
USO DE INDICADORES E MÉTRICAS PARA AVALIAÇÃO DA QUALIDADE DA INFORMAÇÃO

Use of indicators and metrics for assessing information quality

(1) **Josefina Aparecida Soares Guedes**, (2) **Rita de Cássia Fonseca**,
(3) **Faimara do Rocio Strauhs**

Universidade Federal do Paraná, Brasil, josefinaguedes@yahoo.com.br.

Universidade Estadual do Centro Oeste, Brasil, ritadecfonseca@gmail.com.

Universidade Tecnológica Federal do Paraná, Brasil, faimara@utfpr.edu.br.



Resumo

A qualidade de informações fornecidas varia amplamente e a expressiva quantidade de informações acessíveis dificulta o acesso à informação relevante para pesquisadores e outros consumidores de informação. O objetivo deste estudo foi identificar o uso de indicadores e de métricas para avaliar a qualidade de dados e de informações a partir da literatura branca especializada disponível nas bases de dados da Proquest, da Web of Science e da Scopus. O Protocolo de Recomendação Prisma e a Análise de Conteúdo foram utilizados para identificar construtos, métricas e indicadores usados. Os resultados apontaram o uso mais frequente dos indicadores de precisão, de confiabilidade e de completude. Houve discrepância e sobreposição de conceitos na definição dos indicadores, associadas ao tipo de indicador e ao objeto de avaliação, ou seja, dado ou informação. As métricas de avaliação, de contexto e de conteúdo foram utilizadas para avaliar a qualidade de dados, fontes de dados e informação de documentos. Conclui-se que a escolha de indicadores e de métricas para análise da qualidade de informação depende da tarefa e das preferências subjetivas do usuário da informação, sendo preferível usar métricas de avaliação rastreáveis, simples e de fácil compreensão.

Palavras-chave: Indicadores de qualidade da informação; Métricas; Dimensões de qualidade da informação; Informação.

Abstract

The quality of the information provided varies widely and the significant amount of accessible information makes it difficult for researchers to access relevant information. The aim of this study was to identify the use of indicators and metrics to assess the quality of data and information from white specialized literature available in the databases of Proquest, Web of Science and Scopus. The Prisma Recommendation Protocol

and Content Analysis were used to identify the constructs, metrics and indicators used. The results showed the most frequent use of indicators of precision, reliability and completeness. There is a discrepancy and overlap of concepts in the definition of indicators that are associated with the type of indicator and the object of evaluation, that is, data or information. Assessment, context and content metrics were used to assess the quality of data, data sources and document information. The choice of indicators and metrics for analyzing the quality of information depends on the task and the subjective preferences of the information user, being preferable to use traceable, simple and easy-to-understand evaluation metrics.

Keywords: Information quality indicators; Metrics; Dimensions of information quality; Information.

1 Introdução

A disponibilização exponencial de fontes de dados e informações na Web facilita a aquisição destes para criação de conhecimento, entretanto, a qualidade da informação pode se tornar um entrave diante da dificuldade de reconhecer a informação útil que agregue valor (Han et al. 2015; Reznik e Lyshevski 2015; Bibi et al. 2016).

Para serem realmente valiosos, tanto dados como informações devem possuir várias características essenciais, como ser completos, exatos, atualizados e adequados ao propósito para o qual serão utilizados. Essas características se traduzem em indicadores e dimensões que são utilizadas para avaliar a qualidade de dados e informações (Turban et al. 2004).

A qualidade da informação é entendida como “aptidão para uso” (Juran et al. 1974) ou estar “adequada para o uso” (Wang e Strong 1996; Naumann 2002; Mohammadi et al. 2017; Cichy e Rass 2020; Härting e Lewoniewski 2020). Entretanto, esses direcionamentos são subjetivos, pois, uma informação pode ser considerada de qualidade e adequada para um indivíduo em um dado momento e inadequada e dispensável em outras tarefas. Ainda, um segundo usuário pode considerar a qualidade da mesma informação inapropriada para suas tarefas (Bizer e Cyganiak 2009; Mohammadi et al. 2017; Cichy e Rass 2020; Härting e Lewoniewski 2020). Assim, a qualidade da informação é estabelecida a partir do ponto de vista do usuário e a escolha dos indicadores para métricas dependerá da valoração das dimensões da informação em determinado contexto de aplicação (Härting e Lewoniewski 2020).

De outra parte, a categorização da informação em dimensões é um método adotado para classificar a informação. Wang e Strong (1996), Knight e Burn (2005) e Greef et al. (2012) propuseram diferentes construtos para avaliar a qualidade da informação, segundo a categorização.

Dentre as dimensões e indicadores apontados por esses autores figuram: precisão, clareza, concisão, confiabilidade, completude, consistência, relevância, disponibilidade, Integridade, utilidade.

Diante da diversidade de indicadores e de métricas disponíveis torna-se importante saber quais destes seriam mais usuais e que possibilitariam oferecer a qualidade desejada da informação em diversos contextos.

Considerando que a qualidade da informação ainda é pouco explorada (Ayyash 2015; Bibi et al. 2016; Azeroual et al. 2020), este estudo teve por objetivo identificar o uso de indicadores de qualidade da informação e de métricas para análise da qualidade de dados e informações utilizados em um grupo de artigos, recuperados a partir da literatura branca especializada em bases de dados da Web of Science, Proquest e Scopus.

2 Revisão da Literatura

No domínio da Ciência da Informação, estudos classificam a qualidade da informação a partir de duas perspectivas (Nehmy e Pahim 1998; Cichy e Rass 2020): (i) a do produto, que considera a informação como coisa e baseada em dimensões como acessibilidade, atualidade, confiabilidade, objetividade, precisão e validade; e (ii) a do usuário: com base na necessidade de informação, relevância, eficácia e utilidade.

Kopyltsov (2020) registra também que há várias interpretações para a qualidade, com destaque para a qualidade vista como uma categoria filosófica que é basicamente a certeza idêntica de ser e, de outra parte, estar associada ao atendimento de requisitos – ou ser utilizável. No entanto, esse autor afirma que a qualidade em qualquer definição tem um número infinito de gradações quantitativas e não existem qualidades em si, mas apenas coisas que possuem infinitas qualidades tal como os órgãos de sentido humano que permitem distinguir as coisas em um sentido qualitativo.

A noção de qualidade da informação é, portanto, ainda vaga, imprecisa, subjetiva e de potencial discussão (Nehmy e Paim 1998; Valls e Vergueiro 2006; Assis e Moura 2011; Fagundes e Ribeiro Junior 2020). De acordo com a norma ISO 9001, o conceito de qualidade compreenderia medição e controle, melhoria contínua, alcance de metas e objetivos e envolvimento da alta direção

organizacional nos processos gerenciais destas (Associação Brasileira de Normas Técnicas 2008). Então, o conceito de qualidade da informação poderia ser entendido como uma subconcepção específica do domínio do conceito geral de qualidade estabelecido pela referida norma (Bizer e Cyganiak 2009).

Diante da diversidade de dimensões que integram a informação, a qualidade desta pode ser também entendida como uma construção multidimensional, já que pode depender de vários indicadores (Arazy et al. 2017; Bizer 2007; Chen et al. 1998; Han et al. 2015; Cichy e Rass 2020; Härting e Lewoniewski 2020). O estabelecimento de quais dimensões da qualidade são relevantes e quais níveis de qualidade são necessários para cada dimensão é determinado pelos objetivos ou tarefas e pelas preferências subjetivas do consumidor de informações (Calazans 2008; Naumann 2002; Wang e Strong 1996; Cichy e Rass 2020).

Há que se considerar ainda a diferenciação entre dado e informação, a escolha de dimensões e de indicadores para avaliar dados pode não ser adequada para avaliar informações. Isto porque dados são definidos como fatos discretos e objetivos sobre eventos; já informação diz respeito a ordem das coisas, a ações ou fenômenos registrados de forma sistemática e que podem ser correlacionados com outros fenômenos ou ações (Davenport e Prusak 2003). Dados agrupados, classificados e formatados agregam valor e se tornam em informação.

Logo, Indicadores de qualidade de informação deveriam ser usados para auxiliar a identificação de informações relevantes. Lembrando que tais indicadores devem ser fundamentados em dimensões da informação tais como atualidade, precisão, clareza, relevância, entre outros (Knight e Burn 2005). A avaliação da qualidade da informação pode ser feita a partir de um conjunto de indicadores atrelados a uma ou a mais das dimensões da informação (Bizer e Cyganiak 2009). Nesse contexto, as métricas podem ser utilizadas para determinar com mais precisão a qualidade da informação combinando indicadores, heurísticas e algoritmos para quantificar a qualidade da informação utilizada em diferentes contextos.

A informação, justifica-se, permeia todos os processos das organizações, sendo concebida como a matéria prima-prima para gerar desenvolvimento organizacional (Davenport e Prusak 2003; Dante 2008; Paletta e Ueki 2019). Assim, o uso de informação é uma atividade contínua e que auxilia a sobrevivência no mercado competitivo, contribuindo para o desenvolvimento de

inovações em produtos e serviços (Choo 2003). Nesse processo, tanto as informações internas quanto as externas precisam ter qualidade (Sugahara e Jannuzzi 2005), já que as análises com base em valores de dados pobres, podem levar a decisões erradas nas instituições de pesquisa, comprometendo, entre outros aspectos, a reputação das mesmas (Azeroual et al. 2020; Sá e Martins 2016).

Para facilitar a seleção de informações relevantes para usuários e demandantes, e minimizar o problema de sobrecarga de informações, os sistemas de informação devem diferenciar os dados que podem ser seguramente resumidos dos dados que devem ser consultados em sua forma original (Dante 2008). Então, indicadores e métricas podem auxiliar o processo de estabelecer a pertinência e a qualidade de dados e de informações, utilizando-se tanto de métodos quantitativos como qualitativos para uma possível seleção (Rasaiah et al. 2015). O uso de indicadores qualitativos fornece subsídios para o pesquisador identificar a relevância ou confiabilidade da informação. Já as métricas propiciam uma análise aprofundada da qualidade da informação, por incluírem vários indicadores em um sistema de medição, para verificar se a informação em tela satisfaz a necessidade do pesquisador ou do consumidor daquelas informações (Bizer e Cyganiak 2009).

2.1 Indicadores

Na apresentação de medidas, os indicadores são unidades de informação, que sintetizam ou realçam características de um sistema, simplificando seus fenômenos (Van Bellen 2004). A utilização de indicadores contribui no desenvolvimento de tarefas de medição e análise, sendo útil para comunicar ideias, pensamentos e valores que podem auxiliar na tomada de decisão em diferentes contextos (Botelho et al. 2015; Rösch et al. 2017). No entanto, a subjetividade pode estar presente inclusive nas definições dos indicadores, dificultando a avaliação da informação. Além disso, indicadores são interdependentes entre si e normalmente é necessário um subconjunto formado de mais de um indicador para atestar a relevância de uma informação (Bizer e Cyganiak 2009).

Wand e Wang (1996) apresentaram indicadores da informação com base em conceitos ontológicos em que se sobressairam a acurácia, a consistência e a atualidade como sendo os mais citados na literatura. Detalhando alguns indicadores Wang e Strong (1996), destacaram que a

completude compreende até que ponto os dados têm amplitude, profundidade e escopo suficientes para a tarefa em questão. Neste alinhamento;

- o indicador de **precisão** se refere a até que ponto os dados são corretos, confiáveis e certificados;
- o indicador de **disponibilidade** está associado à disponibilização de dados em relação a sua data de publicação e se estes dados são apropriados para a tarefa em questão;
- o indicador de **consistência**, relaciona-se a apresentação do dado em mesmo formato e compatibilidade com os dados anteriores de um contexto analisado;
- o indicador de **acessibilidade**, relaciona-se a disponibilidade das informações ou se estas podem ser recuperadas de forma fácil e rápida.

Para Cichy e Rass (2020), os indicadores de qualidade da informação mais comumente citados incluem: integridade, precisão, atualidade, consistência e acessibilidade. Entretanto, dependendo do contexto e da perspectiva dos dados, pode haver outros indicadores de qualidade que devem ser considerados.

Avaliando os modelos de qualidade de informação para a Web, Parker et al. (2006) observaram que indicadores de acessibilidade e atualidade foram os mais recorrentes. Entretanto, os autores destacaram que além destes, no mínimo um modelo de avaliação deve apresentar as seguintes dimensões: exatidão, relevância, credibilidade, objetividade, completeza, adequação, representação, fonte e compreensividade.

Estudando a qualidade de sistemas de recuperação da informação, Azeroual et al. (2020), destacaram que o fracasso destes está associado a qualidade dos dados e à incompletude, à redundância e à falta de correção e atualização das informações coletadas e armazenadas.

Assis e Moura (2011) ressaltaram que a qualidade da informação na Web deve ser analisada como um processo e não como um recorte arbitrário e operacionalizável para uso em contextos e propósitos específicos sujeitos a desatualizações. Para as autoras, na avaliação da qualidade da informação deve ser levado em conta o ambiente interativo e colaborativo da Web em que o usuário da informação é ativo e estabelece práticas e trocas simbólicas constantes.

Em um estudo sobre os principais fatores que influenciam na determinação da qualidade de páginas colaborativas de dados abertos, Härting e Lewoniewski (2020) destacaram que a qualidade dos dados pode ser definida a partir de vários indicadores, corroborando com outros estudos que consideram a relevância dos indicadores conforme o tipo de seu uso (Arazy et al. 2017; Cichy e Rass 2020; Gigerenzer e Gaissmaier 2011).

A avaliação da relevância e da qualidade da informação consiste, então, no processo de verificar se uma informação satisfaz as necessidades do consumidor de informações em uma situação específica (Naumann 2002; Bizer 2007). A avaliação e a filtragem de informações podem depender de vários indicadores de qualidade, como já afirmado, bem como de metainformações, ou seja, informações sobre informações classificadas. Portanto, as estruturas de filtragem de informações precisam de meios flexíveis para representar informações junto com metainformações relacionadas à qualidade (Bizer e Cyganiak 2009).

2.2 Métricas

A capacidade de medir e de rastrear a qualidade da informação é fundamental para gerenciar e melhorar a qualidade da produção, da distribuição de informações e para a criação do conhecimento. Após a seleção dos indicadores, com base em características ou dimensões apropriadas, deve ser estabelecido um processo de medição desses indicadores (Bizer e Cyganiak 2009; Rasaiah et al. 2015). Esta tarefa é facilitada pelo uso de métricas de avaliação.

No campo da Ciência da Informação, a bibliometria, a cientometria, a informetria, a webometria e, mais recentemente, a altmetria são métricas utilizadas para avaliar documentos científicos. Estas métricas são compostas por indicadores que são avaliados a partir de aplicações matemáticas e estatísticas. Os indicadores bibliométricos e cientométricos, considerados mais tradicionais, possibilitam a análise estatística da produção científica incluindo análise de citações, índice H e fator de impacto de periódicos científicos. Já os indicadores altmétricos ampliam essa percepção de impacto para além das citações com análises de medições de impacto social do conhecimento científico (Oliveira e Araújo 2020, Maatouk 2021).

De acordo com Maatouk (2021), a altmetria aliada à inteligência artificial – AI e tecnologias semânticas podem ser usadas como métricas para análise de metadados e de

informações em casos de grande volume de dados como o *big data* e outras plataformas *on line*. Esse autor desenvolveu o AIPedia, um repositório estruturado a partir da modelagem de dados RDF, metaontologia OWL e mecanismo de inferência SPARQL, com a finalidade de traduzir metadados semanticamente enriquecidos. Os resultados de consultas sofisticadas por meio do SPARQL possibilitam a análise de dados e informações de forma mais aprofundada se comparado com métricas isoladas ou análises manuais (Maatouk 2021).

Para Bizer e Cyganiak (2009), uma métrica de avaliação da qualidade da informação é um procedimento para medir uma das dimensões da qualidade da informação. O cálculo das métricas de avaliação é fundamentado em pontuação estabelecida pelos indicadores de qualidade utilizando uma função de pontuação (Bizer e Cyganiak 2009). As métricas de avaliação são heurísticas projetadas para atender a uma situação de avaliação específica (Arazy et al. 2017).

As heurísticas, segundo Gigerenzer e Gaissmaier (2011), são processos cognitivos eficientes, conscientes ou inconscientes, que ignoram parte da informação com o propósito de obter julgamentos mais precisos e de forma mais rápida e fácil do que a ponderação e a adição de todas as informações em um dado contexto. Segundo esses autores existem três passos cognitivos fundamentais na seleção de uma heurística: (i) procura: as decisões são tomadas entre alternativas e por esse motivo há uma necessidade de procura ativa; (ii) parar de procurar: a procura por alternativas tem que terminar devido as capacidades limitantes da mente humana; (iii) decisão: assim que as alternativas estiverem encontradas e a procura for cessada, um conjunto final de heurísticas são chamadas para que a decisão possa ser tomada.

Métricas de avaliação podem incluir também metainformações sobre as circunstâncias em que as informações foram criadas (Pipino et al. 2005). Bizer e Cyganiak (2009) classificaram as métricas para análise da qualidade da informação em conteúdo, contexto e avaliação. As métricas baseadas em conteúdo usam informações para serem avaliadas como indicadores de qualidade. Essas métricas possibilitam a análise do conteúdo da informação ou a comparação de informações com outras relacionadas. Geralmente uma métrica envolve pontuações de avaliação combinando termos ou frases em um documento e/ou analisando a estrutura do documento.

De acordo com Bizer e Cyganiak (2009), as métricas baseadas em contexto utilizam indicadores de qualidade que fornecem metainformações sobre quem disse o quê, quando e em

quais circunstâncias as informações foram coletadas. Uma metainformação relacionada à credibilidade da informação na Web, por exemplo, pode fornecer informações sobre a identidade do provedor de informações e se o nome dele está contido em uma lista de provedores confiáveis (Bizier e Cyganiak 2009). Já as métricas baseadas em avaliação, segundo esses autores, dependem de classificações explícitas sobre a própria informação, fontes de informação ou provedores de informações. As classificações podem ser originárias do próprio consumidor de informação, de outros consumidores destas ou de especialistas do domínio (Bizer e Cyganiak 2009).

O desenho de um sistema de classificação é comumente efetuado com o auxílio de sistema de computação a partir de classificações de qualidade. O cálculo das pontuações de avaliação é feito a partir de uma função de pontuação. A função de pontuação determina quais classificações são levadas em consideração, atribuindo-se pesos diferentes às classificações (Bizer 2007; Josang et al. 2007). A lógica *fuzzy* e a análise de regressão podem ser utilizadas na construção de métricas (Cichy e Rass 2020).

Estabelecido o viés conceitual e as bases de análise, especialmente os apontamentos de indicadores ou métricas comumente utilizados para assegurar em diferentes contextos da qualidade da informação, apresenta-se sequencialmente a metodologia da pesquisa.

3 Metodologia

Este estudo foi baseado em uma revisão bibliométrica efetuada nas bases de dados da Proquest (Library & Information Science Abstracts, Technology collection, Computer and Information Systems), Scopus e Web of Science que indexam a literatura internacional nas áreas de Ciências Sociais Aplicadas e Ciência da Informação.

A pesquisa em cada base de dados foi efetuada em várias etapas sendo a mais recente de janeiro de 2021, mediante o uso dos termos: qualidade da informação, indicadores e métricas. Os termos de busca foram traduzidos para a língua inglesa e conectados a operadores booleanos, a fim de recuperar materiais relevantes dentro do escopo do estudo. A seguinte estratégia de busca foi utilizada nas bases selecionadas: "*information quality*" and *indicators and metrics*. Como

critério de busca nas bases de dados foram utilizados os seguintes filtros: artigos de periódicos publicados no período de 2012-2021, com texto completo e revisados por pares.

Após a busca nas bases de dados, foi utilizado o *check-list* do Protocolo de Recomendação Prisma para a seleção dos artigos, estabelecimento de critérios de elegibilidade e inclusão dos artigos para síntese qualitativa (Liberati et al. 2009).

O Protocolo PRISMA é largamente usado para condução de revisões sistemáticas e meta-análises na área da saúde, mas pode também ser usado para as demais áreas. É formado por um *check-list* de 27 itens para selecionar a literatura relevante em determinado contexto. Inclui um diagrama do fluxo da informação com quatro diferentes fases de uma revisão sistemática da literatura: 1) Identificação – quantificação dos documentos recuperados em base de dados científicas e em outras fontes; eliminação dos documentos duplicados, seguido pela avaliação dos documentos potencialmente úteis; 2) Seleção – rastreamento e quantificação de documento disponíveis em texto completo e exclusão dos demais; 3) Elegibilidade – Quantificação de documentos encontrados com textos completos elegíveis para o estudo e eliminados aqueles que não estão alinhados como tema em estudo; 4) Inclusão – Número de documentos selecionados para análise quantitativa e qualitativa (Moher et al. 2009).

Neste estudo, conforme a Recomendação PRISMA, os critérios de elegibilidade incluíram: (i) quais indicadores de qualidade foram usados; (ii) quais indicadores foram mais citados; (iii) que conceitos foram atribuídos aos indicadores usados. Para consecução do objetivo do estudo foram recuperados 425 artigos e após excluídos os materiais duplicados entre as bases utilizadas, foram selecionados 18 artigos que sinalizavam os indicadores ou métricas usados para assegurar maior qualidade da informação em diferentes contextos.

Adicionalmente, neste estudo, foi efetuada uma análise de conteúdo incluindo categorias alinhadas com os critérios de elegibilidade estabelecidos pelo Protocolo PRISMA. A análise de conteúdo, proposta por Bardin (2011), é considerada um método misto para avaliação de documentos a partir de categorias que permitem classificar componentes dos documentos. Após a leitura crítica dos documentos, seguem-se a codificação e a categorização do *corpus* de análise. Na fase de codificação se estabelecem as principais informações coletadas que servem de base para a elaboração da categorização, fundamental para a análise dos indicadores levantados. A partir

dessas fases é possível estabelecer as categorias de contexto, categorias de análise e as unidades de contexto e de registro, tendo como base os objetivos do estudo, as palavras-chave usadas na pesquisa, a revisão de literatura e o alinhamento conceitual. Por fim, segue-se a fase de tratamento dos dados possibilitando a descrição, a interpretação e as inferências.

As categorias de contexto desta pesquisa são **indicadores** e **métricas**. As categorias de análise incluíram os indicadores de qualidade da informação, considerando os de precisão, de disponibilidade, de confiabilidade e de consistência entre outros, e as métricas de avaliação, de conteúdo e de contexto. As unidades de registro foram os indicadores e as métricas citadas nos documentos avaliados.

A partir da revisão de literatura, da utilização do Protocolo Prisma e da análise de conteúdo foi possível discutir e descrever os resultados que seguem delineados no próximo capítulo.

4 Resultados e discussões

Contextualizando, dos 425 artigos recuperados, 86 deles foram eliminados por estarem duplicados. Dos 339 artigos restantes foram excluídos 229 que não estavam alinhados com a pesquisa. Após a seleção dos itens elegíveis para o estudo, fizeram parte do *corpus* de pesquisa 18 documentos. Após a leitura sistemática, foram identificados os indicadores e métricas apontados nos documentos.

A partir da análise dos documentos selecionados foi possível correlacionar as métricas e os indicadores, registrando-se a frequência com que estes são visualizados. Os indicadores arrolados nos estudos avaliados, perfazendo um total de 28 registros, são indicados no Quadro 1.

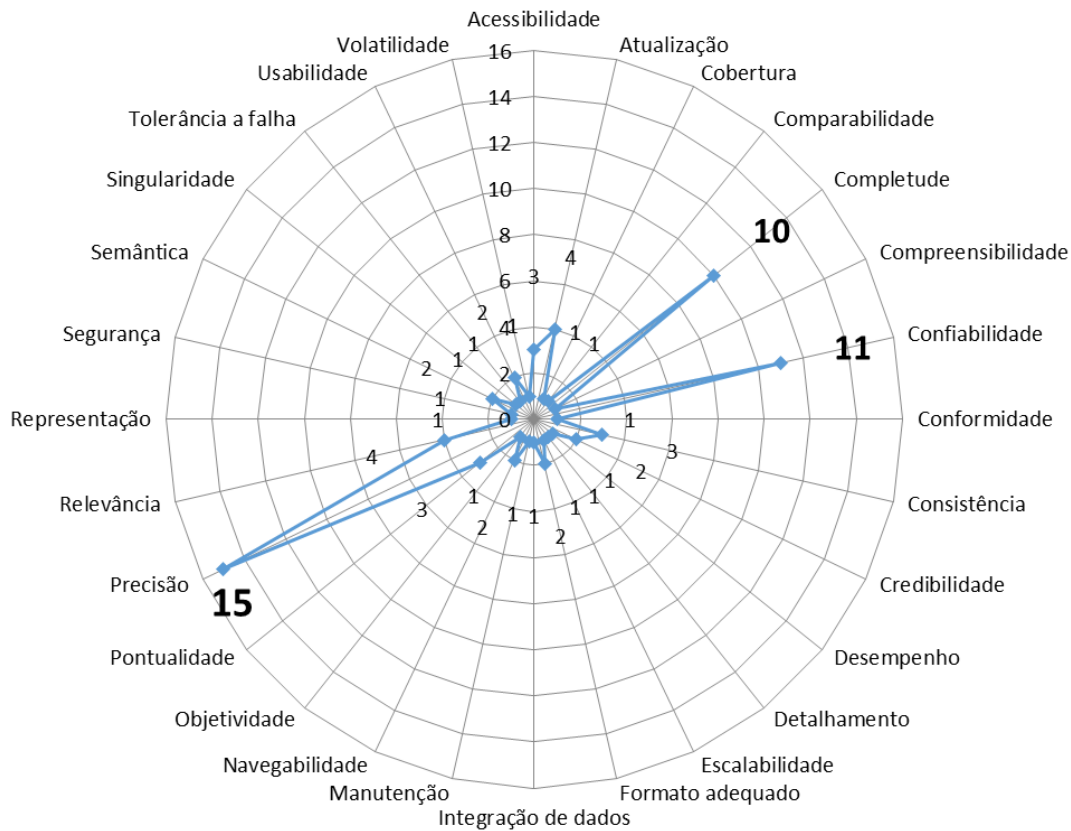
Quadro 1 – Indicadores arrolados nos estudos avaliados.

INDICADORES		Atualização	Cobertura	Comparabilidade	Compleude	Confiabilidade	Conformidade	Consistência	Credibilidade	Desempenho	Detalhamento	Escalabilidade	Formato adequado	Integração de dados	Compreensibilidade	Manutenção	Navegabilidade	Objetividade	Pontualidade	Precisão	Relevância	Representação	Segurança	Semântica	Singularidade	Tolerância a falha	Usabilidade	Volatilidade
AUTORES ANALISADOS	Ayyash 2017				*													*		*	*	*						
	Azeroual et al. 2020				*															*	*	*						
	Behkamal et al. 2014		*		*															*	*							
	Bibi, Jannisar e Inayet, 2016				*	*		*																*	*		*	
	Bizer 2007				*	*		*													*	*			*			
	Harting e Lewoniewski 2020		*		*																*	*			*			
	Josang e Ismail 2006				*	*															*	*			*			
	Knight e Burn 2005				*	*		*	*					*							*	*						
	Manzoor e Truong e Dustdar 2014				*	*															*	*						
	Mohammadi, Abrizah e Nazari 2015				*	*															*	*					*	
	Nehmy e Paim 1998		*	*																	*	*						
	Pinto, Guerreiro Quesada, Granell, 2014	*	*										*								*	*						
	Reiter et al. 2014				*	*	*		*												*	*						
	Reznik e Lyshevski 2015				*	*														*	*							*
	Rukkas e Zholkevych 2015				*	*															*	*						
	Sá e Martins 2016	*			*	*				*		*					*					*	*				*	
Saracevic 1996	*		*	*	*						*	*		*						*	*					*		
Sugahara e Jannuzzi 2005				*	*												*			*	*		*					

Fonte: As autoras

A frequência dos indicadores nos documentos está sinalizada no Gráfico 1, destacando-se a precisão, a confiabilidade e a completude, que tiveram uma ligeira diferença entre eles.

Gráfico 1 – Distribuição dos indicadores nos estudos.



Fonte: As autoras

O indicador **precisão** foi o mais citado ($f=15$) nos estudos avaliados (Ayyash 2015; Azeroual et al. 2020; Behkamal et al. 2014; Bizer 2007; Härting e Lewoniewski 2020; Josang et al. 2007); Knight e Burn 2005; Manzoor et al. 2014; Mohammadi et al. 2017; Nehmy e Paim 1998; Reiter et al. 2012; Reznik e Lyshevski 2015; Rukkas e Zholtkevych 2015; Saracevic 1996; Sugahara e Jannuzzi 2005). Este indicador também foi apontado como frequentemente utilizados em diferentes estudos de avaliação da informação como os de Greef, Freiras e Romanel (2012), Arazy, Kopak e Hadar (2017) e Azeroual et al. (2020).

O indicador **confiabilidade** teve destaque nos estudos avaliados ficando em segundo lugar no cômputo de frequência ($f = 11$) (Bibi e Jannisar e Inayet 2016; Bizer 2007; Josang et al. 2007); Knight e Burn 2005; Manzoor et al. 2014; Mohammadi et at. 2017; Reiter et al. 2012; Reznik e Lyshevski 2015; Rukkas e Zholtkevych 2015; Sá e Martins 2016; Saracevic 1996).

Já **completude** foi o terceiro indicador de destaque ($f=10$) (Ayyash 2015; Azeroual et al. 2020; Behkamal et al. 2014; Bizer 2007; Härting e Lewoniewski 2020; Knight e Burn 2005; Mohammadi et al. 2017; Reznik e Lyshevski 2015, Saracevic 1996; Sugahara e Jannuzzi 2005). A frequência de citação dos indicadores nos estudos avaliados ficou bem próxima, denotando a importância de cada como indicador de qualidade da informação.

A partir da conceitualização dos indicadores e métricas, de acordo com a revisão de literatura (Behkamal et al. 2014, Bizer e Cyganiak 2009; Han et al. 2015, Härting e Lewoniewski 2020; Lu et al. 2015; Rasaiah et al. 2015), registra-se que o conceito dos indicadores de qualidade pode variar de acordo com a tarefa ou as preferências do usuário (Bizer, e Cyganiak 2009). Neste estudo foram observadas algumas sobreposições de conceitos, entendidas nesta pesquisa, como a atribuição de mais de um conceito para um indicador, conforme demonstrado na Figura 1.

Figura 1 - Sobreposição de conceito dos indicadores.



Fonte: As autoras

O indicador de precisão foi conceituado em todos os estudos avaliados como “ausência de erros”. Uma preocupação adicional com a precisão semântica, juntamente com a precisão sintática foi observada nos estudos de Behkamal et al. 2014, Han et al. 2015 e Rasaiah et al. 2015. Por outro

lado, no alinhamento de conceitos, observou-se sobreposição para indicadores de confiabilidade, completude e consistência com maior e menor ocorrência nos estudos.

A confiabilidade, nos estudos de Lu et al. (2015), foi definida em termos de funcionamento correto ou consistente da fonte de dados. Manzoor et al. (2014), conceituaram esse indicador como sendo a crença na correção da informação. Observa-se que para esse indicador houve uma ligeira divergência de conceito que pode estar associada ao objeto de estudo, já que o primeiro avaliou a qualidade de fontes de dados e o segundo, a qualidade de informação de contexto. Observou-se ainda uma sobreposição de conceitos desse indicador com definições associadas à precisão e à tolerância a falhas do sistema.

Para o indicador de completude foram descritos três conceitos diferentes. Para Ayyash (2015), a completude estaria associada ao atendimento da necessidade de informação do cliente; Arazy et al. (2017) relacionaram esse indicador à presença de informações relevantes; já Reiter et al. (2012) correlacionaram completude à amplitude e à profundidade da informação. Essas divergências podem estar relacionadas ao tipo de estudo e indicadores usados, sendo que os dois primeiros avaliaram a qualidade de informações também do ponto de vista do usuário e o último estudo avaliou a qualidade de dados.

Quanto ao indicador de consistência, Wand e Wang (1996), destacaram que este conceito se refere a vários aspectos de representação e de valores de dados relacionados em um determinado contexto. Kopyltsov (2020), corrobora no sentido de que contendo um único simbolismo no mesmo produto ou contexto, a terminologia atende aos requisitos de consistência. Entretanto, neste estudo observou-se que o indicador de consistência apresentou sobreposição de conceitos relacionados à atualização, precisão e taxa de erro (Bizer 2007; Knight e Burn 2005; Bibi et al. 2016).

Os resultados deste estudo corroboram outros que ressaltaram o caráter subjetivo dos indicadores e a dependência da tarefa e das preferências do usuário da informação (Cichy e Rass 2020, Bizer 2007; Bizer e Cyganiak 2009; Calazans 2008; Härting e Lewoniewski 2020; Naumann 2002; Wang e Strong 1996). Isso porque, dependendo do contexto analisado, determinados indicadores de qualidade podem ser considerados importantes para um usuário, mas irrelevantes para outros.

Entretanto, a validade de avaliações da informação, a partir de indicadores, parece estar respaldada no uso de heurísticas que possam ser aplicadas em uma determinada situação e que sejam suficientemente precisas para serem úteis do ponto de vista do consumidor da informação (Arazy et al. 2017; Bizer e Cyganiak 2009; Gigerenzer e Gaissmaier 2011). Por outro lado, uma boa qualidade de dados e de informações é crucial, já que a má qualidade dos dados pode ter impactos sociais e econômicos substanciais, colocando em risco a implementação de muitas estratégias e políticas (Maatouk 2021; Sá e Martins 2016).

Quanto às métricas, de acordo com a classificação de Bizer e Cyganiak (2009), verificou-se que neste estudo, a métrica de avaliação foi a mais usada, seguida pela métrica de contexto e métrica de conteúdo. A aplicação das métricas de avaliação e de contexto foi efetuada com auxílio de aplicativos de *software* para mensurar a qualidade de dados e fontes de dados.

A escolha de métricas de avaliação adequadas para dimensões de qualidade específicas é limitada por vários fatores. Para utilizar uma métrica de avaliação é preciso que os indicadores estejam disponíveis. A avaliação da pontualidade ou atualização da informação, por exemplo, é uma tarefa fácil já que esse indicador está frequentemente disponível. Por outro lado, o acesso a outras dimensões como precisão ou objetividade, muitas vezes é difícil, pois pode envolver o consumidor ou os especialistas da informação, verificando ou classificando informações (Bizer e Cyganiak 2009). Neste estudo as métricas de avaliação foram observados em 11 estudos (Azeroual et al. 2020, Bibi et al. 2016; Bizer 2007; Härting e Lewoniewski 2020, Knight e Burn 2005, Manzoor et al. 2014; Reznik e Lyshevski 2015; Rukkas e Zholtkevych 2015; Sá e Martins 2016; Saracevic 1996; Sugahara e Jannuzzi 2005).

A métrica de contexto foi identificada em cinco estudos analisados (Härting e Lewoniewski 2020; Mohammadi et al. 2017; Nehmy e Paim 1998; Pinto et al. 2014; Reiter et al. 2012), para avaliar a qualidade de metadados e informação de contexto, também utilizando aplicativos. A métrica de conteúdo foi aplicada em quatro estudos (Ayyash 2015; Azeroual et al. 2020; Behkamal et al. 2014; Josang et al. 2007). Em dois destes estudos foram utilizadas métricas de conteúdo, incluindo pontuações de avaliação de usuários para análise de informações de documentos (Ayyash 2015; Azeroual et al. 2020). Dois outros estudos usaram duas métricas em conjunto,

sendo que Härting e Lewoniewski (2020) utilizaram métricas de contexto e de avaliação e Azeroual et al. (2020) usaram métricas de avaliação e de conteúdo, conjuntamente (Quadro 2).

Quadro 2 - Métricas utilizadas.

Métricas	Autores Analisados																	
	Ayyash 2017	Azeroual et al. 2020	Behkamal et al. 2014	Bibi e Jannisar e Inayet 2016	Bizer 2007	Harting e Lewoniewski 2020	Josang e Ismail 2006	Knight e Burn 2005	Manzoor e Truong e Dustdar 2014	Mohammadi e Abrizah e Nazari 2015	Nehmy e Paim 1998	Pinto, Guerreiro Quesada e Granell, 2014	Reiter et al. 2014	Reznik e Lyshevski 2015	Rukkas e Zholtkevych 2015	Sá e Martins 2016	Saracevic 1996	Sugahara e Jannuzzi 2005
Métricas de avaliação		*		*	*	*		*	*					*	*	*	*	*
Métricas de conteúdo	*	*	*				*											
Métricas de contexto						*				*	*	*	*					

Fonte: As autoras

Considerando que a métrica de avaliação foi mais utilizada nos estudos avaliados cabe ressaltar que a escolha por esta métrica pode ser influenciada pela qualidade dos indicadores de qualidade disponíveis. Se um consumidor de informações tiver dúvidas sobre a qualidade de certos indicadores, deverá escolher uma métrica de avaliação que dependa de outros indicadores disponíveis. Deste modo, o fator chave para um consumidor de informação confiar nos resultados da avaliação é a compreensão do processo de avaliação (Bizer e Cyganiak 2009). Isto porque, conforme destacaram Härting e Lewoniewski (2020), a qualidade de um dado ou informação pode ser conceituada como um construto multidimensional e dependente do objetivo do consumidor da informação.

As preferências subjetivas constituem também um limitador da escolha de métricas. O consumidor de informações pode considerar um conjunto de indicadores de qualidade específicos e funções de pontuação mais confiável do que outros conjuntos (Bizer e Cyganiak 2009). Assim,

não existe uma política ideal para uma tarefa específica, pois o conceito de uma boa política de seleção de indicadores e métricas para análise da qualidade da informação difere de usuário para usuário.

5 Conclusões

Dados e informações constituem insumos cruciais para o avanço do conhecimento, mas, para criar valor esses componentes precisam ser de qualidade. Neste estudo, verificou-se que os indicadores mais utilizados foram a precisão, a confiabilidade e a completude, conforme o Quadro 1 e o Gráfico 1.

Foram identificadas divergências na conceitualização dos indicadores usados, além de sobreposições das definições, ilustradas na Figura 1. Tais divergências podem estar relacionadas ao tipo de estudo e indicadores usados, já que a análise da qualidade de dados difere da análise de qualidade de informações.

As métricas de avaliação foram as mais utilizadas envolvendo estudos relacionados à qualidade de dados e fontes de dados. Métricas de contexto foram usadas para avaliar a qualidade de metadados e de informação de contexto. As métricas de conteúdo abarcaram análises de informações de documentos. Na maioria dos estudos as métricas foram aplicadas com auxílio de aplicativos e sistemas computadorizados.

A qualidade da informação é comumente concebida como uma construção multidimensional, podendo ser utilizado um conjunto de indicadores na análise da qualidade de dados e informações. A escolha de indicadores e métricas para análise da qualidade de informação depende do objeto e das preferências subjetivas do usuário da informação, sendo preferível usar métricas de avaliação rastreáveis, relativamente simples e de fácil compreensão.

Referências

Associação Brasileira de Normas Técnicas. *NBR ISO 9001: 2008*. Sistemas de gestão da qualidade – Requisitos. ABNT, 2008.

Guedes, Josefina Aparecida Soares, et al. Uso de indicadores e métricas para avaliação da qualidade da informação. *Brazilian Journal of Information Science: Research trends*, vol. 15, publicação contínua, 2021, e02121 DOI: 10.36311/1981-1640.2021.v15.e02121

- Arazy, Ofer, et al. "Heuristic principles and differential judgments in the assessment of information quality". *Journal of the Association for Information Systems*, vol. 18, no. 5, 2017, pp. 403-432, doi:10.17705/1jais.00458. Acessado 8 maio 2020.
- Assis, Juliana de, e Moura, Maria Aparecida. "A qualidade da informação na Web: uma abordagem semiótica". *Informação & Informação*, vol. 16, no. 3, 2011, pp. 96-117, doi:10.17705/1jais.00458. Acessado 9 maio 2021.
- Ayyash, Mohannad Moufeed. "Identifying information quality dimensions that affect customers satisfaction of e-banking services". *Journal of Theoretical and Applied Information Technology*, vol. 82, 2015, pp. 122-130, doi:10.5433/1981-8920.2011v16n3p96. Acessado 9 maio 2021.
- Azeroual, Otmane, et al. "Data quality as a critical success factor for user acceptance of research information systems". *Data*, vol. 5, no. 2, 2020, pp. 1-13, doi:10.3390/data5020035. Acessado 9 maio 2021.
- Bardin, L. *Análise de conteúdo*. 3 ed., Edições 70, 2011.
- Behkamal, Behshid, et al. "A Metrics-driven approach for quality assessment of linked open data". *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 9, no. 2, 2014, pp. 64-79, doi:10.4067/S0718-18762014000200006. Acessado em 9 maio 2021.
- Bibi, Ruqia., et al. "Quality Implication for Prognoses Success in Web Applications". *International Journal of Modern Education and Computer Science*, vol. 8, no. 3, 2016, pp. 37-44.
- Bizer, Christian. *Quality-driven information filtering in the context of web-based information systems*, 2007. Freie Universit at Berlin, Ph.D. Thesis.
- Bizer, Christian, and Cyganiak, Richard. Quality-driven information filtering using the WIQA policy framework. *Journal of Web Semantics*, v. 7, n. 1, 2009, pp. 1-10, doi:10.1016/j.websem.2008.02.005. Acessado 9 maio 2021.
- Botelho, Kilça Tanaka, et al. "Indicadores de sustentabilidade empresarial: um estudo exploratório". *Divers@!*, vol. 8, no. 2, 2015, doi:10.5380/diver.v8i2.45050. Acessado 15 maio 2021.
- Calazans, Angélica Toffano Seidel. "Qualidade da informação: conceitos e aplicações". *TransInformação*, vol. 20, no. 1, 2008, pp. 29-45.
- Chen, Ying, et al. "Processamento de consultas com controle de qualidade na World Wide Web". *World Wide Web*, vol. 1, no. 4, 1998, pp. 241-255.
- Choo, Chun Wei. *A organização do conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões*. Senac, 2003.

- Cichy, Corinna, and Rass, Stefan. *Fuzzy Expert Systems for Automated Data Quality Assessment and Improvement Processes*. EKAW (Posters & Demos), 2020, pp. 7-21.
- Dante, Gloria Ponjuán. "Gestión de información: precisiones conceptuales a partir de sus orígenes". *Pesquisa Brasileira em Ciência da Informação e Biblioteconomia*, vol. 3, no. 2, 2008.
- Davenport, Thomas H., e Prusak, Lawrence. *Conhecimento empresarial: como as organizações gerenciam o seu capital intelectual*. 14 ed, Campus, 2003.
- Fagundes, Melissa Figueira, e Ribeiro Junior, Divino Ignácio. "Modelo baseado em Frictionless Data aplicado aos dados abertos governamentais". *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, vol. 18, no. 1, 2020, doi:10.20396/rdbci.v18i00.8661528. Acessado 16 maio 2021.
- Gigerenzer, Gerd, and Gaissmaier, Wolfgang. "Heuristic decision making". *Annual review of psychology*, vol. 62, 2011, pp. 451-482, doi:10.1146/annurev-psych-120709-145346. Acessado 12 maio 2020.
- Greef, Ana Carolina, et al. *Lean office: operação, gerenciamento e tecnologias*. 1 ed., Atlas, 2012.
- Han, Jingyu, et al. "Web article quality ranking based on web community knowledge". *Computing*, vol. 97, no. 5, 2015, pp. 509-537, doi:10.1007/s00607-014-0435-4. Acessado 12 maio 2021.
- Härting, Ralf-Christian, and Lewoniewski, Włodzimierz. "Main influencing factors of quality determination of collaborative open data pages". *Information*, vol.11, no. 6, 2020, pp. 1-10, doi:10.3390/info11060283. Acessado 12 maio 2021.
- Josang, Audun, Ismail, Roslan and Boyd, Colinet. "A survey of trust and reputation systems for online service provision". *Decision support systems*, vol. 43, no. 2, 2007, pp. 618-644, doi:10.1016/j.dss.2005.05.019. Acessado 13 maio 2021.
- Juran, Joseph M., et al. *Quality control handbook*. 1 ed., McGraw Hill, 1974.
- Knight, Shirlee-ann, and Burn, Janice. "Developing a framework for assessing information quality on the World Wide Web". *Informing Science*, vol. 8, no. 1, 2005, pp. 159-172.
- Kopyltsov, A. V. "Selection of metrics in software quality evaluation". *Journal of Physics: Conference Series*, vol. 1515, no. 3, 2020, doi:10.1088/1742-6596/1515/3/032018. Acessado 10 maio 2021.
- Liberati, Alessandro, et al. "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration". *Journal of clinical epidemiology*, vol. 62, no.10, 2009, doi:10.1016/j.jclinepi.2009.06.006. Acessado 10 maio 2021.
- Lu, Lihua, et al. "Integre dados inconsistentes e heterogêneos com base no feedback do usuário." *Jornal Internacional de Computação Inteligente e Cibernética*, vol. 8, no. 2, 2015, p. 187-203, doi:10.1108/IJICC-04-2014-0013. Acessado 10 maio 2020.

- Maatouk, Yasser. "Building AIPedia ontology to evaluate research impact in artificial intelligence area". *Academia Letters*, 2021, <https://doi.org/10.20935/AL2781>. Acessado 10 maio 2020.
- Manzoor, Atif, et al. "Quality of context: models and applications for context-aware systems in pervasive environments". *The Knowledge Engineering Review*, vol. 29, no. 2, 2014, pp. 154-170, doi:10.1017/S0269888914000034. Acessado 10 maio 2021.
- Moher, David, et al. "The PRISMA Group preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement". *PLoS Med*, v. 6, n. 7, 2009.
- Mohammadi, Faegheh, et al. "Is the information fit for use? Exploring teachers perceived information quality indicators for Farsi web-based learning resources". *Malaysian Journal of Library & Information Science*, vol. 20, no. 1, 2017, pp. 99-122.
- Naumann, Felix. *Quality-driven query answering for integrated information systems*. Springer, 2002.
- Nehmy, Rosa Maria Quadros, e Paim, Isis. "A desconstrução do conceito de 'qualidade da informação'". *Ciência da Informação*, vol. 27, no. 1, 1998, pp. 36-45, doi:10.1590/S0100-19651998000100005. Acessado 17 maio 2021.
- Oliveira, Dalgiza Andrade de, e Araújo, Ronaldo Ferreira de. "A contribuição das métricas para o campo da ciência da informação". *Perspectivas em Ciência da Informação*, 2020, pp. 300-318.
- Paletta, Francisco Carlos, e Ueki, Greicyene Hamaguchi. "A informação como agente de mudanças nas organizações". *Revista Conhecimento em Ação*, v. 4, 2019, pp. 146-155.
- Parker, Michelle B., et al. "An evaluation of Information quality frameworks for the World Wide Web". *Annual Conference on WWW Applications: South Africa*, 2006.
- Pinto, María, et al. "Dissemination of information and visibility of the european higher education area through the websites of Spanish: a longitudinal metric analysis, 2007-2012". *Scientometrics*, v. 98, no. 2, 2014, pp 1235–1255.
- Pipino, Leo, et al. "Developing measurement. scales for data-quality dimensions". *Information quality*, vol. 1, 2005, pp. 37-521.
- Rasaiah, Barbara A., et al. "Assessing field spectroscopy metadata quality". *Remote Sensing*, vol. 7, no. 4, 2015, pp. 4499-4526.
- Reiter, Michael, et al. "Quality of data driven simulation workflows". *2012 IEEE 8th International Conference on E-Science*. IEEE, 2012. pp. 1-8.
- Reznik, Leon, and Lyshevski, Sergey Edward. "Data quality indicators composition and calculus: Engineering and information systems approaches". *Sensors & Transducers*, vol. 185, no. 2, 2015, pp. 140-148.
-
- Guedes, Josefina Aparecida Soares, et al. Uso de indicadores e métricas para avaliação da qualidade da informação. *Brazilian Journal of Information Science: Research trends*, vol. 15, publicação contínua, 2021, e02121 DOI: 10.36311/1981-1640.2021.v15.e02121

- Rösch, Christine, et al. "Indicator system for the sustainability assessment of the German energy system and its transition". *Energy, Sustainability and Society*, vol. 7, no. 1, 2017, pp. 1-13.
- Rukkas, Kyrylo, and Zholtkevych, Galyna. "Distributed Datastores: Towards Probabilistic Approach for Estimation of Reliability". *ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer*, 2015.
- Sá, Patrícia Moura, and Martins, Rita. "Data quality requirements for water bills". *The TQM Journal*. vol. 28, no. 6, 2016, pp. 933-953.
- Saracevic, Tefko. "Ciência da informação: origem, evolução e relações". *Perspectivas em ciência da informação*, vol. 1, no. 1, 1996, pp. 41- 62.
- Sugahara, Cibele Roberta, and Jannuzzi, Paulo de Martino. "Estudo do uso de fontes de informação para inovação tecnológica na indústria brasileira". *Ciência da Informação*, vol. 34, 2005, pp. 45-56.
- Turban, Efraim, et al. *Tecnologia da informação para gestão*. 3 ed., Bookman, 2004.
- Valls, Valéria Martin, e Vergueiro, Waldomiro de Castro Santos. "A gestão da qualidade em serviços de informação no Brasil: uma nova revisão de literatura, de 1997 a 2006". *Perspectivas em ciência da informação*, vol. 11, no. 1, 2006, pp. 118-137.
- Van Bellen, Hans Michael. "Desenvolvimento sustentável: uma descrição das principais ferramentas de avaliação". *Ambiente & Sociedade*, vol. 7, no. 1, 2004, pp. 67-87.
- Wand, Yair, and Wang, Richard Y. Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, vol. 39, no.11, 1996, pp. 86-95.
- Wang, Richard Y., and Strong, Diane M. "Beyond accuracy: What data quality means to data consumers". *Journal of management information systems*, vol.12, no. 4, 1996, pp. 5-33.

Dados da pesquisa

Os dados da pesquisa estão incluídos no artigo.

Copyright: © 2021 Guedes, Josefina Aparecida Soares. This is an open-access article distributed under the terms of the Creative Commons CC Attribution-ShareAlike (CC BY-SA), which permits use, distribution, and reproduction in any medium, under the identical terms, and provided the original author and source are credited.

Received: 09/07/2021

Accepted: 10/07/2021