

- Lista alfabética de termos e descritores: arquivo texto contendo os termos e a relação de termo autorizado;
- Lista alfabética de termos e respectivos termos gerais: arquivo opcional contendo o termo autorizado e a relação de termo geral;
- Lista pré-estabelecida de palavras vazias no idioma do texto dos documentos, para fins de eliminação das palavras consideradas vazias (*stopwords*), como conectivos e artigos;
- Marcação das partes constituintes do documento a ser indexado;
- Todos os arquivos de entrada, incluindo os textos a serem indexados, devem estar no formato txt.

Assim, os arquivos utilizados no processamento do SISA são: o texto completo (título, resumo e texto); uma lista de palavras vazias (*stoplist*); e uma linguagem documentária.

Durante a análise de texto do documento, o software adota uma metodologia por meio da comparação entre termos do documento constituído por título, resumo e texto e termos de uma linguagem documentária, como por exemplo um tesouro da área do conhecimento dos documentos.

O SISA identifica termos de um vocabulário controlado no texto de um documento, identificando sua ocorrência no título, resumo e no texto completo.

Para que determinado termo do vocabulário controlado seja proposto como termo de indexação, é necessário que ele satisfaça algumas das regras envolvendo critérios preestabelecidos de frequência e posição no texto (BANDIM; CORRÊA, 2018). Além dos termos sugeridos de indexação, o sistema lista também termos candidatos à indexação, que são os termos que não são stopwords e não estão no vocabulário controlado, mas que atendem aos critérios de frequência e posição dos termos.

De acordo com Gil Leiva (2017), termos candidatos são:

“Termos não incluídos no vocabulário controlado, nem na lista de stopwords, e que cumprem requisitos como aparecendo um número mínimo de vezes e em diferentes parágrafos. Assim, a indexação do SISA não depende unicamente da presença ou ausência de um termo do vocabulário, e também é possível realizar um feedback automático do vocabulário.” (GIL LEIVA, 2017, p. 146).

Após o sistema sugerir os termos de indexação para um documento, o sistema permite que o usuário possa escolher entre os termos sugeridos e termos candidatos, mantendo assim a característica de sistema semiautomático de indexação.

De acordo com Narukawa, Leiva e Fujita (2009) a metodologia que envolve o processo de indexação pelo SISA se dá em três etapas:

1. **Pré-processamento**: nesta fase as partes constituintes do documento devem ser marcadas pelo indexador para sua identificação por meio dos símbolos: Título deve ser delimitado por #CTI# e #FTI#; Resumo deve ser delimitado por #CR# e #FR#; e Texto deve ser delimitado por #CTE# e #FTE#. É nesta fase também que o software realiza a eliminação das palavras vazias e horizontalização das frases, desta forma, o total de palavras das fontes título, resumo e texto são computadas;
2. **Análise de conteúdo**: nesta fase ocorre a etapa de análise de conteúdo, onde um algoritmo busca a extração de termos da linguagem documentária que coincidem com termos das fontes;
3. **Valoração e ponderação**: esta última etapa consiste na aplicação de critérios de avaliação dos termos para que o sistema possa selecionar os termos de indexação que representarão o conteúdo do documento. Na seleção e proposição dos termos para indexação são aplicados os seguintes critérios: o termo é apresentado como termo de indexação se, um termo autorizado aparece no título e no resumo, ou se um termo autorizado aparece no título e no texto, ou se um termo autorizado aparece no resumo e no texto do documento. No entanto, os termos não autorizados ou semivazios são apresentados como candidatos à indexação se, a palavra semivazia aparece no título, resumo e texto, ou se aparece no texto dez vezes ou mais, além de aparecer em oito parágrafos diferentes ou mais. Assim, o sistema irá selecionar e propor os termos de indexação de acordo com a frequência e posição das palavras em relação ao documento.

Para concluir a indexação, é necessário que haja a intervenção humana para analisar e decidir os termos de indexação propostos pelo sistema. Por ser um sistema flexível, é possível acrescentar ou suprimir termos, isso como forma de permissão para que o indexador possa tomar a decisão considerando as particularidades do sistema de informação (NARUKAWA, LEIVA e FUJITA, 2009).

2.4.3 MAUI (Multi-purpose Automatic Topic Indexing)

O termo MAUI (*Multi-purpose Automatic Topic Indexing*) pode ser traduzido como indexação automática por tópicos em múltiplos domínios. O software foi desenvolvido por Alyona Medelyan como parte de seu projeto de PhD, sob a supervisão de Ian H. Witten e Eibe Frank, no Departamento de Ciência da Computação da Universidade de Waikato, Nova Zelândia.

O MAUI baseia-se no algoritmo de extração de palavras-chave (do inglês *keyphrase*) do software KEA, e fornece funcionalidades adicionais como atribuição de termos de vocabulário controlado, e opcionalmente a atribuição de tópicos a documentos baseados em termos da Wikipedia usando o Wikipedia Miner. Esses novos recursos possibilitam identificar termos de indexação com mais precisão.

O MAUI extrai automaticamente os principais termos em documentos de texto, e a depender da tarefa, os termos correspondem a tópicos, tags, palavras-chave, frases-chave, termos de vocabulário, descritores, termos de índice ou títulos de artigos da Wikipédia.

Para representar o conteúdo dos documentos, o MAUI utiliza métodos estatísticos e linguísticos para a ponderação de termos, estes métodos levam em consideração a frequência do termo (term frequency), posição do termo (term position), intervalo de ocorrência (range of occurrence), frequência de ocorrência (frequency of the occurrence), frequência de co-ocorrência (frequency of co-occurrence), vocabulários controlados (controlled vocabularies) e a Wikipedia.

Para Milne e Witten (2008) a abordagem baseada na Wikipédia é tão precisa quanto os seres humanos na quantificação da relação entre dois conceitos, sendo possível determinar o parentesco total de uma palavra dada a todas as outras candidatas.

De acordo com Medelyan (2009), o software executa as seguintes tarefas:

[...] atribuição de termos com um vocabulário controlado ou tesauro; indexação de assunto; indexação de tópico com termos da Wikipedia; extração de palavras-chave; extração de terminologia; marcação automática, extração de terminologia e indexação de tópico semiautomática e extração de palavras-chave não apenas do texto, mas também com uma referência a um vocabulário controlado (MEDELYAN, 2009, tradução nossa).

Na prática, ele pode ser usado como uma ferramenta de sugestão de marca que fornece aos usuários tags descrevendo os principais tópicos de documentos recentemente adicionados, que podem ser corrigidos ou aprimorados por marcas pessoais, se necessário. Isto pode melhorar a consistência de uma folksonomia, por exemplo, sem comprometer a sua flexibilidade.

O software apresenta a especificidade da utilização de um algoritmo de aprendizagem de máquina para gerar um modelo de seleção de termos de indexação como base no comportamento humano de indexação, e esta característica o aproxima da indexação manual.

A metodologia e etapas de processamento segundo Medelyan (2009, p 118) são:

1. **Geração de tópicos candidatos:** geração de tópicos candidatos individualmente para cada documento de entrada, com base em seu texto, gerando assim, uma lista de tópicos candidatos que são passados para a próxima etapa, que calcula os recursos;
2. **Cálculo de características** – cálculo de características para os termos candidatos: frequência do termo (term frequency), posição da última ocorrência (position of the last occurrence), propagação (spread), similaridade semântica (semantic similarity), frequência inversa da Wikipedia (inverse Wikipedia frequency) e generalidade (generality);
3. **Construção do modelo de indexação:** treinamento do modelo de indexação levando em conta os termos atribuídos pelos indexadores a cada documento do conjunto de treinamento;
4. **Aplicação do modelo aprendido para selecionar tópicos para outros documentos:** aplicação do modelo de indexação treinado para propor termos de indexação a outros documentos.

A partir destas informações, pode-se descrever as características de cada sistema, levantando informações acerca de como são processados os documentos e de como é feita a extração e seleção dos termos que resultam nos descritores propostos por cada sistema.

2.5 Avaliação da indexação automática

De acordo com Gil Leiva (2008), a avaliação da indexação automática pode ser realizada de forma intrínseca (qualitativa ou quantitativa) ou extrínseca (por meio da interconsistência ou recuperação). Basicamente é realizada pela obtenção de índices de consistência, precisão, revocação e medida F. Tais índices são obtidos por meios de fórmulas e refletem diretamente na qualidade da indexação.

Para a nossa melhor compreensão, os tipos de avaliação podem ser observados no Quadro 9.

Quadro 9: Tipos de Avaliação da Indexação

Intrínseca	Intrínseca qualitativa
	Intrínseca quantitativa
Extrínseca	Extrínseca mediante interconsistência
	Extrínseca mediante recuperação

Fonte: Gil Leiva (2008).

Em relação à avaliação na indexação automática, Gil Leiva (1999) resume que a avaliação intrínseca permite medir o grau de consistência da indexação, enquanto a avaliação extrínseca, o grau de revocação, precisão e medida F na indexação automática.

2.5.1 Avaliação intrínseca

A avaliação intrínseca está relacionada à fatores internos e formada por um conjunto de tarefas voltadas para o resultado da indexação, como por exemplo, descritores, cabeçalhos, subcabeçalhos ou identificadores, com o intuito de conhecer sua qualidade. A avaliação intrínseca da indexação pode ser qualitativa ou quantitativa.

- Avaliação Intrínseca Qualitativa

Na avaliação intrínseca qualitativa, Gil Leiva (2008) afirma que alguns elementos determinam a qualidade do produto resultante da indexação, como: a *Exaustividade* – extração de conceitos caracterizadores do conteúdo dos documentos; a *Especificidade* – relação exata entre as unidades conceituais escolhidas e o termo eleito para representá-lo; a *Correção* – ausência de erros de inclusão e omissão de termos; e a *Perspectiva do usuário* – considerar interesses e necessidades dos usuários.

A análise de todos estes elementos está associada à qualidade do atendimento das necessidades dos usuários da informação, no que tange as buscas por meio do sistema de recuperação de informações.

- Avaliação Intrínseca Quantitativa

A avaliação intrínseca quantitativa consiste na reindexação de um conjunto de documentos, sob as mesmas condições da primeira indexação, para a comparação entre dois indexadores. Desta forma, o grau de consistência revelará a diferença ou semelhança entre os indexadores. O grau de consistência será maior quanto mais semelhantes forem as indexações.

De acordo com Gil Leiva (2008, p. 386), podemos obter o grau de consistência por meio das seguintes fórmulas:

Hooper (1965):

$$\frac{C}{A + B - C}$$

Onde,

C: Termos comuns nas duas indexações;
A: Termos usados na indexação A, mas não na B;
B: Termos usados na indexação B, mas não na A.

Rolling (1981),

$$\frac{2C}{A + B}$$

Onde,

C: Termos comuns nas duas indexações;
A: Termos usados na indexação A;
B: Termos usados na indexação B.

Gil Leiva (1999, 2003 e 2008) se baseia na fórmula de Hooper para expressar o grau de consistência:

$$C_i = \frac{T_{co}}{(A+B)-T_{co}}$$

Onde,

T: Número de termos comuns nas duas indexações

A: Número de termos usados na indexação A;

A: Número de termos usados na indexação B.

A avaliação intrínseca quantitativa realizada por meio da intraconsistência, ou seja, passado um tempo, o mesmo profissional indexa o mesmo documento para verificar se há a necessidade de mudança e/ou adequação na indexação, a fim de satisfazer às necessidades de recuperação de usuários. Este índice varia entre 0 e 1, e a aplicação da fórmula pode ser obtida de duas formas: relaxada e rígida. Segundo Gil Leiva, Rubi e Fujita (2008, p 238):

Relaxada: quando um cabeçalho ou subcabeçalho de assunto de um documento coincide com o assunto de outro documento, considera-se coincidência total (1); quando ocorre somente o cabeçalho ou subcabeçalho, considera-se a metade (0,5) e quando não há nenhuma coincidência, o valor é 0; *Rígida*: quando o assunto determinado coincide completamente.

Sendo assim, o índice de consistência pode ser calculado das duas formas: relaxada e rígida. A relaxada ou flexível como é conhecida, pode variar entre 0, 0,5 e 1, a depender da coincidência entre os assuntos dos documentos. Entretanto a rígida, não há meio termo, ou é igual a 0 ou é igual a 1. O índice de consistência é um fator fundamental para avaliar a qualidade da indexação e bastante útil para comparar indexações periódicas em uma unidade de informação.

2.5.2 Avaliação extrínseca

A avaliação extrínseca consiste na comparação com a indexação de outra unidade de informação que tenha indexado os mesmos documentos, ou com relação ao teste da função da indexação na recuperação. Pode ser obtida por meio da interconsistência ou recuperação (GIL LEIVA, 2017, p. 151).

- Avaliação Extrínseca por meio da Interconsistência

Neste método é aplicada a mesma fórmula da consistência apresentada anteriormente, sendo que, a diferença consiste em comparar as indexações de duas instituições ou sistemas de indexação, que tenham indexado o mesmo documento e dispoendo das mesmas ferramentas de indexação.

Cabe ressaltar que, as principais diferenças entre as indexações consistem no uso de linguagens documentarias distintas, políticas de indexação, bem como a subjetividade inerente à indexação (GIL LEIVA, 2008, p. 391, tradução nossa).

- Avaliação Extrínseca por meio da Recuperação

Este tipo de avaliação permite investigar duas bases de dados que apresentem campos e conteúdos comuns, exceto os campos que armazenam a informação. Normalmente costuma ser considerado um método de alto custo, pois envolve tempo e disponibilidade por parte do avaliador, porém, seus resultados são claros e satisfatórios. Para Gil Leiva (2017), a avaliação extrínseca por meio da recuperação consiste em interrogar duas bases de dados a partir dos índices de revocação, precisão e medida F.

No artigo de Bandim e Corrêa (2019) “Indexação automática por atribuição de artigos científicos em português da área de Ciência da Informação” os autores avaliaram a qualidade na indexação automática por meio dos índices de consistência, precisão, revocação e medida F, tendo como padrão de referência as palavras-chaves dos autores, para tanto utilizaram as seguintes fórmulas:

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

O índice de revocação é o resultado da relação entre os termos relevantes recuperados e o total de termos relevantes existentes para cada documento.

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

O índice de precisão é o resultado da relação entre os termos relevantes recuperados e o total de termos recuperados.

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

De acordo com Bandim e Corrêa (2019), a medida F é definida como:

A média harmônica entre o índice de Precisão e o índice de Revocação, sendo uma maneira de combinar a Precisão e Revocação em um único número. Se nenhum termo relevante for recuperado, a Medida F assume valor zero e, quando todos os termos atribuídos forem relevantes (máxima Precisão) e forem exaustivamente recuperados todos os termos relevantes (máxima Revocação), assume valor igual a um que corresponde a 100%. A medida F só assume valores altos quando os índices de Revocação e Precisão são altos também, refletindo um *equilíbrio* entre os dois índices antagônicos (BANDIM; CORRÊA, 2019, p. 6).

A medida F é bastante utilizada pela comunidade científica para avaliar a indexação automática. Lembrando que esta forma de avaliação foi realizada por Bandim e Corrêa (2018, 2019) para cálculo de precisão, revocação e medida F na comparação entre palavras-chaves do sistema e com as dos autores dos artigos científicos.

Para Papagiannopoulou e Tsoumakas (2019), a indexação automática não é uma tarefa simples, pois apresenta alguns problemas como a subjetividade e a falta de esquemas quanto ao desempenho dos métodos de avaliação. No artigo intitulado “A Review of Keyphrase Extraction”, os autores discorrem acerca da indexação automática e discutem os diferentes tipos de avaliação da indexação automática. Os resultados de sua pesquisa apontam que os métodos estatísticos são mais utilizados pela comunidade científica e que as métricas de avaliação mais conhecidas e utilizadas para avaliar o desempenho da indexação automática são: precisão (precision), revocação (recall) e medida F (F measure), além do ranking de medida de qualidade (ranking quality measure) que compreende a média recíproca (mean reciprocal rank) e a precisão média (mean average precision), medida de frequência binária (binary preference measure) e média das palavras-chave corretamente extraídas (average of correctly extract keyphrases).

Todos os tipos de avaliação acima citados visam contribuir para a eficácia dos sistemas de recuperação da informação nas bases de dados. Contudo, a avaliação da indexação se caracteriza em um instrumento de análise e investigação, no que tange aspectos do processo de

indexação, e representa um papel primordial aos SRI no sentido de adequar seus produtos e serviços para o público a que se destina.

2.5.3 Corpus de referência na avaliação de sistemas de indexação automática

A escolha de um corpus de referência é uma etapa bastante importante durante a execução de uma pesquisa sobre indexação automática. Na Linguística, a elaboração de um corpus é considerada uma etapa fundamental, tendo em vista que é por meio dele que se pode caracterizar publicações de acordo com objetivos e aspectos metodológicos planejados para uma pesquisa.

Trask (2004, p. 68) entende que corpus é "um conjunto de textos escritos ou falados numa língua disponível para análise". Sendo assim, o corpus pode ser compilado de acordo com a natureza da pesquisa e aplicado a qualquer área do conhecimento.

Na área da Ciência da Informação, frequentemente corpus são utilizados e se constituem em uma série de documentos categorizados de acordo com os objetivos e desenvolvimento da pesquisa. Os documentos que compõem um corpus podem ser reunidos a partir de publicações inseridas em bases de dados de periódicos científicos da CI e utilizados como instrumentos da pesquisa para avaliar a qualidade da indexação automática.

No artigo "*Automatic keyphrase extraction: a survey of the state of the art*" os autores Hasan e Ng (2014) apresentam um estudo que aponta os principais erros de extração de palavras-chave dos sistemas de indexação automática: o comprimento das fontes, a consistência estrutural, a mudança de tópicos e a correlação de tópicos. Os autores afirmam ainda que os sistemas de indexação automática foram avaliados em *corpora*² de fontes diferenciadas como publicações científicas, e-mails e resumos.

A fim de efetivar a avaliação de sistemas de indexação automática, Hasan e Ng (2014, p. 1267) classificam o processo em duas etapas: a) o mapeamento das palavras-chave dos corpora, tanto da indexação manual quanto da indexação automática; b) o uso de métricas de avaliação, como precisão (P), revocação (R) e medida-F (F) confrontando as duas formas de indexação.

Em sua tese de doutorado, Renato Rocha Souza (2005) construiu um corpus composto de 60 artigos da área da Ciência da Informação escritos em português. O corpus compreende artigos

² Corpora significa o plural de corpus.

científicos inseridos em bases de dados científicas de revistas científicas como a Datagramazero e revista Ciência da Informação. Os documentos que compõem o corpus, em sua totalidade, abrangem os campos científicos da Biblioteconomia, Documentação e Ciência da Informação, vide anexo A.

A presente pesquisa utiliza-se do corpus documental de Souza (2005) pelo fato do corpus ter sido utilizado anteriormente em algumas pesquisas, a exemplo podemos citar a dissertação de Bandim (2017) que utilizou o corpus para avaliar a qualidade do processo de indexação automática do SISA por meio dos índices de consistência, revocação, precisão e medida F. A pesquisa de Bandim (2017) se assemelha a presente pesquisa por utilizar os mesmos métodos de avaliação e corpus de referência.

Cabe ressaltar que a utilização do corpus em pesquisas anteriores representa um fator de qualidade e credibilidade ao experimento da pesquisa frente à necessidade de validar métodos de avaliação de sistemas de indexação automática.

3 METODOLOGIA

Toda pesquisa necessita de uma metodologia, pois de acordo com Gil (2010, p. 25), isto “possibilita melhor organização dos fatos e conseqüentemente o seu entendimento”. Ou seja, possibilita uma aproximação entre os conceitos estabelecidos e a realidade. Assim, é necessária a compreensão do que pode ser determinado como pesquisa.

De acordo com Gil (2007, p. 15), pesquisa é:

O procedimento racional e sistemático que tem como objetivo proporcionar respostas aos problemas que são propostos. Esta será requerida quando não houver informação suficiente para responder ao problema, ou então quando a informação disponível se encontra em tal estado de desordem que não possa ser adequadamente relacionada ao problema.

Portanto, uma pesquisa tem por objetivo responder aos problemas propostos, além de buscar conhecimento e fornecer informações. Deve ser realizada a partir de uma metodologia em que se comprove os resultados para ser considerada válida.

Conforme Lakatos e Marconi (2010, p. 65):

O método é o conjunto das atividades sistemáticas e racionais que, com maior segurança e economia, permite alcançar o objetivo - conhecimentos válidos e verdadeiros - traçando o caminho a ser seguido, detectando erros e auxiliando as decisões do cientista.

Sendo assim, a metodologia serve para auxiliar ao pesquisador quanto aos procedimentos e instrumentos necessários durante a execução da pesquisa, bem como para validar os resultados encontrados. A pesquisa pode ser classificada quanto aos objetivos da pesquisa, natureza da pesquisa, objeto de estudo, coleta e análise dos dados (OLIVEIRA, 2001. P. 19)

3.1 Classificação da pesquisa

A pesquisa classifica-se como exploratória e empírica, onde o método utilizado para construção deste trabalho é um estudo de caso e a forma de coleta de dados se dá por meio de um experimento de indexação. Conforme os objetivos deste trabalho, trata-se de pesquisa exploratória, visto que este tipo de pesquisa busca “estudar, explorar o problema a fim de torná-lo explícito e possibilita a criação de hipóteses” (CAJUEIRO, 2015, p.16).

A pesquisa exploratória é, normalmente, o passo inicial no processo de pesquisa pela experiência e um auxílio que traz a formulação de hipóteses significativas para posteriores pesquisas (CERVO; BERVIAN; SILVA, 2007).

O caráter teórico deste trabalho é composto por pesquisa bibliográfica e revisão da literatura em Ciência da Informação sobre indexação automática e indexação automática por atribuição. Em termos pragmáticos, a pesquisa classifica-se em experimento e estudo de caso por avaliar e comparar a qualidade da indexação realizada pelos sistemas de indexação automática por atribuição SISA e MAUI, que constituem o universo desta pesquisa.

Este trabalho apresenta como instrumento de pesquisa o corpus composto de 60 artigos científicos identificados na Tese de Doutorado de Souza (2005), que serão submetidos aos sistemas SISA e MAUI fazendo uso do vocabulário controlado TBCI.

3.2 Parâmetros de qualidade da indexação

A qualidade da indexação é medida por meio das métricas de indexação: Consistência, Revocação, Precisão e Medida F. De acordo com Gil Leiva (1997), índices de Revocação, Precisão e Medida F são utilizados para a avaliar a indexação de duas bases de dados por meio da comparação de seus resultados.

Ao que se refere à Consistência, Narukawa (2011, p. 128) aponta que vários estudos sobre a indexação automática por atribuição revelam que a margem de consistência obtida nos resultados de investigações varia entre 10 e 60%. Sendo assim, a indexação é considerada de qualidade quando sua consistência média está entre estes percentuais.

No entanto, Lancaster (2004) afirma que o índice de consistência não é o único aspecto determinante da qualidade da indexação.

Bandim e Correa (2019, p. 5) consideram que a indexação é de boa qualidade “quando os termos de indexação disponibilizados aos usuários representam os itens informacionais atribuindo a totalidade de termos relevantes sem incluir termos pouco relevantes como descritores documentais”.

Bandim e Correa (2019) apontam que os índices de Revocação, Precisão e Medida F são fundamentais para qualificar a indexação. O índice de Revocação é obtido por meio da relação entre os termos relevantes atribuídos e o total de termos relevantes existente para cada artigo. Para se obter o índice de Precisão, calcula-se a relação entre os termos relevantes atribuídos e o total de termos atribuídos para cada artigo. A medida F é a média harmônica entre o índice

de Precisão e o índice de Revocação, sendo uma maneira de combinar a Precisão e Revocação em um único número (BANDIM; CORREA, 2019, p. 5-6).

Assim, esta pesquisa baseia-se nas métricas de Consistência, Precisão, Revocação e Medida F, a fim de analisar, comparar e avaliar a indexação automática dos softwares SISA e MAUI.

3.3 Etapas da pesquisa

No intuito de alcançar ao objetivo desta pesquisa, foram elaboradas etapas relacionadas aos objetivos específicos proporcionando assim, uma melhor visualização e planejamento acerca dos procedimentos necessários. No quadro 10 estão todos os objetivos e etapas, a saber:

Quadro 10: Objetivos e etapas da pesquisa

<p>Objetivo A: Especificar, descrever e realizar o processo de indexação automática por atribuição envolvendo os softwares SISA e MAUI</p>	<p style="text-align: center;">Etapas:</p> <p>1 - Levantar e analisar a literatura acerca da indexação automática por atribuição (sistemas de indexação automática, avaliação de sistemas de indexação automática);</p> <p>2 - Descrever os sistemas de indexação automática por atribuição SISA e MAUI;</p> <p>3 - Preparar os arquivos de entrada contendo o TBCI e os textos dos documentos do corpus da pesquisa para submeter aos sistemas de indexação automática SISA e MAUI de acordo com as especificações e configurações de cada sistema;</p> <p>4 - Indexar o corpus pelos sistemas de indexação automática SISA e MAUI, utilizando o TBCI como linguagem de indexação.</p> <p>5 - Criar uma lista de descritores atribuídos por cada sistema a cada documento do corpus.</p>
---	--

<p>Objetivo B: Analisar os resultados obtidos por ambos os sistemas, avaliando a qualidade da representação dos assuntos dos documentos do corpus pelos sistemas SISA e MAUI</p>	<p style="text-align: center;">Etapas:</p> <p>6 - Calcular os valores dos índices de consistência, revocação, precisão e medida F alcançados por cada sistema na indexação de cada documento;</p> <p>7 - Avaliar a indexação de cada um dos sistemas por meio do resultado dos índices obtidos.</p>
<p>Objetivo C: Analisar comparativamente a indexação automática por atribuição dos sistemas SISA e MAUI</p>	<p style="text-align: center;">Etapas:</p> <p>8 - Comparar os índices de consistência, revocação, precisão e medida F dos sistemas SISA e MAUI;</p> <p>9 - Analisar comparativamente a qualidade dos descritores propostos pelos sistemas para cada artigo.</p>

Fonte: Elaborado pelo autor.

As etapas 1, 2 consistem em identificar, caracterizar e descrever os softwares SISA e MAUI a fim de obter o conhecimento necessário para a realização da indexação automática. Estas etapas compreendem a fundamentação teórica desta pesquisa e abordam os aspectos da indexação automática que são fundamentais para compreensão do estudo que se segue.

As etapas 3, 4 consistem na preparação para a realização do experimento de indexação, onde os arquivos que compõem o corpus da pesquisa serão preparados conforme as especificações do processamento de cada software. Lembrando que, tanto o software SISA quanto o MAUI possuem algumas recomendações que deverão ser obedecidas, como por exemplo o formato dos arquivos de entrada, os arquivos utilizados no processamento e a utilização do vocabulário controlado.

A etapa 5 consiste na obtenção dos resultados provenientes da indexação automática do SISA e do MAUI. Neta etapa será criada uma lista de descritores atribuídos por cada sistema a cada documento do corpus.

As etapas seguintes 6 e 7, consistem na obtenção dos índices de consistência, revocação, precisão e medida F, resultados da avaliação intrínseca e extrínseca da indexação, bem como da análise dos resultados obtidos.

A etapa 8 e 9 consistem na análise dos índices obtidos considerando os aportes teóricos desta pesquisa e na análise comparativa entre os índices obtidos pela indexação de cada software, bem como os fatores que influenciaram a indexação.

Diante do exposto, foram definidas etapas conforme os objetivos específicos postos pela pesquisa. O quadro permite visualizar cada etapa necessária para a realização deste estudo. As seções subsequentes detalharão cada etapa relacionando-as aos objetivos específicos desta pesquisa.

ETAPA 1 - Levantar e analisar a literatura acerca da indexação automática por atribuição (sistemas de indexação automática, avaliação de sistemas de indexação automática)

A etapa 1 da presente pesquisa tem por objetivo identificar e descrever os principais aspectos da indexação automática por atribuição na literatura, levando em consideração os estudos acerca de avaliação da indexação automática, bem como sua utilização e viabilidade na área da Ciência da Informação.

Esta etapa ainda compreende a pesquisa bibliográfica que fundamenta esta Dissertação e encontra-se no capítulo 2 da presente pesquisa. Diversas fontes bibliográficas foram utilizadas na elaboração do referido capítulo, dentre elas livros e artigos impressos e eletrônicos de periódicos científicos da área da Ciência da Informação e Ciência da Computação, disponibilizados em meio digital como Repositórios Institucionais, Biblioteca Digital de Teses e Dissertações (BDTD), Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), além de buscadores online.

O capítulo 2 é constituído de produções científicas de autores renomados na área da Ciência da Informação, Ciência da Computação e áreas afins, permitindo assim, o conhecimento dos aspectos da Organização da Informação, Vocabulário Controlado, Indexação, Indexação

Automática, Avaliação da Indexação, bem como sistemas de Indexação Automática. Os seguintes termos foram utilizados como estratégia de busca: *Indexação*, *Indexação Automática*, *Avaliação da Indexação Automática* e *Software de Indexação*, considerando pesquisas entre o período de 1950 a 2019.

O recorte de tempo para compor as referências acerca da temática se deu no período de maio de 2018 a julho de 2019, período este referente à construção do Referencial teórico e da Metodologia desta pesquisa.

Com relação à seleção e à análise das fontes informacionais da fundamentação teórica desta pesquisa, somente as fontes pertinentes para a análise dos dados da pesquisa foram consideradas.

ETAPA 2 - Descrever os sistemas de indexação automática por atribuição SISA e MAUI

Na literatura da área da Ciência da Informação identificamos com facilidade diversos estudos acerca do software SISA. Em contrapartida, o software MAUI ainda é pouco utilizado, porém sua especificidade em utilizar um algoritmo de aprendizagem de máquina para gerar um modelo de seleção de termos de indexação como base no comportamento humano de indexação despertou o interesse em avaliar sua indexação automática.

Para descrever os softwares, foram utilizados aspectos ou categorias tais como: arquivos utilizados no processamento da indexação; formato dos arquivos de entrada; utilização de linguagem documentária; metodologias e etapas; flexibilidade quanto ao idioma e domínio; ponderação dos termos; operações sobre o texto no processamento de linguagem natural; e abordagem adotada por cada sistema durante a indexação automática por atribuição.

ETAPA 3 - Preparar os arquivos de entrada contendo o TBCI e os textos dos documentos do corpus da pesquisa para submeter aos sistemas de indexação automática SISA e MAUI de acordo com as especificações e configurações de cada sistema

Nesta etapa os arquivos de entrada foram formatados mediante as especificações de cada software, bem como preparar e inserir as listas de palavras vazias e o vocabulário controlado TBCI. Sendo assim, o corpus da pesquisa contendo os 60 artigos da área da Ciência da Informação em português foi formatado e analisado separadamente para cada software.

SISA

Os arquivos de entrada do SISA foram devidamente preparados considerando suas características e conforme seu processamento, destacando assim, as marcações de título, resumo e texto dos artigos e a inserção da Lista de stopwords ou palavras vazias, Lista de descritores e lista de termos gerais da Linguagem de indexação TBCI.

Com relação à lista de palavras vazias é a mesma utilizada por Bandim e Corrêa (2019). A lista foi fornecida como entrada ao SISA, consistindo de uma lista pré-estabelecida com palavras que não possuem significado, e que não são levadas em consideração na indexação, a exemplos podemos citar os conectivos, artigos, entre outros. Portanto, ocorre o confronto entre o documento e as palavras vazias, permitindo que o software aponte os termos da indexação (NARUKAWA, 2011, p. 64). A lista de palavras vazias

Com relação à lista de descritores, no artigo de Bandim e Corrêa (2019) “Indexação automática por atribuição de artigos científicos em português da área de Ciência da Informação”, os autores avaliaram a qualidade na indexação automática por meio dos índices de consistência, precisão, revocação e medida F, levando em consideração as palavras-chaves dos autores. A presente pesquisa utilizou a mesma lista de descritores proveniente desse estudo, por ser derivada da mesma linguagem de indexação, o TBCI. A lista de descritores consiste de uma lista organizada seguindo a ordem alfabética e enumerada, em que foi permitido aos descritores somente a relação de equivalência utilizando a sugestão USE.

Com relação a preparação dos artigos científicos que compõem o corpus desta pesquisa, as partes constituintes dos documentos foram marcadas para sua identificação por meio dos símbolos: *Título* foi delimitado por #CTI# e #FTI#; *Resumo* foi delimitado por #CR# e #FR#; e o *Texto* foi delimitado por #CTE# e #FTE#, cujo significado segue abaixo:

- #CTI# - começo do título;
- #FTI# - fim do título;
- #CR# - começo do resumo
- #FR# - fim do resumo;
- #CTE# - começo do texto;
- #FTE# - fim do texto.

Cabe ressaltar que corpus é composto por 60 publicações das revistas científicas DataGramZero e Ciência da Informação, sendo um total de 29 documentos no formato HTML

e 31 documentos no formato PDF. Por exigência do SISA, os arquivos foram convertidos para o formato TXT, além de passar da fonte minúscula para a maiúscula.

Sendo assim, a preparação dos arquivos utilizados pelo SISA consistiu de três tipos de arquivos: a linguagem de indexação (TBCI), a lista de palavras vazias e os artigos científicos.

MAUI

Da mesma forma, os arquivos de entrada do MAUI foram preparados conforme suas especificações e processamento.

Diferente do SISA, o MAUI analisa o texto completo dos documentos sem marcações. Sendo assim, os textos dos 60 artigos do corpus foram submetidos sem marcações e obedecendo a exigência de inserção da lista de stopwords ou palavras vazias, bem como do uso do TBCI como vocabulário controlado.

O formato dos arquivos de texto exigido pelo MAUI é o TXT para o texto completo, lista de palavras-chave, e lista de palavras vazias, e SKOS para o vocabulário controlado. Nesta fase de preparação das fontes, o conjunto de treinamento envolvendo documentos e respectivas palavras-chave foi utilizado para treinar o modelo de aprendizagem de máquina.

Este treinamento se deu por meio de um conjunto dos 30 primeiros documentos do corpus e de termos de uma folksonomia assistida, esta última, fruto da revisão da folksonomia de Silva e Corrêa (2019) na pesquisa “O processo de construção do corpus de referência em Ciência da Informação”. O estudo citado propõe e aplica um percurso metodológico para a construção de corpus de referência na área de Ciência da Informação e resulta numa folksonomia assistida para o corpus de Souza (2005).

Para Santarém (2010, p. 209), a 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". Para Silva e Corrêa (2019), essa questão é determinante para não descaracterizar a folksonomia, apresentando aos usuários a possibilidade de escolha de termos de uma hierarquia pré-existente (SILVA; CORRÊA, 2019).

A folksonomia assistida acima citada, foi utilizada na presente pesquisa por se tratar de um produto de construção social que visa a classificação da informação de forma mais flexível, permitindo a participação de vários indexadores, tornando a informação mais acessível e democrática. Cabe destacar que a mesma foi revisada levando-se em consideração o TBCI como linguagem de indexação.

ETAPA 4 - Indexar o corpus pelos sistemas de indexação automática SISA e MAUI, utilizando o TBCI como linguagem de indexação

Após a preparação das fontes para cada software, e adicionalmente, do treinamento do modelo de aprendizagem de máquina do MAUI, foi realizada a indexação automática e seu resultado foi salvo em arquivos de texto contendo os descritores atribuídos pelos softwares a cada artigo do corpus.

SISA

Para o processamento da indexação automática do SISA, nas configurações do sistema foram selecionamos os arquivos de Lista de palavras vazias e Lista de descritores, e em seguida, foram abertos os arquivos de texto completo e solicitada sua indexação automática, quando ocorreu a análise de conteúdo e a valoração e ponderação dos termos.

Referente à análise de conteúdo, basicamente o texto foi analisado e um algoritmo buscou a extração de termos da linguagem documentária, ou seja, do TBCI, que coincidiram com termos das fontes.

Com relação à valoração e ponderação, esta etapa consistiu na aplicação de critérios de avaliação dos termos que permite a seleção pelo sistema dos termos de indexação que representam o conteúdo dos documentos.

Na seleção e proposição dos termos para indexação foram aplicados os seguintes critérios: o termo foi apresentado como termo de indexação se o termo autorizado apareceu no título e no resumo, ou um termo autorizado apareceu no título e no texto, ou um termo autorizado apareceu no resumo e no texto do documento. (BANDIM; CORRÊA, 2018)

No entanto, os termos semivazios (termos candidatos a descritores que se apresentam em uma lista e podem ou não ser incluídos como termos de indexação) foram apresentados como candidatos à indexação se, a palavra semivazia apareceu no título, resumo e texto, ou se apareceu no texto dez vezes ou mais, além de ter aparecido em oito parágrafos diferentes ou mais.

Assim, o sistema selecionou e propôs os termos de indexação de acordo com a frequência e posição das palavras em relação ao documento.

MAUI

Para representar o conteúdo dos documentos, o MAUI utiliza métodos estatísticos e linguísticos para a ponderação de termos. Tais métodos consideram características como a frequência de ocorrência do termo, a posição de ocorrência do termo, o tamanho do termo em palavras, e a probabilidade de um termo ser palavra-chave no domínio ou corpus.

Na presente pesquisa, o MAUI se baseou na mesma lista de palavras vazias que o SISA, porém como exigência do software, foi dado como entrada o TBCI no formato SKOS e adicionalmente, foi realizado o treinamento supervisionado do modelo de aprendizagem de máquina.

Após o treinamento supervisionado foi realizado o processo da análise do documento por meio da geração de termos candidatos à indexação, seguido do cálculo de características para os termos candidatos e da aplicação do modelo de indexação treinado para a seleção de termos de indexação.

O processamento do MAUI se deu em várias etapas. A primeira delas é a geração de tópicos candidatos, nesta etapa houve a extração dos termos candidatos. Em seguida ocorreu o cálculo de características, ou seja, as características que envolvem o termo em relação ao texto foram calculadas. A terceira etapa é a construção do modelo de indexação, que considerou os termos que foram atribuídos na folksonomia a cada documento do conjunto do treinamento (30 primeiros documentos do corpus). Na última etapa aconteceu a aplicação do modelo aprendido a fim de selecionar os termos que foram propostos pelo MAUI.

ETAPA 5 - Criar uma lista de descritores atribuídos por cada sistema a cada documento do corpus

Como resultado da indexação automática, o SISA gera informes (relatórios) contendo os termos de indexação para cada documento indexado, tais informes foram copiados para um arquivo de texto.

Tendo sido realizada a indexação automática dos artigos pelo MAUI, os termos de indexação obtidos foram registrados em arquivo de texto.

Após a indexação automática dos softwares, os termos de indexação propostos pelos softwares foram dispostos em listas distintas e salvas em arquivos para possibilitar a comparação e avaliação dos termos posteriormente.

ETAPA 6 - Calcular os valores dos índices de consistência, revocação, precisão e medida F alcançado por cada sistema na indexação de cada documento

Esta etapa se deu por meio do preenchimento de duas planilhas contendo as principais informações acerca dos termos propostos por cada software, como podemos observar nas figuras abaixo.

Figura 2: Informações da planilha do SISA

A	B	C	D	E	F	G	H	I	J
Nº do artigo	Termos do SISA	Termo comum com folksonomia	Número de termos do SISA	Número de termos da folksonomia	Número de termos comuns com a folksonomia	Consistência	Precisão	Revocação	Medida F1

Fonte: Dados da pesquisa.

Figura 3: Informações da planilha do MAUI

A	B	C	D	E	F	G	H	I	J
Nº do artigo	Termos do MAUI	Termo comum com folksonomia	Número de termos do MAUI	Número de termos da folksonomia	Número de termos comuns com a folksonomia	Consistência	Precisão	Revocação	Medida F1

Fonte: Dados da pesquisa.

Após o preenchimento das planilhas, os índices de consistência, precisão, Revocação e Medida F foram obtidos, possibilitando assim, a avaliação e comparação dos softwares.

ETAPA 7 - Avaliar a indexação de cada um dos sistemas por meio do resultado dos índices obtidos

A avaliação da indexação automática dos softwares se deu por meio da avaliação intrínseca e extrínseca. A avaliação intrínseca mede o grau de consistência da indexação e a avaliação extrínseca mede o grau de Precisão, Revocação e Medida F na indexação automática.

Gil Leiva (1997, p. 22) define consistência da indexação como o grau de concordância entre indexadores de um mesmo grupo ou entre indexadores de grupos diferentes, quando da

representação da informação essencial de um documento, por meio de um conjunto de termos de indexação selecionados por esses indexadores.

A consistência média de cada software foi obtida por meio da média da consistência em cada um dos artigos do corpus entre cada sistema e os indexadores via folksonomia assistida. Sendo assim, a indexação automática dos softwares pode ser avaliada. Cabe ressaltar que a pesquisa considera os índices de consistência entre 0 e 100% e, quanto mais próxima de 100%, maior será considerada a qualidade dos termos de indexação propostos pelo software.

Na avaliação extrínseca os graus de precisão, revocação e medida F são obtidos por meio da relação entre o número de termos relevantes recuperados, número total de termos relevantes e o número total de termos recuperados. São considerados recuperados os termos atribuídos pelos sistemas aos documentos, e relevantes os termos que constam na folksonomia assistida para cada artigo.

ETAPA 8 - Comparar os índices de consistência, revocação, precisão e medida F dos sistemas SISA e MAUI

Após a avaliação da indexação automática do SISA e do MAUI por meio da obtenção dos índices de Consistência, Precisão, Revocação e Medida F, foi feita a comparação dos índices citados. Alguns aspectos referentes à indexação e às especificações dos softwares foram analisados e descritos para melhor compreensão dos resultados obtidos.

ETAPA 9 - Analisar comparativamente a qualidade dos descritores propostos pelos sistemas para cada artigo

Esta etapa da pesquisa envolveu a análise qualitativa e comparativa dos termos atribuídos pelos softwares SISA e MAUI a partir do TBCI.

A análise qualitativa considerou os valores dos índices de consistência, precisão, revocação e medida F obtidos por meio da atribuição dos termos propostos pelos softwares, seguido da avaliação intrínseca e extrínseca da indexação automática.

A análise comparativa se deu por meio da comparação dos índices de consistência, precisão, revocação e medida F de cada artigo indexado pelos softwares. Durante esta etapa, os fatores que influenciaram a indexação dos softwares foram investigados, a fim de qualificar a

indexação e levantar informações necessárias para elucidar seus resultados e alcançar o objetivo da pesquisa.

4 ANÁLISE E DISCUSSÃO DE RESULTADOS

Neste capítulo será apresentada a análise dos resultados do experimento acerca da avaliação da qualidade da indexação automática por atribuição por meio dos índices de consistência, revocação, precisão e medida F, alcançando assim, os objetivos deste trabalho.

4.1 Análise comparativa das características dos sistemas

Por meio da análise de conteúdo dos trabalhos selecionados na etapa 1, foi possível analisar e descrever as características do processo de indexação de cada um dos sistemas, levantando informações acerca de como os documentos são processados e de como é realizada a extração e seleção dos termos que resultam nos descritores propostos por cada sistema.

Como resultado, o quadro 11 descreve detalhadamente as características dos softwares estudados:

Quadro 11: Descrição dos softwares SISA e MAUI

CARACTERÍSTICA	SISA	MAUI
Arquivos utilizados no processamento	-Texto completo: texto com marcações de título, resumo e texto; -Lista de stopwords; -Vocabulário controlado formatados em dois arquivos textuais	- Texto completo: texto sem marcações; - Lista de stopwords; - Vocabulário controlado; - Conjunto de treinamento envolvendo documentos e respectivas palavras-chave, utilizado para treinar o modelo de aprendizagem de máquina.
Formato dos arquivos de entrada	TXT	TXT, SKOS (vocabulário controlado).
Utilização de Linguagem Documentária	SIM, obrigatório. O vocabulário controlado deve ser formatado em arquivo de texto de acordo com o padrão exigido pelo SISA.	SIM, obrigatório para a indexação automática por atribuição. Incorpora vocabulários controlados no formato SKOS.
Flexibilidade	Extensível para outros idiomas e outras áreas do conhecimento por meio do vocabulário controlado e lista de stopwords.	Extensível para outros idiomas e outras áreas do conhecimento por meio do vocabulário controlado, lista de stopwords, software radicalizador e conjunto de treinamento. Pode ser ajustado para otimizar seu desempenho para coleções específicas.

Etapas de processamento	O processo de análise do documento é efetuado através das etapas: - Preparação das fontes; - Análise do conteúdo: extração de termos do vocabulário controlado que ocorrem no documento constituído por título, resumo e texto. - Valoração e ponderação dos termos: leva em conta critérios preestabelecidos de frequência e posição dos termos no documento para propor os termos de indexação.	Inicialmente, consiste no treinamento de um modelo de aprendizado de máquina supervisionado para propor termos de indexação. Posteriormente, o processo de análise do documento é efetuado através das etapas: - Geração de termos candidatos à indexação; - Cálculo de características para os termos candidatos. - Aplicação do modelo de indexação treinado na seleção de termos de indexação.
Operações sobre o texto	- Análise léxica; - Stopwords; - Vocabulário controlado	- Análise léxica; - Stopwords; - Radicalização; - Extração de n-gramas; - Vocabulário controlado
Ponderação dos termos	Frequência, posição	Frequência, posição, tamanho, probabilidade de ser palavra-chave e grau de relação semântica.
Abordagem na indexação automática por atribuição	Não-supervisionada por meio de sistema especialista. Abordagem estatística e linguística.	Supervisionada no treinamento de modelo de aprendizagem de máquina. Abordagem estatística e linguística.

Fonte: Dados da pesquisa.

De acordo com o quadro, fica evidente que SISA e MAUI possuem algumas características em comum quanto aos arquivos utilizados no processamento, formato dos arquivos de entrada, utilização de linguagem documentária, flexibilidade, ponderação dos termos e processamento da linguagem natural, porém, o MAUI se destaca por possuir outras características além das que o SISA apresenta nestes critérios.

Com relação aos demais critérios utilizados no quadro, percebe-se que o MAUI se diferencia em relação ao SISA quanto ao critério de abordagem na indexação automática por atribuição e por possuir características distintas quanto as etapas de processamento. Cabe ressaltar que o MAUI apresenta uma característica específica de aprendizado de máquina, que pode ser um fator determinante para uma melhor eficácia na indexação automática.

Por meio da análise comparativa dos dois softwares, foi possível perceber suas características e a forma como cada sistema opera, mediante suas especificidades e exigências. A análise comparativa resultante desta pesquisa permite ao indexador e pesquisadores, o acesso às informações mais relevantes e precisas acerca dos softwares analisados.

Os softwares apresentam características em comum, porém o MAUI se mostra mais completo por conta dos recursos adicionais de processamento de linguagem natural que implementa, e por apresentar a especificidade de um algoritmo de aprendizado de máquina. O uso de aprendizado de máquina promete oferecer melhor eficácia na representação da informação por gerar modelo a partir da indexação manual, considerada a melhor forma de indexação por pesquisadores da área.

4.2. Análise comparativa da qualidade na indexação automática

Analisando-se os resultados do experimento pode-se afirmar que, comparando os termos de indexação automática atribuídos pelo SISA com os termos atribuídos pelo MAUI, verificou-se que o SISA atribuiu termos, na maioria das vezes, simples e genéricos, em contrapartida, o MAUI atribuiu uma grande quantidade de termos de indexação, na maioria das vezes, termos compostos, influenciando diretamente os cálculos dos índices de consistência na indexação e, os índices de precisão, revocação e medida F.

4.2.1 Indexação automática do SISA

Para analisar a indexação automática do SISA, foi construída uma planilha constituída dos seguintes campos: Número de termos do SISA, Número de termos da folksonomia, Número de termos comuns com a folksonomia, Consistência, Precisão, Revocação e Medida F, conforme a tabela 1:

Tabela 1: Resultados do SISA

Nº Artigo	Número de termos do SISA	Número de termos da folksonomia	Número de termos comuns com a folksonomia	Consistência	Precisão	Revocação	Medida F
1	6	11	2	13%	33%	18%	24%
2	4	10	4	40%	100%	40%	57%
3	4	10	2	17%	50%	20%	29%
4	5	10	1	7%	20%	10%	13%

5	3	10	1	8%	33%	10%	15%
6	5	10	3	25%	60%	30%	40%
7	6	11	4	31%	67%	36%	47%
8	3	14	2	13%	67%	14%	24%
9	6	12	3	20%	50%	25%	33%
10	3	12	2	15%	67%	17%	27%
11	5	10	3	25%	60%	30%	40%
12	6	10	5	40%	83%	50%	63%
13	4	13	3	21%	75%	23%	35%
14	8	12	5	33%	63%	42%	50%
15	3	11	2	17%	67%	18%	29%
16	6	11	5	42%	83%	45%	59%
17	6	11	4	31%	67%	36%	47%
18	2	15	2	13%	100%	13%	24%
19	7	13	3	18%	43%	23%	30%
20	7	11	5	38%	71%	45%	56%
21	1	9	1	11%	100%	11%	20%
22	2	7	2	29%	100%	29%	44%
23	1	11	1	9%	100%	9%	17%
24	4	10	2	17%	50%	20%	29%
25	5	13	4	29%	80%	31%	44%
26	5	9	2	17%	40%	22%	29%
27	7	9	2	14%	29%	22%	25%
28	8	9	5	42%	63%	56%	59%
29	5	7	3	33%	60%	43%	50%
30	5	14	2	12%	40%	14%	21%
31	5	9	3	27%	60%	33%	43%
32	8	8	3	23%	38%	38%	38%
33	11	13	5	26%	45%	38%	42%
34	4	13	2	13%	50%	15%	24%
35	9	7	5	45%	56%	71%	63%
36	6	7	2	18%	33%	29%	31%
37	9	12	4	24%	44%	33%	38%

38	7	14	3	17%	43%	21%	29%
39	10	7	4	31%	40%	57%	47%
40	7	12	5	36%	71%	42%	53%
41	10	11	5	31%	50%	45%	48%
42	10	12	9	69%	90%	75%	82%
43	4	8	3	33%	75%	38%	50%
44	4	11	2	15%	50%	18%	27%
45	5	8	3	30%	60%	38%	46%
46	6	7	5	63%	83%	71%	77%
47	6	12	5	38%	83%	42%	56%
48	5	9	4	40%	80%	44%	57%
49	4	11	3	25%	75%	27%	40%
50	5	9	3	27%	60%	33%	43%
51	4	11	3	25%	75%	27%	40%
52	7	9	4	33%	57%	44%	50%
53	3	8	2	22%	67%	25%	36%
54	8	10	3	20%	38%	30%	33%
55	3	9	2	20%	67%	22%	33%
56	9	10	5	36%	56%	50%	53%
57	4	9	3	30%	75%	33%	46%
58	5	7	5	71%	100%	71%	83%
59	5	10	4	36%	80%	40%	53%
60	8	13	7	50%	88%	54%	67%

Fonte: Dados da pesquisa (2019).

Por meio da tabela, podemos observar que no campo *Número de termos do SISA*, a quantidade de termos atribuídos pela indexação automática do software revelou um mínimo de 1 e um máximo de 11 termos, gerando uma média 5,55 termos atribuídos pelo software.

No campo *Número de termos da folksonomia*, a quantidade de termos atribuídos foi de um mínimo de 7 e um máximo de 15 termos, gerando uma média de 10,35 termos.

No campo *Número de termos comuns com a folksonomia*, a quantidade de termos obtida foi de um mínimo 1 e um máximo de 9 termos, gerando uma média de 3,35 termos comuns.

Foi verificada uma média de 28% no índice de Consistência, 63% quanto ao índice de Precisão, 34% de Revocação e a Medida F de 42%.

De acordo com Narukawa (2011), vários estudos acerca da indexação automática por atribuição apontam que os índices de consistência obtidos, estão, aproximadamente, entre 10% e 60%, sendo assim, a média de 28% no índice de consistência para o corpus de 60 artigos indexados, está entre o intervalo de índices esperados, levando-se em consideração os estudos acima citados.

Além do mais, a média encontrada para o índice de consistência de 28% é superior à média de 19% encontrada por Bandim e Corrêa (2018), ao compararem as palavras-chave de autores de artigos científicos da área da Ciência da Informação e os termos propostos pelo SISA a partir do vocabulário controlado TBCI.

Cabe destacar que Bandim e Corrêa (2019) obtiveram a média da precisão, revocação e medida F de 30%, 37% e 30%, respectivamente. Comparando-os com os resultados obtidos na presente pesquisa, os índices foram menos satisfatórios, tendo em vista que os índices de precisão, revocação e medida F desta última, foram de 63%, 34% e 42%, respectivamente.

Narukawa (2011) ressalta que o SISA sofre algumas influências acerca da atribuição de termos pelo fato de o software atribuir termos que somente façam parte do vocabulário controlado, além de compará-los com termos do artigo considerando a presença destes termos no título, resumo e texto do artigo analisado. Como exemplos acerca de influências que o SISA sofre, podemos citar o singular e plural de palavras, além da dificuldade em atribuir termos compostos, termos que possuam aspas, hifens, parênteses e outros.

4.2.2 Indexação automática do MAUI

Para analisar a indexação automática do MAUI foi construída uma planilha constituída dos seguintes campos: Número de termos do MAUI, Número de termos da folksonomia, Número de termos comuns com a folksonomia, Consistência, Precisão, revocação e Medida F, como podemos observar na tabela 2:

Tabela 2: Resultados do MAUI

Nº Artigo	Número de termos do MAUI	Número de termos da folksonomia	Número de termos comuns com a folksonomia	Consistência	Precisão	Revocação	Medida F
1	7	11	5	38%	71%	45%	56%
2	8	10	5	38%	63%	50%	56%
3	8	10	5	38%	63%	50%	56%
4	10	10	5	33%	50%	50%	50%
5	10	10	4	25%	40%	40%	40%
6	10	10	5	33%	50%	50%	50%
7	10	11	5	31%	50%	45%	48%
8	10	14	5	26%	50%	36%	42%
9	10	12	7	47%	70%	58%	64%
10	10	12	5	29%	50%	42%	45%
11	6	10	3	23%	50%	30%	38%
12	10	10	5	33%	50%	50%	50%
13	10	13	7	44%	70%	54%	61%
14	10	12	7	47%	70%	58%	64%
15	10	11	7	50%	70%	64%	67%
16	10	11	7	50%	70%	64%	67%
17	10	11	7	50%	70%	64%	67%
18	10	15	2	9%	20%	13%	16%
19	10	13	6	35%	60%	46%	52%
20	10	11	4	24%	40%	36%	38%
21	6	9	3	25%	50%	33%	40%
22	10	7	6	55%	60%	86%	71%
23	6	11	3	21%	50%	27%	35%
24	7	10	4	31%	57%	40%	47%
25	10	13	6	35%	60%	46%	52%
26	9	9	3	20%	33%	33%	33%
27	10	9	5	36%	50%	56%	53%
28	10	9	4	27%	40%	44%	42%
29	10	7	5	42%	50%	71%	59%

30	10	14	4	20%	40%	29%	33%
31	10	9	3	19%	30%	33%	32%
32	10	8	5	38%	50%	63%	56%
33	10	13	6	35%	60%	46%	52%
34	10	13	6	35%	60%	46%	52%
35	10	7	5	42%	50%	71%	59%
36	10	7	3	21%	30%	43%	35%
37	10	12	8	57%	80%	67%	73%
38	10	14	8	50%	80%	57%	67%
39	10	7	5	42%	50%	71%	59%
40	10	12	8	57%	80%	67%	73%
41	10	11	4	24%	40%	36%	38%
42	10	12	7	47%	70%	58%	64%
43	10	8	6	50%	60%	75%	67%
44	10	11	5	31%	50%	45%	48%
45	10	8	4	29%	40%	50%	44%
46	10	7	6	55%	60%	86%	71%
47	10	12	5	29%	50%	42%	45%
48	10	9	6	46%	60%	67%	63%
49	10	11	5	31%	50%	45%	48%
50	10	9	5	36%	50%	56%	53%
51	10	11	7	50%	70%	64%	67%
52	10	9	6	46%	60%	67%	63%
53	7	8	4	36%	57%	50%	53%
54	10	10	7	54%	70%	70%	70%
55	10	9	4	27%	40%	44%	42%
56	10	10	6	43%	60%	60%	60%
57	10	9	3	19%	30%	33%	32%
58	10	7	4	31%	40%	57%	47%
59	10	10	4	25%	40%	40%	40%
60	10	13	5	28%	50%	38%	43%

Fonte: Dados da pesquisa (2019).

A partir da tabela, podemos verificar que no campo *Número de termos do MAUI*, a quantidade de termos atribuídos pela indexação automática do software revelou um mínimo de 6 e um máximo de 10 termos, gerando uma média 9,57 termos atribuídos pelo software.

No campo *Número de termos da folksonomia*, a quantidade de termos atribuídos foi de um mínimo de 7 e um máximo de 15 termos, gerando uma média de 10,35 termos.

No campo *Número de termos comuns com a folksonomia*, a quantidade de termos obtida foi de um mínimo 2 e um máximo de 8 termos, gerando uma média de 5,15 termos comuns.

Foi verificada uma média de 36% no índice de Consistência, 54% quanto ao índice de Precisão, 51% de Revocação e a Medida F obtida foi em média 52%.

Sendo assim, a média de 36% da consistência obtida pela indexação automática do MAUI neste experimento é considerada boa, pois encontra-se entre o intervalo de índices esperados, levando-se em consideração os estudos sobre indexação automática e de acordo a categorização dos níveis de consistência de Bandim e Corrêa (2018):

0 a 11% - desempenho insatisfatório (corresponde a uma média de 0 a 1 termo em comum); 11 a 25% - desempenho satisfatório (corresponde a uma média de 1 a 2 termos em comum); 25 a 43% - desempenho bom (corresponde a uma média de 2 a 3 termos em comum); 43 a 67% - desempenho ótimo (corresponde a uma média de 3 a 4 termos em comum); 67 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 (BANDIM; CORRÊA, 2018. P. 72).

Sendo assim, os autores basearam-se na fórmula do índice de consistência e nos níveis de desempenho de acordo com as faixas de valores correspondentes.

4.2.3 Análise comparativa da qualidade da indexação automática entre os sistemas SISA e MAUI

Neste experimento, foram obtidos os índices de Consistência, Precisão, Revocação e Medida F, a fim de comparar os termos propostos pelos softwares e qualificar a indexação automática do SISA e do MAUI.

As tabelas a seguir apresentam as médias das métricas do SISA e MAUI, respectivamente:

Tabela 3: Resultado das métricas do SISA

	Número de termos do SISA	Número de termos da folksonomia	Número de termos comuns com a folksonomia	Consistência	Precisão	Revocação	Medida F
Mínimo	1,00	7,00	1,00	7%	20%	9%	13%
Média	5,55	10,35	3,35	28%	63%	34%	42%
Desvio padrão	2,27	2,05	1,53	14%	20%	16%	16%
Máximo	11,0	15,00	9,00	71%	100%	75%	83%

Fonte: Dados da pesquisa.

Tabela 4: Resultado das métricas do MAUI

	Número de termos do MAUI	Número de termos da folksonomia	Número de termos comuns com a folksonomia	Consistência	Precisão	Revocação	Medida F
Mínimo	6,00	7,00	2,00	9%	20%	13%	16%
Média	9,57	10,35	5,15	36%	54%	51%	52%
Desvio padrão	1,11	2,05	1,42	11%	13%	15%	12%
Máximo	10,00	15,00	8,00	57%	80%	86%	73%

Fonte: Dados da pesquisa.

Diante do exposto, as faixas de valores apresentadas e o resultado de 28% e 36%, dos índices de consistência do SISA e do MAUI, respectivamente, nos leva a compreensão de que o desempenho de ambos é bom, tendo em vista que os dois softwares estão na faixa de 25 a 43%.

Com relação ao índice de Precisão, o SISA apresentou uma média de 63%, enquanto o MAUI apresentou uma média de 54%. De acordo com Hlava (2002), o índice de precisão deve ter por volta de 60% para ser considerado bom, sendo assim, o SISA superou o MAUI.

Com relação ao índice de Revocação, o SISA apresentou uma média de 34% e o MAUI de 51%. Sendo assim, o MAUI apresentou uma maior revocação em detrimento de sua precisão ter sido mais baixa que a do SISA.

Os índices citados acima interferem diretamente nos percentuais médios da Medida F, que nos permite calcular a média harmônica entre o índice de precisão e o índice de revocação, sendo uma maneira de combinar a precisão e revocação em um único número. Para se conseguir valores médios altos para medida F é necessário que se tenha valores altos para revocação e precisão (BANDIM; CORRÊA, 2017, p. 67).

As médias da Medida F apresentadas pelo SISA e MAUI foram, respectivamente, de 42% e 52%. De acordo com Silva e Corrêa (2019, p. 100), “como a medida F é a média harmônica entre precisão e revocação, entende-se que quanto mais perto de 100%, maior será a qualidade da indexação, e quando tende a 0% menos qualificado”. Sendo assim, o MAUI se mostrou mais próximo do 100% e, portanto, apresentou maior qualidade na indexação.

Diante do exposto, é importante ressaltar que durante o processo de indexação existiram alguns fatores intervenientes que agiram diretamente na indexação, como por exemplo o contexto, a política e os objetivos da indexação, a diversidade e as características do objeto a ser indexado e até mesmo quanto ao uso do vocabulário controlado (GIL LEIVA, 2008).

Durante a pesquisa verificamos os principais fatores que interferiram na indexação automática dos softwares:

- Atribuição de termos no plural e singular;
- Dificuldade em atribuir termos compostos;
- Diferenças entre as estruturas nos termos de indexação;
- Atribuição automática de termos gerais e específicos.

Com relação aos termos no singular e no plural foram constatadas variações de um mesmo conceito, como os termos *hipertexto* e *hipertextos* propostos pelo MAUI no artigo 22, cabe destacar que ambos são termos autorizados do TBCI para diferentes conceitos.

Com relação à atribuição de termos compostos, o SISA se destacou pois apresentou dificuldades nesse quesito, como podemos verificar em vários arquivos, e a exemplo podemos citar o artigo 3, em que software propõe o termo *Direito*, sendo que o termo aparece no artigo diversas vezes, mas como palavra composta e não isolada.

Outro exemplo de fator interveniente é o caso dos dois softwares não distinguirem alguns termos por conta de suas estruturas, como por exemplo os termos *Pesquisa* e *Pesquisadores*, no artigo 24. Outro exemplo é no artigo 27 em que o MAUI propôs os termos *Índice h* e *Índices* de forma errônea.

Por fim, verificamos que os softwares atribuíram muitos termos de forma genérica e específica. Destacamos o SISA por possuir dificuldade maior que a o MAUI referente à atribuição de termos compostos. De acordo com Bandim e Corrêa (2017), isto acontece porque no processo de análise documentária automática do conteúdo do documento pelo sistema SISA, os vocabulários controlados condicionam a seleção automática dos termos que serão atribuídos para indexação, neste caso o TBCI.

Cabe ressaltar que, o fato de o MAUI propor mais termos compostos não faz dele um sistema mais preciso, pelo contrário. Segundo Narukawa (2011, p. 136) “um dos fatores que tem influência sobre os índices de consistência na indexação é a quantidade de termos atribuídos”. Sendo assim, constatamos que o MAUI proporciona um número elevado de termos de indexação em relação à indexação do SISA. A explicação do elevado número de termos propostos pelo MAUI é a atribuição dos termos gerais e também dos específicos.

No entanto, os índices da indexação automática são influenciados pelos fatores intervenientes, que por sua vez, estão associados às questões morfológicas, linguísticas e semânticas, e ainda à aplicação do TBCI.

Os fatores intervenientes não configuram erros dos sistemas SISA e MAUI, mas sim em conhecimento no intuito de melhorias e adaptações nos algoritmos dos sistemas para que possa selecionar e atribuir mais termos importantes na indexação automática por atribuição com o uso de vocabulário controlado (BANDIM; CORRÊA, 2017, p 69).

Nas tabelas a seguir, podemos visualizar os melhores casos da indexação automática realizada pelo SISA e pelo MAUI, respectivamente:

Tabela 5: Os (7) sete melhores casos da indexação realizada pelo SISA

Artigo	Termos do SISA	Número de termos da Folksonomia	Número de termos comuns com a Folksonomia	Consistência	Precisão	Revocação	Medida F
Artigo 12	6	10	5	40%	83%	50%	63%
Artigo 16	6	11	5	42%	83%	45%	59%
Artigo 35	9	7	5	45%	56%	71%	63%
Artigo 42	10	12	9	69%	90%	75%	82%
Artigo 46	6	7	5	63%	83%	71%	77%
Artigo 58	5	7	5	71%	100%	71%	83%
Artigo 60	8	13	7	50%	88%	54%	67%

Fonte: Dados da pesquisa (2019).

Tabela 6: Os 7 (sete) melhores casos da indexação realizada pelo MAUI

Artigo	Termos do MAUI	Número de termos da Folksonomia	Número de termos comuns com a Folksonomia	Consistência	Precisão	Revocação	Medida F
Artigo 16	10	11	7	50%	70%	64%	67%
Artigo 17	10	11	7	50%	70%	64%	67%
Artigo 22	10	7	6	55%	60%	86%	71%
Artigo 37	10	12	8	57%	80%	67%	73%
Artigo 40	10	12	8	57%	80%	67%	73%
Artigo 46	10	7	6	55%	60%	86%	71%
Artigo 54	10	10	7	54%	70%	70%	70%

Fonte: Dados da pesquisa (2019).

Podemos constatar que, comparando os percentuais obtidos nas tabelas 3 e 4, verifica-se claramente, que os dois softwares apresentaram bons resultados.

Os sete melhores casos do SISA apresentaram o índice de consistência que variou de 42 a 71%; o índice de precisão entre 83 e 100%; índice de revocação que variou de 45 a 71%, resultando assim, na medida F entre 59 a 83%.

Os sete melhores casos do MAUI apresentaram o índice de consistência que variou de 50 a 57%; o índice de precisão entre 60 e 80%; índice de revocação que variou de 64 a 86%, resultando assim, na medida F entre 67 a 73%. Os softwares apresentaram em comum os artigos 16 e 46.

Tabela 7: Os 7(sete) piores casos da indexação realizada pelo SISA

Artigo	Termos do SISA	Número de termos da Folksonomia	Número de termos comuns com a Folksonomia	Consistência	Precisão	Revocação	Medida F
Artigo 4	5	10	1	7%	20%	10%	13%
Artigo 5	3	10	1	8%	33%	10%	15%
Artigo 8	3	14	2	13%	67%	14%	24%
Artigo 10	3	12	2	15%	67%	17%	27%
Artigo 21	1	9	1	11%	100%	11%	20%
Artigo 23	1	11	1	9%	100%	9%	17%
Artigo 27	7	9	2	14%	29%	22%	25%

Fonte: Dados da pesquisa (2019).

Tabela 8: Os sete (7) piores casos da indexação realizada pelo MAUI

Artigo	Termos do SISA	Número de termos da Folksonomia	Número de termos comuns com a Folksonomia	Consistência	Precisão	Revocação	Medida F
Artigo 18	10	15	2	9%	20%	13%	16%
Artigo 23	6	11	3	21%	50%	27%	35%
Artigo 26	9	9	3	20%	33%	33%	33%
Artigo 30	10	14	4	20%	40%	29%	33%
Artigo 31	10	9	3	19%	30%	33%	32%
Artigo 36	10	7	3	21%	30%	43%	35%
Artigo 57	10	9	3	19%	30%	33%	32%

Fonte: Dados da pesquisa (2019).

Podemos constatar que, comparando os percentuais obtidos nas tabelas 5 e 6, verifica-se claramente, que os dois softwares apresentaram resultados parecidos, porém o índice de Medida F evidenciou que o MAUI apontou melhores resultados.

Os sete piores casos do SISA apresentaram o índice de consistência que variou de 7 a 15%, o índice de precisão entre 20 e 100%, índice de revocação que variou de 9 a 22%, resultando assim, na medida F entre 13 a 27%.

Os sete piores casos do MAUI apresentaram o índice de consistência que variou de 9 a 21%, o índice de precisão entre 20 e 50%, índice de revocação que variou de 13 a 43%, resultando assim, na medida F entre 16 a 35%. Os softwares apresentaram em comum o artigo 23.

4.2.4 Análise conjunta da indexação automática dos sistemas

A seguir, analisa-se a indexação automática dos softwares em cada artigo que compõe o corpus da pesquisa.

No artigo 1 foram propostos 6 termos pelo SISA e 7 termos pelo MAUI. A folksonomia, considerada nesta pesquisa como a indexação de referência, possui 11 termos. Dos 6 termos atribuídos pelo SISA, 2 são comuns com a folksonomia. Dos 7 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia.

Constatou-se que, dos 11 termos da folksonomia, 10 eram termos compostos, sendo assim, foi observado que o SISA atribuiu apenas 2 termos compostos e 4 termos simples, sendo que os 2 termos compostos são comuns a folksonomia. Em contrapartida, o MAUI atribuiu 7 termos compostos, sendo 5 deles comuns com a folksonomia.

Os softwares atribuíram 2 termos em comum: *Transferência da Informação e Gestão do Conhecimento*.

Os erros encontrados na indexação automática do SISA estão relacionados à atribuição de termos de forma superficial e genérica, fazendo com que os termos propostos fossem pouco representativos do conteúdo do artigo. Os erros encontrados na indexação automática do MAUI também estão voltados para a abrangência e generalidade dos termos propostos, o que prejudica a recuperação da informação.

No artigo 2 foram propostos 4 termos pelo SISA e 8 termos pelo MAUI. A folksonomia possui 10 termos. Dos 4 termos atribuídos pelo SISA, 4 são comuns com a folksonomia. Dos 8 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia.

Foi observado que o SISA acertou os 4 termos com relação à folksonomia, portanto, não houveram erros. O MAUI acertou 5, porém observou-se que os erros cometidos pelo software foi a inclusão de termos que não representam o conteúdo do artigo: *Cientistas*, *Comunicação* e *Resumos* que são termos frequentes no artigo em questão, porém não representam o conteúdo do artigo.

Os softwares atribuíram 3 termos comuns: *Comunicação Científica*, *Ciência da Informação* e *Notícias*.

No artigo 3 foram propostos 4 termos pelo SISA e 8 termos pelo MAUI. A folksonomia possui 10 termos. Dos 4 termos atribuídos pelo SISA, 2 são comuns a folksonomia. Dos 8 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia.

Constatou-se que o SISA acertou 2 termos com relação à folksonomia, e que os erros na indexação estão voltados para termos genéricos e simples, afetando assim a representação do conteúdo do artigo. O MAUI acertou 5, porém observou-se que os erros cometidos pelo software foi a inclusão do termo *Resumos* que não representa o conteúdo do artigo, bem como a atribuição do termo *Bibliotecas*, que é considerado pelo TBCI como um termo não preferido e a atribuição do termo *Direito* que é muito abrangente para representar o conteúdo do artigo.

Os softwares atribuíram 3 termos comuns: *Direito à Informação*, *Recuperação da Informação* e *Direito*.

No artigo 4 foram propostos 5 termos pelo SISA e 10 termos pelo MAUI. A folksonomia possui 10 termos. Dos 5 termos atribuídos pelo SISA, apenas 1 é comum com a folksonomia. Dos 7 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia.

Constatou-se que, os 10 termos da folksonomia são compostos, o SISA atribuiu apenas 2 termos compostos e 3 termos simples, sendo que dos 2 termos compostos, apenas 1 é comum com a folksonomia. Em contrapartida, o MAUI atribuiu 10 termos, sendo 2 deles simples e 8 compostos, onde 5 são comuns com a folksonomia.

Os erros do SISA estão pautados na atribuição de termos simples que aparecem com frequência. Os erros do MAUI envolvem a atribuição de palavras simples, como por exemplo o termo *Arquivos*, este seria um bom termo caso tivesse sido acompanhado da palavra aberto, ou seja, *Arquivos Abertos*, no entanto, arquivos abertos não é termo do TBCI. Assim como o MAUI, os erros do SISA, na indexação deste artigo, estão pautados na atribuição de termos simples que aparecem com frequência.

Os softwares não atribuíram termo em comum.

No artigo 5 foram propostos 3 termos pelo SISA e 10 termos pelo MAUI. A folksonomia possui 10 termos. Dos 3 termos atribuídos pelo SISA, apenas 1 é comum com a folksonomia. Dos 10 termos atribuídos pelo MAUI, 4 são comuns com a folksonomia. A folksonomia apresenta 10 termos, sendo 2 deles simples.

Constatou-se que o SISA atribuiu 3 termos simples, sendo que apenas 1 é comum a folksonomia. Em contrapartida, o MAUI atribuiu 10 termos, sendo 2 deles simples e 8 compostos, onde 4 são comuns com a folksonomia.

Os erros do SISA estão pautados na atribuição de termos simples e genéricos. Os erros do MAUI incidiram na atribuição de termos irrelevantes para a representação do conteúdo do artigo analisado.

Os softwares atribuíram 1 termo comum: *Pesquisa*. Porém tal termo não conta na lista de termos da folksonomia, logo não seria um bom descritor.

Podemos visualizar todas as informações dos artigos 1 ao 5 no quadro 12:

Quadro 12: Análise comparativa dos termos dos artigos 1 ao 5

Corpus de referência	Quantidade de termos atribuídos		Termos comuns com a folksonomia		Termos comuns entre SISA e MAUI	Erros com relação a Folksonomia de cada Artigo		Acertos com relação a Folksonomia de cada Artigo	
	SISA	MAUI	SISA	MAUI		SISA	MAUI	SISA	MAUI
Artigo 1	6	7	2	5	Transferência da Informação; Gestão do Conhecimento.	Avaliação; Acesso; Descarte; Gestão.	Ciência da Informação; Tecnologias da informação e comunicação.	Transferência da Informação; Gestão do Conhecimento.	Gestão do Conhecimento; Conhecimento nas organizações; Transferência da Informação; Conhecimento Tácito; Processos de Gestão.
Artigo 2	4	8	4	5	Comunicação científica; Ciência da Informação; Notícias.		Cientistas, Comunicação; Resumos	Comunicação Científica; Ciência da Informação; Educação; Notícias.	Divulgação Científica; Ciência da Informação; Comunicação Científica; Notícias; Informação Científica e Tecnológica.
Artigo 3	4	8	2	5	Direito à Informação; Recuperação da Informação; Direito.	Avaliação; Direito.	Direito; Bibliotecas; Resumos.	Direito à Informação; Recuperação da Informação.	Direito à Informação; Sociedade da Informação; Recuperação da Informação; Acesso à Informação; Ciência da Informação.
Artigo 4	5	10	1	5	Acesso Livre	Acesso; Conceitos; Filosofia; Publicações	Direito; Arquivos.	Acesso Livre	Direito Autoral; Acesso Livre; Tecnologias da informação e comunicação; Publicações eletrônicas; Periódicos Científicos.
Artigo 5	3	10	1	4	Pesquisa	Acesso; Pesquisa.	Acesso ao Documento; Fitas Sonoras; Fluxo da Informação; Análise Comparativa Pesquisa; Pesquisadores.	Estudantes	Informação Científica e Tecnológica; Sociedade da Informação; Comunidades Científicas; Educação.

Fonte: Elaborado pelo autor.

No artigo 6 foram propostos 5 termos pelo SISA e 10 termos pelo MAUI. A folksonomia possui 10 termos. Dos 5 termos atribuídos pelo SISA, 3 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia.

Com relação aos erros do SISA, podemos constatar que os termos são genéricos e, portanto, não representam bem o conteúdo do artigo. Os erros do MAUI mostram que os termos propostos são muito abrangentes e não representa bem o conteúdo do artigo.

Os softwares apresentaram 4 termos em comum.

No artigo 7 foram propostos 6 termos pelo SISA e 10 termos pelo MAUI. A folksonomia possui 11 termos. Dos 6 termos atribuídos pelo SISA, 4 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia.

Com relação aos erros, o SISA aponta dois termos: Dados e Gestão. Podemos constatar que o termo Dados não está na folksonomia, porém este termo representa bem a conteúdo do artigo e foi proposto pelos dois softwares. Com relação aos erros do MAUI, o software não considerou o singular de Tomada de Decisão, que representa bem o conteúdo do artigo, além de propor termos irrelevantes para a indexação.

Os softwares apresentaram 4 termos em comum.

No artigo 8 foram propostos 3 termos pelo SISA e 10 termos pelo MAUI. A folksonomia possui 14 termos. Dos termos atribuídos pelo SISA, 2 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 5 são comuns com folksonomia.

Com relação ao erro do SISA, podemos constatar que o termo proposto Pesquisa é muito abrangente para representar o conteúdo do artigo. Com relação aos erros do MAUI, o software também apontou o termo Pesquisa, porém como já explicado, o termo é por demais abrangente. Os demais erros incidem em termos irrelevantes.

Os softwares apresentaram 3 termos em comum.

No artigo 9 foram propostos 6 termos pelo SISA e 10 termos pelo MAUI. A folksonomia possui 12 termos. Dos termos atribuídos pelo SISA, 3 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 7 são comuns com folksonomia.

Com relação aos erros do SISA, podemos constatar que os erros estão voltados para abrangência de termos gerais e não específicos. Com relação ao termo Disseminação Seletiva da Informação, foi observado que ambos os softwares erraram ao propor este termo, pois não está presente na folksonomia, bem como não é um termo preferido pelo TBCI.

Com relação aos erros do MAUI, o software apresentou termos abrangentes para representar o conteúdo do artigo.

Os softwares apresentaram 4 termos em comum, lembrando que o termo Disseminação Seletiva da Informação não está presente na folksonomia, no entanto, os softwares entenderam que o termo representa bem o conteúdo.

No artigo 10 foram propostos 3 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 12 termos. Dos termos atribuídos pelo SISA, 2 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 5 são comuns com folksonomia.

Com relação aos erros do SISA e MAUI, podemos constatar que os erros estão voltados para abrangência de termos gerais e não específicos.

Os softwares apresentaram apenas 1 termo em comum: Economia.

Podemos visualizar todas as informações dos artigos 6 ao 10 no quadro 13:

Quadro 13: Análise comparativa dos termos dos artigos 6 ao 10.

Corpus de referência	Quantidade de termos atribuídos		Termos comuns com a folksonomia		Termos comuns entre SISA e MAUI	Erros com relação a Folksonomia de cada Artigo		Acertos com relação a Folksonomia de cada Artigo	
	SISA	MAUI	SISA	MAUI		SISA	MAUI	SISA	MAUI
Artigo 6	5	10	3	5	Recuperação da Informação; Ciência da Informação; Ciência da Computação; Universidades.	Navegação; Universidades.	Áreas do Conhecimento, Bases de Dados; Centros de Análise de Informação; Ciências Humanas; Universidades.	Ciência da Informação; Ciência da Computação; Recuperação da Informação.	Recuperação da Informação; Ciência da Informação; Ciência da Computação; Tesouros e Sistemas de informação.
Artigo 7	6	4	10	5	Inteligência Competitiva; Gestão do Conhecimento; Gestão da Informação; Dados.	Dados; Gestão.	Sociedade da Informação; Sistemas de Informação; Tecnologias da Informação e Comunicação; Tomada de Decisões; Dados.	Inteligência Competitiva; Competitividade; Gestão do Conhecimento; Gestão da Informação.	Inteligência Competitiva; Gestão do Conhecimento; Recursos de Informação; Ciência da Informação; Gestão da Informação.
Artigo 8	3	10	2	5	Áreas do Conhecimento; Ciência da Informação; Pesquisa.	Pesquisa	Estados da Arte; Processamento de Informações; Comunicação Científica; Artigos de	Áreas do Conhecimento; Ciência da Informação.	Ciência da Informação; Conceitos de Informação; Áreas do Conhecimento; Tipos de Documento;

							Periódico; Pesquisa.		Indústria da Informação.
Artigo 9	6	10	3	7	Ciência da Informação; Fluxo da Informação; Inteligência Competitiva; Disseminação Seletiva da Informação.	Disseminação Seletiva da Informação; Gestão e Relevância.	Sociedade da Informação; Disseminação Seletiva da Informação; Usuários.	Ciência da Informação; Fluxo da informação; Inteligência Competitiva.	Tomada de Decisões, Fluxo da Informação; Comunicação Informal; Inteligência Competitiva, Universidades; Ciência da Informação; Gestão da Informação.
Artigo 10	3	10	2	5	Economia	Pertinência	Processamento de Informações; Pesquisa e Desenvolvimento; Imagens; Computadores Pessoais; Comunicação.	Economia; Levantamentos.	Sociedade da Informação; Sistemas de Informação; Tecnologias da Informação e Comunicação; Probabilidade e Estatística; Economia

Fonte: Elaborado pelo autor.

No artigo 11 foram propostos 5 termos pelo SISA e 6 termos pelo MAUI. A folksonomia apresenta 10 termos. Dos termos atribuídos pelo SISA, 3 são comuns a folksonomia. Dos 3 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia.

Com relação aos erros do SISA e MAUI, podemos constatar que os erros estão voltados para abrangência de termos gerais e não específicos. Foi constatado um erro de inclusão no MAUI referente ao termo Pesquisadores. O software atribuiu este termo na indexação, porém o termo não é relevante.

Os softwares apresentaram apenas 4 termos em comum.

No artigo 12 foram propostos 6 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 10 termos. Dos termos atribuídos pelo SISA, 5 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia.

Com relação ao único erro do SISA, podemos constatar que está voltado para a abrangência do termo genérico Acesso, que não representa bem o conteúdo do artigo. Foi constatado que o MAUI atribuiu para este artigo, termos de pouca relevância para a representar o seu conteúdo.

Os softwares apresentaram apenas 3 termos em comum.

No artigo 13 foram propostos 4 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 13 termos. Dos 4 termos atribuídos pelo SISA, 3 são comuns a folksonomia. Dos 7 termos atribuídos pelo MAUI, 7 são comuns com a folksonomia.

Com relação aos acertos do SISA, constatou-se que o software acertou 3 termos para o artigo 13, o único erro foi com relação ao termo Papel, que por ser muito genérico, acaba prejudicando a representação de seu conteúdo. O MAUI acertou 7 termos, porém observou-se que os erros cometidos pelo software foi a inclusão de termos que não representam o conteúdo do artigo.

Os softwares atribuíram 2 termos comuns: Formação Profissional e Sistemas de Informação.

No artigo 14 foram propostos 8 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 12 termos. Dos 8 termos atribuídos pelo SISA, 5 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 7 são comuns com a folksonomia.

Com relação aos acertos do SISA, constatou-se que o software acertou 5 termos para o artigo em questão, e com relação aos erros, constatou-se que a atribuição de termos muito genéricos comprometeu a indexação.

O MAUI acertou 7 termos, porém observou-se que os erros cometidos pelo software estão relacionados a atribuição de termos que não representam o conteúdo do artigo analisado, devido a sua generalidade.

Observou-se que ambos os softwares atribuíram o termo Gestores de maneira errada. Este erro evidencia que a composição dos termos interferiu na atribuição de termos, pois foi constatado que, se o termo fosse composto, como por exemplo Gestores da Informação, seria um termo mais específico e iria representar melhor o conteúdo deste artigo.

Os softwares atribuíram 5 termos comuns.

No artigo 15 foram propostos 3 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 11 termos. Dos 3 termos atribuídos pelo SISA, 2 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 6 são comuns com a folksonomia.

Com relação aos acertos do SISA, constatou-se que o software acertou 2 termos para o artigo em questão, e com relação ao erro, constatou-se que a atribuição do termo simples Educação comprometeu a indexação, pois foi constatado que se fosse o termo composto Educação Superior seria melhor representado.

O MAUI acertou 7 termos: Profissionais da Informação, Formação Profissional, Educação Superior, Ciência da Informação, Pesquisa, Paradigmas e Sociedade da Informação. Observou-se que houve a inclusão do termo Pesquisadores por erro do software por conta do radical dos termos Pesquisa e Pesquisadores.

Artigo 14	8	10	5	7	Profissionais de Informação; Gestão da Informação; Gestão do Conhecimento; Gestores; Inclusão Digital.	Acesso; Gestores ; gestão.	Necessidades de Informação; Gestores ; Ciência da Informação.	Profissionais de Informação ; Gestão da Informação ; Gestão do Conhecimento ; Inclusão Digital ; Formação Profissional.	Competência em Informação; Profissionais de Informação ; Gestão da Informação ; Engenharia do Conhecimento; Gestão do Conhecimento ; Sociedade da Informação; Inclusão Digital .
Artigo 15	3	10	2	7	Formação Profissional; Educação; Pesquisa.	Educação	Educação ; Pesquisadores; Universidades.	Formação Profissional ; Pesquisa .	Profissionais de Informação; Formação Profissional ; Educação Superior; Ciência da Informação; Pesquisa ; Paradigmas; Sociedade da Informação.

Fonte: Elaborado pelo autor.

No artigo 16 foram propostos 6 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 11 termos. Dos 6 termos atribuídos pelo SISA, 5 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 7 são comuns com a folksonomia.

O erro do SISA foi analisado e constatou-se que a atribuição do termo Educação se deve pela frequência com que a palavra se apresenta no texto, porém foi observado que a palavra composta Educação Superior seria mais adequada para representar o conteúdo do artigo.

Assim como o SISA, o MAUI também atribuiu o termo Educação, e os outros termos não foram relevantes para a representação do conteúdo do artigo, além de serem de baixa frequência, o que torna os erros ainda mais evidentes, totalizando 3 erros. Os softwares atribuíram 6 termos comuns.

No artigo 17 foram propostos 6 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 11 termos. Dos 6 termos atribuídos pelo SISA, 4 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia.

Os erros do SISA foram analisados e constatou-se que a atribuição de termos simples como Autoria e Complexidade foram considerados relevantes, tendo em vista a temática do artigo, além de serem termos genéricos.

O MAUI atribuiu 3 termos compostos errados, e apesar de serem mais específicos dos que os termos errados do SISA, não refletem o conteúdo do artigo.

Os softwares atribuíram 4 termos comuns.

No artigo 18 foram propostos 2 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 15 termos. Dos 2 termos atribuídos pelo SISA, 2 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, apenas 2 são comuns com a folksonomia.

Não houve erro na indexação do SISA, ou seja, todos os termos propostos pelo software estão presentes na folksonomia. Em contrapartida o MAUI apresentou 8 erros em 10 termos propostos. Os erros estão voltados para a ausência de frequência dos termos atribuídos no artigo, além de contemplar 2 dos 8 termos de forma genérica: Mudanças e Pesquisadores.

Os softwares não atribuíram termos comuns.

No artigo 19 foram propostos 7 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 13 termos. Dos 7 termos atribuídos pelo SISA, 3 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, apenas 6 são comuns com a folksonomia.

Os softwares atribuíram um único termo em comum entre si: Ambiente Organizacional. Foi constatado que o mesmo termo foi o único acerto em comum dos softwares.

Com relação aos erros, o SISA demonstrou novamente a questão da generalidade com relação à atribuição dos termos.

Os erros do MAUI evidenciam que o fato de atribuir termos compostos e apresentar uma maior especificidade, os termos não correspondem a termos relevantes que representem os artigos.

No artigo 20 foram atribuídos 7 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 11 termos. Dos 7 termos atribuídos pelo SISA, 5 termos estão presentes na folksonomia. Dos 10 termos do MAUI, 4 termos são comuns com a folksonomia.

Os erros do SISA evidenciam que termos simples e genéricos não resultaram numa boa indexação. Com relação ao MAUI, seus erros apontam que os termos compostos propostos pelo software não são bons a fim de representar o conteúdo do artigo.

Os softwares apresentaram 4 termos em comum.

Podemos visualizar os termos dos artigos 16 ao 20 no quadro 15:

Quadro 15: Análise comparativa dos termos dos artigos 16 ao 20.

Corpus de referência	Quantidade de termos atribuídos		Termos comuns com a folksonomia		Termos comuns entre SISA e MAUI	Erros com relação a Folksonomia de cada Artigo		Acertos com relação a Folksonomia de cada Artigo	
	SISA	MAUI	SISA	MAUI		SISA	MAUI	SISA	MAUI
Artigo 16	6	10	5	7	Ciência da Informação; Formação Profissional; Sociedade da Informação; Educação Superior; Biblioteconomia; Educação.	Educação	Arquitetura da Informação; Arquitetos de informação; Educação.	Ciência da Informação, Biblioteconomia; Educação Superior; Formação Profissional; Sociedade da Informação.	Ciência da Informação; Profissionais de Informação; Formação Profissional; Sociedade da Informação; Educação Superior; Biblioteconomia; Ensino de Biblioteconomia.
Artigo 17	6	10	4	7	Propriedade Intelectual; Controle Bibliográfico; Tecnologias da Informação e Comunicação.	Autoria; Complexidade	Correio Eletrônico; Acesso Universal; Periódicos Científicos.	Controle Bibliográfico; Internet; Propriedade Intelectual; Tecnologias da Informação e Comunicação.	Publicações Eletrônicas; Propriedade Intelectual; Controle Bibliográfico; Tecnologia da Informação e Comunicação; Inovação; Sociedade da Informação; Comunicação Científica.
Artigo 18	2	10	2	2			Dissertações e Teses; Bibliografias; Mudança; Comunidades Científicas; Sistemas de Organização do Conhecimento; Pesquisadores; Resumos; Cenários.	Categorias; Sociologia.	Pesquisa; Construtivismo
Artigo 19	7	10	3	6	Ambiente Organizacional	Avaliação; Bibliotecas.	Bibliotecas Centrais; Centros de Análise e Informação; Frequência de uso; Buscas de Informação.	Ambiente Organizacional; Gestão do Conhecimento; Relevância.	Ambiente Organizacional; Inteligência Competitiva; Monitoramento Ambiental; Gestão da Informação; Conhecimento nas Organizações; Informação para negócios.

Artigo 20	7	10	5	4	Ciência da Informação; Pesquisa; Publicações; Dissertações e Teses.	Avaliação; Pesquisa	Áreas do Conhecimento; Estudos de Caso; Pesquisa , Pesquisadores; Conselhos Editoriais; Anais de Congressos.	Ciência da Informação; Dissertações e Teses ; Programas de Pós-Graduação; Programas de Pós-Graduação em Ciência da Informação; Publicações .	Ciência da Informação ; Projetos de Pesquisa; Publicações; Dissertações e Teses .

Fonte: Elaborado pelo autor.

No artigo 21 foi atribuído 1 termo pelo SISA e 6 termos pelo MAUI. A folksonomia apresenta 9 termos. Sendo assim, o SISA não apresentou erro algum. Dos 6 termos atribuídos pelo MAUI, 3 termos estão presentes na folksonomia.

Com relação aos erros do MAUI, os termos compostos propostos pelo software não são bons a fim de representar o conteúdo do artigo.

Os softwares apresentaram 1 termo em comum: Áreas do Conhecimento.

No artigo 22 foram propostos 2 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 7 termos. Os 2 termos atribuídos pelo SISA foram considerados corretos, pois são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 4 são comuns com a folksonomia.

Como já mencionado, o SISA não apresentou erro. O MAUI apresentou 5 erros, e, constatou-se que a atribuição dos termos não foi relevante para representar o conteúdo do artigo. Porém, observou-se que o MAUI não distinguiu o plural e singular das palavras propondo assim, os termos: Hipertexto e Hipertextos.

Os softwares atribuíram 1 termo comum: Ciência da Informação.

No artigo 23 foi proposto 1 termo pelo SISA e 6 termos pelo MAUI. A folksonomia apresenta 11 termos. Dos 6 termos atribuídos pelo MAUI, 3 são comuns com a folksonomia.

Os softwares atribuíram 1 termo comum: História.

Os erros encontrados na indexação automática do MAUI estão relacionados à atribuição de termos irrelevantes que não retratam o conteúdo do artigo.

No artigo 24 foram propostos 4 termos pelo SISA e 7 termos pelo MAUI. A folksonomia apresenta 10 termos. Dos 4 termos atribuídos pelo SISA, 2 são comuns a folksonomia. Dos 7 termos atribuídos pelo MAUI, 4 são comuns com a folksonomia.

Os erros encontrados na indexação automática do SISA estão relacionados à atribuição dos termos Ecologia e Validade de forma genérica, fazendo com que os termos propostos não representem o conteúdo do artigo.

Os erros encontrados na indexação automática do MAUI também estão voltados para a generalidade, além dos termos demonstrarem baixa frequência no texto do artigo.

Os softwares atribuíram 2 termos comuns entre si: Pesquisa e Pesquisadores.

No artigo 25 foram propostos 5 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 13 termos. Dos 5 termos atribuídos pelo SISA, 4 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 6 são comuns com a folksonomia.

O erro único erro do SISA foi analisado e constatou-se que a atribuição do termo Dados se deve pela alta frequência com que a palavra se apresenta no texto, porém o termo é muito abrangente para representar o conteúdo do artigo analisado.

Assim como o SISA, o MAUI também atribuiu o termo Dados, e os outros termos não foram relevantes para a representação do conteúdo do artigo.

Os softwares atribuíram 4 termos comuns.

Podemos visualizar os termos dos artigos 21 ao 25 no quadro 16:

Quadro 16: Análise comparativa dos termos dos artigos 21 ao 25.

Corpus de referência	Quantidade de termos atribuídos		Termos comuns com a folksonomia		Termos comuns entre SISA e MAUI	Erros com relação a Folksonomia de cada Artigo		Acertos com relação a Folksonomia de cada Artigo	
	SISA	MAUI	SISA	MAUI		SISA	MAUI	SISA	MAUI
Artigo 21	1	6	1	3	Áreas do Conhecimento		Bibliotecas Nacionais; Ciências Humanas; Ciência da Informação.	Áreas do Conhecimento	Áreas do Conhecimento; Inteligência Artificial; Sociologia.
									Ciência da Informação; Paradigmas, Processamento de

Artigo 22	2	10	2	4	Ciência da Informação		World Wide Web; Conceitos de Informação; áreas do Conhecimento; Hipertexto.	Ciência da Informação; Complexidade.	Informações; Tecnologias da Informação e Comunicação; Organização do Conhecimento; Hipertextos.
Artigo 23	1	6	1	3	História		Políticas Públicas; Medicina; Suportes de Informação	História	História; Sociedade da Informação; Conceitos de Informação.
Artigo 24	4	7	2	4	Pesquisa; Pesquisadores	Ecologia; Validade	Pesquisa e Desenvolvimento; Organização do Conhecimento; Uso da Informação.	Pesquisa; Pesquisadores.	Áreas do Conhecimento; Políticas Públicas; Pesquisa; Pesquisadores.
Artigo 25	5	10	4	6	Ciência da Informação; Pesquisa, Pesquisadores; Dados.	Dados	Ciências Sociais Aplicadas; áreas do Conhecimento; Comunidades Científicas; Dados.	Ciência da Informação; Pesquisa Pesquisadores.	Ciência da Informação; Agências de Fomento; Projetos de Pesquisa; Bolsas de Pesquisas; Pesquisa; Pesquisadores.

Fonte: Elaborado pelo autor.

No artigo 26 foram propostos 5 termos pelo SISA e 9 termos pelo MAUI. A folksonomia apresenta 9 termos. Dos 5 termos atribuídos pelo SISA, 2 são comuns a folksonomia. Dos 9 termos atribuídos pelo MAUI, 3 são comuns com a folksonomia.

Os erros do SISA foram analisados e constatou-se que a atribuição do termo Educação se deve pela frequência com que a palavra se apresenta no texto, porém foi observado que a palavra composta Educação Superior seria mais adequada para representar o conteúdo do artigo.

Os erros do SISA e do MAUI constituíram em atribuir termos que não são relevantes para a representação do conteúdo do artigo.

Os softwares atribuíram 4 termos comuns.

No artigo 27 foram propostos 7 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 9 termos. Dos 7 termos atribuídos pelo SISA, 2 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia.

Os erros do SISA foram analisados e constatou-se que a atribuição dos termos Dados, Gestão, Indicadores, Monitoramento e Relevância são abrangentes para representar o conteúdo do artigo.

Com relação aos erros do MAUI, destacou-se a atribuição dos termos Pesquisa e Indicadores como termos muito simples para representar o conteúdo do artigo, mesmo percebendo a alta frequência dos termos no artigo.

Os softwares atribuíram 2 termos comuns.

No artigo 28 foram propostos 8 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 9 termos. Dos 8 termos atribuídos pelo SISA, 5 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 4 são comuns com a folksonomia.

Os dois softwares acertaram os mesmos termos: Sistemas de Informação, Bibliotecas Digitais, Gestão da Informação, Governo Eletrônico e Inovação.

Os erros do SISA se devem ao fato da atribuição de termos simples para representar o conteúdo do artigo, além de um erro semântico do termo Papel.

Os erros do MAUI estão voltados para a pouca relevância na atribuição dos termos.

Os softwares atribuíram 5 termos comuns.

No artigo 29 foram propostos 5 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 7 termos. Dos 5 termos atribuídos pelo SISA, 3 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia.

Os erros do SISA foram analisados e constatou-se que a atribuição do termo Acesso e Gestão são simples e abrangentes para representar o conteúdo do artigo.

Os erros do SISA se pautaram na atribuição de termos irrelevantes para a representação do conteúdo do artigo, e cabe destacar que, assim como o SISA, o MAUI também atribuiu o termo Acesso.

Os softwares atribuíram 4 termos comuns.

No artigo 30 foram propostos 5 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 14 termos. Dos 5 termos atribuídos pelo SISA, 2 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 4 são comuns com a folksonomia.

Com relação aos erros do SISA, constatou-se que a atribuição de termos simples e abrangentes como: Avaliação, Acesso e Periódicos apresentam alta frequência no artigo, porém não retratam o conteúdo do artigo.

Os erros do MAUI apontam que a atribuição de termos se deu por meio de termos simples, e compostos. Os termos simples evidenciaram uma alta frequência no documento analisado, porém não possuem força para representar o conteúdo. Em contrapartida, os termos compostos propostos pelo software não evidenciaram relevância.

Os softwares atribuíram 4 termos comuns.

Podemos visualizar todas as informações dos artigos 26 ao 30 no quadro 17:

Quadro 17: Análise comparativa dos termos dos artigos 26 ao 30

Corpus de referência	Quantidade de termos atribuídos		Termos comuns com a folksonomia		Termos comuns entre SISA e MAUI	Erros com relação a Folksonomia de cada Artigo		Acertos com relação a Folksonomia de cada Artigo	
	SISA	MAUI	SISA	MAUI		SISA	MAUI	SISA	MAUI
Artigo 26	5	9	2	3	Ciência da Informação; Monitoramento; Semântica; Subsídios.	Ciência da Informação; Gestão; Subsídios.	Gestão da Informação, Transferência da Informação, Fluxo da Informação, Instrumentalismo e subsídios.	Monitoramento; Semântica.	Monitoramento, Semântica; Recuperação da Informação.
Artigo 27	7	10	2	5	Base de Dados; Pesquisadores.	Dados; Gestão; Indicadores; Monitoramento; Relevância.	Índice H; índices; Áreas do Conhecimento; Pesquisa; Indicadores.	Bases de Dados; Pesquisadores	Bolsas de Pesquisa; Bases de Dados; Agências de Fomento; Pesquisadores; Indicadores de C&T.
Artigo 28	8	10	5	4	Sistemas de Informação; Gestão da Informação; Bibliotecas Digitais; Governo Eletrônico; Bibliotecas.	Gestão, Bibliotecas; Papel.	Necessidades de Informação; Informação Governamental; Tomada de Decisões; Revisões de Literatura; Bibliotecas; Dissertações e Teses.	Sistemas de Informação; Bibliotecas Digitais; Gestão da Informação; Governo Eletrônico; Inovação.	Sistemas de Informação; Gestão da Informação; Bibliotecas Digitais; Governo Eletrônico.
					Governo Eletrônico;		Sociedade da Informação;	Governo Eletrônico;	Governo Eletrônico;

Artigo 29	5	10	3	5	Informação Governamental; Políticas de Informação; Acesso.	Acesso; Gestão.	Acesso; Políticas Públicas; Gestão do Conhecimento; Ciência da Informação.	Informação Governamental; Políticas de Informação.	Informação Governamental; Gestão da Informação; Políticas de Informação; Serviços de Informação.
Artigo 30	5	10	2	4	Periódicos Eletrônicos; World Wide Web; Acesso; Periódicos; Avaliação.	Avaliação, Acesso; Periódicos.	Acesso; Periódicos; Arquivos; Associações de Elaboração de Normas; Ciência da Informação; Avaliação.	Periódicos Eletrônicos; World Wide Web.	Periódicos Eletrônicos; World Wide Web; Artigos de Periódicos; Usuários.

Fonte: Elaborado pelo autor.

No artigo 31 foram propostos 5 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 9 termos. Dos 5 termos atribuídos pelo SISA, 3 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 3 são comuns com a folksonomia.

Os 2 erros do SISA foram analisados e constatou-se que a atribuição do termo Cenários não é um termo preferido do TBCI, e o termo Economia por ser muito genérico não reflete o conteúdo do artigo.

Assim como o SISA, o MAUI também atribuiu o termo Cenários, e os outros termos propostos não foram relevantes para a representação do conteúdo do artigo. Cabe ressaltar que os termos propostos pelo MAUI são compostos, porém possuem baixa frequência no artigo.

Os softwares atribuíram 2 termos comuns: Sociedade da Informação e Cenários.

No artigo 32 foram propostos 8 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 8 termos. Dos 8 termos atribuídos pelo SISA, 3 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia.

Os erros do SISA foram analisados e constatou-se que a atribuição dos termos simples e genéricos não representam o conteúdo do artigo para fins de indexação.

Os termos propostos pelo MAUI são compostos, porém possuem baixa frequência e não foram relevantes para a representação do conteúdo do artigo. Os softwares atribuíram 2 termos comuns: Sociedade da Informação e Cenários.

No artigo 33 foram propostos 11 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 13 termos. Dos 11 termos atribuídos pelo SISA, 5 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 6 são comuns com a folksonomia.

Os sistemas apresentaram erros comuns: Desenvolvimento de Coleções e Referências Bibliográficas. Nos erros do SISA destaca-se a atribuição do termo Periódicos, o SISA considerou a posição do termo em relação ao texto, ou seja, o termo Periódicos faz parte do título, resumo e palavras-chave do artigo, além de possuir uma alta frequência no texto, mas é um termo genérico que não representa bem o conteúdo do artigo. Os demais termos foram analisados e percebeu-se que a maioria são termos simples e genéricos para fins de representação do conteúdo do artigo. Os termos compostos não foram considerados relevantes.

Com relação aos erros do MAUI, os termos propostos pelo MAUI são compostos, porém possuem baixa frequência no artigo e não refletem bem o conteúdo do artigo em questão.

Os softwares atribuíram 6 termos comuns: Bibliotecas Digitais, Bibliotecas Virtuais, Profissionais de Informação, Bibliotecas e Referências Bibliográficas.

No artigo 34 foram propostos 4 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 13 termos. Dos 4 termos atribuídos pelo SISA, 2 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 6 são comuns com a folksonomia.

Os erros do SISA foram analisados e constatou-se que a atribuição dos termos simples e genéricos não representam o conteúdo do artigo para fins de indexação.

Os termos propostos pelo MAUI são compostos, porém possuem baixa frequência e não foram relevantes para a representação do conteúdo do artigo.

Os softwares atribuíram 3 termos comuns: Ciências da Saúde, Transferência da Informação e Nutrição.

No artigo 35 foram propostos 9 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 7 termos. Dos 9 termos atribuídos pelo SISA, 5 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia.

Os sistemas apresentaram apenas 1 erro comum: Dados. O SISA apresentou 4 erros e foi identificado que são termos simples e genéricos para fins de representação do conteúdo do artigo.

Com relação aos erros do MAUI, notou-se que, apesar de os termos serem compostos, não apresentam relevância para a representação do conteúdo do artigo em questão.

Os softwares atribuíram 5 termos comuns: Linguagens de Marcação, HTML, Ciência da Informação, XML e Dados.

Podemos visualizar todas as informações dos artigos 31 ao 35 no quadro 18:

Quadro 18: Análise comparativa dos termos dos artigos 31 ao 35

Corpus de referência	Quantidade de termos atribuídos		Termos comuns com a folksonomia		Termos comuns entre SISA e MAUI	Erros com relação a Folksonomia de cada Artigo		Acertos com relação a Folksonomia de cada Artigo	
	SISA	MAUI	SISA	MAUI		SISA	MAUI	SISA	MAUI
Artigo 31	5	10	3	3	Sociedade da Informação	Cenários	Comunicação Informal; Informação Científica e Tecnológica; Políticas Públicas; Ciência da Informação; Serviços de Informação; Cenários ; Indústria da Informação.	Economia; Internet; Sociedade da Informação.	Regimes de Informação; Políticas de Informação; Sociedade da Informação.
Artigo 32	8	10	3	5	Bases de Dados; Estratégias de Busca; Recuperação da Informação.	Dados; CD-ROM; Complexidade; Planejamento; Termos.	Planejamento Estratégico; Buscas de Informação; Base de Dados em Linha; Estudo de Uso; Sistemas de Recuperação da Informação.	Bases de Dados; Estratégias de Busca; Recuperação da Informação.	Bases de Dados; Estratégias de Busca; Recuperação da Informação; Linguagens Documentárias; Termos de Indexação.
Artigo 33	11	10	5	6	Bibliotecas Digitais; Bibliotecas Virtuais; Profissionais de Informação; Bibliotecas; Referências Bibliográficas.	Periódicos; Autorias; Coleções; Desenvolvimento de coleções; Planejamento; Referências Bibliográfica.	Desenvolvimento de coleções; Tipos de Documento; Ciência da informação; Referências Bibliográficas.	Bibliotecas Digitais; Bibliotecas Virtuais; Bibliotecas; Profissionais de Informação.	Bibliotecas Digitais; Bibliotecas Virtuais; Profissionais de Informação; Artigos de Periódico; Produtividade de autor; Bibliotecas.
Artigo 34	4	10	2	6	Ciências da Saúde Transferência	Acesso; Ciências da Saúde.	Comunicação Informal; Ciências da Saúde; Necessidades de Informação;	Nutrição; Transferência da Informação.	Informação Científica e Tecnológica; Transferência da Informação; Nutrição; Universidades; Pesquisa e

					da Informação; Nutrição.		Ciência da Informação.		Desenvolvimento; Institutos de Pesquisa.
Artigo 35	9	10	5	5	Linguagens de Marcação; HTML; Ciência da Informação; XML; Dados.	Conceitos; Complexidade; Dados e Usuários.	Padrões de Metadados; Dados ; Áreas do Conhecimento; Recuperação da Informação; Ciência da Computação.	Internet; XML; Ciência da Informação; HTML; Linguagens de Marcação.	Linguagens de Marcação; HTML; Ciência da Informação; XML; SGML

Fonte: Elaborado pelo autor.

No artigo 36 foram propostos 6 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 7 termos. Dos 6 termos atribuídos pelo SISA, 2 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 3 são comuns com a folksonomia.

Os sistemas apresentaram apenas 1 erro em comum: Cartas. O SISA apresentou 4 erros e o MAUI 7 erros. Os softwares apresentaram 2 acertos em comum: Lei de Lotka e Bibliometria.

Os erros do SISA foram identificados e analisados como termos simples e genéricos para fins de representação do conteúdo do artigo.

Com relação aos erros do MAUI, notou-se que, não possuem relevância para a representação do conteúdo do artigo em questão.

Os softwares apresentaram 3 termos em comum: Lei de Lotka, Cartas e Bibliometria.

No artigo 37 foram propostos 9 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 12 termos. Dos 9 termos atribuídos pelo SISA, 4 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 8 são comuns com a folksonomia.

Os sistemas não apresentaram erro em comum.

Os softwares apresentaram 4 termos em comum: Informação para Negócios, “Bases de Dados, Indústria da Informação e Empresas. Constatou-se que estes termos comuns são termos de acertos de ambos.

Os erros do SISA foram identificados e analisados como termos simples e genéricos para fins de representação do conteúdo do artigo. Destaca-se aqui, o termo Dados que foi contabilizado 133 vezes no documento. Porém, o termo Dados por si só é muito genérico. O artigo menciona o termo Bases de dados, portanto, Dados não representa bem seu conteúdo.

Com relação aos erros do MAUI, notou-se que, não possuem relevância para a representação do conteúdo do artigo em questão.

No artigo 38 foram propostos 7 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 14 termos. Dos 7 termos atribuídos pelo SISA, 3 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 8 são comuns com a folksonomia.

Os softwares apresentaram apenas 1 erro em comum: Bibliotecas.

Acertaram 2 termos em comum: Bibliotecas Híbridas e Usuários.

Com relação aos erros do SISA, foi constatado que os termos propostos são irrelevantes e destacamos Papel como um erro semântico, já que o termo não é preferido no TBCI.

Com relação aos erros do MAUI, os termos atribuídos pelo software não são relevantes para a indexação do artigo.

Os softwares apresentaram 3 termos em comum.

No artigo 39 foram propostos 10 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 7 termos. Dos 10 termos atribuídos pelo SISA, 4 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia.

Os softwares não apresentaram erros em comum.

Todos os acertos de SISA são comuns aos acertos do MAUI, cabe destacar que os termos propostos foram todos compostos e relevantes para a representação do conteúdo do artigo.

Os softwares apresentaram 4 termos em comum.

No artigo 40 foram propostos 7 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 12 termos. Dos 7 termos atribuídos pelo SISA, 5 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 8 são comuns com a folksonomia.

Os softwares não apresentaram erro comum.

Os softwares apresentaram 4 acertos em comum.

Constatou-se que os erros do SISA estão voltados para termos simples e genéricos. Com relação aos erros do MAUI, foi verificado que apesar dos termos propostos serem compostos, os mesmos não possuem relevância a fim de representar o conteúdo do artigo analisado.

Os softwares apresentaram 4 termos em comum.

Podemos visualizar todas as informações dos artigos 36 ao 40 no quadro 19:

Quadro 19: Análise comparativa dos termos dos artigos 36 ao 40

Corpus de referência	Quantidade de termos atribuídos		Termos comuns com a folksonomia		Termos comuns entre SISA e MAUI	Erros com relação a Folksonomia de cada Artigo		Acertos com relação a Folksonomia de cada Artigo	
	SISA	MAUI	SISA	MAUI		SISA	MAUI	SISA	MAUI
Artigo 36	6	10	2	3	Lei de Lotka; Cartas; Bibliometria.	Biblioteconomia; Cartas ; Dados; Medicina.	Arquivos Privados; Cartas ; Faculdades de Biblioteconomia; Bases de Dados; Ciência da Informação; Arquivos; Análise de Dados.	Bibliometria; Lei de Lotka.	Lei de Lotka ; Produtividade de Autor; Bibliometria.
Artigo 37	9	10	4	8	Informação para Negócios; Bases de Dados; Indústria da Informação; Empresas.	Dados; Cenários; Categorias; Classificação; Notícias.	Competências no Uso de Computadores; Informação Científica e Tecnológica.	Bases de Dados; Informação para negócios; Empresas; Indústria da Informação.	Informação para Negócios; Bases de Dados ; Informação Financeira; Informação Jurídica; Acesso à Informação; Tomada de Decisões; Indústria da Informação; Empresas.
Artigo 38	7	10	3	8	Bibliotecas Híbridas; Bibliotecas; Usuários.	Educação; Bibliotecas ; Papel; Universidades.	Bibliotecas ; Bases de Dados.	Bibliotecas Híbridas ; Internet; Usuários.	Ensino à Distância; Bibliotecas Híbridas ; Tecnologias da Informação e Comunicação; Bibliotecas Universitária; Usuários ; Serviços de Biblioteca, Bibliotecas Digitais; Necessidades de Informação.
Artigo 39	10	10	4	5	Estratégias de Busca; Bases de Dados; Recuperação da Informação; Sistemas de Recuperação da Informação.	Acesso; Complexidade; Conceitos; Dados; Planejamento; Usuário.	Planejamento Estratégico; Buscas de Informação; Bases de Dados em Linha; Linguagens	Recuperação da Informação; Bases de Dados; Estratégias de Busca; Sistemas de Recuperação da Informação.	Estratégias de Busca; Bases de Dados; Recuperação da Informação; Sistemas de Recuperação da Informação;

							Documentárias; Buscas.		Mecanismos de Busca.
Artigo 40	7	10	5	8	Gestão da Informação; Recursos de Informação; Profissionais de Informação; Ciência da Informação.	Gestão; Avaliação.	Acesso à Informação; Serviços de Informação.	Gestão da Informação; Formação Profissional; Ciência da Informação; Profissionais de Informação; Recursos de Informação.	Gestão da Informação Recursos de Informação; Profissionais de Informação; Tomada de Decisões; Ciência da Informação; Necessidades de Informação; Tecnologias da Informação e Comunicação; Mercado de Trabalho.

Fonte: Elaborado pelo autor.

No artigo 41 foram propostos 10 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 11 termos. Dos 10 termos atribuídos pelo SISA, 5 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 4 são comuns com a folksonomia.

Os softwares não apresentaram erro comum.

Os softwares apresentaram 2 acertos em comum.

Constatou-se que os erros do SISA estão voltados para termos simples e genéricos. Com relação aos erros do MAUI, foi verificado que apesar dos termos propostos serem compostos, os mesmos não possuem relevância a fim de representar o conteúdo do artigo analisado.

Os softwares apresentaram 2 termos em comum: Ciência da Informação e Literatura Cinzenta, que correspondem aos acertos de ambos.

No artigo 42 foram propostos 10 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 12 termos. Dos 10 termos atribuídos pelo SISA, 9 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 7 são comuns com a folksonomia.

Os softwares não apresentaram erro comum.

Os softwares apresentaram 7 acertos em comum.

Foi constatado que todos os acertos do MAUI, também são acertos do SISA, sendo que o SISA acertou 2 termos a mais que o MAUI.

Constatou-se que o único erro do SISA encontra-se na atribuição do termo Gestão, que foi considerado genérico e simples pela análise.

Com relação aos erros do MAUI, foi verificado que apesar dos termos propostos serem compostos, os mesmos possuem pouca frequência e não possuem relevância a fim de representar o conteúdo do artigo analisado.

Os softwares apresentaram 7 termos em comum.

No artigo 43 foram propostos 4 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 8 termos. Dos 4 termos atribuídos pelo SISA, 3 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 6 são comuns com a folksonomia.

Os softwares não apresentaram erro comum.

O SISA acertou 3 termos que também são acertos do MAUI: Necessidades de Informação, Empresas e Inovação. O MAUI ainda acertou Informação Científica e Tecnológica, Tomada de Decisões e Conhecimento nas Organizações, totalizando 6 acertos.

Constatou-se que o único erro do SISA encontra-se na atribuição do termo Pesquisa, que foi considerado genérico e simples pela análise, apesar da sua alta frequência no texto do artigo.

Os softwares apresentaram 3 termos em comum, cabe ressaltar que os termos em comum são todos os acertos do SISA.

No artigo 44 foram propostos 4 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 11 termos. Dos 4 termos atribuídos pelo SISA, 2 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia.

Os softwares apresentaram 3 termos em comum: Gestão do Conhecimento, Gestão e Competitividade.

Apresentaram apenas 1 erro em comum: Gestão, pois o termo é muito genérico e simples para representar o conteúdo do artigo analisado.

Os erros do SISA: Gestão e Papel são muito genéricos e simples para representar o conteúdo do artigo, cabe ressaltar que Papel foi atribuído pelo SISA por um erro semântico, pois o TBCI o reconhece como sendo um termo não preferido.

Os erros do MAUI estão pautados na atribuição de termos simples e gerais como Empresa, Gestão e Resumos, e os termos compostos atribuídos pelo MAUI como Base de Dados e Conhecimento Tácito, foram analisados e verificou-se que, o primeiro é irrelevante para representação do artigo, enquanto o segundo não é um termo preferido do TBCI.

No artigo 45 foram propostos 5 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 8 termos. Dos 5 termos atribuídos pelo SISA, 3 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 4 são comuns com a folksonomia.

Os softwares apresentaram 3 termos em comum, que também foram caracterizados como todos os acertos do SISA.

Os erros do SISA: Avaliação e Internet são muito genéricos para representar o conteúdo do artigo em questão, sendo que Internet não é um termo preferido do TBCI.

Os erros do MAUI estão pautados na atribuição de termos compostos que apresentam baixa frequência no texto, sendo assim, não representam o conteúdo do artigo. Cabe destacar que o termo Áreas do Conhecimento não é um termo preferido do TBCI.

Ainda com relação aos erros do MAUI, houve um erro com relação ao termo simples Comunicação, que evidenciou ser um termo abrangente e de baixa frequência no texto do artigo em questão, sendo assim, não representa o conteúdo do artigo.

Podemos visualizar todas as informações dos artigos 41 ao 45 no quadro 20:

Quadro 20: Análise comparativa dos termos dos artigos 41 ao 45

Corpus de referência	Quantidade de termos atribuídos		Termos comuns com a folksonomia		Termos comuns entre SISA e MAUI	Erros com relação a Folksonomia de cada Artigo		Acertos com relação a Folksonomia de cada Artigo	
	SISA	MAUI	SISA	MAUI		SISA	MAUI	SISA	MAUI
Artigo 41	10	10	5	4	Ciência da Informação; Literatura Cinzenta.	Autoria; Dados; Índice; Pesquisa; Periódicos.	Autoria Individual; Institutos de Pesquisa; Sociedade a Informação; Tipos de Documento; Bases de Dados; Coleta de Dados.	Ciência da Informação; Programas de Pós-Graduação; Programas de Pós-Graduação em Ciência da Informação; Literatura Cinzenta; Publicações.	Ciência da Informação; Artigos de Periódico; Literatura Cinzenta; Comunidades Científicas.
Artigo 42	10	10	9	7	Sistemas de Informação; Informação para Negócios; Gestão do Conhecimento; Bibliotecas Virtuais; Inteligência Competitiva; Gestão da Informação; Profissionais de Informação.	Gestão	Centro de Análise de Informação; Bibliotecas de Empresas; Acesso à Informação.	Informação Para Negócios; Bibliotecas; Bibliotecas Virtuais; Empresas; Gestão do Conhecimento; Gestão da Informação; Inteligência Competitiva; Profissionais de Informação; Sistemas de Informação.	Sistemas de Informação; Informação para Negócios; Gestão do Conhecimento; Bibliotecas Virtuais; Inteligência Competitiva; Gestão da Informação; Profissionais de Informação.
Artigo 43	4	10	3	6	Necessidades de Informação; Empresas; Inovação.	Pesquisa	Estados da Arte; Estudos de Usuários; Estudos de Caso; Ciência da Informação.	Necessidades de Informação; Empresas; Inovação.	Necessidades de Informação; Informação Científica e Tecnológica; Tomada de Decisões; Conhecimento nas Organizações; Empresas; Inovação.

Artigo 44	4	10	2	5	Gestão do Conhecimento; Gestão; Competitividade.	Gestão e Papel	Empresas; Gestão ; Bases de Dados; Conhecimento Tácito; Resumos.	Competitividade; Gestão do Conhecimento.	Gestão do Conhecimento; Conhecimento nas Organizações; Tecnologias da Informação e Comunicação; Competitividade; Tomada de Decisões.
Artigo 45	5	10	3	4	Ciência da Informação; Bibliometria; Informetria.	Avaliação; Internet.	Fluxo de Informação; Áreas do Conhecimento; World, Wide Web; Produtividade de Autor; Tomada de Decisões; Comunicação.	Ciência da Informação; Bibliometria; Informetria	Ciência da Informação; Bibliometria; Bases de Dados; Informetria.

Fonte: Elaborado pelo autor.

No artigo 46 foram propostos 6 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 7 termos. Dos 5 termos atribuídos pelo SISA, 5 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 6 são comuns com a folksonomia.

Os softwares apresentaram 6 termos em comum onde 5 deles são termos corretos comuns.

O único erro do SISA: Pesquisa é abrangente demais para representar o conteúdo do artigo, além de não ser um termo preferido do TBCI.

Os erros do MAUI estão pautados na atribuição de termos compostos que apresentam baixa frequência no texto e são termos não preferidos do TBCI, e sendo assim, não representam o conteúdo do artigo.

No artigo 47 foram propostos 6 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 12 termos. Dos 6 termos atribuídos pelo SISA, 5 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia.

Os softwares apresentaram 6 termos em comum, 5 deles são acertos comuns. O termo Periódicos é um erro comum aos softwares por não ser um termo preferido do TBCI, além de sua generalidade para representar o artigo analisado.

Os erros do MAUI estão pautados na atribuição de termos compostos que apresentam baixa frequência no texto, sendo assim, não representam o conteúdo do artigo. Cabe destacar que todos são termo não preferidos do TBCI.

O artigo 48 foram propostos 5 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 9 termos. Dos 5 termos atribuídos pelo SISA, 4 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 6 são comuns com a folksonomia.

Com relação ao único erro do SISA: Economia foi considerado um termo muito simples e genérico para a representação do artigo analisado.

Com relação aos erros do MAUI, o software também atribuiu o erro Economia, além dos termos compostos: Processamento de Informações, Tomada de Decisões e Sistemas de Informação, que não possuem representatividade frente ao conteúdo do artigo analisado.

Os softwares apresentaram 4 termos em comum.

No artigo 49 foram propostos 4 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 11 termos. Dos 4 termos atribuídos pelo SISA, 3 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia.

Os softwares apresentaram 3 termos em comum: Descritores, Periódicos e Monitoramento.

O SISA apresentou apenas um erro: Pesquisa, que foi considerado como simples e genérico para fins de representação do conteúdo do artigo, além de não ser um termo preferido pelo TBCI.

O MAUI por sua vez, apresentou 5 erros: Informação Científica e Tecnológica, Tomada de Decisões, Títulos de Documentos, Buscas e Mapas. Com relação aos termos compostos, Informação Científica e Tecnológica e Títulos de Documentos não são termos preferidos do TBCI. O termo Tomada de Decisões apresenta ausência de frequência no texto do artigo, constatando-se assim, irrelevante para representar o conteúdo do artigo.

No artigo 50 foram propostos 5 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 9 termos. Dos 5 termos atribuídos pelo SISA, 3 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia.

Os softwares apresentaram apenas um erro em comum: Papel, que foi considerado genérico para representar o conteúdo do artigo analisado.

Com relação aos erros do SISA, constatou-se que Cooperação e Papel são muito simples e abrangentes para representar o conteúdo do artigo analisado.

Com relação aos erros do MAUI, destacou-se a atribuição do termo Tecnologias da Informação e Comunicação, pois este termo não é preferido pelo TBCI.

Os softwares apresentaram 4 termos em comum.

Podemos visualizar todas as informações dos artigos 46 ao 50 no quadro 21:

Quadro 21: Análise comparativa dos termos dos artigos 46 ao 50

Corpus de referência	Quantidade de termos atribuídos		Termos comuns com a folksonomia		Termos comuns entre SISA e MAUI	Erros com relação a Folksonomia de cada Artigo		Acertos com relação a Folksonomia de cada Artigo	
	SISA	MAUI	SISA	MAUI		SISA	MAUI	SISA	MAUI
Artigo 46	6	10	5	6	Teoria do Caos; Lei de Bradford; Ciência da Informação; Bibliometria; Periódicos.	Pesquisa	Revisões de Literatura; Faculdades de Biblioteconomia; Publicações Seriadas; Unidades de Informação.	Bibliometria; Ciência da Informação; Lei de Bradford; Periódicos; Teoria do Caos.	Teoria do Caos; Lei de Bradford; Ciência da Informação; Bibliometria; Modelos Matemáticos; Periódicos.
Artigo 47	6	10	5	5	Periódicos Eletrônicos; Periódicos Científicos; Ciência da Informação; World Wide Web; Periódicos; Usuários.	Periódicos	Periódicos; Obras de Referência; Artigos de periódico; Acesso; Monitores de Vídeo.	Periódicos Eletrônicos; Usuários; Ciência da Informação; Periódicos Científicos; World Wide Web.	Periódicos Eletrônicos; Periódicos Científicos; Ciência da Informação; World Wide Web; Usuários
Artigo 48	5	10	4	6	Economia da Informação; Economia; Empresas; Custos.	Economia	Economia; Processamento de Informações; Tomada de Decisões; Sistemas de Informação.	Economia da Informação; Custos; Empresas; Inovação.	Economia da Informação; Usos da Informação; Informação para Negócios; Empresas; Custos; Tecnologias da Informação; Comunicação.
Artigo 49	4	10	3	5	Descritores; Periódicos; Monitoramento.	Pesquisa	Informação Científica e Tecnológica; Tomada de Decisões; Títulos de Documentos; Buscas; Mapas.	Descritores; Periódicos; Monitoramento.	Descritores; Bases de Dados; Periódicos; Monitoramento; Inteligência Competitiva.
Artigo 50	5	10	3	5	Formação Profissional; Educação; Bibliotecários; Papel.	Cooperação; Papel.	Tomada de Decisões; Ciência da Informação; Tecnologias da Informação e Comunicação; Educação Básica; Papel.	Formação Profissional; Bibliotecários; Educação.	Formação Profissional; Sociedade da Informação; Educação; Bibliotecários; Profissionais de Informação.

Fonte: Elaborado pelo autor.

No artigo 51 foram propostos 4 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 11 termos. Dos 4 termos atribuídos pelo SISA, 3 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 7 são comuns com a folksonomia.

O único erro do SISA apontado pela análise foi Categorias, que é um termo genérico e não representa bem o conteúdo do artigo em questão.

Com relação aos erros do MAUI: Sítios Web, Suportes de Informação e Sociedade da Informação, constatou-se que os termos são preferidos não preferidos do TBCI.

Todos os acertos do SISA estão contidos nos acertos do MAUI.

Os softwares apresentaram 3 termos em comum.

No artigo 52 foram propostos 7 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 9 termos. Dos 7 termos atribuídos pelo SISA, 4 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 6 são comuns com a folksonomia.

Os softwares apresentaram 4 termos comum.

Os softwares apresentaram 3 acertos em comum e 1 erro em comum.

O SISA apresentou apenas o termo Papel como erro. Este erro foi devido ao software ter atribuído o termo de forma simples e genérica, não traduzindo assim, o conteúdo do artigo em questão.

Com relação aos erros do MAUI, os termos Políticas de Informação e Transferência da Informação foram atribuídos pelo software, mas são termos não preferidos do TBCI. O termo Comunicação Informal não representa o conteúdo do artigo e esta afirmação pode ser evidenciada pela ausência do termo no texto do artigo analisado. O termo Papel, assim como o SISA, também errou ao atribuir o termo por questões de generalidade.

No artigo 53 foram propostos 3 termos pelo SISA e 7 termos pelo MAUI. A folksonomia apresenta 8 termos. Dos 3 termos atribuídos pelo SISA, 2 são comuns a folksonomia. Dos 7 termos atribuídos pelo MAUI, 4 são comuns com a folksonomia.

Os softwares apresentaram 3 termos comuns Recuperação da Informação e Usuários.

Os softwares apresentaram 3 acertos em comum e nenhum erro em comum.

O SISA apontou o termo Professores como erro e foi constatado que o erro se deu devido a generalidade do termo frente à representação do conteúdo do artigo.

Com relação aos erros do SISA, os termos Buscas, Booleanas e Acesso são termos simples e genéricos que prejudicam a representação do conteúdo do artigo. Cabe ressaltar que o termo Acesso é um termo não preferido do TBCI.

Com relação aos erros do MAUI, foi constatada a atribuição de termos simples e genéricos, fazendo com que não relevantes para a representação do conteúdo do artigo.

Os softwares apresentaram 2 termos em comum.

No artigo 54 foram propostos 8 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 10 termos. Dos 8 termos atribuídos pelo SISA, 3 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 7 são comuns com a folksonomia.

Os softwares apresentaram 3 termos em comum.

Com relação aos erros do SISA, os termos atribuídos foram analisados e dentre os 5 termos, 4 são simples, o que faz com que estes termos sejam considerados genéricos ao ponto de representar o conteúdo do artigo. O termo errado composto é Competência em Informação, e o erro está pautado no fato do TBCI reconhecê-lo como não preferido.

Com relação aos erros do MAUI, o software atribuiu termos compostos, mas que não são relevantes para a representação do conteúdo do artigo analisado, tendo em vista a generalidade atribuída aos termos.

No artigo 55 foram propostos 3 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 9 termos. Dos 3 termos atribuídos pelo SISA, 2 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 4 são comuns a folksonomia.

Os softwares apresentaram 2 termos em comum: Mercado de Trabalho e Empresas.

Podemos visualizar todas as informações dos artigos 51 ao 55 no quadro 22:

Quadro 22: Análise comparativa dos termos dos artigos 51 ao 55

Corpus de referência	Quantidade de termos atribuídos		Termos comuns com a folksonomia		Termos comuns entre SISA e MAUI	Erros com relação a Folksonomia de cada Artigo		Acertos com relação a Folksonomia de cada Artigo	
	SISA	MAUI	SISA	MAUI		SISA	MAUI	SISA	MAUI
							Sítios Web; Suportes de Informação; Sociedade da Informação.	Acessibilidade; Bibliotecas; Usuários.	Acessibilidade; Digitalização; Tecnologias da Informação e Comunicação; Acesso à

Artigo 51	4	10	3	7	Acessibilidade; Bibliotecas; Usuários.	Categorias			Informação; Bibliotecas; Usuários; Sistemas de Informação.
Artigo 52	7	10	4	6	Ciência da Informação; Estudos de Usuários; Educação; Papel.	Educação; Pesquisa; Papel.	Políticas de Informação; Projetos de Pesquisa; Transferência da Informação; Comunicação Informal; Papel.	Estudos de Usuários; Ciência da Informação; Internet.	Ciência da Informação; Estudos de Usuários; Projetos de Pesquisa; Educação; Fluxo da Informação; Hipertextos.
Artigo 53	3	7	2	4	Recuperação da Informação; Usuários.	Professores	Buscas; Booleanas; Acesso.	Recuperação da Informação; Usuários	Recuperação da Informação; comportamento do Usuário; Mecanismos de Busca; Usuário.
Artigo 54	8	10	3	7	Competência em Informação; Bibliotecários; Educação.	Filosofia; Acesso; Avaliação; Competência em Informação; Papel.	Acesso à Informação; Tomada de Decisões; Serviços de Informação.	Bibliotecas; Bibliotecários; Educação.	Competência em Informação; Sociedade da Informação; Buscas de Informação; Profissionais de Informação; Bibliotecários; Educação; Usos da Informação.
Artigo 55	3	10	2	4	Mercado de Trabalho; Empresas.	Empresas	Gestão do Conhecimento; Gestão da Informação; Empresas; Processos de Gestão; Tomada de Decisões; Sociedade da Informação.	Mercado de Trabalho; Formação Profissional.	Profissionais de Informação; Mercado de Trabalho; Ciência da Informação; Conhecimento nas Organizações.

Fonte: Elaborado pelo autor.

No artigo 56 foram propostos 9 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 10 termos. Dos 9 termos atribuídos pelo SISA, 5 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 6 são comuns com a folksonomia.

Os softwares apresentaram 2 erros em comum: Cientistas e História.

Os softwares apresentaram 5 acertos em comum.

Com relação aos erros do SISA, constatou-se que os 4 erros são termos simples e de pouca relevância para a representação do conteúdo do artigo. Cabe ressaltar que em Pesquisa, o termo é considerado pelo TBCI como não preferido.

Com relação aos erros do MAUI, constatou-se que 2 são termos simples e 2 compostos. Os termos Cientistas e História são termos genéricos para representar o conteúdo do artigo em questão. No caso dos termos compostos: Comunicação Informal e Conceitos de Informação, constatou-se que apesar de serem termos compostos, possuem baixa frequência no texto do artigo analisado e não correspondem a termos relevantes.

Os softwares apresentaram 7 termos em comum.

No artigo 57 foram propostos 4 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 9 termos. Dos 4 termos atribuídos pelo SISA, 3 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 3 são comuns com a folksonomia.

Os softwares não apresentaram erros em comum.

Os softwares apresentaram 2 acertos em comum: Ciência da Informação e Recuperação da Informação.

Com relação ao erro do SISA, constatou-se que o termo Avaliação atribuído pelo sistema não possui relevância para representar o conteúdo do artigo analisado, devido a sua generalidade.

Com relação aos 7 erros do MAUI, em 5 termos foram constatados termos que são considerados pelo TBCI como não preferidos. E os demais erros estão voltados para atribuição de termos genéricos que não possuem força para representar o conteúdo do artigo analisado.

Os softwares apresentaram 2 termos em comum: Ciência da Informação e Recuperação da informação.

No artigo 58 foram propostos 5 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 7 termos. Dos 5 termos atribuídos pelo SISA, 5 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 4 são comuns com a folksonomia.

Os softwares não apresentaram erros em comum.

Os softwares apresentaram 3 acertos em comum: Ciência da Informação e Recuperação da Informação.

O SISA não apresentou erros.

Com relação ao erro do SISA, constatou-se que o termo “ Avaliação” atribuído pelo sistema não possui relevância para representar o conteúdo do artigo analisado, devido a sua generalidade.

Com relação aos erros do MAUI, constatou-se que o software atribuiu 6 termos compostos e dentre eles, 3 termos são considerados pelo TBCI como não preferidos, e os demais erros estão voltados para atribuição de termos genéricos que não possuem força para representar o conteúdo do artigo analisado, além de baixa frequência no texto.

Os softwares apresentaram 3 termos em comum.

No este artigo 59 foram propostos 5 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 10 termos. Dos 5 termos atribuídos pelo SISA, 4 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 4 são comuns com a folksonomia.

Os softwares não apresentaram erros em comum.

Os softwares apresentaram 2 acertos em comum: Educação e Bibliometria.

O SISA apresentou um único erro: Pesquisa, foi constatado que o termo é genérico e simples para representar o conteúdo do artigo em questão.

Com relação aos erros do MAUI, constatou-se que os termos são todos compostos, e que apresentam pouca relevância para representar o conteúdo do artigo analisado devido a sua generalidade. Foram constatados ainda, dois termos que são considerados pelo TBCI como não preferidos: Áreas do Conhecimento e Periódicos Científicos.

Os softwares apresentaram 2 termos em comum.

No artigo 60 foram propostos 8 termos pelo SISA e 10 termos pelo MAUI. A folksonomia apresenta 13 termos. Dos 8 termos atribuídos pelo SISA, 7 são comuns a folksonomia. Dos 10 termos atribuídos pelo MAUI, 5 são comuns com a folksonomia. Os softwares não apresentaram erros em comum.

Os softwares apresentaram 4 acertos em comum: Inteligência Competitiva, Agentes Inteligentes, Gestão da Informação e Estudo de Caso.

Com relação ao único erro do SISA, constatou-se que o termo Gestão é considerado pelo TBCI como termo não preferido.

Com relação aos 5 erros do MAUI, os termos Gestão do Conhecimento, Tecnologias da Informação e Comunicação e Conceitos de Informação são considerados pelo TBCI como não preferidos. Os demais erros estão voltados para atribuição de termos genéricos que não possuem força para representar o conteúdo do artigo analisado.

Os softwares apresentaram 4 termos em comum.

Podemos visualizar todas as informações dos artigos 55 ao 60 no quadro 23:

Artigo 58	5	10	5	4	Recuperação da Informação; Ciência da Informação; Ciência Cognitiva.		Sistemas de Informação; Buscas de Informação; Ciência da Computação; Necessidades de Informação; Teoria da Informação; Aplicações do Computador.	Ciência da Informação; Ciência Cognitiva; Indexação; Interação Homem-Computador; Recuperação da Informação.	Recuperação da Informação; Processamento de Informações; Ciência da Informação; Ciência Cognitiva.
Artigo 59	5	10	4	4	Educação; Bibliometria.	Pesquisa	Ciência da Informação; Referências Bibliográficas; Áreas do Conhecimento; Literatura Científica; Periódicos Científicos; Páginas da Web.	Educação, Bibliometria; Cientometria; Fonte de Pesquisa.	Frente de Pesquisa; Educação; Bibliometria; Comunicação.
Artigo 60	8	10	5	5	Inteligência Competitiva; Agentes Inteligentes; Gestão da Informação; Estudos de Caso.	Gestão	Engenharia de Produção; Gestão do Conhecimento; Tomada de Decisões; Tecnologias da Informação e Comunicação; Conceitos de Informação.	Agentes Inteligentes; Inteligência Competitiva; Internet; Estudos de Caso; Gestão da Informação; Inovação; Monitoramento.	Inteligência Competitiva; Agentes Inteligentes; Usos da Informação; Gestão da Informação; estudos de Caso.

Fonte: Elaborado pelo autor.

De acordo com a análise dos quadros pode-se visualizar os erros e acertos dos softwares estudados, bem como os termos comuns atribuídos pelo SISA e pelo MAUI por meio da indexação automática por atribuição a partir da Linguagem de Indexação TBCI.

5 CONCLUSÃO

Esta pesquisa foi desenvolvida a partir da verificação de que a temática da indexação automática necessita de conhecimentos acerca de softwares para sua aplicação e utilização. Nesse contexto, buscamos estudar os softwares de indexação automática por atribuição.

De acordo com o objetivo geral da pesquisa, podemos constatar que a base teórico-metodológica selecionada e construída ao longo do trabalho se mostrou extremamente necessária, os objetivos específicos provenientes do objetivo geral da pesquisa foram alcançados, bem como o problema de pesquisa foi devidamente respondido.

O intuito da pesquisa visa contribuir para o desenvolvimento da temática ao levantar informações que nos permitiram identificar aspectos da indexação automática dos dois softwares estudados. Sendo assim, os três objetivos específicos foram alcançados por meio da metodologia proposta pela pesquisa.

Quanto ao primeiro objetivo específico da pesquisa, cuja a ideia perpassava pela especificação e descrição referente ao processo da indexação de ambos os softwares, podemos afirmar que o objetivo foi alcançado por meio do levantamento acerca da literatura da indexação automática por atribuição. Os dois softwares foram devidamente analisados e descritos minuciosamente num quadro, para proporcionar uma melhor visualização.

O objetivo ainda demandava a realização do processo de indexação automática pelo SISA e pelo MAUI. Quanto a este processo, podemos afirmar que, a fundamentação teórica encontrada no capítulo 2 da presente pesquisa foi extremamente necessária, tendo em vista que MAUI é um software pouco conhecido e pela primeira vez utilizado no Brasil na área da CI.

Quanto ao segundo objetivo específico da pesquisa, em que a análise dos resultados obtidos por ambos os sistemas, bem como a avaliação da qualidade da representação dos assuntos dos documentos do corpus pelos sistemas SISA e MAUI esteve em evidência, podemos afirmar que, assim como o objetivo específico anterior, o levantamento da literatura pertinente norteou os caminhos da pesquisa para que os respectivos valores dos índices de consistência, revocação, precisão e medida F alcançados por cada sistema na indexação de cada documento fossem obtidos e analisados.

Quanto ao terceiro objetivo específico, que consistiu na comparação da indexação automática dos sistemas SISA e MAUI, podemos afirmar que a partir dos resultados do experimento de indexação e da comparação dos termos de indexação automática atribuídos pelo SISA e dos termos atribuídos pelo MAUI, e, levando-se em consideração as especificações de cada software, verificamos que o MAUI atribuiu uma grande quantidade de termos de

indexação, enquanto o SISA, atribuiu menos termos que o MAUI, porém os termos possuem uma maior precisão. Cabe ressaltar que, a quantidade de termos atribuídos pelos softwares é apenas uma das variáveis influenciadoras no desempenho de cada sistema.

Quanto à problemática da pesquisa: Qual a qualidade da indexação dos sistemas SISA e MAUI? Como se dá o processo de indexação em ambos os sistemas? Qual o sistema que apresenta melhor desempenho? Quais os tipos de erros cometidos pelos sistemas na indexação?

Podemos constatar que tanto o SISA quanto MAUI, estão longe de substituir a indexação humana, pois a indexação por se tratar de um processo subjetivo, o homem sempre será considerado o melhor indexador. No entanto, tendo em vista o cenário da sociedade atual, bem como o montante de informações científicas que é produzido pela ciência e a necessidade de organizar, armazenar, acessar e disponibilizar informações, os softwares de indexação automática podem contribuir para o processo de indexação, no sentido de otimizar o tempo do indexador.

Os resultados obtidos por ambos os softwares foram considerados bons. O SISA obteve as médias de 28% no índice de consistência, 63% no índice de precisão, 34% no índice de revocação e 42% no índice da medida F, enquanto que o MAUI obteve a média de 36% no índice de consistência, 54% no índice de precisão, 51% no índice de revocação e 52% no índice de medida F. Assim, podemos concluir que o MAUI apresentou o melhor desempenho por meio das métricas.

Alguns fatores intervenientes foram observados no processo de indexação de ambos os softwares: atribuição de plural e singular dos termos propostos, dificuldade em atribuir termos compostos, diferenças entre as estruturas nos termos de indexação, atribuição automática de termos gerais e específicos.

A maior limitação da pesquisa foi a ausência de trabalhos acerca do MAUI na literatura brasileira na área da Ciência da Informação. Este fato limitou a pesquisa no sentido estritamente teórico, porém, tendo em vista que o MAUI é bastante utilizado na literatura estrangeira, e que este é o primeiro trabalho de aplicação na área da CI no Brasil, esperamos que outros possam vir a surgir para que o MAUI possa ser cada vez mais conhecido e/ou aprimorado.

Considerando a importância em dar continuidade aos estudos acerca da indexação automática por atribuição, a pesquisa sugere novos experimentos com o MAUI, a fim de aprofundar os conhecimentos relacionados ao software, e permitir assim, que o MAUI seja mais utilizado e conseqüentemente aprimorado. Cabe ressaltar que, o MAUI possui a consistência próxima à indexação humana, portanto seria interessante comparar a indexação automática do software com a indexação semiautomática.

No tocante ao SISA, a pesquisa sugere experimentos com versões mais recentes do software que englobe aspectos morfológicos e semânticos, em detrimento da frequência relativa dos termos no texto dos documentos.

A pesquisa possui impacto tecnológico por tratar das temáticas da Indexação Automática e da Avaliação de Sistemas de Indexação Automática, sendo assim, seus resultados são de suma importância, tendo em vista a escassez de trabalhos na Literatura da área. Conhecer aspectos teóricos e práticos acerca da Indexação automática e Avaliação da Indexação Automática é fundamental para desenvolver a área da CI, frente à preocupação com o aumento, acesso e disponibilização do volume de informações gerado a todo momento pelas comunidades acadêmica e científica, bem como ao uso da tecnologia.

Esperamos, com a reflexão dos resultados desta pesquisa, a compreensão de que softwares de indexação automática podem contribuir nos processos de organização da informação, mas acima de tudo, a avaliação destes softwares se constitui numa ferramenta para medir a qualidade da indexação automática, e sobretudo, no aprimoramento de recursos que podem ser implementados aos softwares, inclusive, melhorar o desempenho dos sistemas.

Por fim, a pesquisa buscou fornecer informações acerca dos softwares SISA e MAUI, com o intuito de disponibilizar informações relacionadas ao processo de indexação automática por atribuição, bem como descrever informações referentes às características de cada software. A pesquisa visa contribuir para o desenvolvimento da área da Ciência da Informação e das temáticas indexação automática e avaliação de sistemas de indexação automática por atribuição numa perspectiva de minimizar a escassez de trabalhos de mesma natureza. Este trabalho representa ainda o resultado dos estudos acerca da indexação automática do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco (PPGCI-UFPE), tendo como área de concentração *Informação, Memória e Tecnologia*, discutido na linha de pesquisa *Comunicação e visualização da memória*.

REFERÊNCIAS

- ALMEIDA, T. de. **O vocabulário controlado como instrumento de organização e representação da informação na FINEP**. Rio de Janeiro: UFRJ / IBICT, 2011. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal do Rio de Janeiro; Instituto Brasileiro de Informação em Ciência e Tecnologia, 2011.
- ANDERSON, J. D. Standards for indexing: revising the American National Standard Guidelines Z39.4. **Journal of the American Society for Information**, v. 45, n. 8, 1994.
- ANDREEWSKY, A.; S RUAS, V. **Indexação automática baseada em métodos linguísticos e estatísticos e sua aplicabilidade à língua portuguesa**. Rio de Janeiro: PUC, 1982.
- ARAÚJO JUNIOR, R. H. de. Precisão no processo de busca e recuperação da informação. Brasília: Thesaurus, 2007. BARDIN, L. **Análise de conteúdo**. Lisboa: Edições 70, 1991. p. 225.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12676**: métodos para análise de documentos – determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro, 1992. 4 p.
- BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de informação**: conceitos e tecnologia das máquinas de busca. 2. ed. Porto Alegre: Bookman, 2013.
- BANDIM, M. A. S.; CORRÊA, R. F. A consistência na indexação automática por atribuição de artigos científicos na área de Ciência da Informação. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 23, n. 53, p.64-77, set. 2018. DOI: <https://doi.org/10.5007/1518-2924.2018v23n53p64>
- BANDIM, M. A. S.; CORRÊA, R. F. Indexação automática por atribuição de artigos científicos em português da área de Ciência da Informação. **Transinformação**, v. 31, p. 1-12, 2019. DOI: <https://doi.org/10.1590/2318-0889201931e180004>
- BARDONI, L.; PAZIENZA, M. T. Documents automatic indexing in na environmental domain. **Internation Forum on Information and Documentation**, v. 22, n. 1, 1997.
- BARITÉ, M. Control de vocabulario: orígenes, evolución y proyección. **Ciência da Informação**, v. 43, n. 1, 2014.
- BERTRAND-GASTALDY, S. PAGOLA. G. L'analyse du contenu textuel em vue de la contruction de thesaurus et l'indexation assistiees par ordinateus; applications possibles avec sato. **Documentation et bibliothèques**, v. 38, n. 2, 1992.
- BLAIR, D. C. **Language and representation in information retrieval**. Amsterdam, Elsevier, 1990.
- BONURA, L. S. The art of indexing. Nueva York: Jon Wiley, 1994.
- BORGES, G. S. B.; MACULAN, B. C. M. S.; LIMA, G. A. B. O. **Indexação automática e semântica**: estudo de análise do conteúdo de teses e dissertações. **Informação & Sociedade: Estudos**, João Pessoa, v.18, n.2, p. 181-193, maio/ago. 2008.

- BORKO, H.; BERNIER, C. **Indexing concepts and methods**. New York: Academic Press, 1978.
- BOYCE, B. & LOCKARD, M. Automatic and manual indexing performance in a small file of medical literature. *Bulletin of Medical Library Association*, 63 (4):378-85, Oct. 1975
- BOYCE, B.; LOCKARD, M. Automatic and manual indexing performance in a small file of medical literature. *Bulletin of Medical Library Association*, 63 (4):378-85, Oct. 1975.
- BRÄSCHER, M; CAFÉ, L. Organização da informação ou organização do conhecimento? In: LARA, Marilda Lopes Ginez de; SMIT, Johanna Wilhelmina (Orgs.). **Temas de Pesquisa em Ciência da Informação no Brasil**. São Paulo: Escola de Comunicações e Artes/USP, 2010. p. 87-103.
- CAFÉ, L. M. A ; SALES, R. de. Organização da informação: conceitos básicos e breve fundamentação teórica. In: ROBREDO, Jaime; BRÄSCHER, Marisa (Orgs.). **Passeios pelo bosque da informação: estudos sobre representação e organização da informação e do conhecimento**. Brasília: IBICT, 2010. p. 115-129.
- CAMPOS, M. L. A.; GOMES, H. E. Metodologia de elaboração de tesouro conceitual: a categorização como princípio norteador. **Perspectiva em Ciência da Informação**, Belo Horizonte, v. 11, n. 3, p. 348-359, 2006.
- CARNEIRO, M.C. Diretrizes para uma política de indexação. **Revista da Escola de Biblioteconomia: UFMG**, Belo Horizonte, p. 187-220, set. 1985.
- CARROLL, J. M; ROELOFFS, R. Computer selection of keywords using word-frequency analysis. *American Documentation*, 20 (3):227-33, July 1969.
- CASTELLS, M. **A Sociedade em Rede**. v.1, 6. ed. ver. ampl. São Paulo: Paz e Terra, 2002.
- CESARINO, M. A.; PINTO, M. C. M. F. Cabeçalhos de assunto como linguagem de indexação. **Revista da Escola de Biblioteconomia: UFMG**, Belo Horizonte, v. 7, n. 2, p. 268-288, set. 1978.
- CHAUMIER, J. Indexação: conceito, etapas e instrumentos. Trad. José Augusto Chaves Guimarães. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v.21, n.1/2, p.63-79, jan./jun. 1988.
- CINTRA, A. M. M. Elementos de lingüística para estudos de indexação. **Ciência da Informação**, v. 12, n. 1, 1983.
- CINTRA, A. M. M. et al. **Para entender as linguagens documentárias**. 2. ed. rev. atual. São Paulo: Polis, 2002. 96 p.
- CINTRA, A. M. et al. **Para entender as linguagens documentárias**. São Paulo: Polis; APB, 1994. 72p.
- COOPER, W. S. Indexing documents by gedanken experimentation, **Jornal of the American Society for Information Science**, 1978.

DAHLBERG, I. Teoria do conceito. **Ciência da informação**. Rio de Janeiro, v. 7, n. 2, p. 101-107, jul./dez. 1978.

DAMERAU, F. J. An experiment in automatic indexing. **American Documentation**, v.16, n. 4, 1965.

DIAS, E. W.; NAVES, M. L. **Análise de assunto**: teoria e prática. Brasília: Teshaurus, 2007. 116 p.

DODEBEI, V. L. D. **Tesauro**: linguagem de representação da memória documentária. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002. 119 p.

DOWBOR, L. **Sociedade do conhecimento**. São Paulo, Pioneira, 1994.

FERNEDA, E. **Ontologia como recurso de padronização terminológica de um sistema de recuperação de informação**. 96 f. Relatório de Pesquisa (Pós-Doutorado em Ciência da Informação) – Universidade Federal da Paraíba, João Pessoa, 2013.

FERNEDA, E. **Recuperação de informação**: análise da contribuição da ciência da computação para a ciência da informação. São Paulo, 2003. 147 p. Tese (Doutorado em Ciência da Informação e Documentação) – Universidade de São Paulo. São Paulo, 2003.

FUJITA, M. S. L. (Org.). **A indexação de livros**: a percepção de catalogadores e usuários de bibliotecas universitárias. Um estudo de observação do contexto sociocognitivo com protocolos verbais. São Paulo: Cultura Acadêmica, 2009.

FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para indexação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 1, n. 1, p. 60-90, jul./dez. 2003.

FUJITA, M. S. L. A leitura documentária na perspectiva de suas variáveis: leitor-contexto. **DataGramZero**, v. 5, n. 4, ago. 2004.

FUJITA, M. S. L. A política de indexação para representação e recuperação da informação. In: GIL LEIVA, I.; FUJITA, M. S. L. (Ed.). **Política de indexação**. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012. p. 17-28.

FUJITA, M. S. L. A representação documentária no processo de indexação com o modelo de leitura documentária para textos científicos e livros: uma abordagem cognitiva com protocolo verbal. **Ponto de Acesso**, Salvador, v. 7, n. 1, p. 42-66, abr. 2013.

FUJITA, M. S. L. Avaliação da eficácia de recuperação do sistema de indexação Preci. **Ci. Inf.**, Brasília, v. 18, n. 2, p. 120-134, jul./dez. 1989.

FUJITA, M. S. L. **Precisão na língua portuguesa**: teoria e prática de indexação. Brasília: Ed. UnB, 1988.

FUJITA, M. S. L.; GIL-LEIVA, I. **As linguagens de indexação em bibliotecas nacionais, arquivos nacionais e sistemas de informação na América Latina**. Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2010.

FUJITA, M. S. L.; CRUZ, M. C. A.; PATRÍCIO, B. O. M. A construção de tesouros na perspectiva dos manuais de indexação. **Encontro Nacional de Pesquisa em Ciência da Informação**, n. XVIII ENANCIB, 2017.

GARCÍA GUTIÉRREZ, A. L. **Lingüística documental**. Barcelona, Mitre, 1984.

GIL LEIVA, I, RUBI, M. P., FUJITA, M. S. L. Consistência na indexação em bibliotecas universitárias brasileiras. **Transinformação**, Campinas, v. 20, p. 233-54, 2008.

GIL LEIVA, I. **La automatización de la indización de documentos**. Gijón: Trea, 1999.

GIL LEIVA, I. Prefácio. In: FUJITA, M. S. L. (Org.). **Política de indexação para bibliotecas: elaboração, avaliação e implantação**. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2016. p. 11-13.

Gil Leiva, I. SISA – Automatic Indexing System for Scientific Articles: Experiments with location heuristics rules versus TF-IDF rules. **Knowledge Organization**, v.44, n.3, p.139-162, 2017.

GIL LEIVA, I. Sistema para la indización automática (SISA) de artículos de revista de Biblioteconomía y documentación. **II Jornadas de Tratamiento y Recuperación de Information**. Madrid: 2003.

GIL-LEIVA, I. **La automatización de la indización de documentos**. Gijón: Trea, 1999.

GIL-LEIVA, I. **Manual de indización: teoría y práctica**. Gijón: Trea, 2009.

GIL-LEIVA, I. SISA – Automatic indexing system for scientific articles: Experiments with location heuristics rules versus TF-IDF Rules. **Knowledge Organization**, v.44, n. 3, p. 139-162, 2017.

GODBY, C. J; WATSON, B. C. Experimental linguistic indexing for information retrieval: ELIXIR, **Journal of Library Administration**. v. 34, n. ½. 2001.

GUINCHAT, C.; MENO, M. **Introdução geral às ciências e técnicas da informação e documentação**. 2. ed. corr. aum. Brasília: IBICT, 1994.

HALLER, J. Análise automática de textos em sistemas de informação. Revista de Biblioteconomia de Brasília, 1983.

HALLER, J. Processamento de textos em linguagem natural. In: **Congresso Nacional de Informática**, 15., Rio de Janeiro, 1982.

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 dez. 2019.

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

HMEIDI, I; KANAAN, G; EVENS, M. Design and implementation of automatic indexing for information retrieval with Arabia documents. **Journal of American Society for Information Science**, v. 48. N. 10, 1997.

INÁCIO, M. de O. **Avaliação da indexação em bibliotecas universitárias: uma aplicação em catálogos online**. 2012. 157 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Informação, Universidade Estadual Paulista (Unesp), Marília 2012.

KARETNYK, D; KARLSSON, F; SMART, G. Knowledge based indexing of-morpho-syntactically analysed language. **Expert Systems for Information Management**, vol. 4. n, 1, 1991.

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

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

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. Brasília: Briquet de Lemos, 1993.

LANCASTER, F. W. **Information retrieval systems: characteristics, testing and evaluation**. 2.ed. New York : Wiley, 1979.

LAPA, R. **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.

LAPA, R.; CORRÊA, R. F. Indexação Automática no Âmbito da Ciência da Informação no Brasil. **Informação & Tecnologia (itec)**. Marília/João Pessoa, v. 2, n. 1, p.1-18, 1 jul. 2014. Semestral.

LE COADIC, Y. **A ciência da informação**. 2 ed. Brasília briquet de lemos 2004.

LUHN, H. P. A statistical approach to mechanized encoding and searching of literary information. **IBM Journal of Research and Development**, v. 1, n. 4, p. 309-317, 1957.

MARCONI, M. A.; LAKATOS, E. M. Fundamentos de metodologia científica. 7. ed. São Paulo: Atlas, 2010.

MATTELART, A. Para que “Nova Sociedade Mundial da Informação”. In: MORAES, Denias de. (Org.). **Sociedade Mdiatizada**. Rio de Janeiro: Mauad, 2006.

MEDELYAN, O. **Human-competitive automatic topic indexing**. PhD Thesis. University of Waikato, New Zealand, 2009.

MEDELYAN, O.; FRANK, E.; WITTEN, I. H. Human-competitive tagging using automatic keyphrase extraction. In: **Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing**, 2009.

MIINE, D., I. H. WITTEN. (2008). Learning to link with Wikipedia. In: *Proc. ACM Conf. on Information and Knowledge Management*. New York, NY: ACM Press, pp. 509–518.

MILSTEAD, J. L. Needs for research in indexing. **Journal of the American Society for Information Science**, v. 45, n. 8, 1994.

MOOERS, C. Zatocoding applied to mechanical organization of knowledge. *American Documentation*, v. 2, n. 1, 1951.

MOOERS, C. N. Zato coding applied to mechanical Organization of Knowledge. **American Documentation**, v. 2, n. 1, p. 20-32, 1951.

MOREIRO GONZÁLEZ, J. A. **El contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural**. Gijón (Asturias): Trea, 2004. 291 p.

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

NARUKAWA, C. M.; FUJITA, M. S. L. O uso de linguagens documentárias alfabéticas na indexação automatizada. In: **Seminário Nacional de Bibliotecas Universitárias**, 16., 2010, Rio de Janeiro. Anais..., Rio de Janeiro: ENANCIB, 2010.

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

NARUKAWA, C. M.; GIL-LEIVA, I.; FUJITA, M. S. L. Indexação automatizada de artigos de periódicos científicos: análise da aplicação do software sisa com uso da terminologia decs na área de odontologia. **Informação & Sociedade: Estudos**, v. 19, n. 2, 2009.

OLIVEIRA, L. de. Política de Indexação: concepções acerca do conceito e percepções em torno de sua elaboração. **Ciência da Informação em Revista**, n. 2, v. 4, p. 39-58, 2017.

ORTEGA, C. D. Fundamentos da organização da informação frente à produção de documentos. **Transinformação**, Campinas, v. 20, n. 1, p. 7-15, jan./abr. 2008.

PAGANI, R. N.; KOVALESKI, J. L.; RESENDE, L. M. Tics na composição da methodi ordinatio: construção de portfólio bibliográfico sobre modelos de transferência de tecnologia. **Ciência da Informação**, v. 46, n. 2, 2017. DOI: 10.18225/ci.inf.v47i1.1886

PAPAGIANNPOULOU, E; TSOUMAKAS, G. **A Review of keyphrase extraction**. Aristotle University of Thessalonik, 2019. Disponível em: <https://arxiv.org/pdf/1905.05044.pdf>, acesso em: 10 de jun, 2019.

PINHEIRO, L. V. R. **Medidas de consistência da indexação: interconsistência**. *Ciência da Informação*, Rio de Janeiro, v.7, n.2, p.109-114, 1978.

PINHEIRO, L. V. R.; FERREZ, H. D. **Tesouro Brasileiro de Ciência da Informação**. Rio de Janeiro; Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), 2014.

PINTO MOLINA. M. C. F. El análisis de contenido: la indización de documentos. In: J. López Yepes: **Fundamentos de información y documentación**, Madrid: 1989.

PINTO, MOLINA . C. M. F. Análise e representação de assuntos em sistemas de recuperação de informação; linguagens de indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 14, n. 2, p. 169-186, set. 1985.

RAMALHO, R. A. S. Representação skos da categoria tecnologia da informação e comunicação do tesouro brasileiro de ciência da informação: um estudo preliminar. **Encontro Nacional de Pesquisa em Ciência da Informação**, n. XVIII ENANCIB, 2017.

ROBREDO, J. **Documentação de hoje e de amanhã**. 4. ed. rev. ampl. Brasília, DF: Ed. Do Autor, 2005.

ROBREDO, J. Indexação automática de textos: uma abordagem otimizada e simples, **Ciência da Informação**, Brasília, v. 20, n. 2, p. 130-136, jul./dez. 1991

RUBI, M. P. **Política de indexação para construção de catálogos coletivos em bibliotecas universitárias**. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2008.

RUBI, M. P. Política de indexação. In: GIL LEIVA, I.; FUJITA, M. S. L. (Ed.). **Política de indexação**. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012. p. 107-120.

RUBI, M. P.; FUJITA, M. S. L. Elementos de política de indexação em manuais de indexação de sistemas de informação especializados. **Perspect. cienc. inf.**, Belo Horizonte, v. 8, n. 1, p.66-77, jan./jun. 2003.

SALTON, G. A comparison between manual and automatic indexing systems, *Computing Reviews*, 10 ,(6):274, June. 1969.

SALTON, G. A new comparison between conventional indexing and automatic text processing. **Journal of the American Society for Information Science**, 23 (2):75-84, Mar./Apr. 1972.

SALTON, G. **Automatic text analysis: automatic document indexing and classification methods are examined and their effectiveness assessed**. *Science*, 168 (3929):335-43, 17 Apr. 1970.

SALTON, G.; BUCKLEY, C. TermWeighting Approaches in Automatic Text Retrieval. *Information Processing and Management*, v. 24, n. 5, 1988. SARACEVIC, T. Information Science. **Journal of the American Society for Information Science**, v. 50, n. 12, 1999

SALTON, G.; MCGILL, J. M. **Introduction to Modern Information Retrieval**. New York: McGraw-Hill, 1983.

SALTON, G; YANG, C. S. On the specification of term values in automatic indexing. **Journal of Documentator**, 1973.

SARACEVIC, T. A natureza interdisciplinar da ciência da informação. **Ciência da Informação**, v. 24, n. 1, 1995.

- SARACEVIC, T. Individual differences in organizing, searching and retrieving information. **Proceedings of the 5th Annual Meeting of the American Society of Information Science**, 1987.
- SARACEVIC, T. Information Science. **Journal of the American Society for Information Science**, v. 50, n. 12, 1999.
- SARACEVIC, T. Information science: origins, evolution and relations. In: VAKKARI, P., CRONIN, B. (Ed.) **Conceptions of library and information science**. London: Taylor Graham, 1992.
- SILVA, B. F. de M. **A folksonomia assistida aplicada na construção de Corpus de Referência para avaliação de sistemas de indexação automática**. 188 f. Dissertação (Mestrado) - Universidade Federal de Pernambuco: Recife, 2019.
- SILVA, M. R.; FUJITA, M. S. L. A prática de indexação: análise da evolução de tendências teóricas e metodológicas, **Transinformação**, Campinas, v. 16, n. 2, p. 133-161, maio/ago. 2004.
- SILVA, R. E. da; SANTOS, P. L.V. A. da C; FERNEDA, E. Modelos de recuperação de informação e web semântica: a questão da relevância. **Informação & Informação**, v. 18, n. 3, p. 27-44, 2013
- SOUZA, R. R. **Uma proposta de metodologia para escolha automática de descritores utilizando sintagmas nominais**. Belo Horizonte, 2005. 215f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação da Universidade Federal de Minas Gerais. Belo Horizonte, 2005.
- SPARCK JONES, K. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28 |1):11-21, Mar. 1972.
- SPARK JONES, K. Getting started in computerized indexing. **The indexer**, v. 15. n. 1, 1986.
- TRASK, R. L. **Dicionário de Linguagem e Lingüística**. São Paulo: Contexto. 2004, 364 p.
- UNISIST. Princípios de indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v.10, n.1, p.83-94, mar. 1981.
- VALE, E. A. Linguagens de indexação. In. : SMIT, J.W. (coord.). **Análise documentária: a análise da síntese**. Brasília: IBICT, 1997. 2. ed. p.13-28.
- VAN SLYPE, G. **Lenguages de indización: concepción, contrucción y utilización en los sistemas documentales**. Trad. Pedro Hípola e Félix de Moya. Madrid: Fundación Germán Sánchez Ruipérez, 1991.
- VIEIRA, S. B. Indexação automática e manual: revisão de literatura. **Ci. Inf.**, Brasília, v. 17, n. 1, p. 43-57, jan./jun. 1988.
- VOGEL, M. J. **A noção de estrutura lingüística e de processo de estruturação e sua influência no conceito e na elaboração de linguagens documentárias**. São Paulo, 2007.

Dissertação (Mestrado em Ciência da Informação) – Universidade de São Paulo, Escola de Comunicações e Artes, São Paulo, 2007.

VOGEL, M. J. M. A influência da jean-claude gardin e a linha francesa na evolução do conceito de linguagem documentária. **Perspectivas em Ciência da Informação**, n. Especial, p. 80-92, 2009.

WARD, M . L. The future of the human indexer. **Journal of the American Society for Information Science**, v. 28 n. 4, 1996.

WELLISCH, H. H. The art of indexing and some fallacies of its automatic. **Logos**. v. 3, n. 2, 1992.

ZIPF, G.K. **Human Behavior and the Principle of Least Effort**. Cambridge, MA: Addison-Wesley, 1949.

APÊNDICES

APÊNDICE A - Lista de termos atribuídos pelo SISA

Artigo 01

TÉRMINOS DE INDIZACIÓN.

- 1.- TRANSFERENCIA DA INFORMACÃO
- 2.- AVALIAÇÃO
- 3.- ACESSO
- 4.- DESCARTE
- 5.- GESTÃO
- 6.- GESTÃO DO CONHECIMENTO

Artigo 02

TÉRMINOS DE INDIZACIÓN.

- 1.- COMUNICAÇÃO CIENTIFICA
- 2.- CIENCIA DA INFORMACÃO
- 3.- EDUCACÃO
- 4.- NOTICIAS

Artigo 03

TÉRMINOS DE INDIZACIÓN.

- 1.- AVALIAÇÃO
- 2.- DIREITO
- 3.- DIREITO A INFORMACÃO
- 4.- RECUPERAÇÃO DA INFORMACÃO

Artigo 04

TÉRMINOS DE INDIZACIÓN.

- 1.- ACESSO
- 2.- ACESSO LIVRE
- 3.- CONCEITOS
- 4.- FILOSOFIA
- 5.- PUBLICAÇÕES

Artigo 05

TÉRMINOS DE INDIZACIÓN.

- 1.- ACESSO
- 2.- ESTUDANTES
- 3.- PESQUISA

Artigo 06

TÉRMINOS DE INDIZACIÓN.

- 1.- CIENCIA DA INFORMACÃO
- 2.- CIENCIA DA COMPUTAÇÃO
- 3.- NAVEGAÇÃO
- 4.- RECUPERAÇÃO DA INFORMACÃO
- 5.- UNIVERSIDADES

Artigo 07

TÉRMINOS DE INDIZACIÓN.

- 1.- INTELIGENCIA COMPETITIVA
- 2.- COMPETITIVIDADE
- 3.- DADOS
- 4.- GESTÃO
- 5.- GESTÃO DO CONHECIMENTO
- 6.- GESTÃO DA INFORMACÃO

Artigo 08

TÉRMINOS DE INDIZACIÓN.

- 1.- AREAS DO CONHECIMENTO
- 2.- CIENCIA DA INFORMACÃO
- 3.- PESQUISA

Artigo 09

TÉRMINOS DE INDIZACIÓN.

- 1.- CIENCIA DA INFORMACÃO
- 2.- DISSEMINAÇÃO SELETIVA DA INFORMACÃO
- 3.- FLUXO DA INFORMACÃO
- 4.- GESTÃO
- 5.- INTELIGENCIA COMPETITIVA
- 6.- RELEVANCIA

Artigo 10

TÉRMINOS DE INDIZACIÓN.

- 1.- ECONOMIA
- 2.- LEVANTAMENTOS
- 3.- PERTINENCIA

Artigo 11

TÉRMINOS DE INDIZACIÓN.

- 1.- CIENCIA DA INFORMACÃO
- 2.- PESQUISA
- 3.- CLASSIFICAÇÃO
- 4.- DADOS
- 5.- PROPRIEDADE INTELECTUAL

Artigo 12

TÉRMINOS DE INDIZACIÓN.

- 1.- CIENCIA DA INFORMACÃO
- 2.- PESQUISA
- 3.- ACESSO
- 4.- ACESSO A INFORMACÃO
- 5.- ARQUIVOLOGIA
- 6.- BIBLIOTECONOMIA

Artigo 13

TÉRMINOS DE INDIZACIÓN.

- 1.- FORMAÇÃO PROFISSIONAL
- 2.- ECONOMIA
- 3.- PAPEL
- 4.- SISTEMAS DE INFORMAÇÃO

Artigo 14
TÉRMINOS DE INDIZACIÓN.

- 1.- PROFISSIONAIS DE INFORMAÇÃO
- 2.- ACESSO
- 3.- GESTORES
- 4.- GESTÃO
- 5.- GESTÃO DA INFORMAÇÃO
- 6.- GESTÃO DO CONHECIMENTO
- 7.- INCLUSÃO DIGITAL
- 8.- FORMAÇÃO PROFISSIONAL

Artigo 15
TÉRMINOS DE INDIZACIÓN.

- 1.- FORMAÇÃO PROFISSIONAL
- 2.- EDUCAÇÃO
- 3.- PESQUISA

Artigo 16
TÉRMINOS DE INDIZACIÓN.

- 1.- CIENCIA DA INFORMAÇÃO
- 2.- EDUCAÇÃO
- 3.- BIBLIOTECONOMIA
- 4.- EDUCAÇÃO SUPERIOR
- 5.- FORMAÇÃO PROFISSIONAL
- 6.- SOCIEDADE DA INFORMAÇÃO

Artigo 17
TÉRMINOS DE INDIZACIÓN.

- 1.- AUTORIA
- 2.- COMPLEXIDADE
- 3.- CONTROLE BIBLIOGRAFICO
- 4.- INTERNET
- 5.- PROPRIEDADE INTELECTUAL
- 6.- TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO

Artigo 18
TÉRMINOS DE INDIZACIÓN.

- 1.- CATEGORIAS
- 2.- SOCIOLOGIA

Artigo 19

TÉRMINOS DE INDIZACIÓN.

- 1.- AVALIAÇÃO
- 2.- AMBIENTE ORGANIZACIONAL
- 3.- BIBLIOTECAS

- 4.- GESTÃO
- 5.- GESTÃO DO CONHECIMENTO
- 6.- PESQUISA
- 7.- RELEVANCIA

Artigo 20
TÉRMINOS DE INDIZACIÓN.

- 1.- CIENCIA DA INFORMAÇÃO
- 2.- AVALIAÇÃO
- 3.- DISSERTAÇÕES E TESES
- 4.- PROGRAMAS DE POS-GRADUAÇÃO
- 5.- PROGRAMAS DE POS-GRADUAÇÃO EM CIENCIA DA INFORMAÇÃO
- 6.- PUBLICAÇÕES
- 7.- PESQUISA

Artigo 21
TÉRMINOS DE INDIZACIÓN.

- 1.- AREAS DO CONHECIMENTO

Artigo 22
TÉRMINOS DE INDIZACIÓN.

- 1.- CIENCIA DA INFORMAÇÃO
- 2.- COMPLEXIDADE

Artigo 23
TÉRMINOS DE INDIZACIÓN.

- 1.- HISTORIA

Artigo 24
TÉRMINOS DE INDIZACIÓN.

- 1.- PESQUISA
- 2.- ECOLOGIA
- 3.- PESQUISADORES
- 4.- VALIDADE

Artigo 25
TÉRMINOS DE INDIZACIÓN.

- 1.- CIENCIA DA INFORMAÇÃO
- 2.- PESQUISA
- 3.- DADOS
- 4.- ORÇAMENTO
- 5.- PESQUISADORES

Artigo 26
TÉRMINOS DE INDIZACIÓN.

- 1.- MONITORAMENTO
- 2.- SEMANTICA
- 3.- CIENCIA DA INFORMAÇÃO
- 4.- GESTÃO
- 5.- SUBSIDIOS

Artigo 27

TÉRMINOS DE INDIZACIÓN.

- 1.- DADOS
- 2.- BASES DE DADOS
- 3.- GESTÃO
- 4.- INDICADORES
- 5.- MONITORAMENTO
- 6.- PESQUISADORES
- 7.- RELEVANCIA

Artigo 28

TÉRMINOS DE INDIZACIÓN.

- 1.- GESTÃO
- 2.- SISTEMAS DE INFORMAÇÃO
- 3.- BIBLIOTECAS
- 4.- BIBLIOTECAS DIGITAIS
- 5.- GESTÃO DA INFORMAÇÃO
- 6.- GOVERNO ELETRONICO
- 7.- INOVAÇÃO
- 8.- PAPEL

Artigo 29

TÉRMINOS DE INDIZACIÓN.

- 1.- GOVERNO ELETRONICO
- 2.- INFORMAÇÃO GOVERNAMENTAL
- 3.- POLITICAS DE INFORMAÇÃO
- 4.- ACESSO
- 5.- GESTÃO

Artigo 30

TÉRMINOS DE INDIZACIÓN.

- 1.- AVALIAÇÃO
- 2.- ACESSO
- 3.- PERIODICOS
- 4.- PERIODICOS ELETRONICOS
- 5.- WORLD WIDE WEB

Artigo 31

TÉRMINOS DE INDIZACIÓN.

- 1.- CENARIOS
- 2.- ECONOMIA
- 3.- INTERNET
- 4.- PAPEL
- 5.- SOCIEDADE DA INFORMAÇÃO

Artigo 32

TÉRMINOS DE INDIZACIÓN.

- 1.- BASES DE DADOS
- 2.- DADOS
- 3.- CD-ROM

4.- COMPLEXIDADE

5.- ESTRATEGIAS DE BUSCA

6.- PLANEJAMENTO

7.- RECUPERAÇÃO DA INFORMAÇÃO

8.- TERMOS

Artigo 33

TÉRMINOS DE INDIZACIÓN.

- 1.- BIBLIOTECAS
- 2.- BIBLIOTECAS VIRTUAIS
- 3.- PERIODICOS
- 4.- AUTORIA
- 5.- BIBLIOTECAS DIGITAIS
- 6.- COLEÇÕES
- 7.- DESENVOLVIMENTO DE COLEÇÕES
- 8.- INTERNET
- 9.- PLANEJAMENTO
- 10.- PROFISSIONAIS DE INFORMAÇÃO
- 11.- REFERENCIAS BIBLIOGRAFICAS

Artigo 34

TÉRMINOS DE INDIZACIÓN.

- 1.- ACESSO
- 2.- CIENCIAS DA SAUDE
- 3.- NUTRIÇÃO
- 4.- TRANSFERENCIA DA INFORMAÇÃO

Artigo 35

TÉRMINOS DE INDIZACIÓN.

- 1.- CONCEITOS
- 2.- INTERNET
- 3.- XML
- 4.- COMPLEXIDADE
- 5.- CIENCIA DA INFORMAÇÃO
- 6.- DADOS
- 7.- HTML
- 8.- LINGUAGENS DE MARCAÇÃO
- 9.- USUARIOS

Artigo 36

TÉRMINOS DE INDIZACIÓN.

- 1.- BIBLIOMETRIA
- 2.- LEI DE LOTKA
- 3.- BIBLIOTECONOMIA
- 4.- CARTAS
- 5.- DADOS
- 6.- MEDICINA

Artigo 37

TÉRMINOS DE INDIZACIÓN.

- 1.- BASES DE DADOS
- 2.- DADOS
- 3.- INFORMAÇÃO PARA NEGOCIOS
- 4.- CENARIOS
- 5.- CATEGORIAS

- 6.- CLASSIFICAÇÃO
- 7.- EMPRESAS
- 8.- INDÚSTRIA DA INFORMAÇÃO
- 9.- NOTÍCIAS

Artigo 38

TÉRMINOS DE INDIZACIÓN.

- 1.- EDUCAÇÃO
- 2.- BIBLIOTECAS
- 3.- BIBLIOTECAS HÍBRIDAS
- 4.- INTERNET
- 5.- PAPEL
- 6.- UNIVERSIDADES
- 7.- USUÁRIOS

Artigo 39

TÉRMINOS DE INDIZACIÓN.

- 1.- RECUPERAÇÃO DA INFORMAÇÃO
- 2.- ACESSO
- 3.- BASES DE DADOS
- 4.- COMPLEXIDADE
- 5.- CONCEITOS
- 6.- DADOS
- 7.- ESTRATÉGIAS DE BUSCA
- 8.- PLANEJAMENTO
- 9.- SISTEMAS DE RECUPERAÇÃO DA INFORMAÇÃO
- 10.- USUÁRIOS

Artigo 40

TÉRMINOS DE INDIZACIÓN.

- 1.- GESTÃO
- 2.- GESTÃO DA INFORMAÇÃO
- 3.- FORMAÇÃO PROFISSIONAL
- 4.- AVALIAÇÃO
- 5.- CIÊNCIA DA INFORMAÇÃO
- 6.- PROFISSIONAIS DE INFORMAÇÃO
- 7.- RECURSOS DE INFORMAÇÃO

Artigo 41

TÉRMINOS DE INDIZACIÓN.

- 1.- CIÊNCIA DA INFORMAÇÃO
- 2.- PROGRAMAS DE POS-GRADUAÇÃO
- 3.- PROGRAMAS DE POS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
- 4.- AUTORIA
- 5.- DADOS
- 6.- ÍNDICE
- 7.- LITERATURA CINZENTA
- 8.- PESQUISA
- 9.- PUBLICAÇÕES
- 10.- PERIÓDICOS

Artigo 42

TÉRMINOS DE INDIZACIÓN.

- 1.- GESTÃO
- 2.- INFORMAÇÃO PARA NEGÓCIOS
- 3.- BIBLIOTECAS
- 4.- BIBLIOTECAS VIRTUAIS
- 5.- EMPRESAS
- 6.- GESTÃO DO CONHECIMENTO
- 7.- GESTÃO DA INFORMAÇÃO
- 8.- INTELIGÊNCIA COMPETITIVA
- 9.- PROFISSIONAIS DE INFORMAÇÃO
- 10.- SISTEMAS DE INFORMAÇÃO

Artigo 43

TÉRMINOS DE INDIZACIÓN.

- 1.- NECESSIDADES DE INFORMAÇÃO
- 2.- EMPRESAS
- 3.- INOVAÇÃO
- 4.- PESQUISA

Artigo 44

TÉRMINOS DE INDIZACIÓN.

- 1.- COMPETITIVIDADE
- 2.- GESTÃO
- 3.- GESTÃO DO CONHECIMENTO
- 4.- PAPEL

Artigo 45

TÉRMINOS DE INDIZACIÓN.

- 1.- BIBLIOMETRIA
- 2.- AVALIAÇÃO
- 3.- CIÊNCIA DA INFORMAÇÃO
- 4.- INFORMETRIA
- 5.- INTERNET

Artigo 46

TÉRMINOS DE INDIZACIÓN.

- 1.- BIBLIOMETRIA
- 2.- PESQUISA
- 3.- CIÊNCIA DA INFORMAÇÃO
- 4.- LEI DE BRADFORD
- 5.- PERIÓDICOS
- 6.- TEORIA DO CAOS

Artigo 47

TÉRMINOS DE INDIZACIÓN.

- 1.- PERIÓDICOS
- 2.- PERIÓDICOS ELETRONICOS
- 3.- USUÁRIOS
- 4.- CIÊNCIA DA INFORMAÇÃO
- 5.- PERIÓDICOS CIENTÍFICOS
- 6.- WORLD WIDE WEB

Artigo 48

TÉRMINOS DE INDIZACIÓN.

- 1.- ECONOMIA
- 2.- ECONOMIA DA INFORMAÇÃO
- 3.- CUSTOS
- 4.- EMPRESAS
- 5.- INOVAÇÃO

Artigo 49
TÉRMINOS DE INDIZACIÓN.

- 1.- MONITORAMENTO
- 2.- DESCRITORES
- 3.- PESQUISA
- 4.- PERIODICOS

Artigo 50
TÉRMINOS DE INDIZACIÓN.

- 1.- FORMAÇÃO PROFISSIONAL
- 2.- BIBLIOTECARIOS
- 3.- COOPERAÇÃO
- 4.- EDUCAÇÃO
- 5.- PAPEL

Artigo 51
TÉRMINOS DE INDIZACIÓN.

- 1.- ACESSIBILIDADE
- 2.- BIBLIOTECAS
- 3.- CATEGORIAS
- 4.- USUARIOS

Artigo 52
TÉRMINOS DE INDIZACIÓN.

- 1.- ESTUDOS DE USUARIOS
- 2.- USUARIOS
- 3.- CIENCIA DA INFORMAÇÃO
- 4.- EDUCAÇÃO
- 5.- INTERNET
- 6.- PESQUISA
- 7.- PAPEL

Artigo53
TÉRMINOS DE INDIZACIÓN.

- 1.- PROFESSORES
- 2.- RECUPERAÇÃO DA INFORMAÇÃO
- 3.- USUARIOS

Artigo 54
TÉRMINOS DE INDIZACIÓN.

- 1.- FILOSOFIA
- 2.- ACESSO
- 3.- AVALIAÇÃO
- 4.- BIBLIOTECAS
- 5.- BIBLIOTECARIOS

- 6.- COMPETENCIA EM INFORMAÇÃO
- 7.- EDUCAÇÃO
- 8.- PAPEL

Artigo 55
TÉRMINOS DE INDIZACIÓN.

- 1.- MERCADO DE TRABALHO
- 2.- FORMAÇÃO PROFISSIONAL
- 3.- EMPRESAS

Artigo 56
TÉRMINOS DE INDIZACIÓN.

- 1.- CIENCIA DA INFORMAÇÃO
- 2.- CIENTISTAS
- 3.- CIENTISTAS DA INFORMAÇÃO
- 4.- COMUNICAÇÃO CIENTIFICA
- 5.- HISTORIA
- 6.- INFORMAÇÃO CIENTIFICA E TECNOLÓGICA
- 7.- PESQUISA
- 8.- RELEVANCIA
- 9.- SOCIOLOGIA

Artigo 57
TÉRMINOS DE INDIZACIÓN.

- 1.- AVALIAÇÃO
- 2.- CIENCIA DA INFORMAÇÃO
- 3.- PESQUISA
- 4.- RECUPERAÇÃO DA INFORMAÇÃO

Artigo 58
TÉRMINOS DE INDIZACIÓN.

- 1.- CIENCIA DA INFORMAÇÃO
- 2.- CIENCIA COGNITIVA
- 3.- INDEXAÇÃO
- 4.- INTERAÇÃO HOMEM-COMPUTADOR
- 5.- RECUPERAÇÃO DA INFORMAÇÃO

Artigo 59
TÉRMINOS DE INDIZACIÓN.

- 1.- EDUCAÇÃO
- 2.- BIBLIOMETRIA
- 3.- CIENTOMETRIA
- 4.- FRENTE DE PESQUISA
- 5.- PESQUISA

Artigo 60
TÉRMINOS DE INDIZACIÓN.

- 1.- AGENTES INTELIGENTES
- 2.- INTELIGENCIA COMPETITIVA
- 3.- INTERNET
- 4.- ESTUDOS DE CASO
- 5.- GESTÃO

6.- GESTÃO DA INFORMAÇÃO
7.- INOVAÇÃO

8.- MONITORAMENTO

APÊNDICE B - Lista de termos atribuídos pelo MAUI

Artigo 1

- 1.- gestão do conhecimento
- 2.- conhecimento nas organizações
- 3.- transferência da informação
- 4.- ciência da informação
- 5.- conhecimento tácito
- 6.- processos de gestão
- 7.- tecnologias da informação e comunicação

Artigo 2

- 1.- divulgação científica
- 2.- ciência da informação
- 3.- comunicação científica
- 4.- notícias
- 5.- cientistas
- 6.- comunicação
- 7.- informação científica e tecnológica
- 8.- resumos

Artigo 3

- 1.- direito à informação
- 2.- sociedade da informação
- 3.- recuperação da informação
- 4.- acesso à informação
- 5.- direito
- 6.- bibliotecas
- 7.- ciência da informação
- 8.- resumos

Artigo 4

- 1.- direito autoral
- 2.- acesso livre
- 3.- tecnologias da informação e comunicação
- 4.- publicações eletrônicas
- 5.- grupos de discussão
- 6.- direito
- 7.- acesso ao documento
- 8.- publicações impressas
- 9.- periódicos científicos
- 10.- arquivos

Artigo 5

- 1.- acesso ao documento
- 2.- informação científica e tecnológica

- 3.- sociedade da informação
- 4.- fitas sonoras
- 5.- fluxo da informação
- 6.- análise comparativa
- 7.- comunidades científicas
- 8.- educação
- 9.- pesquisa
- 10.- pesquisadores

Artigo 6

- 1.- recuperação da informação
- 2.- ciência da informação
- 3.- ciência da computação
- 4.- áreas do conhecimento
- 5.- tesouros
- 6.- bases de dados
- 7.- centros de análise de informação
- 8.- ciências humanas
- 9.- sistemas de informação
- 10.- universidades

Artigo 7

- 1.- inteligência competitiva
- 2.- gestão do conhecimento
- 3.- sociedade da informação
- 4.- recursos de informação
- 5.- ciência da informação
- 6.- gestão da informação
- 7.- sistemas de informação
- 8.- tecnologias da informação e comunicação
- 9.- tomada de decisões
- 10.- dados

Artigo 8

- 1.- estados da arte
- 2.- ciência da informação
- 3.- conceitos de informação
- 4.- áreas do conhecimento
- 5.- tipos de documento
- 6.- indústria da informação
- 7.- processamento de informações
- 8.- comunicação científica
- 9.- artigos de periódico
- 10.- pesquisa

Artigo 9

- 1.- tomada de decisões
- 2.- fluxo da informação
- 3.- comunicação informal
- 4.- inteligência competitiva
- 5.- universidades
- 6.- ciência da informação
- 7.- sociedade da informação
- 8.- disseminação seletiva da informação
- 9.- gestão da informação
- 10.- usuários

Artigo 10

- 1.- sociedade da informação
- 2.- sistemas de informação
- 3.- tecnologias da informação e comunicação
- 4.- probabilidade e estatística
- 5.- economia
- 6.- processamento de informações
- 7.- pesquisa e desenvolvimento
- 8.- imagens
- 9.- computadores pessoais
- 10.- comunicação

Artigo 11

- 1.- ciência da informação
- 2.- propriedade intelectual
- 3.- protocolos de comunicação
- 4.- dados
- 5.- pesquisa
- 6.- pesquisadores

Artigo 12

- 1.- ciência da informação
- 2.- profissionais de informação
- 3.- acesso à informação
- 4.- informação governamental
- 5.- recuperação da informação
- 6.- áreas do conhecimento
- 7.- sistemas de informação
- 8.- biblioteconomia
- 9.- disseminação da informação
- 10.- buscas de informação

Artigo 13

- 1.- profissionais de informação
- 2.- formação profissional
- 3.- sistemas de informação
- 4.- ciência da informação
- 5.- disseminação da informação
- 6.- acesso à informação
- 7.- tecnologias da informação e comunicação
- 8.- periódicos científicos
- 9.- sociedade da informação
- 10.- competências profissionais

Artigo 14

- 1.- competência em informação
- 2.- profissionais de informação
- 3.- gestão da informação
- 4.- engenharia do conhecimento
- 5.- gestão do conhecimento
- 6.- necessidades de informação
- 7.- sociedade da informação
- 8.- gestores
- 9.- inclusão digital
- 10.- ciência da informação

Artigo 15

- 1.- profissionais de informação
- 2.- formação profissional
- 3.- educação superior
- 4.- ciência da informação
- 5.- educação
- 6.- pesquisa
- 7.- pesquisadores
- 8.- universidades
- 9.- paradigmas
- 10.- sociedade da informação

Artigo 16

- 1.- ciência da informação
- 2.- profissionais de informação
- 3.- formação profissional
- 4.- sociedade da informação
- 5.- educação superior
- 6.- biblioteconomia
- 7.- arquitetura de informação
- 8.- arquitetos de informação
- 9.- ensino de biblioteconomia
- 10.- educação

Artigo 17

- 1.- publicações eletrônicas
- 2.- propriedade intelectual
- 3.- controle bibliográfico
- 4.- tecnologias da informação e comunicação
- 5.- inovação
- 6.- sociedade da informação
- 7.- correio eletrônico
- 8.- comunicação científica
- 9.- acesso universal
- 10.- periódicos científicos

Artigo 18

- 1.- dissertações e teses
- 2.- bibliografias
- 3.- mudança
- 4.- comunidades científicas
- 5.- sistemas de organização do conhecimento
- 6.- pesquisa
- 7.- pesquisadores
- 8.- construtivismo
- 9.- resumos
- 10.- cenários

Artigo 19

- 1.- ambiente organizacional
- 2.- inteligência competitiva
- 3.- monitoramento ambiental
- 4.- gestão da informação
- 5.- bibliotecas centrais
- 6.- conhecimento nas organizações
- 7.- centros de análise de informação
- 8.- informação para negócios
- 9.- frequência de uso
- 10.- buscas de informação

Artigo 20

- 1.- ciência da informação
- 2.- áreas do conhecimento
- 3.- projetos de pesquisa
- 4.- estudos de caso
- 5.- pesquisa
- 6.- pesquisadores
- 7.- conselhos editoriais
- 8.- publicações

- 9.- dissertações e teses
- 10.- anais de congressos

Artigo 21

- 1.- áreas do conhecimento
- 2.- inteligência artificial
- 3.- bibliotecas nacionais
- 4.- sociologia
- 5.- ciências humanas
- 6.- ciência da informação

Artigo 22

1. ciência da informação
2. paradigmas
- 3.- processamento de informações
- 4.- World Wide Web
- 5.- tecnologias da informação e comunicação
- 6.- conceitos de informação
- 7.- áreas do conhecimento
- 8.- organização do conhecimento
- 9.- hipertextos
- 10.- hipertexto

Artigo 23

- 1.- história
- 2.- sociedade da informação
- 3.- conceitos de informação
- 4.- políticas públicas
- 5.- medicina
- 6.- suportes de informação

Artigo 24

- 1.- áreas do conhecimento
- 2.- pesquisa e desenvolvimento
- 3.- políticas públicas
- 4.- pesquisa
- 5.- pesquisadores
- 6.- organização do conhecimento
- 7.- usos da informação

Artigo 25

- 1.- ciência da informação
- 2.- agências de fomento
- 3.- ciências sociais aplicadas
- 4.- projetos de pesquisa

- 5.- bolsas de pesquisa
- 6.- pesquisa
- 7.- pesquisadores
- 8.- áreas do conhecimento
- 9.- comunidades científicas
- 10.- dados

Artigo 26

- 1.- ciência da informação
- 2.- monitoramento
- 3.- gestão da informação
- 4.- semântica
- 5.- transferência da informação
- 6.- recuperação da informação
- 7.- fluxo da informação
- 8.- instrumentalismo
- 9.- subsídios

Artigo 27

- 1.- bolsas de pesquisa
- 2.- bases de dados
- 3.- agências de fomento
- 4.- índices
- 5.- índices
- 6.- áreas do conhecimento
- 7.- pesquisadores
- 8.- pesquisa
- 9.- indicadores
- 10.- indicadores de C&T

Artigo 28

- 1.- sistemas de informação
- 2.- gestão da informação
- 3.- bibliotecas digitais
- 4.- necessidades de informação
- 5.- governo eletrônico
- 6.- informação governamental
- 7.- tomada de decisões
- 8.- revisões de literatura
- 9.- bibliotecas
- 10. dissertações e teses

Artigo 29

- 1.- governo eletrônico
- 2.- informação governamental
- 3.- gestão da informação
- 4.- políticas de informação

- 5.- sociedade da informação
- 6.- serviços de informação
- 7.- acesso
- 8.- políticas públicas
- 9.- gestão do conhecimento
- 10.- ciência da informação

Artigo 30

- 1.- periódicos eletrônicos
- 2.- World Wide Web
- 3.- acesso
- 4.- periódicos
- 5.- arquivos
- 6.- associações de elaboração de normas
- 7.- ciência da informação
- 8.- artigos de periódico
- 9.- usuários
- 10.- avaliação

Artigo 31

- 1.- regimes de informação
- 2.- políticas de informação
- 3.- sociedade da informação
- 4.- comunicação informal
- 5.- informação científica e tecnológica
- 6.- políticas públicas
- 7.- ciência da informação
- 8.- serviços de informação
- 9.- cenários
- 10.- indústria da informação

Artigo 32

- 1.- bases de dados
- 2.- estratégias de busca
- 3.- recuperação da informação
- 4.- linguagens documentárias
- 5.- planejamento estratégico
- 6.- buscas de informação
- 7.- bases de dados em linha
- 8.- estudos de uso
- 9.- sistemas de recuperação da informação
- 10.- termos de indexação

Artigo 33

- 1.- bibliotecas digitais
- 2.- bibliotecas virtuais

- 3.- profissionais de informação
- 4.- artigos de periódico
- 5.- desenvolvimento de coleções
- 6.- produtividade de autor
- 7.- tipos de documento
- 8.- bibliotecas
- 9.- ciência da informação
- 10.- referências bibliográficas

Artigo 34

- 1.- comunicação informal
- 2.- ciências da saúde
- 3.- informação científica e tecnológica
- 4.- transferência da informação
- 5.- necessidades de informação
- 6.- nutrição
- 7.- universidades
- 8.- pesquisa e desenvolvimento
- 9.- ciência da informação
- 10.- institutos de pesquisa

Artigo 35

- 1.- linguagens de marcação
- 2.- HTML
- 3.- ciência da informação
- 4.- padrões de metadados
- 5.- XML
- 6.- dados
- 7.- áreas do conhecimento
- 8.- SGML
- 9.- recuperação da informação
- 10.- ciência da computação

Artigo 36

- 1.- lei de Lotka
- 2.- arquivos privados
- 3.- produtividade de autor
- 4.- cartas
- 5.- bibliometria
- 6.- faculdades de biblioteconomia
- 7.- bases de dados
- 8.- ciência da informação
- 9.- arquivos
- 10.- análise de dados

Artigo 37

- 1.- informação para negócios
- 2.- bases de dados
- 3.- informação financeira
- 4.- informação jurídica
- 5.- acesso à informação
- 6.- competência no uso de computadores
- 7.- informação científica e tecnológica
- 8.- tomada de decisões
- 9.- indústria da informação
- 10.- empresas

Artigo 38

- 1.- ensino a distância
- 2.- bibliotecas híbridas
- 3.- bibliotecas
- 4.- tecnologias da informação e comunicação
- 5.- bibliotecas universitárias
- 6.- usuários
- 7.- serviços de biblioteca
- 8.- bibliotecas digitais
- 9.- necessidades de informação
- 10.- bases de dados

Artigo 39

- 1.- estratégias de busca
- 2.- bases de dados
- 3.- recuperação da informação
- 4.- planejamento estratégico
- 5.- sistemas de recuperação da informação
- 6.- buscas de informação
- 7.- bases de dados em linha
- 8.- linguagens documentárias
- 9.- mecanismos de busca
- 10.- buscas

Artigo 40

- 1.- gestão da informação
- 2.- recursos de informação
- 3.- profissionais de informação
- 4.- acesso à informação
- 5.- tomada de decisões
- 6.- ciência da informação

- 7.- necessidades de informação
- 8.- tecnologias da informação e comunicação
- 9.- mercado de trabalho
- 10.- serviços de informação

Artigo 41

- 1.- ciência da informação
- 2.- artigos de periódico
- 3.- literatura cinzenta
- 4.- autoria individual
- 5.- institutos de pesquisa
- 6.- sociedade da informação
- 7.- comunidades científicas
- 8.- tipos de documento
- 9.- bases de dados
- 10.- coleta de dados

Artigo 42

- 1.- sistemas de informação
- 2.- informação para negócios
- 3.- gestão do conhecimento
- 4.- centros de análise de informação
- 5.- bibliotecas virtuais
- 6.- inteligência competitiva
- 7.- bibliotecas de empresas
- 8.- acesso à informação
- 9.- gestão da informação
- 10.- profissionais de informação

Artigo 43

- 1.- necessidades de informação
- 2.- informação científica e tecnológica
- 3.- tomada de decisões
- 4.- conhecimento nas organizações
- 5.- estados da arte
- 6.- estudos de usuários
- 7.- empresas
- 8.- inovação
- 9.- estudos de caso
- 10.- ciência da informação

Artigo 44

- 1.- gestão do conhecimento

- 2.- conhecimento nas organizações
- 3.- tecnologias da informação e comunicação
- 4.- empresas
- gestão
- 6.- competitividade
- 7.- tomada de decisões
- 8.- bases de dados
- 9.- conhecimento tácito
- 10.- resumos

Artigo 45

- 1.- ciência da informação
- 2.- fluxo da informação
- 3.- bibliometria
- 4.- bases de dados
- 5.- áreas do conhecimento
- 6.- World Wide Web
- 7.- produtividade de autor
- 8.- informetria
- 9.- tomada de decisões
- 10.- comunicação

Artigo 46

- 1.- teoria do caos
- 2.- lei de Bradford
- 3.- ciência da informação
- 4.- revisões de literatura
- 5.- bibliometria
- 6.- faculdades de biblioteconomia
- 7.- publicações seriadas
- 8.- modelos matemáticos
- 9.- unidades de informação
- 10.- periódicos

Artigo 47

- 1.- periódicos eletrônicos
- 2.- periódicos científicos
- 3.- ciência da informação
- 4.- World Wide Web
- 5.- periódicos
- 6.- usuários
- 7.- obras de referência
- 8.- artigos de periódico
- 9.- acesso
- 10.- monitores de vídeo

Artigo 48

- 1.- economia da informação
- 2.- usos da informação
- 3.- economia
- 4.- informação para negócios
- 5.- processamento de informações
- 6.- tomada de decisões
- 7.- empresas
- 8.- sistemas de informação
- 9.- custos
- 10.- tecnologias da informação e comunicação

Artigo 49

- 1.- descritores
- 2.- bases de dados
- 3.- periódicos
- 4.- monitoramento
- 5.- informação científica e tecnológica
- 6.- tomada de decisões
- 7.- inteligência competitiva
- 8.- títulos de documentos
- 9.- buscas
- 10.- mapas

Artigo 50

- 1.- formação profissional
- 2.- sociedade da informação
- 3.- educação
- 4.- tomada de decisões
- 5.- ciência da informação
- 6.- bibliotecários
- 7.- profissionais de informação
- 8.- tecnologias da informação e comunicação
- 9.- educação básica
- 10.- papel

Artigo 51

- 1.- acessibilidade
- 2.- digitalização
- 3.- tecnologias da informação e comunicação
- 4.- acesso à informação
- 5.- sítios web
- 6.- bibliotecas
- 7.- usuários
- 8.- suportes de informação
- 9.- sociedade da informação

- 10.- sistemas de informação

Artigo 52

- 1.- ciência da informação
- estudos de usuários
- 3.- políticas de informação
- 4.- projetos de pesquisa
- 5.- educação
- 6.- fluxo da informação
- 7.- transferência da informação
- 8.- comunicação informal
- 9.- papel
- 10.- hipertextos

Artigo 53

- 1.- recuperação da informação
- 2.- comportamento do usuário
- 3.- mecanismos de busca
- 4.- buscas
- 5.- usuários
- 6.- buscas booleanas
- 7.- acesso

Artigo 54

- 1.- competência em informação
- 2.- sociedade da informação
- 3.- acesso à informação
- 4.- buscas de informação
- tomada de decisões
- profissionais de informação
- bibliotecários
- educação
- usos da informação
- serviços de informação

Artigo 55

- 1.- profissionais de informação
- 2.- mercado de trabalho
- 3.- ciência da informação
- 4.- conhecimento nas organizações
- 5.- gestão do conhecimento
- 6.- gestão da informação
- 7.- empresas
- 8.- processos de gestão
- 9.- tomada de decisões

10.- sociedade da informação

Artigo 56

- 1.- ciência da informação
- 2.- informação científica e tecnológica
- 3.- cientistas da informação
- 4.- comunicação informal
- 5.- teorias na ciência da informação
- 6.- comunicação científica
- 7.- cientistas
- 8.- conceitos de informação
- 9.- sociologia
- 10.- história

Artigo 57

- 1.- ciência da informação
- 2.- recuperação da informação
- 3.- sociedade da informação
- 4.- políticas de informação
- 5.- comunicação informal
- 6.- informação científica e tecnológica
- 7.- tecnologias da informação e comunicação
- 8.- usos da informação
- 9.- buscas de informação
- 10.- transferência da informação

Artigo 58

- 1.- recuperação da informação
- 2.- processamento de informações
- 3.- ciência da informação

- 4.- ciência cognitiva
- 5.- sistemas de informação
- 6.- buscas de informação
- 7.- ciência da computação
- 8.- necessidades de informação
- 9.- teoria da informação
- 10.- aplicações de computador

Artigo 59

- 1.- frente de pesquisa
- 2.- ciência da informação
- 3.- educação
- 4.- referências bibliográficas
- 5.- áreas do conhecimento
- 6.- literatura científica
- 7.- bibliometria
- 8.- periódicos científicos
- 9.- comunicação
- 10.- páginas da web

Artigo 60

- 1.- inteligência competitiva
- 2.- engenharia de produção
- 3.- agentes inteligentes
- 4.- usos da informação
- 5.- gestão da informação
- 6.- estudos de caso
- 7.- gestão do conhecimento
- 8.- tomada de decisões
- 9.- tecnologias da informação e comunicação
- 10.- conceitos de informação

APÊNDICE C - Folksonomia assistida utilizada na pesquisa

Artigo 1

1. Transferência da informação
2. Gestão do conhecimento
3. Conhecimento nas organizações
4. Conhecimento tácito
5. Processos de gestão
6. Gestão da informação
7. Acesso à informação
8. Competitividade
9. Gestão de documentos
10. Repositórios digitais
11. Disseminação da informação

Artigo 2

1. Divulgação Científica
2. Comunicação Científica
3. Ciência da Informação
4. Notícias
5. Educação
6. Informação Científica e Tecnológica
7. Acesso à Informação
8. Áreas do Conhecimento
9. Gestão da Informação
10. Sociedade da Informação

Artigo 3

1. Direito à Informação
2. Recuperação da Informação
3. Sociedade da Informação
4. Acesso à Informação
5. Ciência da Informação
6. Atributos da Informação
7. Ética na Informação
8. Biblioteconomia
9. História
10. Liberdade de Pensamento

Artigo 4

1. Acesso Livre
2. Autoarquivamento de Documentos
3. Comunicação Científica
4. Publicações Científicas
5. Publicações Eletrônicas
6. Periódicos Científicos
7. Direito Autoral
8. Literatura Científica
9. Repositórios Digitais
10. Tecnologias da Informação e Comunicação

Artigo 5

1. Divulgação Científica
2. Comunicação Científica
3. Informação Científica e Tecnológica
4. Acesso à Informação
5. Sociedade da Informação
6. Educação
7. Estudantes
8. Áreas do Conhecimento
9. Comunidades Científicas
10. Dados Científicos

Artigo 6

1. Interdisciplinaridade
2. Ciência da Computação
3. Ciência da Informação
4. Manutenção de Tesouros
5. Tesouros
6. Elaboração de Linguagens Documentárias
7. Recuperação da Informação
8. Interação Homem-Computador
9. Sistemas de Informação
10. Mercado de Trabalho

Artigo 7

1. Inteligência Competitiva
2. Gestão do Conhecimento
3. Gestão da Informação
4. Fluxo da Informação
5. Transferência da Informação
6. Competitividade
7. Ciência da Informação
8. Economia da Informação
9. Informação para negócios
10. Conhecimento nas organizações
11. Recursos de Informação

Artigo 8

1. Conceitos de Informação
2. Ciência da Informação
3. Áreas do Conhecimento
4. Tecnologias da Informação e Comunicação
5. Indústria da Informação
6. Conteúdos da Informação
7. Disseminação da Informação
8. Explosão da Informação
9. Gestão de Documentos
10. Gestão do Conhecimento
11. Modelagem do Conhecimento
12. Registros Bibliográficos
13. Sociedade da Informação
14. Tipos de Documento

Artigo 9

1. Fluxo da Informação
2. Universidades
3. Inteligência Competitiva
4. Gestão da Informação
5. Tomada de Decisões
6. Ciência da Informação
7. Tecnologias da Informação e Comunicação
8. Comunicação nas Organizações
9. Cultura Organizacional
10. Informação Estratégica
11. Obsolescência da Literatura
12. Comunicação Informal

Artigo 10

1. Economia
2. Métodos Matemáticos e Estatísticos
3. Probabilidade e Estatística
4. Levantamentos
5. Dados Numéricos
6. Informação Governamental
7. Sociedade da Informação
8. Tecnologias da Informação E Comunicação
9. Atributos da Informação
10. Inteligência Competitiva
11. Sociedade do Conhecimento
12. Sistemas de Informação

Artigo 11

1. Ciência da Informação
2. Ensino de Ciência da Informação
3. Recuperação da Informação
4. Armazenamento de Dados
5. Educação
6. Pesquisa
7. Gestão da Informação
8. Interdisciplinaridade
9. Propriedade Intelectual
10. Unidades de Informação

Artigo 12

1. Ciência da Informação
2. Biblioteconomia
3. Arquivologia
4. Sistemas de Informação
5. Ensino de Ciência da Informação
6. Profissionais de Informação
7. Acesso à Informação
8. Educação
9. Pesquisa
10. Teoria da Informação

Artigo 13

1. Formação Profissional
2. Profissionais de Informação
3. Ciência da Informação
4. Competências Profissionais

5. Conhecimento nas Organizações
6. Disseminação da Informação
7. Economia
8. Gestão do Conhecimento
9. Sociedade da Informação
10. Sistemas de Informação
11. Bibliotecas
12. Biblioteconomia
13. Profissão e Mercado de Trabalho

Artigo 14

1. Profissionais de Informação
2. Formação Profissional
3. Competências Profissionais
4. Gestão da Informação
5. Gestão do Conhecimento
6. Sociedade da Informação
7. Mercado de Trabalho
8. Fluxo da Informação
9. Inclusão Digital
10. Competência em Informação
11. Engenharia do Conhecimento
12. Profissão e Mercado de Trabalho

Artigo 15

1. Profissionais de Informação
2. Formação Profissional
3. Competências Profissionais
4. Ensino de Ciência da Informação
5. Ensino e Pesquisa em Ciência da Informação e Áreas Afins
6. Educação Superior
7. Pesquisa
8. Ciência da Informação
9. Paradigmas
10. Instituições de Ensino e Pesquisa
11. Sociedade do Conhecimento

Artigo 16

1. Ciência da Informação
2. Educação Superior

3. Formação Profissional
4. Biblioteconomia
5. Sociedade da Informação
6. Ensino de Ciência da Informação
7. Ensino de Biblioteconomia
8. Ensino e Pesquisa em Ciência Da Informação e Áreas Afins
9. Profissionais de Informação
10. Competências Profissionais
11. Cientistas da Informação

Artigo 17

1. Comunicação Científica
2. Publicações Eletrônicas
3. Internet
4. Tecnologias da Informação e Comunicação
5. Inovação
6. Sociedade da Informação
7. Produtividade Científica
8. Armazenamento de Dados
9. Controle Bibliográfico
10. Propriedade Intelectual
11. Segurança da Informação

Artigo 18

1. Sociedade da Informação
2. Tecnologias da Informação e Comunicação
3. Pesquisa
4. Divulgação Científica
5. Categorias
6. Construtivismo
7. Taxonomias
8. Ciência da Informação
9. Economia da Informação
10. Escolas e Correntes Filosóficas
11. Produtividade Científica
12. Relações entre Conceitos
13. Sociedades Científicas
14. Sociologia do Conhecimento
15. Sociologia

Artigo 19

1. Inteligência Competitiva
2. Monitoramento Ambiental
3. Gestão da Informação
4. Gestão do Conhecimento
5. Ambiente Organizacional
6. Conhecimento nas Organizações
7. Informação para Negócios
8. Recursos De Informação
9. Bases de Dados
10. Bibliotecas Especializadas
11. Competitividade
12. Relevância
13. Tecnologias da Informação e Comunicação

Artigo 20

1. Ciência Da Informação
2. Programas De Pós-Graduação Em Ciência Da Informação
3. Pesquisa Em Ciência Da Informação
4. Projetos De Pesquisa
5. Avaliação De Publicações Científicas
6. Dissertações e Teses
7. Ensino de Ciência da Informação
8. Profissionais de Informação
9. Programas de Pós-Graduação
10. Agências de Fomento
11. Publicações

Artigo 21

1. Sociedade da Informação
2. Promoção do Livro e da Leitura
3. Ciência Cognitiva
4. Inteligência Artificial
5. Linguística
6. Sociologia
7. Multidisciplinaridade
8. Áreas do Conhecimento
9. Usuários

Artigo 22

1. Paradigmas
2. Ciência da Informação
3. Tecnologias da Informação e Comunicação
4. Hipertextos
5. Complexidade
6. Organização da Informação
7. Processamento de Informações

Artigo 23

1. Disseminação da Informação
2. Ciência da Informação
3. Documentação
4. História
5. Sociedade da Informação
6. Interdisciplinaridade
7. Relações entre Conceitos
8. Teoria da Informação
9. Tipos de Documento
10. Conceitos de Informação
11. Bibliotecas de Prisões

Artigo 24

1. Pesquisa
2. Pesquisadores
3. Políticas Públicas
4. Comunicação Científica
5. Informação Científica e Tecnológica
6. Institutos de Pesquisa
7. Universidades
8. Paradigmas
9. Transdisciplinaridade
10. Áreas do Conhecimento

Artigo 25

1. Agências de Fomento
2. Ciência da Informação
3. Bolsas de Pesquisa
4. Projetos de Pesquisa
5. Gestão de Recursos Humanos
6. Pesquisa
7. Pesquisadores
8. Pesquisa e Desenvolvimento
9. Programas De Pós-Graduação em Ciência da Informação

10. Levantamentos
11. Dados Numéricos
12. Informação Financeira
13. Orçamento

Artigo 26

1. Monitoramento
2. Palavras-Chaves
3. Semântica
4. Compressão de Dados
5. Informação Estratégica
6. Recuperação da Informação
7. Processamento de Textos
8. Processamento da Linguagem Natural
9. Relevância

Artigo 27

1. Indicadores de C&T
2. Bolsas de Pesquisa
3. Agências de Fomento
4. Pesquisadores
5. Bases de Dados
6. Bases de Dados Referenciais
7. Análise de Dados
8. Produtividade Científica
9. Projetos de Pesquisa

Artigo 28

1. Sistemas de Informação
2. Governo Eletrônico
3. Serviços de Informação
4. Gestão da Informação
5. Unidades de Informação
6. Bibliotecas Digitais
7. Inovação
8. Tecnologias Da Informação E Comunicação
9. Interoperabilidade

Artigo 29

1. Governo Eletrônico
2. Políticas de Informação
3. Informação Governamental
4. Acesso à Informação
5. Gestão da Informação
6. Gestão de Conteúdos na Web

7. Serviços de Informação

Artigo 30

1. Análise de Dados
2. Periódicos Eletrônicos
3. Periódicos Científicos
4. Artigos de Periódico
5. World Wide Web
6. Registros de Uso
7. Coleta de Dados
8. Estudos de Usuários
9. Perfil do Usuário
10. Usuários
11. Disseminação da Informação
12. Processos de Gestão
13. Publicações de Acesso Livre
14. Recuperação da Informação

Artigo 31

1. Políticas de Informação
2. Sociedade da Informação
3. Internet
4. Infraestrutura de Informação
5. Economia
6. Economia da Informação
7. Governo Eletrônico
8. Regimes de Informação
9. Relações entre Conceitos

Artigo 32

1. Bases de Dados
2. Estratégias de Busca
3. Linguagens Documentárias
4. Recuperação da Informação
5. Revisões de Literatura
6. Controle de Vocabulário
7. Processamento da Linguagem Natural
8. Termos de Indexação

Artigo 33

1. Bibliotecas Digitais
2. Bibliotecas Virtuais
3. Artigos de Periódico

4. Citações Bibliográficas
5. Informação Científica e Tecnológica
6. Produtividade Científica
7. Produtividade de Periódicos
8. Produtividade de Autor
9. Revisões de Literatura
10. Tecnologias da Informação e Comunicação
11. Profissionais de Informação
12. Bibliotecas
13. Internet

Artigo 34

1. Informação Científica e Tecnológica
2. Transferência da Informação
3. Inovação
4. Acesso à Informação
5. Ciência e Tecnologia de Alimentos
6. Nutrição
7. Comunicação Científica
8. Fluxo da Informação
9. Interação Universidade-Empresa
10. Pesquisa e Desenvolvimento
11. Tecnologias da Informação e Comunicação
12. Institutos de Pesquisa
13. Universidades

Artigo 35

1. Linguagens de Marcação
2. Internet
3. Intranetes
4. Xml
5. Html
6. Sgml
7. Ciência da Informação

Artigo 36

1. Bibliometria
2. Lei de Lotka
3. Produtividade de Autor
4. Disseminação da Informação
5. Literatura Científica
6. Métodos Matemáticos e Estatísticos

7. Bibliografias

Artigo 37

1. Informação para Negócios
2. Bases de Dados
3. Tecnologias da Informação E Comunicação
4. Tomada de Decisões
5. Acesso à Informação
6. Empresas
7. Recursos de Informação
8. Sociedade da Informação
9. Informação Financeira
10. Informação Jurídica
11. Bases de Dados Factuais
12. Indústria da Informação

Artigo 38

1. Bibliotecas Híbridas
2. Ensino a Distância
3. Usuários
4. Serviços de Informação
5. Acesso à Informação
6. Necessidades de Informação
7. Tecnologias da Informação e Comunicação
8. Internet
9. Bibliotecas Universitárias
10. Bibliotecas Digitais
11. Recursos de Informação
12. Serviços de Biblioteca
13. Suportes de Informação
14. Usos da Informação

Artigo 39

1. Estratégias de Busca
2. Recuperação da Informação
3. Bases de Dados
4. Revisões de Literatura
5. Sistemas de Recuperação da Informação
6. Mecanismos de Busca
7. Interdisciplinaridade

Artigo 40

1. Gestão da Informação
2. Ciência da Informação
3. Profissionais de Informação
4. Formação Profissional
5. Recursos de Informação
6. Mercado de Trabalho
7. Competências Profissionais
8. Tecnologias da Informação e Comunicação
9. Tomada de Decisões
10. Explosão da Informação
11. Gestão do Conhecimento
12. Necessidades de Informação

Artigo 41

1. Ciência da Informação
2. Programas de Pós-Graduação
3. Programas de Pós-Graduação Em Ciência da Informação
4. Literatura Científica
5. Literatura Cinzenta
6. Artigos de Periódico
7. Bibliometria
8. Produtividade Científica
9. Comunidades Científicas
10. História da Ciência da Informação
11. Publicações

Artigo 42

1. Informação para Negócios
2. Gestão do Conhecimento
3. Gestão da Informação
4. Sistemas de Informação
5. Consultores de Informação
6. Inteligência Competitiva
7. Bibliotecas
8. Bibliotecas Virtuais
9. Profissionais de Informação
10. Serviços de Biblioteca
11. Unidades de Informação
12. Empresas

Artigo 43

1. Informação Científica e Tecnológica
2. Inovação
3. Necessidades de Informação
4. Tomada de Decisões
5. Conhecimento nas Organizações
6. Empresas
7. Análise de Informação na Indústria
8. Cinco Forças de Porter

Artigo 44

1. Competitividade
2. Conhecimento nas Organizações
3. Gestão do Conhecimento
4. Informação Estratégica
5. Processos de Gestão
6. Ambiente Organizacional
7. Inovação
8. Tomada de Decisões
9. Tecnologias da Informação e Comunicação
10. Métodos de Análise na Inteligência Competitiva
11. Ensino e Pesquisa em Ciência Da Informação e Áreas Afins

Artigo 45

1. Bibliometria
2. Informetria
3. Cientometria
4. Webmetria
5. Análise Quantitativa
6. Métodos Matemáticos e Estatísticos
7. Bases de Dados
8. Ciência da Informação

Artigo 46

1. Bibliometria
2. Lei de Bradford
3. Ciência da Informação
4. Periódicos

5. Teoria do Caos
6. Modelos Matemáticos
7. Métodos Matemáticos e Estatísticos

Artigo 47

1. Periódicos Eletrônicos
2. Periódicos Científicos
3. Usuários
4. World Wide Web
5. Ciência da Informação
6. Usabilidade
7. Estudos de Usuários
8. Buscas de Informação
9. Buscas Em Linha
10. Indexação Automática
11. Usos da Informação
12. Economia da Informação

Artigo 48

1. Usos da Informação
2. Economia da Informação
3. Inovação
4. Inteligência Competitiva
5. Custos
6. Empresas
7. Gestão da Informação
8. Informação para Negócios
9. Tecnologias da Informação e Comunicação

Artigo 49

1. Monitoramento
2. Bibliometria
3. Biblioteconomia
4. Ciência da Informação
5. Periódicos
6. Descritores
7. Pesquisadores
8. Inteligência Competitiva
9. Gestão do Conhecimento
10. Bases de Dados
11. Periódicos Científicos

Artigo 50

1. Ensino de Biblioteconomia
2. Profissionais de Informação
3. Bibliotecários
4. Biblioteconomia
5. Formação Profissional
6. Educação
7. Sociedade da Informação
8. Gestão da Informação
9. Gestão do Conhecimento

Artigo 51

1. Acessibilidade
2. Bibliotecas
3. Acesso à Informação
4. Usuários
5. Perfil do Usuário
6. Comunicação Mediada por
7. Computador
8. Sistemas de Informação
9. Tecnologias da Informação e Comunicação
10. Inclusão Digital
11. Internet
12. Digitalização

Artigo 52

1. Estudos de Usuários
2. Ciência da Informação
3. Usuários
4. Internet
5. Projetos de Pesquisa
6. Educação
7. Hipertextos
8. Disseminação da Informação
9. Fluxo da Informação

Artigo 53

1. Comportamento do Usuário
2. Estratégias de Busca
3. Mecanismos de Busca

4. Recuperação da Informação
5. Linguagens Documentárias
6. Buscas de Informação
7. Serviços de Referência
8. Usuários

Artigo 54

1. Competência Em Informação
2. Bibliotecários
3. Bibliotecas
4. Sociedade da Informação
5. Educação
6. Buscas de Informação
7. Recursos de Informação
8. Gestão do Conhecimento
9. Usos da Informação
10. Profissionais de Informação

Artigo 55

1. Competências Profissionais
2. Profissionais de Informação
3. Mercado de Trabalho
4. Formação Profissional
5. Ensino de Ciência da Informação
6. Inteligência Competitiva
7. Ciência da Informação
8. Sistemas de Informação
9. Conhecimento das Organizações

Artigo 56

1. Teorias da Ciência da Informação
2. História da Ciência da Informação
3. Comunicação Científica
4. Ciência da Informação
5. Cientistas da Informação
6. Biblioteconomia
7. Informação Científica e Tecnológica
8. Interdisciplinaridade
9. Sociologia
10. Epistemologia da Ciência da Informação

Artigo 57

1. Ciência da Informação
2. Recuperação da Informação
3. Sociedade da Informação
4. Paradigmas
5. Regimes de Informação
6. Pesquisa
7. Áreas do Conhecimento
8. Ensino e Pesquisa em Ciência da Informação e Áreas Afins
9. Epistemologia da Ciência da Informação

Artigo 58

1. Ciência da Informação
2. Ciência Cognitiva
3. Processamento de Informações
4. Indexação
5. Recuperação da Informação
6. Interação Homem-Computador
7. Categorização Automática de Textos

Artigo 59

1. Comunicação Científica
2. Bibliometria
3. Cientometria
4. Análise de Citação
5. Educação
6. Comunicação
7. Frente de Pesquisa
8. Padrões de Comunicação Científica
9. Epistemologia Arqueológica
10. Produtividade Científica

Artigo 60

1. Agentes Inteligentes
2. Inteligência Competitiva
3. Internet
4. Monitoramento
5. Gestão da Informação
6. Acesso à Informação
7. Aplicações de Computador
8. Programas de Computador
9. Estudos de Caso

10. Inovação
11. Usos da Informação
12. Disseminação da Informação
13. Competitividade

Nº do artigo	Termos do SISA	Termo comum com folksonomia	Número de termos do SISA	Número de termos da folksonomia	Número de termos comuns com a folksonomia	Consistência	Precisão	Revocação	Medida F1
Artigo 1	TRANSFERÊNCIA DA INFORMAÇÃO GESTÃO DO CONHECIMENTO AVALIAÇÃO ACESSO DESCARTE	1 1 0 0 0	6	11	2	13%	33%	18%	24%
Artigo 2	COMUNICAÇÃO CIENTÍFICA CIÊNCIA DA INFORMAÇÃO EDUCAÇÃO NOTÍCIAS	1 1 1 1	4	10	4	40%	100%	40%	57%
Artigo 3	DIREITO A INFORMAÇÃO RECUPERAÇÃO DA INFORMAÇÃO AVALIAÇÃO DIREITO	1 1 0 0	4	10	2	17%	50%	20%	29%
Artigo 4	ACESSO LIVRE CIÊNCIA DA INFORMAÇÃO CONCEITOS FILOSOFIA PUBLICAÇÕES	1 0 0 0 0	5	10	1	7%	20%	10%	13%
Artigo 5	ESTUDANTES ACESSO PESQUISA	1 0 0	3	10	1	8%	33%	10%	15%
Artigo 6	CIÊNCIA DA INFORMAÇÃO CIÊNCIA DA COMPUTAÇÃO RECUPERAÇÃO DA INFORMAÇÃO NAVEGAÇÃO UNIVERSIDADES	1 1 1 0 0	5	10	3	25%	60%	30%	40%
Artigo 7	INTELIGÊNCIA COMPETITIVA COMPETITIVIDADE GESTÃO DO CONHECIMENTO GESTÃO DA INFORMAÇÃO DADOS	1 1 1 1 0	6	11	4	31%	67%	36%	47%
Artigo 8	AREAS DO CONHECIMENTO CIÊNCIA DA INFORMAÇÃO PESQUISA	1 1 0	3	14	2	13%	67%	14%	24%
Artigo 9	CIÊNCIA DA INFORMAÇÃO FLUXO DA INFORMAÇÃO INTELIGÊNCIA COMPETITIVA DISSEMINAÇÃO SELETIVA DA INFOF GESTÃO	1 1 1 0 0	6	12	3	20%	50%	25%	33%
Artigo 10	ECONOMIA LEVANTAMENTOS PERTINÊNCIA	1 1 0	3	12	2	15%	67%	17%	27%
Artigo 11	CIÊNCIA DA INFORMAÇÃO PESQUISA PROPRIEDADE INTELECTUAL CLASSIFICAÇÃO DADOS	1 1 1 0 0	5	10	3	25%	60%	30%	40%
Artigo 12	CIÊNCIA DA INFORMAÇÃO PESQUISA ACESSO À INFORMAÇÃO ARQUIVOLOGIA BIBLIOTECONOMIA	1 1 1 1 1	6	10	5	45%	83%	50%	63%
Artigo 13	FORMAÇÃO PROFISSIONAL ECONOMIA SISTEMAS DE INFORMAÇÃO PAPEL	1 1 1 0	4	13	3	21%	75%	23%	35%
Artigo 14	PROFISSIONAIS DE INFORMAÇÃO GESTÃO DA INFORMAÇÃO GESTÃO DO CONHECIMENTO INCLUSÃO DIGITAL FORMAÇÃO PROFISSIONAL	1 1 1 1 1	8	12	5	33%	63%	42%	50%
Artigo 15	FORMAÇÃO PROFISSIONAL PESQUISA EDUCAÇÃO	1 1 0	3	11	2	17%	67%	18%	29%
Artigo 16	CIÊNCIA DA INFORMAÇÃO BIBLIOTECONOMIA EDUCAÇÃO SUPERIOR FORMAÇÃO PROFISSIONAL SOCIEDADE DA INFORMAÇÃO	1 1 1 1 1	6	11	5	42%	83%	45%	59%
Artigo 17	CONTROLE BIBLIOGRÁFICO INTERNET PROPRIEDADE INTELECTUAL TECNOLOGIAS DA INFORMAÇÃO E C AUTORIA	1 1 1 1 0	6	11	4	31%	67%	36%	47%
Artigo 18	CATEGORIAS SOCIOLOGIA	1 1	2	15	2	13%	100%	13%	24%
Artigo 19	AMBIENTE ORGANIZACIONAL GESTÃO DO CONHECIMENTO RELEVÂNCIA AVALIAÇÃO BIBLIOTECAS	1 1 1 0 0	7	13	3	18%	43%	23%	30%
Artigo 20	CIÊNCIA DA INFORMAÇÃO DISSERTAÇÕES E TESES PROGRAMAS DE PÓS-GRADUAÇÃO PROGRAMAS DE PÓS-GRADUAÇÃO EM	1 1 1 1	7	11	5	38%	71%	45%	56%

Nº do artigo	Termos do SISA	Termo comum com folksonomia	Número de termos do SISA	Número de termos da folksonomia	Número de termos comuns com a folksonomia	Consistência	Precisão	Revocação	Medida F1
	PUBLICAÇÕES	1							
	ÁREAS DO CONHECIMENTO	1							
Artigo 21			1	9	1	11%	100%	11%	20%
Artigo 22	CIÊNCIA DA INFORMAÇÃO COMPLEXIDADE	1 1	2	7	2	29%	100%	29%	44%
Artigo 23	HISTÓRIA	1	1	11	1	9%	100%	9%	17%
Artigo 24	PESQUISA PESQUISADORES ECOLOGIA VALIDADE	1 1 0 0	4	10	2	17%	50%	20%	29%
Artigo 25	CIENCIA DA INFORMAÇÃO PESQUISA ORÇAMENTO PESQUISADORES DADOS	1 1 1 1 0	5	13	4	29%	80%	31%	44%
Artigo 26	MONITORAMENTO SEMANTICA CIENCIA DA INFORMAÇÃO GESTÃO SUBSIDIOS	1 1 0 0 0	5	9	2	17%	40%	22%	29%
Artigo 27	BASE DE DADOS PESQUISADORES DADOS GESTÃO INDICADORES	1 1 0 0 0	7	9	2	14%	29%	22%	25%
Artigo 28	SISTEMAS DE INFORMAÇÃO BIBLIOTECAS DIGITAIS GESTÃO DA INFORMAÇÃO GOVERNO ELETRÔNICO INOVAÇÃO	1 1 1 1 1	8	9	5	42%	63%	56%	59%
Artigo 29	GOVERNO ELETRÔNICO INFORMAÇÃO GOVERNAMENTAL POLITICAS DE INFORMAÇÃO ACESSO GESTÃO	1 1 1 0 0	5	7	3	33%	60%	43%	50%
Artigo 30	PERIODICOS ELETRONICOS WORLD WIDE WEB AVALIAÇÃO ACESSO PERIODICOS	1 1 0 0 0	5	14	2	12%	40%	14%	21%
Artigo 31	ECONOMIA INTERNET SOCIEDADE DA INFORMAÇÃO CENARIOS ECONOMIA	1 1 1 0 0	5	9	3	27%	60%	33%	43%
Artigo 32	BASES DE DADOS ESTRATEGIAS DE BUSCA RECUPERAÇÃO DA INFORMAÇÃO DADOS CD-ROM	1 1 1 0 0	8	8	3	23%	38%	38%	38%
Artigo 33	BIBLIOTECAS DIGITAIS BIBLIOTECAS VIRTUAIS BIBLIOTECAS INTERNET PROFISSIONAIS DE INFORMAÇÃO	1 1 1 1 1	11	13	5	26%	45%	38%	42%
Artigo 34	NUTRIÇÃO TRANSFERENCIA DA INFORMAÇÃO ACESSO CIENCIAS DA SAUDE	1 1 0 0	4	13	2	13%	50%	15%	24%
Artigo 35	INTERNET XML CIENCIA DA INFORMAÇÃO HTML LINGUAGENS DE MARCAÇÃO	1 1 1 1 1	9	7	5	45%	56%	71%	63%
Artigo 36	BIBLIOMETRIA LEI DE LOTKA BIBLIOTECONOMIA CARTAS DADOS	1 1 0 0 0	6	7	2	18%	33%	29%	31%
Artigo 37	BASES DE DADOS INFORMAÇÃO PARA NEGOCIOS EMPRESAS INDUSTRIA DA INFORMAÇÃO DADOS	1 1 1 1 0	9	12	4	24%	44%	33%	38%
Artigo 38	BIBLIOTECAS HIBRIDAS INTERNET USUARIOS EDUCAÇÃO BIBLIOTECAS HIBRIDAS	1 1 1 0 0	7	14	3	17%	43%	21%	29%
Artigo 39	RECUPERAÇÃO DA INFORMAÇÃO BASES DE DADOS ESTRATÉGIAS DE BUSCA SISTEMAS DE RECUPERAÇÃO DA INFOR ACESSO	1 1 1 1 0	10	7	4	31%	40%	57%	47%
	GESTÃO DA INFORMAÇÃO FORMAÇÃO PROFISSIONAL	1 1							

Nº do artigo	Termos do SISA	Termo comum com folksonomia	Número de termos do SISA	Número de termos da folksonomia	Número de termos comuns com a folksonomia	Consistência	Precisão	Revocação	Medida F1	
Artigo 40	CIENCIA DA INFORMACÃO PROFISSIONAIS DE INFORMACÃO RECURSOS DE INFORMACÃO	1 1 1	7	12	5	36%	71%	42%	53%	
Artigo 41	CIENCIA DA INFORMACÃO PROGRAMA DE PÓS-GRADUAÇÃO PROGRAMA DE PÓS-GRADUAÇÃO EM C LITERATURA CINZENTA PUBLICAÇÕES	1 1 1 1 1	10	11	5	31%	50%	45%	48%	
Artigo 42	INFORMACÃO PARA NEGOCIOS BIBLIOTECAS BIBLIOTECAS VIRTUAIS EMPRESAS GESTÃO DO CONHECIMENTO	1 1 1 1 1	10	12	9	69%	90%	75%	82%	*Outros termos comuns: 7.- GESTÃO DA INFORMACÃO 8.- INTELIGENCIA COMPETITIVA 9.- PROFISSIONAIS DE INFORMACÃO 10.- SISTEMAS DE INFORMACÃO
Artigo 43	NECESSIDADES DE INFORMACÃO EMPRESAS INOVACÃO PESQUISA	1 1 1 0	4	8	3	33%	75%	38%	50%	
Artigo 44	COMPETITIVIDADE GESTÃO DO CONHECIMENTO GESTÃO PAPEL	1 1 0 0	4	11	2	15%	50%	18%	27%	
Artigo 45	BIBLIOMETRIA CIENCIA DA INFORMACÃO INFORMETRIA AVALIACÃO INTERNET	1 1 1 0 0	5	8	3	30%	60%	38%	46%	
Artigo 46	BIBLIOMETRIA CIENCIA DA INFORMACÃO LEI DE BRADFORD PERIODICOS TEORIA DO CAOS	1 1 1 1 1	6	7	5	63%	83%	71%	77%	
Artigo 47	PERIODICOS ELETRONICOS USUARIOS CIENCIA DA INFORMACÃO PERIODICOS CIENTIFICOS WORD WIDE WEB	1 1 1 1 1	6	12	5	38%	83%	42%	56%	
Artigo 48	ECONOMIA DA INFORMACÃO CUSTOS EMPRESAS INOVACÃO ECONOMIA	1 1 1 1 0	5	9	4	40%	80%	44%	57%	
Artigo 49	MONITORAMENTO DESCRITORES PERIODICOS PESQUISA	1 1 1 0	4	11	3	25%	75%	27%	40%	
Artigo 50	FORMACÃO PROFISSIONAL BIBLIOTECARIOS EDUCACÃO COOPERACÃO PAPEL	1 1 1 0 0	5	9	3	27%	60%	33%	43%	
Artigo 51	ACESSIBILIDADE BIBLIOTECAS USUARIOS CATEGORIAS	1 1 1 0	4	11	3	25%	75%	27%	40%	
Artigo 52	ESTUDOS DE USUARIOS USUARIOS CIENCIA DA INFORMACÃO INTERNET EDUCACÃO	1 1 1 1 0	7	9	4	33%	57%	44%	50%	
Artigo 53	RECUPERACÃO DA INFORMACÃO USUARIOS PROFESSORES	1 1 0	3	8	2	22%	67%	25%	36%	
Artigo 54	BIBLIOTECAS BIBLIOTECARIOS EDUCACÃO FILOSOFIA ACESSO	1 1 1 0 0	8	10	3	20%	38%	30%	33%	
Artigo 55	MERCADO DE TRABALHO FORMACÃO PROFISSIONAL EMPRESAS	1 1 0	3	9	2	20%	67%	22%	33%	
Artigo 56	CIENCIA DA INFORMACÃO CIENTISTAS DA INFORMACÃO COMUNICACÃO CIENTIFICA INFORMACÃO CIENTIFICA E TECNOLOG SOCIOLOGIA	1 1 1 1 1	9	10	5	36%	56%	50%	53%	
Artigo 57	CIENCIA DA INFORMACÃO PESQUISA RECUPERACÃO DA INFORMACÃO AVALIACÃO	1 1 1 0	4	9	3	30%	75%	33%	46%	
Artigo 58	CIENCIA DA INFORMACÃO CIENCIA COGNITIVA INDEXACÃO INTERACÃO HOMEM-COMPUTADOR RECUPERACÃO DA INFORMACÃO	1 1 1 1 1	5	7	5	71%	100%	71%	83%	
Artigo 59	EDUCACÃO BIBLIOMETRIA CIENOMETRIA FONTE DE PESQUISA PESQUISA	1 1 1 1 0	5	10	4	36%	80%	40%	53%	

Nº do artigo	Termos do SISA	Termo comum com folksonomia	Número de termos do SISA	Número de termos da folksonomia	Número de termos comuns com a folksonomia	Consistência	Precisão	Revocação	Medida F1	
Artigo 60	AGENTES INTELIGENTES	1	8	13	7	50%	88%	54%	67%	*Outros termos comuns: 7.- INOVAÇÃO 8.- MONITORAMENTO
	INTELIGENCIA COMPETITIVA	1								
	INTERNET	1								
	ESTUDOS DE CASO	1								
GESTÃO DA INFORMAÇÃO	1									
	Mínimo		1,00	7,00	1,00	7%	20%	9%	13%	
	Média		5,55	10,35	3,35	28%	63%	34%	42%	
	Desvio padrão		2,27	2,05	9,00	14%	20%	16%	16%	
	Máximo		11,00	15,00		71%	100%	75%	83%	
	Número de zeros		0	0	0					
	Número de uns		2	0	4					
	Número de dois		2	0	16					
	Número de três		6	0	16					
	Número de quatro		10	0	9					
	Número de cinco		13	0	13					
	Número de seis ou		27	60	2					
			60	60	60					

Nº do artigo	Termos do MAUI	Termo comum com folksonomia	Número de termos do MAUI	Número de termos da folksonomia	Número de termos comuns com a folksonomia	Consistência	Precisão	Revocação	Medida F1
Artigo 01	gestão do conhecimento	1	7	11	5	38%	71%	45%	56%
	conhecimento nas organizações	1							
	transferência da informação	1							
	conhecimento tácito	1							
Artigo 02	processos de gestão	1	8	10	5	38%	63%	50%	56%
	divulgação científica	1							
	ciência da informação	1							
	comunicação científica	1							
Artigo 03	notícias	1	8	10	5	38%	63%	50%	56%
	informação científica e tecnológica	1							
	direito à informação	1							
	sociedade da informação	1							
Artigo 04	recuperação da informação	1	10	10	5	33%	50%	50%	50%
	acesso à informação	1							
	ciência da informação	1							
	direito autorai	1							
Artigo 05	tecnologias da informação e publicações eletrônicas	1	10	10	4	25%	40%	40%	40%
	periódicos científicos	1							
	informação científica e tecnológica	1							
	sociedade da informação	1							
Artigo 06	comunidades científicas	1	10	10	5	33%	50%	50%	50%
	educação	1							
	educação	0							
	recuperação da informação	1							
Artigo 07	ciência da informação	1	10	10	5	33%	50%	50%	50%
	ciência da computação	1							
	tesauros	1							
	recuperação da informação	1							
Artigo 08	inteligência competitiva	1	10	11	5	31%	50%	45%	48%
	gestão do conhecimento	1							
	recursos de informação	1							
	ciência da informação	1							
Artigo 09	gestão da informação	1	10	14	5	26%	50%	36%	42%
	ciência da informação	1							
	conceitos de informação	1							
	áreas do conhecimento	1							
Artigo 10	tipos de documento	1	10	12	7	47%	70%	58%	64%
	indústria da informação	1							
	tomada de decisões	1							
	fluxo da informação	1							
Artigo 11	comunicação informal	1	10	12	5	29%	50%	42%	45%
	inteligência competitiva	1							
	universidades	1							
	sociedade da informação	1							
Artigo 12	tecnologias da informação e comunicações	1	10	10	3	23%	50%	30%	38%
	probabilidade e estatística	1							
	economia	1							
	ciência da informação	1							
Artigo 13	propriedade intelectual	1	10	10	5	33%	50%	50%	50%
	pesquisa	1							
	protocolos de comunicação	0							
	dados	0							
Artigo 14	ciência da informação	1	10	13	7	44%	70%	54%	61%
	profissionais de informação	1							
	acesso à informação	1							
	sistemas de informação	1							
Artigo 15	biblioteconomia	1	10	12	7	47%	70%	58%	64%
	profissionais de informação	1							
	formação profissional	1							
	sistemas de informação	1							
Artigo 16	disseminação da informação	1	10	11	7	50%	70%	64%	67%
	competência em informação	1							
	profissionais de informação	1							
	gestão da informação	1							
Artigo 17	engenharia do conhecimento	1	10	11	7	50%	70%	64%	67%
	gestão do conhecimento	1							
	profissionais de informação	1							
	formação profissional	1							
Artigo 18	educação superior	1	10	15	2	9%	20%	13%	16%
	ciência da informação	1							
	educação superior	1							
	pesquisa	1							
Artigo 19	publicações eletrônicas	1	10	11	7	50%	70%	64%	67%
	propriedade intelectual	1							
	controle bibliográfico	1							
	tecnologias da informação e comunicações	1							
Artigo 20	inovação	1	10	15	2	9%	20%	13%	16%
	pesquisa	1							
	construtivismo	1							
	dissertações e teses	0							
Artigo 21	bibliografias	0	10	13	6	35%	60%	46%	52%
	mudança	0							
	ambiente organizacional	1							
	inteligência competitiva	1							
Artigo 22	monitoramento ambiental	1	10	11	4	24%	40%	36%	38%
	gestão da informação	1							
	conhecimento nas organizações	1							
	ciência da informação	1							
Artigo 23	projetos de pesquisa	1	10	11	4	24%	40%	36%	38%
	publicações	1							
	dissertações e teses	1							
	áreas do conhecimento	0							
Artigo 24	áreas do conhecimento	1	10	11	4	24%	40%	36%	38%
	inteligência artificial	1							

*Outros termos comuns:
6 - ciência da informação
9 - gestão da informação

*Outros termos comuns:
9 - sociedade da informação
10 - competências profissionais

*Outros termos comuns:
7 - sociedade da informação
9 - inclusão digital

*Outros termos comuns:
8 - paradigmas
10 - sociedade da informação

*Outros termos comuns:
6 - biblioteconomia
9 - ensino de biblioteconomia

*Outros termos comuns:
6 - sociedade da informação
8 - comunicação científica

*Outros termos comuns:
8 - informação para negócios

Nº do artigo	Termos do MAUI	Termo comum com folksonomia	Número de termos do MAUI	Número de termos da folksonomia	Número de termos comuns com a folksonomia	Consistência	Precisão	Revocação	Medida F1	
Artigo 21	sociologia bibliotecas nacionais ciências humanas	1 0 0	6	9	3	25%	50%	33%	40%	
Artigo 22	ciência da informação paradigmas processamento de informações tecnologias da informação e comuni organização do conhecimento	1 1 1 1	10	7	6	55%	60%	86%	71%	*Outros termos comuns: 9 - hipertextos
Artigo 23	história sociedade da informação conceitos de informação políticas públicas medicina	1 1 1 0 0	6	11	3	21%	50%	27%	35%	
Artigo 24	áreas do conhecimento políticas públicas pesquisa pesquisadores pesquisa e desenvolvimento	1 1 1 1 0	7	10	4	31%	57%	40%	47%	
Artigo 25	ciência da informação agências de fomento projetos de pesquisa bolsas de pesquisa pesquisa	1 1 1 1 1	10	13	6	35%	60%	46%	52%	*Outros termos comuns: 7 - pesquisadores
Artigo 26	monitoramento semântica recuperação da informação ciência da informação gestão da informação	1 1 1 0 0	9	9	3	20%	33%	33%	33%	
Artigo 27	bolsas de pesquisa bases de dados agências de fomento pesquisadores indicadores de C&T	1 1 1 1 1	10	9	5	36%	50%	56%	53%	
Artigo 28	sistemas de informação gestão da informação bibliotecas digitais governo eletrônico necessidades de informação	1 1 1 1 0	10	9	4	27%	40%	44%	42%	
Artigo 29	governo eletrônico informação governamental gestão da informação políticas de informação serviços de informação	1 1 1 1 1	10	7	5	42%	50%	71%	59%	
Artigo 30	periódicos eletrônicos World Wide Web artigos de periódico usuários acesso	1 1 1 1 0	10	14	4	20%	40%	29%	33%	
Artigo 31	regimes de informação políticas de informação sociedade da informação comunicação informal informação científica e tecnológica	1 1 1 0 0	10	9	3	19%	30%	33%	32%	
Artigo 32	bases de dados estratégias de busca recuperação da informação linguagens documentárias termos de indexação	1 1 1 1 1	10	8	5	38%	50%	63%	56%	
Artigo 33	bibliotecas digitais bibliotecas virtuais profissionais de informação artigos de periódico produtividade de autor	1 1 1 1 1	10	13	6	35%	60%	46%	52%	*Outros termos comuns: 8 - bibliotecas
Artigo 34	informação científica e tecnológica transferência da informação nutrição universidades pesquisa e desenvolvimento	1 1 1 1 1	10	13	6	35%	60%	46%	52%	*Outros termos comuns: 10 - institutos de pesquisa
Artigo 35	linguagens de marcação HTML ciência da informação XML SGML	1 1 1 1 1	10	7	5	42%	50%	71%	59%	
Artigo 36	lei de Lotka produtividade de autor bibliometria arquivos privados cartas	1 1 1 0 0	10	7	3	21%	30%	43%	35%	
Artigo 37	informação para negócios bases de dados informação financeira informação jurídica acesso à informação	1 1 1 1 1	10	12	8	57%	80%	67%	73%	*Outros termos comuns: 8 - tomada de decisões 9 - indústria da informação 10 - empresas
Artigo 38	ensino a distância bibliotecas híbridas tecnologias da informação e comuni bibliotecas universitárias usuários	1 1 1 1 1	10	14	8	50%	80%	57%	67%	*Outros termos comuns: 7 - serviços de bibliotecas 8 - bibliotecas digitais 9 - necessidades de informação
Artigo 39	estratégias de busca bases de dados recuperação da informação sistemas de recuperação da informa mecanismos de busca	1 1 1 1 1	10	7	5	42%	50%	71%	59%	
Artigo 40	gestão da informação recursos de informação profissionais de informação tomada de decisões ciência da informação	1 1 1 1 1	10	12	8	57%	80%	67%	73%	*Outros termos comuns: 7 - necessidades de informação 8 - tecnologias da informação e comunicação 9 - mercado de trabalho
Artigo 41	ciência da informação artigos de periódico literatura cinzenta comunidades científicas	1 1 1 1	10	11	4	24%	40%	36%	38%	

Nº do artigo	Termos do MAUI	Termo comum com folksonomia	Número de termos do MAUI	Número de termos da folksonomia	Número de termos comuns com a folksonomia	Consistência	Precisão	Revocação	Medida F1	
	autoria individual		0							
Artigo 42	sistemas de informação		1							*Outros termos comuns: 9.- gestão da informação 10.- profissionais de informação
	informação para negócios		1							
	gestão do conhecimento		10	12	7	47%	70%	58%	64%	
	bibliotecas virtuais		1							
Artigo 43	inteligência competitiva		1							*Outros termos comuns: 8.- inovação
	necessidades de informação		1							
	informação científica e tecnológica		10	8	6	50%	60%	75%	67%	
	tomada de decisões		1							
Artigo 44	conhecimento nas organizações		1							
	empresas		1							
	gestão do conhecimento		1							
	conhecimento nas organizações		1							
Artigo 45	tecnologias da informação e comuni		1							
	competitividade		1							
	tomada de decisões		1							
	ciência da informação		1							
Artigo 46	bibliometria		1							*Outros termos comuns: 10.- periódicos
	bases de dados		10	8	4	29%	40%	50%	44%	
	informetria		1							
	fluxo da informação		0							
Artigo 47	teoria do caos		1							
	lei de Bradford		1							
	ciência da informação		10	7	6	55%	60%	86%	71%	
	bibliometria		1							
Artigo 48	modelos matemáticos		1							*Outros termos comuns: 10.- tecnologias da informação e comunicação
	periódicos eletrônicos		1							
	periódicos científicos		1							
	ciência da informação		10	12	5	29%	50%	42%	45%	
Artigo 49	World Wide Web		1							
	usuários		1							
	economia da informação		1							
	usos da informação		1							
Artigo 50	informação para negócios		1							
	empresas		1							
	custos		1							
	descritores		1							
Artigo 51	bases de dados		1							*Outros termos comuns: 7.- usuários 10.- sistemas de informação
	periódicos		1							
	monitoramento		1							
	inteligência competitiva		1							
Artigo 52	formação profissional		1							*Outros termos comuns: 10.- hipertextos
	sociedade da informação		1							
	educação		1							
	bibliotecários		1							
Artigo 53	profissionais de informação		1							
	acessibilidade		1							
	digitalização		1							
	tecnologias da informação e comuni		1							
Artigo 54	acesso à informação		1							*Outros termos comuns: 8.- educação 9.- usos da informação
	bibliotecas		1							
	ciência da informação		1							
	estudos de usuários		1							
Artigo 55	projetos de pesquisa		1							
	educação		1							
	fluxo da informação		1							
	recuperação da informação		1							
Artigo 56	comportamento do usuário		1							*Outros termos comuns: 9.- sociologia
	mecanismos de busca		1							
	usuários		1							
	buscas		0							
Artigo 57	competência em informação		1							
	sociedade da informação		1							
	buscas de informação		1							
	profissionais de informação		1							
Artigo 58	bibliotecários		1							
	mercado de trabalho		1							
	ciência da informação		1							
	conhecimento nas organizações		1							
Artigo 59	gestão do conhecimento		0							
	ciência da informação		1							
	informação científica e tecnológica		1							
	cientistas da informação		1							
Artigo 60	teorias na ciência da informação		1							*Outros termos comuns: 9.- sociologia
	comunicação científica		1							
	ciência da informação		1							
	recuperação da informação		1							
Artigo 61	sociedade da informação		1							
	políticas de informação		1							
	comunicação informal		0							
	recuperação da informação		1							
Artigo 62	processamento de informações		1							
	ciência da informação		1							
	ciência cognitiva		1							
	sistemas de informação		0							
Artigo 63	frente de pesquisa		1							
	educação		1							
	bibliometria		1							
	comunicação		1							
Artigo 64	ciência da informação		0							
	inteligência competitiva		1							
	agentes inteligentes		1							
	usos da informação		1							
Artigo 65	gestão da informação		1							
	estudos de casos		1							
	Mínimo		6,00	7,00	2,00	9%	20%	13%	16%	
	Média		9,57	10,35	5,15	36%	54%	51%	52%	
Artigo 66	Desvio padrão		1,11	2,05	1,42	11%	13%	15%	12%	
	Máximo		10,00	15,00	8,00	57%	80%	86%	73%	
	Número de zeros		0	0	0					

Nº do artigo	Termos do MAUI	Termo comum com folksonomia	Número de termos do MAUI	Número de termos da folksonomia	Número de termos comuns com a folksonomia	Consistência	Precisão	Revocação	Medida F1
		Número de uns	0	0	0				
		Número de dois	0	0	1				
		Número de três	0	0	7				
		Número de quatro	0	0	11				
		Número de cinco	0	0	19				
		Número de seis ou	60	60	22				
			60	60	60				

APÊNDICE F- Análise dos termos de indexação do SISA e MAUI

Corpus de referência	Quantidade de termos atribuídos		Termos comuns com a folksonomia		Termos comuns entre SISA e MAUI	Erros com relação a Folksonomia de cada Artigo		Acertos com relação a Folksonomia de cada Artigo	
	SISA	MAUI	SISA	MAUI		SISA	MAUI	SISA	MAUI
Artigo 1	6	7	2	5	Transferência da Informação; Gestão do Conhecimento.	Avaliação; Acesso; Descarte; Gestão.	Ciência da Informação; Tecnologias da informação e comunicação.	Transferência da Informação; Gestão do Conhecimento.	Gestão do Conhecimento; Conhecimento nas organizações; Transferência da Informação; Conhecimento Tácito; Processos de Gestão.
Artigo 2	4	8	4	5	Comunicação científica; Ciência da Informação; Notícias.		Cientistas, Comunicação; resumos	Comunicação Científica; Ciência da Informação; Educação; Notícias.	Divulgação Científica; Ciência da Informação; Comunicação Científica; Notícias; Informação Científica e Tecnológica.
Artigo 3	4	8	2	5	Direito à Informação; Recuperação da Informação; Direito.	Avaliação; Direito.	Direito; Bibliotecas; Resumos.	Direito à Informação; Recuperação da Informação.	Direito à Informação; Sociedade da Informação; Recuperação da Informação; Acesso à Informação; Ciência da Informação.
Artigo 4	5	10	1	5	Acesso Livre	Acesso; Conceitos; Filosofia; Publicações	Direito; Arquivos.	Acesso Livre	Direito Autoral; Acesso Livre; Tecnologias da informação e comunicação; Publicações eletrônicas; Periódicos Científicos.

Artigo 5	3	10	1	4	Pesquisa	Acesso; Pesquisa .	Acesso ao Documento; Fitas Sonoras; Fluxo da Informação; Análise Comparativa Pesquisa ; Pesquisadores.	Estudantes	Informação Científica e Tecnológica; Sociedade da Informação; Comunidades Científicas; Educação.
Artigo 6	5	10	3	5	Recuperação da Informação; Ciência da Informação; Ciência da Computação; Universidades.	Navegação; Universidades .	Áreas do Conhecimento, Bases de Dados; Centros de Análise de Informação; Ciências Humanas; Universidades .	Ciência da Informação; Ciência da Computação; Recuperação da Informação.	Recuperação da Informação; Ciência da Informação; Ciência da Computação; Tesouros e Sistemas de informação.
Artigo 7	6	4	10	5	Inteligência Competitiva; Gestão do Conhecimento; Gestão da Informação; Dados.	Dados; Gestão.	Sociedade da Informação; Sistemas de Informação; Tecnologias da Informação e Comunicação; Tomada de Decisões; Dados .	Inteligência Competitiva; Competitividade; Gestão do Conhecimento; Gestão da Informação.	Inteligência Competitiva; Gestão do Conhecimento; Recursos de Informação; Ciência da Informação; Gestão da Informação.
Artigo 8	3	10	2	5	Áreas do Conhecimento; Ciência da Informação; Pesquisa.	Pesquisa	Estados da Arte; Processamento de Informações; Comunicação Científica; Artigos de Periódico; Pesquisa .	Áreas do Conhecimento; Ciência da Informação.	Ciência da Informação; Conceitos de Informação; Áreas do Conhecimento; Tipos de Documento; Indústria da Informação.
Artigo 9	6	10	3	7	Ciência da Informação; Fluxo da Informação; Inteligência Competitiva; Disseminação Seletiva da Informação.	Disseminação Seletiva da Informação; Gestão e Relevância.	Sociedade da Informação; Disseminação Seletiva da Informação; Usuários.	Ciência da Informação; Fluxo da informação; Inteligência Competitiva.	Tomada de Decisões, Fluxo da Informação; Comunicação Informal; Inteligência Competitiva, Universidades; Ciência da Informação; Gestão da Informação.

Artigo 10	3	10	2	5	Economia	Pertinência	Processamento de Informações; Pesquisa e Desenvolvimento; Imagens; Computadores Pessoais; Comunicação.	Economia; Levantamentos.	Sociedade da Informação; Sistemas de Informação; Tecnologias da Informação e Comunicação; Probabilidade e Estatística; Economia.
Artigo 11	5	6	3	4	Ciência da Informação; Propriedade Intelectual; Dados e Pesquisa.	Classificação; Dados.	Protocolos de Comunicação; Dados, Pesquisadores.	Ciência da Informação; Pesquisa; Propriedade Intelectual.	Ciência da Informação; Propriedade Intelectual; Pesquisa.
Artigo 12	6	10	5	5	Ciência da Informação; Acesso à Informação; Biblioteconomia	Acesso	Informação Governamental; Recuperação da Informação; Áreas do Conhecimento; Disseminação da Informação; Buscas de Informação.	Ciência da Informação; Pesquisa; Acesso à Informação; Arquivologia; Biblioteconomia	Ciência da Informação; Profissionais de Informação; Acesso à Informação; Sistemas de Informação; Biblioteconomia.
Artigo 13	4	10	3	7	Formação Profissional; Sistemas de Informação	Papel	Acesso à Informação; Tecnologias da Informação e Comunicação; Periódicos Científicos.	Formação Profissional; Economia; Sistemas de Informação.	Profissionais de informação; Formação Profissional, sistemas de Informação; Ciência da Informação; Disseminação da Informação; Sociedade da Informação; Competências profissionais.
					Profissionais de Informação; Gestão da Informação;		Necessidades de Informação;	Profissionais de Informação; Gestão da Informação; Gestão do Conhecimento;	Competência em Informação; Profissionais de Informação; Gestão da Informação; Engenharia do Conhecimento; Gestão do

Artigo 14	8	10	5	7	Gestão do Conhecimento; Gestores; Inclusão Digital.	Acesso; Gestores ; gestão.	Gestores ; Ciência da Informação.	Inclusão Digital ; Formação Profissional.	Conhecimento ; Sociedade da Informação; Inclusão Digital .
Artigo 15	3	10	2	7	Formação Profissional; Educação; Pesquisa.	Educação	Educação ; Pesquisadores; Universidades.	Formação Profissional ; Pesquisa .	Profissionais de Informação; Formação Profissional ; Educação Superior; Ciência da Informação; Pesquisa ; Paradigmas; Sociedade da Informação.
Artigo 16	6	10	5	7	Ciência da Informação; Formação Profissional; Sociedade da Informação; Educação Superior; Biblioteconomia; Educação.	Educação	Arquitetura da Informação; Arquitetos de informação; Educação .	Ciência da Informação, Biblioteconomia ; Educação Superior ; Formação Profissional ; Sociedade da Informação .	Ciência da Informação ; Profissionais de Informação; Formação Profissional ; Sociedade da Informação ; Educação Superior ; Biblioteconomia ; Ensino de Biblioteconomia.
Artigo 17	6	10	4	7	Propriedade Intelectual; Controle Bibliográfico; Tecnologias da Informação e Comunicação.	Autoria; Complexidade	Correio Eletrônico; Acesso Universal; Periódicos Científicos.	Controle Bibliográfico ; Internet ; Propriedade Intelectual ; Tecnologias da Informação e Comunicação .	Publicações Eletrônicas; Propriedade Intelectual ; Controle Bibliográfico ; Tecnologia das Informação e Comunicação ; Inovação; Sociedade da Informação; Comunicação Científica.
Artigo 18	2	10	2	2			Dissertações e Teses; Bibliografias; Mudança; Comunidades Científicas; Sistemas de Organização do Conhecimento; Pesquisadores; Resumos; Cenários.	Categorias; Sociologia.	Pesquisa; Construtivismo

Artigo 24	4	7	2	4	Pesquisa; Pesquisadores	Ecologia; Validade	Pesquisa e Desenvolvimento; Organização do Conhecimento; Uso da Informação.	Pesquisa; Pesquisadores.	Áreas do Conhecimento; Políticas Públicas; Pesquisa; Pesquisadores.
Artigo 25	5	10	4	6	Ciência da Informação; Pesquisa, Pesquisadores; Dados.	Dados	Ciências Sociais Aplicadas; áreas do Conhecimento; Comunidades Científicas; Dados.	Ciência da Informação; Pesquisa Orçamento; Pesquisadores.	Ciência da Informação; Agências de Fomento; Projetos de Pesquisa; Bolsas de Pesquisas; Pesquisa; Pesquisadores.
Artigo 26	5	9	2	3	Ciência da Informação; Monitoramento; Semântica; Subsídios.	Ciência da Informação; Gestão; Subsídios.	Gestão da Informação, Transferência da Informação, Fluxo da Informação, Instrumentalismo e subsídios.	Monitoramento; Semântica.	Monitoramento, Semântica; Recuperação da Informação.
Artigo 27	7	10	2	5	Base de Dados; Pesquisadores.	Dados; Gestão; Indicadores; Monitoramento; Relevância.	Índice H; índices; Áreas do Conhecimento; Pesquisa; Indicadores.	Bases de Dados; Pesquisadores	Bolsas de Pesquisa; Bases de Dados; Agências de Fomento; Pesquisadores; Indicadores de C&T.
Artigo 28	8	10	5	4	Sistemas de Informação; Gestão da Informação; Bibliotecas Digitais; Governo Eletrônico; Bibliotecas.	Gestão, Bibliotecas; Papel.	Necessidades de Informação; Informação Governamental; Tomada de Decisões; Revisões de Literatura; Bibliotecas; Dissertações e Teses.	Sistemas de Informação; Bibliotecas Digitais; Gestão da Informação; Governo Eletrônico; Inovação.	Sistemas de Informação; Gestão da Informação; Bibliotecas Digitais; Governo Eletrônico.
Artigo 29	5	10	3	5	Governo Eletrônico; Informação Governamental; Políticas de	Acesso; Gestão.	Sociedade da Informação; Acesso; Políticas Públicas; Gestão do	Governo Eletrônico; Informação Governamental;	Governo Eletrônico; Informação Governamental; Gestão da Informação; Políticas de Informação;

					Informação; Acesso.		Conhecimento; Ciência da Informação.	Políticas de Informação.	Serviços de Informação.
Artigo 30	5	10	2	4	Periódicos Eletrônicos; World Wide Web; Acesso; Periódicos; Avaliação.	Avaliação, Acesso; Periódicos.	Acesso; Periódicos; Arquivos; Associações de Elaboração de Normas; Ciência da Informação; Avaliação.	Periódicos Eletrônicos; World Wide Web.	Periódicos Eletrônicos; World Wide Web; Artigos de Periódicos; Usuários.
Artigo 31	5	10	3	3	Sociedade da Informação	Cenários; Economia.	Comunicação Informal; Informação Científica e Tecnológica; Políticas Públicas; Ciência da Informação; Serviços de Informação; Cenários; Indústria da Informação.	Economia; Internet; Sociedade da Informação.	Regimes de Informação; Políticas de Informação; Sociedade da Informação.
Artigo 32	8	10	3	5	Bases de Dados; Estratégias de Busca; Recuperação da Informação.	Dados; CD-ROM; Complexidade; Planejamento; Termos.	Planejamento Estratégico; Buscas de Informação; Base de Dados em Linha; Estudo de Uso; Sistemas de Recuperação da Informação.	Bases de Dados; Estratégias de Busca; Recuperação da Informação.	Bases de Dados; Estratégias de Busca; Recuperação da Informação; Linguagens Documentárias; Termos de Indexação.
Artigo 33	11	10	5	6	Bibliotecas Digitais; Bibliotecas Virtuais; Profissionais de Informação; Bibliotecas; Referências Bibliográficas.	Periódicos; Autorias; Coleções; Desenvolvimento de coleções; Planejamento; Referências Bibliográfica.	Desenvolvimento de coleções; Tipos de Documento; Ciência da informação; Referências Bibliográficas.	Bibliotecas Digitais; Bibliotecas Virtuais; Bibliotecas; Internet; Profissionais de Informação.	Bibliotecas Digitais; Bibliotecas Virtuais; Profissionais de Informação; Artigos de Periódico; Produtividade de autor; Bibliotecas.
							Comunicação Informal;		Informação Científica e

Artigo 34	4	10	2	6	Ciências da Saúde Transferência da Informação; Nutrição.	Acesso; Ciências da Saúde.	Ciências da Saúde; Necessidades de Informação; Ciência da Informação.	Nutrição; Transferência da Informação.	Tecnológica; Transferência da Informação; Nutrição; Universidades; Pesquisa e Desenvolvimento; Institutos de Pesquisa.
Artigo 35	9	10	5	5	Linguagens de Marcação; HTML; Ciência da Informação; XML; Dados.	Conceitos; Complexidade; Dados e Usuários.	Padrões de Metadados; Dados; Áreas do Conhecimento; Recuperação da Informação; Ciência da Computação.	Internet; XML; Ciência da Informação; HTML; Linguagens de Marcação.	Linguagens de Marcação; HTML; Ciência da Informação; XML; SGML
Artigo 36	6	10	2	3	Lei de Lotka; Cartas; Bibliometria.	Biblioteconomia; Cartas; Dados; Medicina.	Arquivos Privados; Cartas; Faculdades de Biblioteconomia; Bases de Dados; Ciência da Informação; Arquivos; Análise de Dados.	Bibliometria; Lei de Lotka.	Lei de Lotka; Produtividade de Autor; Bibliometria.
Artigo 37	9	10	4	8	Informação para Negócios; Bases de Dados; Indústria da Informação; Empresas.	Dados; Cenários; Categorias; Classificação; Notícias.	Competências no Uso de Computadores; Informação Científica e Tecnológica.	Bases de Dados; Informação para negócios; Empresas; Indústria da Informação.	Informação para Negócios; Bases de Dados; Informação Financeira; Informação Jurídica; Acesso à Informação; Tomada de Decisões; Indústria da Informação; Empresas.
Artigo 38	7	10	3	8	Bibliotecas Híbridas; Bibliotecas; Usuários.	Educação; Bibliotecas; Papel; Universidades.	Bibliotecas; Bases de Dados.	Bibliotecas Híbridas; Internet; Usuários.	Ensino à Distância; Bibliotecas Híbridas; Tecnologias da Informação e Comunicação; Bibliotecas Universitária; Usuários; Serviços de Biblioteca, Bibliotecas Digitais; Necessidades de Informação.
Artigo 39	10	10	4	5	Estratégias de Busca; Bases de Dados; Recuperação da Informação; Sistemas de Recuperação da Informação.	Acesso; Complexidade; Conceitos; Dados; Planejamento; Usuário.	Planejamento Estratégico; Buscas de Informação; Bases de Dados em Linha; Linguagens Documentárias; Buscas.	Recuperação da Informação; Bases de Dados; Estratégias de Busca; Sistemas de Recuperação da Informação.	Estratégias de Busca; Bases de Dados; Recuperação da Informação; Sistemas de Recuperação da Informação; Mecanismos de Busca.

Artigo 40	7	10	5	8	Gestão da Informação; Recursos de Informação; Profissionais de Informação; Ciência da Informação.	Gestão; Avaliação.	Acesso à Informação; Serviços de Informação.	Gestão da Informação; Formação Profissional; Ciência da Informação; Profissionais de Informação; Recursos de Informação.	Gestão da Informação Recursos de Informação; Profissionais de Informação; Tomada de Decisões; Ciência da Informação; Necessidades de Informação; Tecnologias da Informação e Comunicação; Mercado de Trabalho.
Artigo 41	10	10	5	4	Ciência da Informação; Literatura Cinzenta.	Autoria; Dados; Índice; Pesquisa; Periódicos.	Autoria Individual; Institutos de Pesquisa; Sociedade a Informação; Tipos de Documento; Bases de Dados; Coleta de Dados.	Ciência da Informação; Programas de Pós-Graduação; Programas de Pós-Graduação em Ciência da Informação; Literatura Cinzenta; Publicações.	Ciência da Informação; Artigos de Periódico; Literatura Cinzenta; Comunidades Científicas.
Artigo 42	10	10	9	7	Sistemas de Informação; Informação para Negócios; Gestão do Conhecimento; Bibliotecas Virtuais; Inteligência Competitiva; Gestão da Informação; Profissionais de Informação.	Gestão	Centro de Análise de Informação; Bibliotecas de Empresas; Acesso à Informação.	Informação Para Negócios; Bibliotecas; Bibliotecas Virtuais; Empresas; Gestão do Conhecimento; Gestão da Informação; Inteligência Competitiva; Profissionais de Informação; Sistemas de Informação.	Sistemas de Informação; Informação para Negócios; Gestão do Conhecimento; Bibliotecas Virtuais; Inteligência Competitiva; Gestão da Informação; Profissionais de Informação.
Artigo 43	4	10	3	6	Necessidades de Informação; Empresas; Inovação.	Pesquisa	Estados da Arte; Estudos de Usuários; Estudos de Caso; Ciência da Informação.	Necessidades de Informação; Empresas; Inovação.	Necessidades de Informação; Informação Científica e Tecnológica; Tomada de Decisões; Conhecimento nas Organizações; Empresas; Inovação.
Artigo 44	4	10	2	5	Gestão do Conhecimento; Gestão; Competitividade.	Gestão e Papel	Empresas; Gestão; Bases de Dados; Conhecimento Tácito; Resumos.	Competitividade; Gestão do Conhecimento.	Gestão do Conhecimento; Conhecimento nas Organizações; Tecnologias da Informação e Comunicação; Competitividade; Tomada de Decisões.

Artigo 45	5	10	3	4	Ciência da Informação; Bibliometria; Informetria.	Avaliação; Internet.	Fluxo de Informação; Áreas do Conhecimento; World, Wide Web; Produtividade de Autor; Tomada de Decisões; Comunicação.	Ciência da Informação; Bibliometria; Informetria	Ciência da Informação; Bibliometria; Bases de Dados; Informetria.
Artigo 46	6	10	5	6	Teoria do Caos; Lei de Bradford; Ciência da Informação; Bibliometria; Periódicos.	Pesquisa	Revisões de Literatura; Faculdades de Biblioteconomia; Publicações Seriadas; Unidades de Informação.	Bibliometria; Ciência da Informação; Lei de Bradford; Periódicos; Teoria do Caos.	Teoria do Caos; Lei de Bradford; Ciência da Informação; Bibliometria; Modelos Matemáticos; Periódicos.
Artigo 47	6	10	5	5	Periódicos Eletrônicos; Periódicos Científicos; Ciência da Informação; World Wide Web; Periódicos; Usuários.	Periódicos	Periódicos; Obras de Referência; Artigos de periódico; Acesso; Monitores de Vídeo.	Periódicos Eletrônicos; Usuários; Ciência da Informação; Periódicos Científicos; World Wide Web.	Periódicos Eletrônicos; Periódicos Científicos; Ciência da Informação; World Wide Web; Usuários
Artigo 48	5	10	4	6	Economia da Informação; Economia; Empresas; Custos.	Economia	Economia; Processamento de Informações; Tomada de Decisões; Sistemas de Informação.	Economia da Informação; Custos; Empresas; Inovação.	Economia da Informação; Usos da Informação; Informação para Negócios; Empresas; Custos; Tecnologias da Informação; Comunicação.
Artigo 49	4	10	3	5	Descritores; Periódicos; Monitoramento.	Pesquisa	Informação Científica e Tecnológica; Tomada de Decisões; Títulos de Documentos; Buscas; Mapas.	Descritores; Periódicos; Monitoramento.	Descritores; Bases de Dados; Periódicos; Monitoramento; Inteligência Competitiva.

Artigo 50	5	10	3	5	Formação Profissional; Educação; Bibliotecários; Papel.	Cooperação; Papel.	Tomada de Decisões; Ciência da Informação; Tecnologias da Informação e Comunicação; Educação Básica; Papel.	Formação Profissional; Bibliotecários; Educação.	Formação Profissional; Sociedade da Informação; Educação; Bibliotecários; Profissionais de Informação.
Artigo 51	4	10	3	7	Acessibilidade; Bibliotecas; Usuários.	Categorias	Sítios Web; Suportes de Informação; Sociedade da Informação.	Acessibilidade; Bibliotecas; Usuários.	Acessibilidade; Digitalização; Tecnologias da Informação e Comunicação; Acesso à Informação; Bibliotecas; Usuários; Sistemas de Informação.
Artigo 52	7	10	4	6	Ciência da Informação; Estudos de Usuários; Educação; Papel.	Educação; Pesquisa; Papel.	Políticas de Informação; Projetos de Pesquisa; Transferência da Informação; Comunicação Informal; Papel.	Estudos de Usuários; Usuários; Ciência da Informação; Internet.	Ciência da Informação; Estudos de Usuários; Projetos de Pesquisa; Educação; Fluxo da Informação; Hipertextos.
Artigo 53	3	7	2	4	Recuperação da Informação; Usuários.	Professores	Buscas; Booleanas; Acesso.	Recuperação da Informação; Usuários	Recuperação da Informação; comportamento do Usuário; Mecanismos de Busca; Usuário.
Artigo 54	8	10	3	7	Competência em Informação; Bibliotecários; Educação.	Filosofia; Acesso; Avaliação; Competência em Informação; Papel.	Acesso à Informação; Tomada de Decisões; Serviços de Informação.	Bibliotecas; Bibliotecários; Educação.	Competência em Informação; Sociedade da Informação; Buscas de Informação; Profissionais de Informação; Bibliotecários; Educação; Usos da Informação.
						Empresas	Gestão do Conhecimento; Gestão da Informação; Empresas; Processos de Gestão; Tomada de Decisões;	Mercado de Trabalho; Formação Profissional.	Profissionais de Informação; Mercado de Trabalho; Ciência da Informação;

Artigo 55	3	10	2	4	Mercado de Trabalho; Empresas.		Sociedade da Informação.		Conhecimento nas Organizações.
Artigo 56	9	10	5	6	Ciência da Informação; Informação Científica e Tecnológica; Cientistas da Informação; Comunicação Científica; Cientistas; Sociologia; História.	Cientistas; História; Pesquisa; Relevância.	Comunicação Informal; Cientistas; Conceitos de Informação; História.	Ciência da Informação; Cientistas da Informação; Comunicação Científica; Informação Científica e Tecnológica; Sociologia.	Ciência da Informação; Informação Científica e Tecnológica; Cientistas da Informação; Teorias na Ciência da Informação; Comunicação Científica; Sociologia.
Artigo 57	4	10	3	3	Ciência da Informação; Recuperação da Informação.	Avaliação	Políticas de Informação; Comunicação Informal; Informação Científica e Tecnológica; Tecnologias da Informação e Comunicação; Usos da Informação; Buscas de Informação; Transferência da Informação.	Ciência da Informação; Pesquisa; Recuperação da Informação	Ciência da Informação; recuperação da informação; Sociedade da Informação.
Artigo 58	5	10	5	4	Recuperação da Informação; Ciência da Informação; Ciência Cognitiva.		Sistemas de Informação; Buscas de Informação; Ciência da Computação; Necessidades de Informação; Teoria da Informação; Aplicações do Computador.	Ciência da Informação; Ciência Cognitiva; Indexação; Interação Homem-Computador; Recuperação da Informação.	Recuperação da Informação; Processamento de Informações; Ciência da Informação; Ciência Cognitiva.
Artigo 59	5	10	4	4	Educação; Bibliometria.	Pesquisa	Ciência da Informação; Referências Bibliográficas; Áreas do Conhecimento; Literatura Científica; Periódicos Científicos; Páginas da Web.	Educação, Bibliometria; Cientometria; Fonte de Pesquisa.	Frente de Pesquisa; Educação; Bibliometria; Comunicação.

Artigo 60	8	10	5	5	Inteligência Competitiva; Agentes Inteligentes; Gestão da Informação; Estudos de Caso.	Gestão	Engenharia de Produção; Gestão do Conhecimento; Tomada de Decisões; Tecnologias da Informação e Comunicação; Conceitos de Informação.	Agentes Inteligentes; Inteligência Competitiva; Internet; Estudos de Caso; Gestão da Informação; Inovação; Monitoramento.	Inteligência Competitiva; Agentes Inteligentes; Usos da Informação; Gestão da Informação; estudos de Caso.
------------------	---	----	---	---	--	--------	---	--	---

ANEXO

ANEXO A - Corpus de Souza (2005) utilizado no experimento

Referências dos 60 artigos científicos
1 - SANTOS, P. C. L. V. A. C.; SANT´ANA, R. C. S. G. A. Transferência da informação: análise para valoração de unidades de conhecimento, Datagramazero: Revista de Ciência da Informação , v. 3, n. 2, abr.2002.
2 - MUELLER, Suzana P. M.. Popularização do Conhecimento Científico. Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 2, abr. 2002.
3 - CASTRO, Ana Lúcia Siaines de. O Valor da Informação: um desafio permanente. Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 3, jun. 2002.
4 - CAFÉ, Lígia; LAGE, Marcia Basílio. Auto-arquivamento: uma opção inovadora para a produção científica. Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 3, jun. 2002.
5 - BURNHAM, Teresinha Fróes. Análise contrastiva: memória da construção de uma metodologia para investigar a tradução de conhecimento científico em conhecimento público. Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 3, jun. 2002.
6 - LEVACOV, Marília et al. O Tesouro Eletrônico do Mundo do Trabalho: produto de um esforço interdisciplinar. Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 4, ago. 2002.
7 - VALENTIM, Marta Lígia Pomim. Inteligência competitiva em organizações: dado, informação e conhecimento. Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 4, ago. 2002.
8 - MIRANDA, Antonio; SIMEÃO, Elmira. A conceituação de massa documental e o ciclo de interação entre tecnologia e o registro do conhecimento. Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 4, ago. 2002.
9 - STAREC, C. Informação e universidade: os pecados informacionais e barreiras na comunicação da informação para a tomada de decisão na universidade. DataGramazero: Revista de Ciência da Informação , v. 3, n. 4, p. 0-0, 2002.
10 - PORCARO, Rosa Maria. Implicações da 'nova economia' para a mensuração estatística: desajustes conceituais e metodológicos. Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 4, ago. 2002.
11 - BRITO, Luiz Carlos. Por uma nova Ciência da Informação: ensino, pesquisa e formação. Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 5, out. 2002.
12 - DIAS, Eduardo Wense. Ensino e pesquisa em Ciência da Informação. Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 5, out. 2002.
13 - CARVALHO, Kátia de. O Profissional da Informação: o humano multifacetado.

Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 5, out. 2002.
14 - TARAPANOFF, Kira; SUAIDEN, Emir; OLIVEIRA, Cecília Leite. Funções Sociais e Oportunidades para Profissionais da Informação. Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 5, out. 2002.
15 - RODRIGUES, Mara Eliane Fonseca. Relação Ensino-Pesquisa: em discussão a formação do Profissional da Informação. Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 5, out. 2002.
16 - CARDOSO, Ana Maria Pereira. Educação para a Informação:: desafios contemporâneos para a Ciência da Informação. Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 5, out. 2002.
17 - GRAÇAS, Maria das. Novas tecnologias e produção científica: uma relação de causa e efeito ou uma relação de muitos efeitos?. Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 6, dez. 2002.
18 - DAGNINO, Renato. Enfoques sobre a relação Ciência, Tecnologia e Sociedade: Neutralidade e Determinismo. Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 6, dez. 2002.
19 - RODRIGUES, Ricardo. Inteligência empresarial: uma avaliação de fontes de informação sobre o ambiente organizacional externo. Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 6, dez. 2002.
20 - SMIT, Johanna W.; DIAS, Eduardo Wense; SOUZA, Rosali Fernandez de. Contribuição da Pós-graduação para a Ciência da Informação no Brasil: uma visão. Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 6, dez. 2002.
21 - DUMONT, Lígia Maria Moreira. Os múltiplos aspectos e interfaces da leitura. Datagramazero: Revista de Ciência da Informação , Brasília, v. 3, n. 6, dez. 2002.
22 - SANTOS, Nilton Bahlis dos. A Informação e o Paradigma Holográfico: a Utopia de Vannevar Bush. Ciência da Informação , Brasília, v. 3, n. 6, p.15-25, dez. 2002.
23 - COSTA, Icléia Thiesen Magalhães. Informação, Memória e Espaço Prisional no Rio de Janeiro. Datagramazero: Revista de Ciência da Informação , Brasília, v. 4, n. 1, fev. 2003.
24 - GÓMEZ, González de; NÉLIDA, Maria. O Contrato Social da Pesquisa:: em busca de uma nova equação entre a autonomia epistêmica e autonomia política. Datagramazero: Revista da Ciência da Informação , Brasília, v. 4, n. 1, fev. 2003.
25 - MUELLER, Suzana Pinheiro Machado; SANTANA, Maria Gorette. A Ciência da Informação no CNPq: fomento à formação de recursos humanos e à pesquisa entre 1994-2002. DataGramZero: Revista de Ciência da Informação , v. 4, n. 1, fev. 2003.
26 - BARRETO, Aldo de Albuquerque. Políticas de Monitoramento da Informação por Compressão Semântica dos seus Estoques. Datagramazero: Revista da Ciência da Informação , Brasília, v. 4, n. 2, abr. 2003.

<p>27 - OLINTO, Gilda. Bolsas de Pesquisador do CNPq: informações sobre política de C&T a partir da base que contém os dados cadastrais dos bolsistas. Ciência da Informação, Brasília, v. 4, n. 2, p.23-39, abr. 2003.</p>
<p>28 - PACHECO, Roberto; KERN, Vinícius. Arquitetura conceitual e resultados da integração de sistemas de informação e gestão da Ciência e tecnologia. Ciência da Informação, Brasília, v. 4, n. 2, p.56-71, abr. 2003.</p>
<p>29 - MARCONDES, Carlos Henrique; JARDIM, José Maria. Políticas de Informação Governamental:: a construção de Governo Eletrônico na Administração Federal do Brasil. Datagrama zero: Revista da Ciência da Informação, Brasília, v. 4, n. 2, abr. 2003.</p>
<p>30 - DIAS, Guilherme Ataíde. Avaliação do acesso a periódicos eletrônicos na web pela análise do arquivo de log de acesso. Ciência da Informação, Brasília, v. 31, n. 1, jan. abr. 2002.</p>
<p>31 - GÓMEZ, González de; NÉLIDA, Maria. Novos cenários políticos para a informação. Ciência da Informação, Brasília, v. 31, n. 1, jan. 2002.</p>
<p>32 - LOPES, Ilza Leite. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. Ciência da Informação, Brasília, v. 31, n. 1, p.41-52, 1 jan-abril. 2002.</p>
<p>33 - OHIRA, Maria Lourdes Blatt. Bibliotecas virtuais e digitais: análise de artigos de periódicos brasileiros (1995/2000). Ciência da Informação, Brasília, v. 31, n. 1, p.61-74, jan-abril. 2002.</p>
<p>34 - PRYSTHON, Cecília; SCHMIDT, Susana. Experiência do Leal/UFPE na produção e transferência de tecnologia. Ciência da Informação, Brasília, v. 31, n. 1, p.84-90, jan-abril, 2002.</p>
<p>35 - ALMEIDA, Maurício Barcellos. Uma introdução ao XML, sua utilização na Internet e alguns conceitos complementares. Ciência da Informação, Brasília, v. 31, n. 2, p.5-13, maio-agosto, 2002.</p>
<p>36 - ALVARADO, Rubén Urbizagástegui. A Lei de Lotka na bibliometria brasileira. Ciência da Informação, Brasília, v. 31, n. 2, p.14-20, maio-agosto, 2002.</p>
<p>37 - CENDÓN, Beatriz Valadares. Bases de dados de informação para negócios. Ciência da Informação, Brasília, v. 31, n. 2, p.30-43, maio-agosto, 2002.</p>
<p>38 - GARCEZ, Eliane Maria Stuart. Biblioteca híbrida: um novo enfoque no suporte à educação a distância. Ciência da Informação, Brasília, v. 31, n. 2, p.44-51, maio-agosto, 2002.</p>
<p>39 - LOPES, Ilza Leite. Estratégia de busca na recuperação da informação: revisão da literatura. Ciência da Informação, Brasília, v. 31, n. 2, p.60-71, maio-agosto, 2002.</p>
<p>40 - MARCHIORI, Patricia Zeni. A Ciência e a gestão da informação: compatibilidades no espaço profissional. Ciência da Informação, Brasília, v. 31, n. 2, p.72-79, maio-agosto, 2002.</p>

<p>41 - POBLACIÓN, Dinah Aguiar; NORONHA, Daisy Pires. Produção das literaturas “branca” e “cinzenta” pelos docentes/doutores dos programas de pós-graduação em Ciência da informação no Brasil. Ciência da Informação, Brasília, v. 31, n. 2, maio 2002.</p>
<p>42 - REZENDE, Yara. Informação para negócios: os novos agentes do conhecimento e a gestão do capital intelectual. Ciência da Informação, Brasília, v. 31, n. 1, p.75-83, jan-abril 2002.</p>
<p>43 - SILVA, Janete Fernandes; FERREIRA, Marta Araújo Tavares; BORGES, Mônica Erichsen Nassif. Análise metodológica dos estudos de necessidades de informação sobre setores industriais brasileiros: proposições: proposições. Ciência da Informação, Brasília, v. 31, n. 2, p.129-141, maio-agosto 2002.</p>
<p>44 - SILVA, Sergio Luis da. Informação e competitividade: a contextualização da gestão do conhecimento nos processos organizacionais. Ciência da Informação, Brasília, v. 31, n. 2, p.142-151, maio-agosto, 2002.</p>
<p>45 - VANTI, Nadia Aurora Peres. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. Ciência da Informação, Brasília, v. 31, n. 2, p.152-162, maio-agosto, 2002.</p>
<p>46 - BORGES, Paulo César Rodrigues. Métodos quantitativos de apoio à bibliometria: a pesquisa operacional pode ser uma alternativa?. Ciência da Informação, Brasília, v. 31, n. 3, set. - dez., 2002.</p>
<p>47 - DIAS, Guilherme Ataíde. Periódicos eletrônicos: considerações relativas à aceitação deste recurso pelos usuários. Ciência da Informação, Brasília, v. 31, n. 3, set. - dez., 2002.</p>
<p>48 - COHEN, Max F.. Alguns aspectos do uso da informação na economia da informação. Ciência da Informação, Brasília, v. 31, n. 3, set. - dez., 2002.</p>
<p>49 - ORTIZ, Lúcia Cunha; ORTIZ, Wilson Aires; SILVA, Sergio Luis da. Ferramentas alternativas para monitoramento e mapeamento automatizado do conhecimento. Ciência da Informação, Brasília, v. 31, n. 3, p.66-76, set. dez.,2002.</p>
<p>50 - SILVA, Edna Lúcia da; CUNHA, Riam Vieira da. A formação profissional no século XXI: desafios e dilemas. Ciência da Informação, Brasília, v. 31, n. 3, p.77-82, set. dez., 2002.</p>
<p>51 - TORRES, Elisabeth Fátima; MAZZONI, Alberto Angel; ALVES, João Bosco da Mota. A acessibilidade à informação no espaço digital. Ciência da Informação, Brasília, v. 31, n. 3, p.83-91, set. dez., 2002.</p>
<p>52 - FREIRE, Isa Maria et al. Estudos de usuários: o padrão que une três abordagens. Ciência da Informação, Brasília, v. 31, n. 3, set. dez., 2002.</p>
<p>53 - AIRES, R. V. N. X.; ALUÍSIO, S. M. Como incrementar a qualidade dos resultados das máquinas de busca: da análise de logs à interação em português. Ciência da Informação, v. 32, n. 1, p. 5-16, 2003.</p>
<p>54 - DUDZIAK, Elisabeth Adriana. Information literacy:: princípios, filosofia e</p>

prática. **Ciência da Informação**, Brasília, v. 32, n. 1, jan. 2003.

55 - FERREIRA, Danielle Thiago. Profissional da informação:: perfil de habilidades demandadas pelo mercado de trabalho. **Ciência da Informação**, Brasília, v. 32, n. 1, jan. abr., 2003.

56 - FREIRE, Isa Maria. O olhar da consCiência possível sobre o campo científico. **Ciência da Informação**, Brasília, v. 32, n. 1, jan. abr., 2003.

57 - GÓMEZ, Maria Nélide González de. As relações entre Ciência, Estado e sociedade:: um domínio de visibilidade para as questões da informação. **Ciência da Informação**, Brasília, v. 32, n. 1, jan. abr., 2003.

58 - LIMA, G. N. B. M. O. Interfaces entre a Ciência da informação e a Ciência cognitiva. **Ciência da Informação**, v. 32, n. 1, jan. abr., 2003.

59 - MOSTAFA, Solange Puntel; MÁXIMO, Luis Fernando. A produção científica da Anped e da Intercom no GT da Educação e Comunicação. **Ciência da Informação**, Brasília, v. 32, n. 1, jan. abr., 2003.

60 - FREIRE, A Maria. SILVA, Helena Pereira da. Inteligência competitiva na Internet:: um processo otimizado por agentes inteligentes. **Ciência da Informação**, Brasília, v. 32, n. 1, jan. abr., 2003.