

AVALIAÇÃO DIRETA E CONJUNTA DE SISTEMAS DE INDEXAÇÃO AUTOMÁTICA POR ATRIBUIÇÃO

DIRECT AND JOINT ASSESSMENT OF AUTOMATIC INDEXING SYSTEMS BY ASSIGNMENT

Sâmela Rouse de Brito Silva¹
Renato Fernandes Correa²
Isidoro Gil-Leiva³

Resumo

Propõe um percurso metodológico para a avaliação direta e conjunta de sistemas de indexação automática por atribuição, bem como aplica o percurso proposto na avaliação da qualidade da indexação dos sistemas SISA (Sistema de Indización Semi-Automático) e MAUI (Multi-purpose Automatic Topic Indexing) na indexação automática por atribuição de artigos científicos escritos em português da área da Ciência da Informação. A pesquisa é exploratória e empírica, envolvendo estudo de caso e coleta de dados por meio de um experimento de indexação. O experimento constitui-se da indexação automática dos artigos científicos de um corpus de referência composto por 60 artigos de periódicos indexados intelectualmente usando o Tesouro Brasileiro de Ciência da Informação como linguagem de indexação. Conclui-se que o percurso metodológico se mostrou adequado e que os sistemas apresentaram bons resultados, porém o MAUI se destacou por apresentar melhores índices de desempenho.

Palavras-chave: Sistemas de indexação automática. Indexação automática por atribuição. Ciência da Informação. Avaliação da indexação automática. SISA.

Abstract

This work proposes a methodological pathway for the direct and joint evaluation of automatic indexing systems by assignment, as well as applying the proposed methodological pathway in the evaluation of the quality of indexing of the SISA (Sistema de Indización Semi-Automático) and MAUI (Multi-purpose Automatic Topic Indexing) for automatic indexing by assignment of scientific articles written in Portuguese in the area of Information Science. The research is exploratory and empirical, involving a case study and data collection through an indexing experiment. The experiment consists of the automatic indexing of scientific articles from a reference corpus composed of 60 articles from scientific journals indexed intellectually using the Brazilian Thesaurus for Information Science as an indexing language. It concludes that the methodological pathway proved to be adequate and that both systems presented good results, however MAUI stood out for presenting better performance indexes.

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

Artigo submetido em 02/12/2020 e aceito para publicação em 12/12/2020

- 1 Mestre em Ciência da Informação pela Universidade Federal de Pernambuco, Brasil. Bibliotecária na Universidade Federal de Alagoas, Brasil. ORCID: <https://orcid.org/0000-0002-9694-2187>. E-mail: ssamela.brito@hotmail.com
- 2 Docente no Programa de Pós-Graduação em Ciência da Informação. Universidade Federal de Pernambuco, Brasil. ORCID: <https://orcid.org/0000-0002-9880-8678>. E-mail: renato.correa@ufpe.br
- 3 Doutor em Filologia Hispânica pela Universidad de Murcia. Docente na Universidad de Murcia, Espanha. ORCID: <https://orcid.org/0000-0002-7175-3099>. E-mail: isgil@um.es

1 INTRODUÇÃO

Na tentativa de contribuir para a disponibilização e recuperação da informação, a indexação automática se constitui numa ferramenta tecnológica para os profissionais indexadores, permitindo poupar tempo e trabalho na indexação, e possibilitando uma melhor uniformidade e homogeneidade com relação aos termos de indexação.

Lancaster (2004) distingue 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 no texto de um 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, como por exemplo um vocabulário controlado (LANCASTER, 2004).

Porém, como qualquer serviço pertencente à unidade informacional, a indexação automática necessita ser avaliada quanto à qualidade na indexação, visando a tomada de decisão quanto ao uso operacional ou necessidade de aprimoramento.

A avaliação da indexação é compreendida como um processo para medir a qualidade da indexação nas unidades de informação. A qualidade da indexação é um conceito subjetivo e pode variar de acordo com os aspectos da indexação a serem analisados e com os objetivos da unidade de informação, porém se baseia na análise do resultado da indexação, isto é, dos termos de indexação.

De acordo com Gil Leiva (2008), a avaliação da indexação pode ser realizada de forma intrínseca pelo indexador durante o processo de indexação (de forma qualitativa, ou quantitativa por meio da intraconsistência) ou extrínseca por grupo de indexadores comparando o resultado de suas indexações. A avaliação extrínseca da indexação ocorre de forma quantitativa por meio da interconsistência ou na recuperação da informação.

A avaliação da indexação automática ocorre por meio da avaliação extrínseca, tendo os termos da indexação intelectual como padrão de referência da qualidade ou correteza na indexação. Por meio da comparação dos termos das indexações intelectual e automática, são calculados valores para métricas como índice de consistência, precisão, revocação e medida F. Tais índices são obtidos por meio de fórmulas e refletem diretamente ou indiretamente a qualidade da indexação automática.

A avaliação direta da indexação automática ocorre quando são comparados os termos de indexação e calculadas as métricas. A avaliação indireta da indexação automática ocorre quando são comparados os resultados de buscas em sistema de recuperação de informação com dois índices

diferentes, um contendo os termos da indexação intelectual e outro contendo os termos da indexação automática.

Golub et al. (2016) propuseram uma estrutura abrangente (do inglês *framework*) para avaliação da indexação automática envolvendo três formas principais: avaliar a qualidade da indexação diretamente por meio da avaliação por um avaliador ou por comparação com um padrão de referência (ou padrão ouro); avaliar a qualidade da indexação assistida por computador diretamente no contexto de um fluxo de trabalho de indexação; e avaliar a qualidade da indexação indiretamente por meio da análise do desempenho na recuperação da informação. Entretanto, tal estrutura proposta precisa ser instanciada e ser testada empiricamente em pesquisas posteriores, voltadas para o desenvolvimento de métodos ou percursos metodológicos de avaliação apropriados (GOLUB et al., 2016).

O desempenho de diferentes sistemas de indexação automática quando a qualidade na indexação automática pode ser aferida e comparada dada a utilização na avaliação da indexação automática de um mesmo percurso metodológico para avaliação da indexação automática.

O percurso metodológico para avaliação da indexação automática envolve a padronização de uso de sistemas, instrumentos, recursos, e métodos de avaliação. Um recurso primordial é um corpus de referência ou padrão ouro, isto é um conjunto de documentos e respectivos termos da indexação intelectual. Adicionalmente, envolve também a padronização do método para avaliação da qualidade da indexação automática.

Uma vez realizada a avaliação da indexação automática para cada sistema de indexação automática por meio de um percurso metodológico de avaliação direta e conjunta de tais sistemas, se torna possível a comparação da qualidade na indexação automática de dois ou mais sistemas por meio da comparação e análise dos valores obtidos para métricas como índice de consistência, precisão, revocação e medida F.

Portanto, este artigo tem como objetivo a proposição e aplicação de um processo de avaliação direta e conjunta de sistemas de indexação automática por atribuição na representação do conteúdo de artigos científicos da área da Ciência da Informação mediante a utilização do Tesouro Brasileiro de Ciência da Informação (TBCI) como linguagem de indexação. Serão avaliados os sistemas: *Sistema de Indización Semi-Automático* (SISA) e o *Multi-Purpose Automatic Topic Indexing* (MAUI).

O presente artigo justifica-se pela necessidade de avaliação direta e conjunta de sistemas de indexação automática, e pelo desconhecimento de um percurso metodológico bem definido para realizar tal avaliação no contexto de sistemas de indexação automática por atribuição.

Para atingir aos objetivos, o percurso metodológico delineado neste artigo foi aplicado em um experimento de indexação em que os sistemas SISA e MAUI realizaram a indexação automática de um corpus constituído por 60 artigos de periódicos científicos da área de Ciência da Informação que foram indexados intelectualmente utilizando o TBCI como linguagem de indexação. O corpus de referência trabalhado neste artigo, é o mesmo utilizado por Silva, B. e Corrêa, R. (2019, 2020).

2 FUNDAMENTAÇÃO TEÓRICA

Esta seção aborda o referencial teórico do percurso metodológico de avaliação da indexação automática. Na subseção 2.1 discute-se o processo de indexação automática e os sistemas de indexação automática. Na seção 2.2 discute-se a importância da avaliação da indexação automática, bem como sua aplicação.

2.1 Sistema de indexação automática

Representar conteúdo dentro de um sistema de recuperação da informação 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 indexação foi amplamente estudada. Alguns trabalhos valiosos sobre a teoria e prática da indexação são os de Frohmann (1990), Farrow (1991), Fugmann, (1993), Hjørland (1997), Mai (2000), Anderson e Perez-Carballo (2001), e Lancaster (2004), entre outros.

O padrão ISO 5963-1985 define indexação como “o ato de descrever ou identificar um documento em termos de seu conteúdo”. Essa definição pode ser completada apontando que às vezes os conceitos identificados e extraídos são normalizados e controlados por meio de um vocabulário controlado e, então, obtém-se uma indexação em linguagem controlada, caso contrário, obtém-se uma indexação em linguagem natural (GIL LEIVA, 2017).

Por outro lado, é necessário acrescentar que a indexação é um processo documental executado tanto em documentos (indexação de documentos) que produz um conjunto de palavras-chave ou termos padronizados que representam todo o conteúdo dos documentos, quanto também é executado (ciente ou inconscientemente) nas questões ou necessidades de informação (indexação das necessidades de informação) originando uma ou várias palavras em linguagem natural que também podem ser traduzidas para um vocabulário controlado e assim ser capaz de interrogar uma base de dados.

Assim, o objetivo geral da indexação é o armazenamento e recuperação da informação, visto que intervém tanto no processo de armazenamento da informação como na sua posterior recuperação nas bases de dados (GIL LEIVA, 2008).

O ato de indexar requer que o documento seja analisado pelo profissional indexador e que as melhores estratégias de busca sejam estudadas a fim de suprir as necessidades de informação dos usuários daquela determinada unidade informacional para qual a indexação está direcionada.

Na indexação, a aplicação da tecnologia é delineada 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”.

A automatização da indexação se deu início na segunda metade do século XX. Remonta aos últimos anos da década de 50 com os trabalhos de Luhn (1957). 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).

Sendo assim, a indexação automática é realizada por meio de programas de computador que podem contribuir para agilizar o processo de indexação, economizando tempo e trabalho ao profissional indexador.

A indexação automática é amplamente adotada pelos sistemas de recuperação da informação, utilizados em bibliotecas, centros de documentação e bases de dados para a construção de índices para busca. Entretanto, 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 sistemas 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).

Para 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).

Os autores nos levam a compreensão de que a automatização da indexação se apresenta como uma ferramenta de auxílio na prática do profissional da informação, e atua contribuindo para que a indexação possa ser realizada de maneira otimizada. Portanto, agilizar o processo intelectual realizado pelos profissionais da área é principal objetivo da indexação automática.

O sistema ou programa de computador que realiza o processo de indexação automática é denominado sistema de indexação automática. Os sistemas de indexação automática podem realizar a indexação automática por extração ou por atribuição, propondo respectivamente palavras-chave do texto do documento ou descritores de vocabulário controlado como termos de indexação.

Nas subseções que se seguem são descritos os respectivos sistemas de indexação automática que serão avaliados no presente artigo. Uma discussão mais detalhada acerca das características desses sistemas pode ser encontrada no trabalho de Silva, S. e Corrêa, R. (2020).

2.1.1 SISA

O sistema SISA (do espanhol *Sistema de Indización Semi-Automático*) foi desenvolvido no ano de 1999-2008 pelo professor e pesquisador Isidoro Gil Leiva, da Universidade de Múrcia, na Espanha. O SISA foi proposto originalmente, em sua tese de doutorado em 1997, como um sistema de indexação automático 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).

Os requisitos de entrada de dados do SISA são descritos a seguir:

- 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.

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.

De acordo com Narukawa, Gil 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#;
- 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 ocorrência dos termos nas fontes para que o sistema possa selecionar os termos de indexação que representarão o conteúdo do documento.

O SISA realiza a indexação automática de um texto e lista os descritores atribuídos automaticamente. Caso se deseje realizar a indexação semiautomática, é necessário que haja a intervenção humana para analisar e editar 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).

2.1.2 MAUI

O sistema MAUI (do inglês *Multi-purpose Automatic Topic Indexing*) foi desenvolvido por Alyona Medelyan como parte de seu projeto de doutorado, sob a supervisão de Ian H. Witten e Eibe Frank, no Departamento de Ciência da Computação da Universidade de Waikato, Nova Zelândia em 2009.

Quanto a representação do conteúdo dos documentos, o software faz uso de métodos estatísticos e linguísticos para a ponderação de termos com base no cálculo de características para cada termo baseados na: Ocorrência, Comprimento, Probabilidade de ser palavra-chave, e Relações semânticas.

O software permite executar as seguintes tarefas (MEDELYAN, 2009): 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.

O MAUI apresenta ainda, 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. Entretanto, exige que seja fornecido como entrada um conjunto de treinamento, que consiste em documentos e respectivos termos da indexação intelectual.

Os requisitos de entrada de dados do MAUI são descritos a seguir:

- Tesouro no formato SKOS: arquivo texto contendo os termos autorizados e seus termos não preferidos, bem como as relações semânticas entre termos autorizados;
- Radicalizador de palavras para o idioma do texto dos documentos: componente de software que reduz as palavras a uma aproximação do radical das mesmas;
- Lista pré-estabelecida de palavras vazias do idioma do texto dos documentos: arquivo texto contendo palavras vazias (stopwords) para fins de eliminação, como conectivos e artigos;
- Um conjunto de treinamento: um conjunto de documentos e respectivos termos de indexação para treinamento de um modelo de aprendizagem de máquina.
- Todos os arquivos de entrada, incluindo os textos a serem indexados, devem estar no formato txt.

Após o treinamento do modelo, os arquivos utilizados no processamento do MAUI são: o texto completo; o modelo treinado; uma lista de palavras vazias; o radicalizador de palavras; e um tesouro.

O MAUI realiza a indexação seguindo as seguintes etapas de processamento, de acordo com Medelyan (2009):

1. Geração de tópicos candidatos – extração de termos candidatos à indexação;
2. Cálculo de características – cálculo de características para os termos candidatos;
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 qualidade do conjunto de treinamento fornecido para o algoritmo de aprendizado de máquina envolvido é a chave para melhor qualidade na indexação automática, seja pela extração automática de palavras-chave ou atribuição de descritores de vocabulário controlado.

2.2 Avaliação de Sistema de Indexação Automática

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).

Para Papagiannopoulou e Tsoumakas (2019), a avaliação da extração de frases-chave ou indexação automática não é uma tarefa simples, pois apresenta dificuldades como a subjetividade e a falta de padronização dos métodos de avaliação. Os autores discorrem acerca da indexação automática e discutem os diferentes tipos de avaliação da indexação automática reportadas na literatura científica.

Uma estrutura abrangente de avaliação da indexação automática proposta por Golub et al. (2016) envolve três formas principais, das quais no presente trabalho, é especializada a primeira: avaliar a qualidade da indexação diretamente por meio da avaliação de um avaliador ou por meio da comparação com um padrão ouro (termos da indexação intelectual).

O presente artigo visa particularizar e adequar a estrutura abrangente de avaliação da indexação automática, bem como testá-la empiricamente, no contexto específico da avaliação de sistemas de indexação automática por atribuição.

De acordo com Gil Leiva (2008), a avaliação da indexação pode ser realizada de forma intrínseca (qualitativa ou quantitativa) ou extrínseca (de forma quantitativa por meio da interconsistência ou recuperação da informação).

A avaliação intrínseca quantitativa é 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.

Na avaliação extrínseca, cuja natureza é quantitativa, se comparam os resultados de indexações realizadas por indexadores diferentes, consistindo na comparação da indexação de dois indexadores ou unidades de informação que tenham indexado os mesmos documentos.

A avaliação da indexação automática ocorre por meio da avaliação extrínseca, onde são comparados termos da indexação automática com os termos da indexação intelectual.

Na avaliação da indexação automática, a avaliação extrínseca pode ocorrer de forma direta ou de forma indireta. De forma direta, se mede o casamento dos termos da indexação automática com termos da indexação intelectual, isto é, a recuperação de termos da indexação intelectual

pelo sistema de indexação automática. De forma indireta, se mede a qualidade na recuperação da informação com relação ao teste da função da indexação na recuperação de documentos. Portanto, a avaliação extrínseca direta, pode ser obtida por meio da interconsistência e recuperação de termos (GOLUB et al., 2016), e de forma indireta por meio da recuperação da informação (GIL LEIVA, 2017).

A avaliação extrínseca direta consiste na obtenção de valores para métricas como índice de consistência, precisão, revocação e medida F. Tais índices são obtidos por meio de fórmulas e buscam mensurar a qualidade da indexação automática em relação a indexação intelectual.

Para Gil Leiva (2017a), a avaliação extrínseca indireta, isto é, por meio da recuperação da informação, consiste em interrogar duas bases de dados ou índices de busca, submetendo um conjunto de expressões de busca, sendo os índices construídos respectivamente com os termos de duas indexações geradas por indexadores diferentes para os mesmos documentos. Depois, em examinar os resultados das buscas quanto a recuperação de documentos relevantes, e calcular os índices de revocação, precisão e medida F.

As subseções a seguir tratam da avaliação extrínseca direta da indexação automática, por ser a modalidade adotada na composição do percurso metodológico proposta neste artigo.

2.2.1 Avaliação extrínseca direta da indexação automática por meio da consistência

De acordo com Gil Leiva (2008, p. 386), pode-se obter o grau de consistência por meio das fórmulas de Hooper (1965) e Rolling (1981). Gil Leiva (1999, 2003 e 2008) se baseia na fórmula de Hooper (1965) 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;

B: Número de termos usados na indexação B.

Este índice varia entre 0 e 1, e a aplicação da fórmula pode ser obtida de duas formas: relaxada e rígida. No presente artigo utiliza-se a comparação rígida entre termos de indexação, em detrimento da comparação relaxada ou parcial dos termos de indexação.

O índice de consistência é uma medida relativa da qualidade na indexação e bastante útil para comparar indexações periódicas de um indexador em uma unidade de informação (intraconsistência) ou indexações de diferentes indexadores (interconsistência). Ele reflete a semelhança ou diferença entre várias indexações. Embora a indexação possa ser consistentemente ruim e consistentemente boa quando se avalia a indexação intelectual, esta métrica reflete melhor a qualidade na indexação automática, dada a garantia de boa qualidade da indexação intelectual usada como padrão de referência.

Na avaliação extrínseca por meio da interconsistência é aplicada a mesma fórmula da consistência, sendo que, a diferença consiste em comparar as indexações de dois indexadores, que tenham indexado o mesmo documento e dispo de as 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).

A consistência foi utilizada na avaliação extrínseca direta da indexação automática nos trabalhos de Narukawa, Gil Leiva e Fujita (2009) e Bandim e Correa (2018, 2019). Narukawa (2011, p. 128) aponta que a maior parte dos estudos sobre consistência na indexação alcançam valores de consistência entre 10% e 60%.

O desempenho de sistemas de indexação automática pode ser categorizado mediante o intervalo dos valores médios de consistência obtidos, de acordo a categorização dos níveis de consistência por 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 para categorizar o desempenho de sistemas de indexação automática.

2.2.1 Avaliação extrínseca direta da indexação automática por meio da recuperação de termos da indexação intelectual

A indexação automática pode ser avaliada quanto à atribuição ou recuperação de termos relevantes (termos também presentes na indexação intelectual) na indexação de um documento, podendo nessa perspectiva serem calculadas as métricas de revocação (R), precisão (P) e medida F.

No artigo de Bandim e Correa (2019), os autores avaliaram a qualidade na indexação automática de forma direta por meio da recuperação de termos da indexação intelectual, calculando os índices de consistência, precisão, revocação e medida F, segundo as seguintes fórmulas:

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

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

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

Bandim e Correa (2019, p. 5) consideram que a indexação automática 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”. Os autores apontam que os índices de revocação, precisão e medida F são fundamentais para qualificar a indexação automática.

O índice de revocação é obtido por meio da relação entre os termos relevantes atribuídos pelo sistema 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 pelo sistema 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).

Silva, B. e Correa, R. (2020) também utilizaram tais fórmulas para medir a qualidade da indexação realizada por indexadores-usuários em relação à indexação realizada pelos autores de artigos científicos.

Esta forma de cálculo da precisão, revocação e medida F é adotada como padrão em conferências de avaliação da indexação automática ou extração de palavras-chave (KIM et al. 2013) (HASAN; NG, 2014).

3 METODOLOGIA

A pesquisa classifica-se como exploratória e empírica, envolvendo um estudo de caso e coleta de dados por meio de um experimento de indexação.

Quanto aos objetivos 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), trazendo a formulação de hipóteses significativas para pesquisas posteriores (CERVO; BERVIAN; SILVA, 2007).

O caráter teórico deste artigo é composto por pesquisa bibliográfica e revisão da literatura em Ciência da Informação sobre avaliação da indexação automática e indexação automática por atribuição. Em termos pragmáticos, a pesquisa classifica-se em experimento e estudo de caso visando propor e aplicar um percurso metodológico para avaliar de forma direta e conjunta a qualidade da indexação realizada pelos sistemas de indexação automática por atribuição SISA e MAUI.

Este trabalho utiliza como instrumentos de pesquisa: o corpus de referência em Ciência da Informação composto de 60 artigos científicos selecionados na tese de doutorado de Souza (2005), cujos textos dos artigos serão submetidos aos sistemas SISA e MAUI; o Tesauro Brasileiro de Ciência da Informação (TBCI) usado como linguagem de indexação; e o percurso metodológico proposto para avaliação direta e conjunta da indexação automática por atribuição.

3.1 Percurso metodológico para avaliação direta e conjunta de sistemas de indexação automática por atribuição

A qualidade da indexação é medida por meio do cálculo das métricas de consistência, revocação, precisão e medida F.

Tais métricas serão calculadas no viés de uma avaliação extrínseca direta da indexação automática. O padrão de qualidade ou referência adotado consiste dos termos da indexação intelectual para um conjunto de documentos, denominado de corpus de referência.

Cada sistema de indexação automática será avaliado na indexação automática por atribuição do corpus de referência, e fará uso da mesma linguagem de indexação usada na indexação intelectual.

Tal configuração do percurso metodológico de avaliação direta e conjunta de sistemas de indexação automática por atribuição permite que os valores das métricas obtidos pelos sistemas possam ser diretamente comparadas.

O percurso metodológico proposto foi dividido em sete etapas, proporcionando assim uma melhor visualização e planejamento acerca dos procedimentos necessários. As etapas são descritas a seguir:

- **Etapa 1** – Selecionar os sistemas de indexação automática por atribuição a serem avaliados, bem como um tesouro do domínio dos textos dos documentos a serem indexados automaticamente, e construção ou seleção de um corpus de referência contendo os documentos e respectivos termos das indexações intelectuais;
- **Etapa 2** - Preparar os arquivos de entrada contendo o tesouro e os textos dos documentos do corpus de referência para submeter como entrada aos sistemas de indexação automática de acordo com as especificações e configurações de cada sistema;
- **Etapa 3** – Realizar o processo de indexação automática por atribuição por meio de cada sistema, indexando de forma automática os documentos do corpus pelos sistemas de indexação automática por atribuição utilizando o tesouro como linguagem de indexação;
- **Etapa 4** - Criar uma lista de descritores atribuídos por cada sistema a cada documento do corpus;
- **Etapa 5** - Calcular os valores dos índices de consistência, revocação, precisão e medida F alcançados por cada sistema na indexação de cada documento, tendo como referência os descritores da indexação intelectual, por meio de uma avaliação extrínseca direta da qualidade da indexação automática;
- **Etapa 6** - Avaliar e comparar o desempenho na qualidade da indexação automática de cada um dos sistemas por meio do resultado dos índices médios obtidos de consistência, revocação, precisão, e medida F;
- **Etapa 7** - Analisar comparativamente a qualidade dos descritores propostos pelos sistemas para cada artigo.

Na parte empírica da pesquisa, o percurso metodológico proposto será aplicado no cálculo das métricas de consistência, precisão, revocação e medida F, a fim de avaliar e comparar a qualidade na indexação automática dos sistemas SISA e MAUI. Os sistemas farão uso do Tesouro Brasileiro

de Ciência da Informação (TBCI) como linguagem de indexação. O corpus de referência utilizado no experimento foi construído por Silva, B. e Corrêa, R. (2020), por meio da indexação intelectual de usuários-indexadores que utilizaram o TBCI como linguagem de indexação.

Na definição do processo de indexação para os sistemas SISA e MAUI, foram utilizados os seguintes arquivos em comum, embora com formatação específica para cada software, contendo: texto completo de cada artigo, lista de palavras vazias, e tesouro. O MAUI utilizou como entrada adicionalmente: um radicalizador de palavras para o idioma português do Brasil e o conjunto de treinamento do modelo.

O conjunto de treinamento do MAUI foi definido como sendo os 30 primeiros documentos do corpus de referência, estando esse valor próximo ao sugerido por Medelyan (2009). Consequentemente, o conjunto de teste foi definido como sendo os 30 últimos documentos do corpus de referência. A necessidade de definição do conjunto de treinamento para o software MAUI, traz influência na etapa 6 do percurso metodológico, pois o desempenho no subconjunto de teste é considerado mais representativo do desempenho dos sistemas em ambiente operativo, sendo este ponto de discussão retomado na subseção 4.3.

3.2 Tesouro Brasileiro de Ciência da Informação

Nesta pesquisa utiliza-se como linguagem de indexação, o Tesouro Brasileiro de Ciência da Informação (TBCI). O TBCI será aplicado aos sistemas de indexação automática por atribuição SISA e MAUI, portanto, torna-se necessário discorrer acerca de seu uso como instrumento para o controle de vocabulário.

O TBCI é resultado de pesquisas acerca do domínio da Ciência da Informação, visando oferecer aos profissionais da informação um instrumento fundamental para a consistência de terminologia e de vocabulário na área de Ciência da Informação (PINHEIRO; FERREZ, 2014).

O tesouro compreende, em torno 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 na área de Ciência da Informação no Brasil e em outros países. Com relação ao uso, o tesouro é direcionado aos indexadores, professores, pesquisadores e qualquer profissional da informação.

O TBCI será aplicado como linguagem de indexação no processo de indexação automática por atribuição do corpus de referência.

3.3 Corpus de referência em Ciência da Informação

Este artigo utiliza um corpus de referência para fins de avaliação de sistemas de indexação automática.

O corpus de referência se constitui em um conjunto de documentos categorizados ou indexados de forma intelectual, visando ser fonte para estudos analíticos ou de avaliação de qualidade na indexação automática. Os documentos que compõem um corpus podem ser reunidos a partir de publicações de natureza diversas, de caráter geral como notícias de jornais, páginas web, e-mails, etc., ou caráter especializado como teses e dissertações, anais de congresso e artigos de periódicos científicos de uma área de especialidade.

Nos artigos de Silva, B. e Corrêa, R. (2019, 2020), os autores propuseram a aplicação de um percurso metodológico para a construção de corpus de referência na área de Ciência da Informação. Os resultados da pesquisa apontaram que o percurso metodológico possui eficácia para construção do corpus de referência em Ciência da Informação, e pode ser adaptado para construção de corpus de referência em outras áreas do conhecimento. Nesta pesquisa, será adotado o corpus de referência resultante do trabalho de Silva, B. e Corrêa, R. (2020).

O corpus de referência utilizado neste artigo, consistem dos artigos de periódicos selecionados por Souza (2005), composto de 60 artigos da área da Ciência da Informação escritos em português. O corpus compreende artigos das revistas científicas *Datagramazero* e *Ciência da Informação*. Os documentos que compõem o corpus, em sua totalidade, abrangem os campos científicos da Biblioteconomia, Documentação e Ciência da Informação. Os documentos do corpus de referência foram indexados por usuários-leitores usando o TBCI como linguagem de indexação, formando uma Folksonomia.

Os documentos do corpus foram utilizados em algumas pesquisas, por instância os trabalhos de Bandim e Correa (2018, 2019) que utilizaram os documentos do corpus para avaliar a qualidade do processo de indexação automática por atribuição do SISA por meio dos índices de consistência, revocação, precisão e medida F. Esses trabalhos anteriores se assemelham a presente pesquisa por realizarem a indexação automática por atribuição dos mesmos textos e fazerem uso das mesmas métricas de avaliação, porém se diferencia pois esta pesquisa faz uso de um corpus de referência onde os documentos foram indexados por usuários-leitores usando o TBCI, ao invés do uso das palavras-chaves dos autores. Este trabalho, também se diferencia por avaliar de forma direta e conjunta mais de um sistema de indexação automática por atribuição.

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 o percurso metodológico de avaliação de sistemas de indexação automática.

4 RESULTADOS

A seguir apresenta-se a descrição dos resultados obtidos pelos sistemas SISA e MAUI. Em seguida, apresenta-se a comparação dos resultados da indexação automática na avaliação conjunta do desempenho dos sistemas.

4.1 Indexação automática do SISA

Para analisar a indexação automática do SISA, foi construída uma planilha constituída dos seguintes campos dispostos na Tabela 1: Número de termos do SISA – número de termos atribuídos automaticamente a cada artigo pelo SISA; Número de termos da Folksonomia – número de termos da indexação intelectual atribuídos a cada artigo; Número de termos comuns com a Folksonomia – quantidade de termos em comum nas indexações; e valores das métricas de consistência, precisão, revocação e medida F.

Tabela 1 - Resultados do SISA

Nº Artigo	Número de termos do SISA	Número de termos da folksonomia	Número de termos comuns	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 (2020).

Por meio da Tabela 1, 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, isto é, 5 a 6 termos atribuídos em média pelo software a cada documento.

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, isto é, 10 ou 11 termos em média para cada documento.

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, isto é 3 ou 4 termos em comum em média por documento entre as indexações intelectual e automática.

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 consistência na indexação apontam que os índices de consistência obtidos, estão 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 valor esperado.

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, tais valores foram menos satisfatórios, tendo em vista que os índices médios de precisão, revocação e medida F no presente artigo, foram de 64%, 34% e 42%, respectivamente.

4.2 Indexação automática do MAUI

Para analisar a indexação automática do MAUI foi construída uma planilha constituída dos seguintes campos dispostos na Tabela 2, nos mesmos moldes da Tabela 1: Número de termos do MAUI; Número de termos da Folksonomia; Número de termos comuns com a Folksonomia; Consistência; Precisão; Revocação; e Medida F.

Tabela 2 - Resultados do MAUI

Nº Artigo	Número de termos do MAUI	Número de termos da folksonomia	Número de termos comuns	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%

AVALIAÇÃO DIRETA E CONJUNTA DE SISTEMAS DE INDEXAÇÃO AUTOMÁTICA POR ATRIBUIÇÃO

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 (2020).

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, que corresponde a 9 a 10 termos atribuídos em média pelo software por documento.

No campo Número de termos da folksonomia, a quantidade de termos atribuídos foi de um mínimo de 7 a um máximo de 15 termos, gerando uma média de 10,35 termos, que corresponde de 10 a 11 termos de indexação por documento.

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, ou seja, em média 5 a 6 termos comuns entre as indexações por documento.

Foi verificada uma média de 36% no índice de Consistência, 54% no 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 de acordo a categorização dos níveis de consistência de Bandim e Corrêa (2018).

4.3 Avaliação conjunta da indexação automática do SISA e MAUI

Neste experimento, foram obtidos os índices médios de consistência, precisão, revocação e medida F, a fim de qualificar e comparar o desempenho na indexação automática do SISA e do MAUI. Os valores médios das métricas foram calculados para o conjunto de teste e para todo o corpus de referência.

Os valores médios das métricas obtidos para o conjunto de teste do corpus de referência por cada sistema podem ser observados na Tabela 3.

Tabela 3 - Resultados médios no conjunto de teste obtidos por SISA e MAUI (30 documentos)

Sistema	Termos comuns	Consistência	Precisão	Revocação	Medida F
SISA	3,87	32%	63%	40%	48%
MAUI	5,33	38%	54%	55%	54%

Fonte: Dados da pesquisa (2020).

De acordo com os valores médios apresentados pelos sistemas para as métricas no conjunto de teste, isto é, nos últimos 30 documentos do corpus de referência que não foram utilizados na aprendizagem de máquina, observa-se uma melhor qualidade na indexação automática do software MAUI, tendo maiores médias para termos comuns com a indexação intelectual, maior consistência média com a indexação intelectual, maior média de revocação e medida F na atribuição de termos da indexação intelectual. O software SISA apresenta maior média na métrica de precisão, chegando a alcançar 9% a mais na média de precisão que o MAUI, entretanto apresenta uma média de revocação 15% mais baixa que o MAUI, levando o mesmo a obter também uma menor medida F, além de menor médias do índice de consistência e número de termos comuns.

Assim, avaliando de forma direta e conjunta a qualidade na indexação automática, percebe-se o sistema MAUI com melhor eficácia que o sistema SISA. Bem como é possível observar uma significativa tendência do SISA em ser mais preciso na atribuição de termos de indexação, e uma expressiva tendência do MAUI em ser mais exaustivo na atribuição de termos da indexação intelectual.

Embora a análise do desempenho na indexação automática no conjunto de teste seja suficiente para comparar de forma imparcial os sistemas MAUI e SISA, reporta-se a seguir uma análise do comportamento dos sistemas em todo o corpus de referência, isto para fins de maior conhecimento sobre a natureza do corpus de referência e a comparação com resultados de trabalhos anteriores. A Tabela 4 mostra os valores médios das métricas obtidos por cada sistema em todo o corpus.

Tabela 4 - Resultados médios em todo o corpus obtidos por SISA e MAUI (60 documentos)

Sistema	Termos comuns	Consistência	Precisão	Revocação	Medida F
SISA	3,35	28%	64%	33%	42%
MAUI	5,15	36%	54%	51%	52%

Fonte: Dados da pesquisa (2020).

Comparando as Tabelas 3 e 4, percebe-se que os valores médios são próximos, embora inferiores para os dois sistemas em todo o corpus, isto é, médias menores para as métricas calculadas em todo o corpus. Tal fato mostra que os primeiros 30 documentos do corpus são um pouco mais difíceis de indexar, e que apesar do MAUI ter sido treinado nesse subconjunto, o desempenho desse em todo o corpus alcança uma qualidade próxima, embora um pouco inferior em média. Apesar de serem um pouco mais difíceis de indexar, o treinamento do modelo via aprendizagem de máquina no MAUI permitiu gerar modelo que generaliza os exemplos de treinamento, levando o MAUI a obter um desempenho melhor na indexação automática dos documentos no conjunto de teste, documentos não apresentados anteriormente.

Os valores médios alcançados de 28% e 36% dos índices de consistência pelo SISA e MAUI respectivamente em todo o corpus, 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% na categorização de Bandim e Corrêa (2018) do desempenho quanto aos níveis de consistência.

Com relação ao índice de precisão, o SISA apresentou uma média de 64%, 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 na precisão. A precisão mede a capacidade do sistema em atribuir univocamente ou exclusivamente os termos da indexação intelectual.

Com relação ao índice de revocação, o SISA apresentou uma média de 33% e o MAUI de 51%. Sendo assim, o MAUI apresentou uma revocação média maior dos termos da indexação intelectual, recuperando metade dos termos de indexação, em detrimento de uma menor precisão média que a do SISA. A revocação mede a capacidade do sistema em atribuir todos os termos da indexação intelectual.

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. Para se conseguir valores médios altos para medida F é necessário que se tenha valores altos para revocação e precisão (BANDIM; CORRÊA, 2019).

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), a medida F mais perto de 100% indica maior qualidade da indexação, sendo assim, o MAUI se mostrou mais próximo do 100% e, portanto, apresentou maior qualidade na indexação automática.

5 CONCLUSÃO

Os valores médios obtidos por ambos os sistemas em todos os documentos do corpus de referência foram considerados bons. O SISA obteve as médias de 28% no índice de consistência, 64% no índice de precisão, 33% 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.

O percurso metodológico proposto se mostrou adequado quando aplicado no experimento de avaliação direta e conjunta dos sistemas de indexação automática MAUI e SISA na indexação automática por atribuição de artigos científicos da área de Ciência da Informação.

Adicionalmente, a pesquisa buscou fornecer informações acerca dos sistemas SISA e MAUI, com o intuito de disponibilizar informações relacionadas ao processo de avaliação da indexação automática por atribuição, bem como contribuir para a temática na área da Ciência da Informação.

Considerando a importância em dar continuidade aos estudos acerca da indexação automática por atribuição, a pesquisa sugere novos experimentos com a avaliação extrínseca indireta dos sistemas, bem como a avaliação em outro corpus de referência, a fim de aprofundar os conhecimentos relacionados aos softwares e seu aprimoramento. Cabe ressaltar que, o MAUI possui a consistência próxima à indexação humana, portanto seria interessante comparar também a indexação automática do software com a indexação semiautomática usando o mesmo.

No tocante ao SISA, a pesquisa sugere experimentos com versões mais recentes do software que englobe aspectos morfológicos e semânticos dos termos.

Considerando que a pesquisa tem como foco principal avaliar de forma direta e conjunta os sistemas de indexação automática, esperamos que seus resultados contribuam para a área da CI, no sentido de elucidar meios de avaliar a qualidade da indexação automática e refletir, com base nos resultados obtidos, quanto a aplicação ou não dos sistemas avaliados nas unidades de informação, além

da compreensão de que a avaliação dos sistemas é imprescindível para aplicação e o aprimoramento dos sistemas.

REFERÊNCIAS

- ANDERSON, J.D.; PEREZ-CARBALLO, J. The Nature of Indexing: How Humans and Machines Analyze Messages and Texts for Retrieval. Part I: Research and the Nature of Human Indexing. **Information Processing & Management**, v. 37, p. 231-54, 2001.
- BANDIM, M. A. S.; CORREA, 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: 10.5007/1518-2924.2018v23n53p64.
- BANDIM, M. A. S.; CORREA, R. F. Indexação automática por atribuição de artigos científicos em português da área de Ciência da Informação. **Transinformação**, v. 31, p. 1-12, 2019. DOI: 10.1590/2318-0889201931e180004.
- CAJUEIRO, R. L. P. **Manual para elaboração de trabalhos acadêmicos**: guia prático do estudante. 3. ed. Petrópolis: Vozes, 2015. 110 p.
- FARROW, J.F. A Cognitive Process Model of Document Indexing. **Journal of Documentation**, v. 47, p. 149-66, 1991.
- FROHMANN, B. Rules of Indexing: A Critique of Mentalism in Information Retrieval Theory. **Journal of Documentation**, v. 46, n.2, p. 81-101, 1990.
- FUGMANN, R. **Subject Analysis and Indexing**: Theoretical Foundation and Practical Advice. Frankfurt/Main: Indeks Verlag, 1993.
- CERVO, A. L.; BERVIAN, P. A.; SILVA, R. **Metodologia científica**. 6. ed. São Paulo: Pearson Prentice Hall, 2007. 162 p.
- GIL LEIVA, I. La automatización de la indización de documentos. Gijón: Trea, 1999.
- GIL LEIVA, I. Sistema para la indización automática (SISA) de artículos de revista de Biblioteconomia y documentación. **II Jornadas de Tratamiento y Recuperación de Information**. Madrid: 2003.
- GIL LEIVA, I. **Manual de indización**: teoría y práctica. Gijón: Trea, 2008.
- 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. 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.

GOLUB, K., et al. A Framework for Evaluating Automatic Indexing or Classification in the Context of Retrieval. **Journal of the Association for Information Science and Technology**, v. 67, n. 1, p. 3-16, 2016.

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

HJORLAND, B. **Information Seeking and Subject Representation: An Activity-Theoretical Approach to Information Science**. Westport, CT: Greenwood Press, 1997.

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

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.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 5963-1985: Documentation -- Methods for Examining Documents, Determining Their Subjects, and Selecting Indexing Terms**. Geneva: ISO, 1985.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. 2. ed. ver. atual. 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, p. 309-17, 1957.

MAI, Jens-Erik. Deconstructing the Indexing Process. **Advances in Librarianship**, v. 23, p. 269-98, 2000.

MEDELYAN, O. **Human-competitive automatic topic indexing**. PhD Thesis. University of Waikato, New Zealand, 2009.

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

NARUKAWA, C. M.; GIL LEIVA, I.; FUJITA, M. S. L. Indexação automatizada de artigos de periódicos científicos: análise da aplicação do software sisa com uso da terminologia decs na área de odontologia. **Informação & Sociedade: Estudos**, v. 19, n. 2, 2009.

PAPAGIANNPOULOU, E; TSOUMAKAS, G. A Review of keyphrase extraction. **ArXiv**, abs/1905.05044, 2019. Disponível em: <https://arxiv.org/pdf/1905.05044.pdf> , acesso em: 10 de jun, 2019.

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

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.

SILVA, B. F. M; CORREA, R. F. O processo de construção de corpus de referência em Ciência da Informação. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, v. 24, p. 1-27, 2019. DOI: 10.5007/1518-2924.2019.e65166.

SILVA, B. F. M; CORREA, R. F. A aplicação da folksonomia assistida na construção de corpus de referência em Ciência da Informação. **Em Questão**, v. 26, n. 2, p. 413-436, 2020. DOI: 10.19132/1808-5245262.413-436.

SILVA, S. R. B; CORREA, R. F. Sistemas de indexação automática por atribuição: uma análise comparativa. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, v. 25, p. 1-25, 2020. DOI: 10.5007/1518-2924.2020.e70740. Acesso em: 16 jul. 2020.

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.

KIM, S. N; MEDELYAN, O.; KAN, M.Y.; BALDWIN, T. 2013. Automatic keyphrase extraction from scientific articles. **Language Resources and Evaluation**, v.47, n.3, p. 723–742.