



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE ARTES E COMUNICAÇÃO
DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO
CURSO DE GRADUAÇÃO EM GESTÃO DA INFORMAÇÃO

EDILTON DE ALBUQUERQUE CAVALCANTI JUNIOR

**CONVERGÊNCIAS ENTRE WEB SEMÂNTICA E LINKED DATA NA
ORGANIZAÇÃO DA INFORMAÇÃO: um estudo bibliométrico por meio da Scielo e
da Web of Science**

RECIFE

2022

EDILTON DE ALBUQUERQUE CAVALCANTI JUNIOR

**CONVERGÊNCIAS ENTRE WEB SEMÂNTICA E LINKED DATA NA
ORGANIZAÇÃO DA INFORMAÇÃO: um estudo bibliométrico por meio da Scielo e
da Web of Science.**

Trabalho de Conclusão apresentado ao Departamento de Ciência da Informação como requisito parcial para obtenção do grau de Bacharel em Gestão da Informação.

Orientadora: Prof^a. Dr^a. Thais Helen do Nascimento Santos.

RECIFE

2022

Catálogo na fonte
Biblioteca Joaquim Cardozo – Centro de Artes e Comunicação

C377c Cavalcanti Junior, Edilton de Albuquerque

Convergências entre web semântica e linked data na organização da informação: um estudo bibliométrico por meio da Scielo e da Web of Science / Edilton de Albuquerque Cavalcanti Junior. – Recife, 2022.
117f.: il. fig., tab.

Sob orientação de Thais Helen do Nascimento Santos.

Trabalho de Conclusão de Curso (Graduação) – Universidade Federal de Pernambuco. Centro de Artes e Comunicação. Departamento de Ciência da Informação. Curso de Gestão da Informação, 2022.

Inclui referências.

1. Gestão da Informação. 2. Web semântica. 3. Linked data. 4. Bases de dados. 5. Análise bibliométrica. I. Santos, Thais Helen do Nascimento (Orientação). II. Título.

020 CDD (22. ed.)

UFPE (CAC 2022-105)



FOLHA DE APROVAÇÃO

CONVERGÊNCIAS ENTRE WEB SEMÂNTICA E LINKED DATA NA ORGANIZAÇÃO DA INFORMAÇÃO: um estudo bibliométrico por meio da Scielo e da Web of Science.

Edilton de Albuquerque Cavalcanti Junior

Trabalho de Conclusão de Curso submetido à Banca Examinadora, apresentado e aprovado de modo remoto (online), conforme autorizado pelo PROACAD/UFPE em Ata de Reunião Virtual dos Coordenadores de Graduação do dia 12 de maio de 2020, pelo Curso de Gestão da Informação, do Departamento de Ciência da Informação, da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Gestão da Informação.

TCC aprovado 19 de maio de 2022.

Banca Examinadora:

Orientadora – Prof^a. Dr^a. Thais Helen do Nascimento Santos.
DCI/Universidade Federal de Pernambuco.

Examinador 1 – Prof. Dr. Fábio Mascarenhas e Silva.
DCI/Universidade Federal de Pernambuco.

Examinador 2 – Prof^a. MSc. Marcela Lino da Silva.
DCI/Universidade Federal de Pernambuco.

Dedicação: As diversas possibilidades que a vida
proporcionou para tornar-me melhor a cada dia!

AGRADECIMENTOS

Aos docentes da Universidade Federal de Pernambuco – UFPE que fazem parte do DCI – Departamento de Ciência da Informação, pelos conhecimentos repassados e apoio que sempre foram proporcionados. Ressalto meus agradecimentos a Prof^a. Dr^a. Thais Helen do Nascimento Santos, pelas informações que contribuíram na elaboração deste trabalho. Estendo meus agradecimentos aos servidores e prestadores de serviços terceirizados dessa Instituição pelos serviços diários, os quais possibilitaram aos discentes um ambiente favorável ao aprendizado. A Deise Silva pelo carinho e atenção em todos os dias e pelos momentos alegres cheios de amor que fazem parte da minha vida, meus sinceros agradecimentos.

RESUMO

As novas tecnologias ocasionam profundas transformações na sociedade, visto que o objeto informação, hoje, encontra-se disponível de forma dinâmica, em grandes volumes, em diversos formatos e meios. A Ciência da Informação desempenha papel fundamental no gerenciamento da informação em diversos ambientes, sendo a Internet um campo repleto de heterogeneidade e caótico para busca de informações. Nesse sentido, Berners-Lee propôs a Web Semântica, em que computadores possam interpretar as informações e realizar inferências por meio da linguagem *Extensible Markup Language (XML)*, *Resource Description Framework (RDF)* e uso de *Ontology Web Language (OWL)* que formam a base para estruturar a Web de significado. Outra proposta é o Linked Data, que visa à interligação de conjuntos dados estruturados na Web de forma legível e interpretável por máquinas. Nesse contexto, o objetivo dessa pesquisa é analisar as publicações internacionais e nacionais que convergem a Web Semântica e Linked Data, na organização da informação. Metodologicamente, a pesquisa quanto aos fins enquadra-se como exploratória; quanto aos dados quali-quantitativa; e quanto aos meios uma pesquisa bibliográfica. Para análise quantitativa utilizou-se métodos estatísticos descritivos e para análise qualitativa método de análise de conteúdo. Os resultados apontam que no aspecto quantitativo há convergência entre as produções nacionais e internacionais nos pontos de baixa produtividade de artigos sobre os temas em conjunto; as escolhas de periódicos para publicação são para aqueles com maiores *qualis* de classificação; existe forte relação de produção de artigos com autoria e coautoria e formação de clusters de colaboração. Há divergência no quantitativo de citações, sendo uma média baixa para artigos nacionais e mais elevada para os internacionais. No aspecto qualitativo foram delineadas 13 categorias temáticas, sendo a categoria conversão e compartilhamento de bases de dados para padrões da Web Semântica e Linked Data a que concentra o maior quantitativo de publicações.

Palavras-chave: web semântica; linked data; gestão da informação; bases de dados; análise bibliométrica.

ABSTRACT

The new technologies cause profound transformations in society, since the information object, today, is dynamically available, in large volumes, in different formats and means. Information Science plays a fundamental role in the management of information in different environments, the Internet being a field full of heterogeneity and chaotic to search for information. In this sense, Berners-Lee proposed the Semantic Web, in which computers can interpret information and make inferences through the *Extensible Markup Language (XML)*, *Resource Description Framework (RDF)* and the use of *Ontology Web Language (OWL)* that form the basis for structuring the Web of meaning. Another proposal is Linked Data, which aims to link structured data sets on the Web in a machine-readable and interpretable way. In this context, the objective of this research is to analyze the international and national publications that converge the Semantic Web and Linked Data, in the organization of information. Methodologically, the research regarding the ends fits as exploratory; as for the qualitative data; and as for the means, a bibliographic research. For quantitative analysis, descriptive statistical methods were used and for qualitative analysis, a content analysis method. The results indicate that in the quantitative aspect, there is convergence between national and international productions in points of low productivity of articles on the themes together; journal choices for publication are for those with the highest ratings; there is a strong relationship between the production of articles with authorship and co-authorship and the formation of collaboration clusters. There is divergence in the number of citations, with a low average for national articles and a higher average for international ones. In the qualitative aspect, 13 thematic categories were outlined, with the category of conversion and sharing of databases for Semantic Web and Linked Data standards being the one that concentrates the largest number of publications.

Keywords: semantic web; linked data; information science; data base; bibliometric analysis.

LISTA DE FIGURAS

Figura 1 – Fases de desenvolvimento de tecnologias e construção da Web 1.0 até 4.0.....	32
Figura 2 – Diagrama da Web Semântica.....	39
Figura 3 – Representação de informação em código HTML de uma página Web.	42
Figura 4 – Resultado do código HTML apresentado em um navegador.....	43
Figura 5 – Transcrição de código HTML para linguagem XML.	43
Figura 6 – Grafo de tripla de um livro descrevendo seu título.....	51
Figura 7 – Grafo genérico de uma tripla (Sujeito, predicado e valor).....	51
Figura 8 – Grafo RDF com três triplas de um mesmo recurso com propriedade diferente. ...	51
Figura 9 – Grafo RDF composto de relações entre as triplas com outros recursos referentes a um mesmo publicador para dois livros.....	52
Figura 10 – Representação de URI interligando um endereço a uma pessoa no formato RDF.	52
Figura 11 – Utilização de vocabulário FOAF/name para descrever pessoas com uso de URI.	53
Figura 12 – Gráfico em RDF composto com 5 nós e URIs.....	54
Figura 13 – Ciclo de Vida do Linked Data.....	66
Figura 14 – Linked Open Data Cloud Diagram	68
Figura 15 – Mapa de coautoria com formação de 1 cluster – <i>Scielo</i>	80
Figura 16 – Mapa de coautores sem a rede de interligações totalmente interligada de quatros clusters.	81
Figura 17 – Mapa de coautoria gerado dos 1252 registros recuperados na <i>Web of Science</i> . ..	84
Figura 18 – Mapa de coautoria com formação de 2 clusters – <i>Web of Science</i>	90
Figura 19 – Mapa de coautores sem a rede de interligações totalmente interligada de seis clusters.	91

LISTA DE QUADROS

Quadro 1 – Representação das informações de grafo RDF.....	54
Quadro 2 – Representação do código OWL em sintaxe XML	61
Quadro 3 – Critérios de maturidade de dados abertos (Linked Open Data – LOD) ...	67
Quadro 4 – Periódicos que publicaram os artigos entre 2016 e 2020 – <i>Scielo</i> Brasil ..	78
Quadro 5 – Relação de autores e coautores em publicação dos artigos sobre Web Semântica e Linked Data – <i>Scielo</i> Brasil.....	79
Quadro 6 – Quantidade de citações por artigo – <i>Scielo</i> Brasil	82
Quadro 7 – Periódicos utilizados para publicações - 2010 a 2021 – <i>Web of Science</i>	87
Quadro 8 – Autores com publicação individualizada	89
Quadro 9 – Autores mais citados sobre os temas Web Semântica e Linked Data entre os anos de 2010 a 2021 na <i>Web of Science</i>	93
Quadro 10 – Categorias de classificação temática.....	98
Quadro 11 – Classificação dos títulos de artigos de acordo com as classes temáticas..	98

LISTA DE TABELAS

Tabela 1 – Lista de periódicos para as publicações sobre Web Semântica e Linked Data - Base de Dados <i>Scielo</i> Brasil no período de 2016 a 2020	78
Tabela 2 – Grau de colaboração de autores de produção científica entre 2016 e 2020 sobre os temas Web Semântica e Linked Data - <i>Scielo</i> Brasil	82
Tabela 3 – Quantidade de publicações nos periódicos entre 2010 e 2021 referente a Web Semântica e Linked Data – <i>Web of Scicnce</i>	86
Tabela 4 – Grau de colaboração de autores de produção científica entre 2010 e 2021 sobre os temas Web Semântica e Linked Data – <i>Web of Science</i>	92
Tabela 5 – Quantificação das categorias por ano de 2010 a 2021 das bases de dados <i>Scielo</i> e <i>Web of Science</i>	103

LISTA DE GRÁFICOS

Gráfico 1 – Quantidade de artigos publicados na Base de Dados da <i>Scielo</i> Brasil sobre os temas: Web Semântica e Linked Data entre 2016 e 2020.....	77
Gráfico 2 – Quantidade de artigos publicados na Base de Dados da <i>Web of Science</i> sobre os temas: Web Semântica e Linked Data entre 2010 e 2021	85
Gráfico 3 – Percentual das publicações nos periódicos entre 2010 e 2021 referente a Web Semântica e Linked Data – <i>Web of Science</i>	86
Gráfico 4 – Quantidade de autores com maior colaboração na produção de artigos sobre Web Semântica e Linked Data – <i>Web of Science</i>	89

LISTA DE SIGLAS

BRAPCI	Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CDU	Classificação Decimal Universal
CI	Ciência da Informação
CSS	Cascading Style Sheets
DAML DARPA	Agent Markup Language DARPA
DATASET	Conjunto de Dados
DC	Dublin Core
DTD	Document Type Definitions
ENDPOINT	Ponto de Acesso
FOAF	Friend of a Friend
HTML	Hypertext Markup Language
HTTP	Hipertext Transfer Protocol
LM	Linguagem de Marcação
LOD	Linked Open Data
NCSA	National Center for Supercomputing Applications
OCLC	Online Computer Library Center
OIL	Ontology Inference Layer
OWL	Ontologic Web Language
OWL DL	Ontologic Web Language Description Logic
OWL Full	Ontologic Web Language Full
OWL Lite	Ontologic Web Language Lite
RDF	Resource Description Framework

RDF Schema	Resource Description Framework Schema
RDFS	Resource Description Framework Schema
SCIELO	Scientific Electronic Library Online
SCOPUS	Base de Dados de Pesquisa Científica e Citações da Elsevier
SGML	Standard Generalized Markup Language
SHOE	Simple HTML Ontology Extension
SPARQL	SPARQL Protocol and RDF Query Language
SQL	Standard Query Language
SRI	Sistema de Recuperação de Informação
TAGS	Etiquetas de Marcação
TRIPLE STORE	Bases de Dados de Triplas
TOVE	Toronto Virtual Enterprise
USP	Universidade Estadual Paulista
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
WOS	Web of Science
W3C	World Wide Web Consortium
XML	Extensible Markup Language
XML Schema	Extensible Markup Language Schema
XSD	XML Schema Definition
XSL	Extensible Stylesheet Language

SUMÁRIO

1 INTRODUÇÃO	16
1.1 CONTEXTO DA PESQUISA.....	20
1.1.1 Cenário	20
1.1.2 Definição do problema	21
1.1.3 Objetivos.....	22
1.1.3.1 Objetivo geral.....	22
1.1.3.2 Objetivos específicos.....	22
1.1.4 Justificativa	22
2 CIÊNCIA DA INFORMAÇÃO E GESTÃO DA INFORMAÇÃO	24
2.1 A CIÊNCIA DA INFORMAÇÃO	24
2.2 A GESTÃO DA INFORMAÇÃO.....	27
3 A INFORMAÇÃO NA INTERNET	31
3.1 A ESTRUTURA DA WEB	31
3.2 LINGUAGEM DE MARCAÇÃO	33
3.3 METADADOS	34
3.3.1 O padrão Dublin Core.....	36
3.4 REPRESENTAÇÃO DO CONHECIMENTO.....	37
4 WEB SEMÂNTICA	38
4.1 A PROPOSTA DA WEB SEMÂNTICA.....	38
4.2 XML – EXTENSIBLE MARKUP LANGUAGE.....	40
4.3 DTD – DOCUMENT TYPE DEFINITIONS	44
4.4 XML SCHEMA.....	45
4.5 RDF – RESOURCE DESCRIPTION FRAMEWORK	48
4.5.1 Modelo de representação dos dados em rdf	50
4.6 RDFS – RESOURCE DESCRIPTION FRAMEWORK SCHEMA.....	56
4.7 ONTOLOGIA.....	57

4.7.1 Engenharia de ontologia	59
4.7.2 Ontologic web language – owl	59
4.8 SPARQL - SIMPLE PROTOCOL AND RDF QUERY LANGUAGE.....	62
5 LINKED DATA	64
5.1 LINKED DATA (DADOS VINCULADOS).....	64
5.2 CICLO DE VIDA DO LINKED DATA	65
6 PROCEDIMENTOS METODOLÓGICOS	69
7 ANÁLISE BIBLIOMÉTRICA DAS PUBLICAÇÕES SOBRE A CONVERGÊNCIA DA WEB SEMÂNTICA E LINKED DATA.....	76
7.1 ASPECTOS QUANTITATIVOS.....	76
7.1.1 Das publicações recuperadas na base de dados da <i>Scielo</i>	76
7.1.2 Das publicações recuperadas na base de dados da <i>Web of Science</i>	83
7.2 ASPECTOS QUALITATIVOS.....	97
8 DISCUSSÃO	104
9 CONCLUSÕES.....	108
REFERÊNCIAS	111

1 INTRODUÇÃO

O ato de comunicar sempre fez parte da sociedade em seu processo evolutivo, visto que, costumes, ideias e aprendizagem foram passados e sistematizados como objetos informacionais para gerações seguintes. A informação tornou-se um elemento essencial para a construção de conhecimento, dessa forma, necessitando ser documentada por meios de registros e servindo como fonte para a consolidação de diversas ciências. Diante da necessidade de sistematização das informações, recuperação, guarda gerenciamento e compartilhamento, a Ciência da Informação (CI) destacou-se como forma de propagação do conhecimento para a sociedade, além do que, pode propor soluções viáveis diante de problemas informacionais, assim, garantindo um maior desenvolvimento social (ANNA, 2018, p. 346; LE COADIC, 2004, p. 2). As diversas mudanças da sociedade impuseram a Ciência da Informação associações com outros saberes, nesse sentido, segundo Araújo (2011, p. 122), a Ciência da Informação é apontada como uma ciência multidisciplinar, haja vista usos basilares da Biblioteconomia, Arquivologia, Museologia e Gestão da Informação. Em vista disso, essa ciência possui como objetivo investigar as propriedades, uso, transmissão, processamento, armazenamento, recuperação e os fluxos da informação para dispor uma maior acessibilidade e usabilidade ao conteúdo informacional. Assim, decorre uma preocupação com o corpo de conhecimento e pesquisas referente à representação da informação em diversos sistemas, bem como nas aplicações direcionadas a computação (BORKO, 1968, p. 1).

O desenvolvimento tecnológico ocasionou profundas transformações na sociedade e a informação destacou-se como principal ativo valorado. Nesse contexto, tem sido utilizada na Tecnologia da Informação e Comunicação (TICs) com a finalidade de proporcionar melhores acessos aos diferentes objetos informacionais (ROZA, 2018, p. 178). Assim, o papel de gerenciamento das informações ganha fundamental importância, visto que, a informação como um montante, precisa ser utilizada de forma a garantir uma maior acurácia para a geração de conhecimento. Maimone e Silveira (2007, p. 65) apontam que o gerenciamento da informação possibilita a sociedade o acesso a informações relevantes para satisfazer as necessidades de cada indivíduo, logo, os profissionais da Ciência da Informação adquiriram um *status* fundamental no processo de gestão do conteúdo informacional.

Machado *et al.* (2016, p. 155) decorrem que a Gestão da Informação (GI) desempenha o papel de identificar pontos relevantes de eficiência e eficácia dos processos de uso da

informação dentro das organizações, além do que, visa organizar o fluxo informacional para identificar possíveis ajustes. Percebe-se que a GI tem fundamental importância dentro das organizações, pois possibilita uma melhor qualidade das atividades e uso dos recursos, geração de conhecimento e, em especial, com as novas tecnologias, ferramentas computacionais e acesso à Internet.

Diante das novas tendências computacionais que estão à disposição das organizações e da sociedade, a Tecnologia da Informação e Comunicação (TICs) tem se destacado como apoio no sentido de oferecer infraestrutura computacional, assim, garantindo a interoperabilidade entre sistemas para a troca de informações. Hedler *et al.* (2016, p. 190) destacam que sistemas de computação nas nuvens têm possibilitado meios de computação distribuída, além do que, a criação de novos produtos e serviços.

Acrescenta-se a importância que tem sido dada aos sistemas operacionais embarcados para acesso à Internet, visto que as aplicações em dispositivos móveis têm possibilitado acessos a vários serviços e a um grande volume de dados, além do que, o poder de processamento, hoje, encontra-se na mão dos usuários em qualquer lugar e a qualquer hora.

A grande rede mundial, desde a sua concepção por Tim Berners-Lee, em 1989, tem passado por grandes transformações. Segundo Laufer (2015, p. 10), a concepção original da Web era fazer uso de recursos de consultas a documentos por meios de estruturas de hipertextos em uma arquitetura cliente/servidor utilizando URL (*Uniform Resource Locator*), linguagem de marcação *HTML (Hypertext Markup Language)* e um protocolo de comunicação *HTTP (Hypertext Transfer Protocol)*.

Nessa estrutura, a compreensão do conteúdo dos documentos disponibilizados na Web fica a cargo dos usuários, não havendo possibilidades de compreensão por máquinas, nem tampouco uso de inferências sobre as informações; assim, apresentava-se uma Web de documentos e não direcionada para informações (DIAS; SANTOS, 2003, p. 79). Tendo em vista esse ambiente e a necessidade de facilitar processos interpretativos ao conteúdo informacional, novas propostas foram lançadas como: linguagens de marcação XML (*Extensible Markup Language*), incorporação de metadados como Dublin Core, utilização de recurso de identificação URI (*Uniform Resource Identifier*) que possibilitaram descrever novos recursos na Web; contudo, ainda havia uma Web direcionada para sintaxe e não para semântica dos documentos. Trabalho publicado por Berners-Lee, Lassila, Hendler (2001, p. 1, tradução nossa), sugere uma nova visão, sendo a Web Semântica onde a “[...] estrutura ao

conteúdo significativo das páginas da Web, criam um ambiente onde os agentes de software que transitam de uma página para outra possam facilmente realizar tarefas sofisticadas para os usuários”.¹ Logo, essa concepção sugere que os dados da Web tenham significado, possibilitando a interpretação do conteúdo informacional por meios de máquinas.

Para alcançar tais objetivos, evidentemente, que novos recursos deveriam ser implementados. Para tanto, além da linguagem de marcação XML que trouxe uma maior flexibilidade na construção do conteúdo das páginas em relação à linguagem HTML, foram desenvolvidos o RDF (*Resource Description Framework*), RDF schema e uso de ontologias (SANTAREM SEGUNDO; CONEGLIAN, 2016, p. 220).

Para a descrição das informações com significado, o uso da estrutura proposta pelo RDF foi baseado para descrever quaisquer coisas na Web por meios de identificadores únicos as URIs (LIMA; CARVALHO, 2005, p. 24). O RDF tem três princípios essenciais: sujeito predicado e o objeto; assim, para que as informações na Web Semântica tenham significado é necessário a formatação em triplas (BREITMAN, 2005, p. 50; LAUFER, 2005, p. 34). Entretanto, o RDF não possui recursos para inferências lógicas do conteúdo informacional, logo, para suprir tal necessidade faz-se uso das ontologias (BREITMAN, 2005, p. 53).

Nesse sentido, a World Wide Web Consortium (2013) (on-line) propôs uma linguagem de ontologia para a Web Semântica: a OWL, com o objetivo de representar o conhecimento sobre as coisas e suas relações podendo ser exploradas em programas de computadores. Assim, criam-se possibilidades de agentes computacionais (softwares) terem acesso a um maior número de documentos, além de facilitar a troca de informações com outros agentes computacionais criando uma rede de interligações de dados conectados: o Linked Data (BERNERS-LEE *et al.*, 2001, p. 3).

Linked Data faz menção ao conjunto de dados estruturados que são publicados e disponibilizados na Web atendendo aos requisitos da Web Semântica e que possam ser interligados a outros conjuntos de dados externos, legíveis e interpretáveis por máquinas (MENDONÇA, 2013, p. 33). Uma nova proposta referente aos dados disponíveis na Web é o Linked Open Data (LOD), visto que além das exigências do Linked Data, os dados deverão

¹ [...] structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users (BERNERS-LEE, LASSILA, HENDLER, 2001, p. 1).

atender as possibilidades de distribuição e reutilização de forma livre, sem que haja restrições de propriedades de uso (MENDONÇA, 2013, p. 34; RAUTENBERG *et al.*, 2017, p. 124).

Os estudos bibliométricos destacam-se como uma ferramenta dentro da Ciência da Informação, a fim de mensurar a produção científica e as relações de produtividade no campo de pesquisa. Dessa forma, sua utilização em análise aos temas Web Semântica e Linked Data possibilitará uma visão de como estão sendo tratados nas diversas áreas das ciências e quais os pontos focais dos pesquisadores diante de suas publicações. Logo, possibilitará um direcionamento para novas pesquisas e campos temáticos, além, de identificar as principais instituições de pesquisa e apoio que têm estimulado estudos sobre esses temas.

De acordo como Tavares e Celerino (2018, p. 11), a bibliometria tem uma característica de multidisciplinaridade para identificar o comportamento evolutivo da produção literária em certos períodos temporais e dentro de dado contexto, por meios de métodos estatísticos e matemáticos utilizando-se de análise quantitativas de produção. Possibilitando, assim, conhecer tendências e a compreensão sobre determinado domínio do saber e os graus de crescimento e interesse sobre determinados temas (CAMARGO; BARBOSA, 2018, p. 114). Para tanto, faz-se uso de métricas que possam quantificar as produções científicas, entre elas destacam-se: análise de citações, fator de impacto e cocitações (GUEDES, 2012, p. 95). Segundo Grácio (2016, p. 83), é salutar destacar a relevância dos estudos das citações, no sentido de averiguar os sistemas de comunicação, estrutura e interações dentro de determinado domínio de estudo.

A proposta dessa pesquisa é direcionada à análise da articulação entre a Web Semântica e Linked Data no processo da organização da informação, por meio de um estudo bibliométrico baseado em periódicos científicos indexados em bases de dados especializadas, seja de âmbito nacional, seja a nível internacional.

Com base nas informações obtidas nas bases de dados *Scielo* e *da Web of Science* o trabalho apresenta-se estruturado em seções e subseções, sendo as seções principais: contexto da pesquisa, fundamentação teórica, caracterização da pesquisa, procedimento metodológicos, análise dos resultados, conclusões e as referências.

1.1 CONTEXTO DA PESQUISA

1.1.1 Cenário

A Ciência da Informação (CI) tem desempenhado ações efetivas diante da informação como seu objeto de trabalho. De acordo com Carvalho e Crippa (2013, p. 242), os estudos de Paul Otlet e La Fontaine, em 1893, em que empreenderam esforços na construção de repositórios bibliográficos e a criação de Sistema de Classificação Decimal Universal (CDU), desde então, a CI tem evoluído para adequar-se aos contextos de transformação que a sociedade tem passado no sentido de disponibilizar a informação de forma eficiente com a finalidade de gerar conhecimento. Fato que, segundo Saracevic (1996, p.2), com o advento da Segunda Guerra Mundial emergiu novas formas de gerenciamento informacionais devido ao grande volume de dados. Nesse sentido, ressalta Le Coadic (2004, p. 206), a Ciência da Informação, apesar de um caráter no campo da Ciência Social, demandou a necessidade de aproximação com a Ciência e Tecnologia.

Diante dos cenários de transformações na sociedade, a Gestão da Informação (GI) mereceu destaque no sentido de organizar a informação e sistematizá-la. Conforme ressalta Davenport (1998, p. 170), a GI possibilitou tratar de um conjunto de atividades estruturadas de modo que as empresas pudessem distribuir e usar a informação e o conhecimento adquirido. Dentro desse contexto, a Tecnologia da Informação passou a integrar o ambiente organizacional, como referendam Nassif e Resende (2016, p. 112), em que diversas aplicações de gestão foram utilizadas para a geração de soluções, em especial, com o advento da Internet.

Nesse aspecto, um novo cenário de conectividade apresentou-se a partir da proposta da implementação da Web, por Berners-Lee, em 1989, onde os documentos poderiam ser acessados por meios links e visualizados em navegadores apenas utilizando certas tecnologias para construção de páginas Web e seu compartilhamento. Entretanto, a estrutura dos documentos estava direcionada para sua apresentação na Web. Nos dizeres de Dias e Santos (2003, p. 79), essa estrutura demandava esforços do usuário para a interpretação do conteúdo e não possibilitando a interpretação e inferência por máquinas. Diante disso, Tim Berners-Lee, Hendler James e Lassila Ora publicaram o artigo “*The Semantic Web*”, na *Scientific American*, em 2001, descrevendo uma nova estrutura de acesso às informações: a Web semântica, onde a Web sintática (atual) passaria para a Web Semântica, ou seja, direcionada para o significado dos documentos, dessa forma, possibilitaria a interpretação e inferência por

máquina (SANTAREM SEGUNDO; CONEGLIAN, 2016, p. 220). Em acréscimo, apontam os autores que diversas tecnologias foram sugeridas para a construção da Web Semântica e que ainda estão sendo aperfeiçoadas, como a linguagem XML, RDF (*Resource Description Framework*) e ontologias.

Devido às tecnologias emergentes e recursos computacionais numerosos, onde possibilita tudo estar conectado na grande rede mundial, um novo cenário é desenhado: O Linked Data, descritos por Bizer, Berners-Lee, Heath (2009, p. 2) como dados vinculados que são usados para criar links digitais entre diferentes fontes e legíveis e interpretáveis por máquinas.

Nesse sentido, abrem-se novas possibilidades de conexão e uso da informação, onde qualquer coisa pode ser referenciada no ambiente Web, portanto, um grande volume de informações que ainda não estão disponíveis, logo poderão ser acessadas. Nesse aspecto, estudos direcionados para a Web Semântica e Linked Data são de fundamental importância, visto que a Ciência da Informação tem um papel em proporcionar condições para novas formas de acesso à informação e geração de conhecimento.

Acrescenta-se que o processo de aprendizagem há de passar por enormes transformações diante das novas possibilidades de acesso remoto e interligação de plataformas que possuem grandes heterogeneidades de dados, além dos já existentes na Internet. Dessa forma, há a necessidade de uma melhor estruturação da informação na Web com a finalidade de que máquinas possam auxiliar em descobertas de novos conteúdos informacionais. Assim, ficando a cargo dos agentes computacionais realizarem busca com maior acurácia e dispor o conteúdo informacional para organizações e usuários.

Dessa feita, torna-se primordial entender o que estão sendo produzidos em artigos científicos e disponibilizados em bases de dados especializadas, sobre esses assuntos e se há uma convergência ou estão sendo tratados de forma individualizada no campo da Ciência da Informação.

1.1.2 Definição do problema

Diante das profundas transformações que a sociedade tem passado devido ao surgimento das novas tecnologias, a Ciência da Informação deve valer-se de um contínuo aprofundamento no campo de pesquisa de seu objeto de trabalho, a informação. Em especial, no ambiente heterogêneo como a Internet onde disponibiliza diversos tipos de informações,

assim, a Ciência da Informação tem um desafio em adequar os dados que já existem com as novas formas de estruturação semântica e criação de redes de significados.

Portanto, as análises nos campos de pesquisa científica podem indicar as áreas de maiores interesses, bem como estimular novos trabalhos científicos para preencher lacunas com campo da Web Semântica e Linked Data dentro da Ciência da Informação, assim, estimulando novas pesquisas dentro da academia.

Para tanto, necessita indagar questões quanto ao nível quantitativo dessa produtividade científica, assim, o presente trabalho toma como ponto de problematização o seguinte: como está ocorrendo a produção científica acerca das convergências entre Web Semântica e Linked Data na organização da informação, por meios das bases de dados especializadas *Scielo e da Web of Science*?

1.1.3 Objetivos

1.1.3.1 Objetivo geral

Analisar as publicações internacionais e nacionais que convergem a Web Semântica e Linked Data, na organização da informação.

1.1.3.2 Objetivos específicos

- Caracterizar a Web Semântica e o Linked Data nos processos de gerenciamento da informação;
- Levantar as publicações sobre o tema em bases de dados nacionais e internacionais;
- Desenvolver análise bibliométrica das publicações relevantes.

1.1.4 Justificativa

O desenvolvimento dessa pesquisa encontra apoio no objeto informacional como fonte de estudo da Ciência da Informação, visto que, é por meio da informação, que se torna possível a geração do conhecimento.

A Ciência da Informação caracteriza-se por seu aspecto multidisciplinar no estudo e análise da informação, de modo que Borko (1968, p. 4) relata o objetivo dessa ciência com a investigação das propriedades, o comportamento, bem como seu uso, meios de transmissão, processamento, armazenamento e recuperação do conteúdo informacional.

Nesse sentido, a informação como objeto de estudo, é designada de acordo com Correia e Zandonade (2018, p. 98), recuperando o pensamento de Barreto (2002), como “estruturas simbólicas significantes com a competência e a intenção de gerar conhecimento no indivíduo, em grupos e na sociedade”. Nessa visão, Buckland (1991, p. 352) destaca que a informação pode ser concebida como um processo no sentido do ato de informar, ou seja, ação de comunicar algo; a informação como conhecimento, ou seja, ligada a algum fato, assunto ou evento a ser transmitido; a informação com coisa atribuída a objetos descritos ou representados em algum suporte físico.

Portanto, a pesquisa justifica-se no estudo da informação diante das novas tecnologias emergentes e acesso para a geração de novos conhecimentos em ambientes digitais, em especial na Web, em virtude do ambiente com grandes fontes de informações heterogêneas, logo, demanda estudos aprofundados de pesquisa para que possa facilitar a gestão da informação e geração de novos conhecimentos. Acrescenta-se, também, que a análise da produção científica em determinada área do saber possibilitará traçar novos rumos de pesquisa.

Some-se a isso o fato de que no campo social as pesquisas sobre Web Semântica e Linked Data irão possibilitar acesso mais rápido ao conjunto de informações, pois máquinas poderão interpretá-las e criar inferências sobre os dados, assim, diminuindo tempo de busca e acurácia no retorno das informações. As novas tecnologias embarcadas e à Internet das coisas, em que tudo estará conectado, ampliará novos horizontes para a recuperação e uso dos objetos informacionais.

No lado pessoal, a escolha por esse tema para o TCC decorre da necessidade de novas descobertas e visão para o futuro no uso dos recursos computacionais e da Internet, bem como ações efetivas como Gestor da Informação em estimular modificações sociais e acesso à informação de forma mais igualitária e democrática para a sociedade.

2 CIÊNCIA DA INFORMAÇÃO E GESTÃO DA INFORMAÇÃO

Nessa seção são apresentados conceitos referentes à Ciência da Informação (CI) na visão de diversos pesquisadores e pontos históricos do surgimento da CI. Também é abordada a importância da informação como objeto de trabalho e seu uso na construção do conhecimento. Acrescenta-se a interação da CI com a Tecnologia da Informação e Comunicação (TIC).

Em seguida são descritos os conceitos sobre a Gestão da Informação (GI) na visão de diversos pesquisadores, posteriormente, relata-se sobre a informação dentro da Gestão da Informação e como ela é apresentada para geração de conhecimento e do fluxo informacional nas organizações. Por fim, é feita referência às tecnologias emergentes e ao papel da TI na Gestão da Informação.

2.1 A CIÊNCIA DA INFORMAÇÃO

O processo de comunicação é inerente ao ser humano para que possa expressar seus costumes e ideias por meios de informações. Fato que no desenvolvimento da sociedade a informação foi adquirindo um papel importante, logo, necessitando ser documentada em forma de registros. Desse modo veio a possibilitar o acesso para futuras gerações e, assim, impulsionando o surgimento de diversas ciências, inclusive, para o gerenciamento da informação e comunicação.

Com efeito, se tornou processo de estudo quanto à forma de registro, recuperação e divulgação, a fim de gerar conhecimento para a sociedade, assim, as iniciativas de sistematizar e gerenciar o objeto informacional e os estudos de vários pesquisadores resultou na Ciência da Informação – (CI) (ANNA, 2018, p. 344). Nesse sentido, a Ciência da Informação obteve um papel essencial como forma de propagar conhecimento, visto que, ao propor métodos, teorias e técnicas aplicadas para o gerenciamento e solução de problemas informacionais, contribuiu para o desenvolvimento social e econômico ocupando, assim, uma posição de Ciência Social (ANNA, 2018, p. 346). Decerto, ao se preocupar em solucionar problemas relacionados à necessidade de acesso à informação para a sociedade, há uma aproximação às questões sociais e culturais (LE COADIC, 2004, p. 21).

A necessidade de solucionar os diversos tipos de problemas demandaram enfoques interdisciplinares com soluções múltiplas, por conseguinte, a CI abarcou diferentes contribuições da Biblioteconomia, Arquivologia e Museologia (ARAÚJO, 2011, p. 122).

Visto que, a Ciência da Informação teve sua base proveniente do estudo e da pesquisa desses problemas sobre o conhecimento registrado, nos quais foram desenvolvidos nas ciências citadas e da Documentação (CORREIA; ZANDONADE, 2018, p. 93).

Como marcos históricos, destacam-se os trabalhos referentes à documentação que trouxeram uma significativa contribuição para a Ciência da Informação, tendo como relevância os estudos de Paul Otlet e Henri La Fontaine, em 1893 (CARVALHO; GRIPPA, 2013, p. 242), onde buscaram a construção de repositórios bibliográfico universal, a normalização de fichas catalográficas e a criação de uma classificação decimal universal (ARAÚJO, 2011, p. 112). As revoluções científicas e as técnicas que surgiram com o advento da Segunda Guerra Mundial, também, contribuíram para o surgimento da CI, some-se a isto, de ser impulsionado pela necessidade de gerenciar o crescimento exponencial das informações nesse período pós-guerra (SARACEVIC, 1996, p. 42). A partir do pós-guerra, a Biblioteconomia, como disciplina acadêmica, dividiu-se dando origem à disciplina Biblioteconomia e Ciência da Informação (CORREIA; ZANDONADE, 2018, p. 85).

Sendo firmada como uma ciência nova, em comparação com as ciências classificadas como clássicas, a Ciência da Informação destaca-se no aspecto da interdisciplinaridade, visto que, mantém relações de associação com outras disciplinas tais como: a Psicologia, Sociologia, Informática, Matemática, Lógica, Estatística, Direito, Filosofia, Política, entre outras (LE COADIC, 1996, p. 22). Nesse contexto, Bicalho e Oliveira (2011, p. 51) esclarecem que a interdisciplinaridade é obtida quando duas ou mais disciplinas realizam intercâmbios com a finalidade de um enriquecimento mútuo. Afirmam as autoras que os esforços são empregados fora do domínio de cada disciplina e de sua linguagem, portanto, cada disciplina lança-se em um novo domínio de ação em construção sem que haja, necessariamente, propriedades exclusivas entre elas.

Acrescenta-se que a Ciência da Informação, apesar de seu caráter no campo da Ciência Social, não deixou à parte os aspectos tecnológicos. De fato, esses desempenham uma posição imprescindível no meio do gerenciamento da informação, portanto, tem aproximado a ciência e tecnologia, visto que “[...] não existe ciência sem tecnologia e nem tecnologia sem ciência” (LE COADIC, 2004, p. 206). Pesquisadores, ao longo do tempo, têm ressaltado a importância dessa ciência e empregado esforços com objetivo de conceituá-la. Segundo Borko (1968, p. 3), a Ciência da Informação como disciplina tem o objetivo de investigar as propriedades,

comportamento da informação, uso, transmissão, processamento, armazenamento e recuperação sendo conceituada como:

A Ciência da Informação é a disciplina que investiga as propriedades e o comportamento informacional, as forças que governam os fluxos de informação, e os significados do processamento da informação, visando à acessibilidade e a usabilidade ótima. A Ciência da Informação está preocupada com o corpo de conhecimentos relacionados à origem, coleção, organização, armazenamento, recuperação, interpretação, transmissão, transformação, e utilização da informação. Isto inclui a pesquisa sobre a representação da informação em ambos os sistemas, tanto naturais quanto artificiais, o uso de códigos para a transmissão eficiente da mensagem, bem como o estudo do processamento e de técnicas aplicadas aos computadores e seus sistemas de programação (BORKO, 1968, p. 1).

Na visão de Saracevic (1996, p. 41), a Ciência da Informação abrange o campo da pesquisa científica, quanto à prática profissional para solucionar problemas informacionais, por meios dos métodos disponíveis a essa ciência. No ponto de vista de Le Coadic (2004, p. 206), a CI é referendada com uma aproximação da ciência e a tecnologia; segundo seus argumentos, a cultura informacional necessita da aproximação da cultura tecnológica, como também da científica. Nesse sentido, o autor aponta cinco princípios científicos norteadores para essa aproximação: o princípio produtivista, o princípio interacionista, o princípio do consumo, o princípio métrico e o princípio eletrodigital.

As novas tecnologias proporcionaram profundas transformações na sociedade e a informação tornou-se o um ativo principal, somando-se ao conjunto de recursos computacionais têm facilitado melhores acessos e a Tecnologia de Informação e Comunicação (TICs) têm desempenhado papéis essenciais nesse cenário (ROZA, 2018, p. 178). O termo informação e seu significado têm sido usados em vários domínios, além da Biblioteconomia e da Ciência da Informação, com efeito, tem causado dificuldade em seu entendimento diante de vários conceitos.

A Ciência da Informação tem como objeto de trabalho a própria informação. Contudo, definir um conceito único para o que seja informação não é uma tarefa fácil, em virtude de sua abrangência e as associações que ora são feitas no campo mais científico utilizando-se as bases do conhecimento científico ou direcionada para o usuário e, por vezes, focada para o campo epistemológico (SILVA e GOMES, 2015, p. 146). Todavia, faz-se necessária a compreensão do que seja informação diante da dimensão como fonte de pesquisa para essa ciência.

Em um sentido amplo, Le Coadic (1996, p. 5) afirma que a informação é “um conhecimento escrito (gravado) sob a forma escrita (impressa ou numérica), oral ou

audiovisual”. O autor sugere que a informação comporta um elemento de sentido, ou seja, há um significado a ser transmitido, por um ser consciente, que utiliza um meio ou um suporte para representar uma mensagem, essa formada por um conjunto de signos. Como descrito por Correia e Zandonade (2018, p. 93), no âmbito da CI, a informação é compreendida como sendo o conhecimento registrado. Como forma exemplificativa de uma representação dos diversos conceitos, Correia e Zandonade (2018, p. 98) apontam algumas das contribuições dos pesquisadores:

À forma material da existência do conhecimento chamamos de informação. Em outras palavras, a informação é um elemento definido do conhecimento, expresso por meio da linguagem natural ou de outros sistemas de signos, apreendidos pelos órgãos dos sentidos (FOG, 1979 apud CORREIA e ZANDONADE, 2018, p. 96).

Estruturas simbolicamente significante com a competência e a intenção de gerar conhecimento no indivíduo, em seu grupo e na sociedade (BARRETO, 2002 apud CORREIA e ZANDONADE, 2018, p. 96).

Conceito social de informação no âmbito da análise de domínio e comunidades discursivas (HJØRLAND, 2002 apud CORREIA e ZANDONADE, 2018, p. 96).

Conjunto estruturado de representações mentais codificadas (símbolos significantes) socialmente contextualizadas e passíveis de serem registradas em qualquer suporte material (papel, filme, banda magnética, disco compacto, etc.) e, portanto, comunicadas de forma assíncrona e multidirecionada (SILVA, 2002 apud CORREIA e ZANDONADE, 2018, p. 97).

Em Buckland (1991, p. 352), encontram-se três esclarecimentos quanto à concepção de informação: i) a informação como processo, onde é tida como o ato de informar, ação de comunicar um conhecimento ou novidade, ou seja, informar algum fato; ii) em seguida a informação tratada como conhecimento, visto que, o conhecimento comunicado está ligado a algum fato, assunto ou evento a ser transmitida, por sua vez, a informação é intangível; iii) e por último, a informação como coisa, nesse sentido, atribuída para o objeto onde pode ser expressa, descrita ou representada em algum suporte físico, assim, a informação tem uma representatividade material, logo, sendo tangível.

A informação como coisa, materializada, possibilita uma representação documental, visto que, essa é resultante de um processamento e de uma organização. A informação coletada, identificada, organizada e selecionada permite ao usuário o acesso à informação desejada para que seja assimilada e, assim, possibilitar a obtenção de novos conhecimentos (MAIMONE; SILVEIRA, 2007, p. 64).

2.2 A GESTÃO DA INFORMAÇÃO

Constata-se que a informação é a fonte básica para a geração de conhecimento. Nesse aspecto, o seu gerenciamento é primordial para que possibilite à sociedade o acesso relevante

de forma que possam satisfazer as necessidades de cada indivíduo. Consequentemente, os profissionais da área da Ciência da Informação desempenham papel importante em sua organização e disponibilização (MAIMONE; SILVEIRA, 2007, p. 65). O gerir a informação torna-se necessário pelo seu valor para as tomadas de decisões, assim, é necessário sistematizá-la para obter informações de qualidade e disponível. A escassez ou o excesso de informação pode trazer consequências, pois o seu valor não é dado, apenas, pela entrada e saída da informação na organização, mas, sobretudo, como ela é usada para o bem estar da humanidade (MARQUES, 2017, p. 65).

Segundo Marques (2017, p. 65), a Gestão da Informação (GI), no pressuposto de sua designação, tem sido conceituada por diversos pesquisadores desde seu surgimento no século passado, além de servir de objeto de estudo nas áreas como: Tecnologia da Informação e Comunicação, Gestão e a Ciência da Informação. O autor aponta as seguintes contribuições de pesquisadores quanto aos conceitos de GI:

[...] um processo que trata de um conjunto estruturado de atividades que incluem o modo como as empresas obtêm, distribuem e usam a informação e o conhecimento (DAVENPORT, 1998, p. 170).

Conjunto de atividades interligadas e encadeadas, relacionadas com a produção/aquisição, o tratamento, o registo e guarda, a comunicação e uso da informação (quer em suportes tradicionais quer em sistemas de informação baseados em computador), a fim de sustentar as operações e funções de gestão da organização, e o controlo na organização. Modelo processual de Gestão de Informação devia abarcar toda a cadeia de valor da informação, começando pela identificação das necessidades de informação, passando pela aquisição, organização e armazenamento, produtos e serviços, distribuição da informação e fechando o ciclo com a utilização da Informação (CHOO, 2003 apud MARQUES 2017, p. 66).

Machado *et al.* (2016, p. 158) explicam que a informação, a partir de seus objetivos intrínsecos, é tratada pela Gestão da Informação nos pontos relevantes para identificar níveis de eficiência e eficácia dos processos de informação dentro das organizações. Nessa linha de pensamento, aprimorando-os para um desempenho satisfatório a Gestão da Informação nasce da demanda de organizar o fluxo informacional. Mediante essa necessidade existem diversos modelos de Gestão da Informação, como exemplo: O modelo da teoria de comunicação onde baseiam em três níveis da informação: o nível técnico voltado para precisão e eficácia; nível semântico que analisa a capacidade de transferir a mensagem; e por último, o nível de eficácia que mensura o impacto da informação no receptor.

Em decorrência dos estudos das organizações, a Gestão da Informação ganha importância, pois, possibilita uma melhor qualidade das atividades buscando uma alta eficiência. Um dos temas importantes dentro da GI é o processo de compartilhamento da

informação. Entretanto, há desafios para compreender sua dimensão e quais os resultados trazem para a organização. Um dos desafios é apontado no sentido das dificuldades das pessoas em compartilharem as informações, em virtude de resistência pessoal ou por meios de processos de seletividade entre pessoas que terão a informação de forma mais acessíveis ou não (NASSIF; RESENDE, 2016, p. 112).

Para diluir as dificuldades de acesso informacional, a Gestão da Informação lança mão de recursos que possam simplificar o acesso dentro das organizações. Nesse sentido, a Tecnologia da Informação (TI) passa a integrar o ambiente organizacional, disponibilizando diversas aplicações de gestão para geração de soluções e, com o advento da Internet, emergem diversas possibilidades de comunicação para facilitar tal ação (NASSIF; RESENDE, 2016, p. 112). A partir da disponibilidade da informação e seu compartilhamento, a Gestão da Informação encontra apoio na Tecnologia da Informação e Comunicação (TIC), a qual exige mudanças dentro da organização em sua forma de atuar e da necessidade de profissionais qualificados (NASSIF; RESENDE, 2016, p. 113).

A Tecnologia da Informação e Comunicação tem como objetivo oferecer infraestrutura computacional e de comunicação para que as organizações possam interligar sistemas, processos e serviços, a fim de um compartilhamento de informações (AGUIAR; MENDES, 2016, p. 139). Com avanço das tecnologias e meios de comunicação propiciados pela Internet, a integração de sistemas tem migrado para um novo patamar, no qual as informações são armazenadas em servidores remotos e os serviços oferecidos às organizações são disponibilizados por demanda (HEDLER *et al.*, 2016, p. 189).

Essa nova maneira de acesso às informações é denominada computação nas nuvens (na Internet), onde possibilita uma forma de processamento distribuído, o que permite a criação de novos produtos e serviços para as organizações e, conseqüentemente, demanda para os usuários uma concepção diferente de trabalho e colaboração (HEDLER *et al.*, 2016, p. 190). Nessa perspectiva, segundo os autores, os trabalhos realizados em computação nas nuvens utilizam de equipamentos digitais, a fim de ter acesso às informações em tempo real em qualquer lugar e hora. Acrescenta-se que os softwares necessários, em geral, são os navegadores padrões que acessam servidores de serviços disponíveis na Internet. Os acessos aos dados ocorrem de forma remota e a necessidade de processamento é disponibilizada por demanda.

Com a propagação de dispositivos móveis e a criação de diversas aplicações, com uso de sistemas operacionais embarcados para acesso a web, os poderes de processamento das informações estão presentes nas mãos dos usuários (HEDLER *et al.*, 2016, p. 195). Logo, uma revolução no acesso à informação vem sendo disponibilizada, cada vez mais, diante do avanço das tecnologias emergentes, e a Ciência da Informação, bem como a Gestão da Informação têm papel fundamental para o gerenciamento deste volume de dados crescente a cada dia.

3 A INFORMAÇÃO NA INTERNET

Nessa seção são abordados diversos conceitos referentes à estrutura da Web e sua concepção por Berners-Lee, além das tecnologias para construção de informações em páginas Web, com linguagens de marcação e *tags*, identificadores de documentos ou de nós na Internet, além de protocolos de comunicação. Destaque também é dado aos metadados na representação da informação na Web, a exemplo o padrão Dublin Core. Em seguida foca-se na importância da representação do conhecimento como processo na organização da informação.

3.1 A ESTRUTURA DA WEB

O acesso ao conteúdo informacional obteve uma rápida expansão com o surgimento da Web. Desde seu surgimento, a grande rede mundial tem passado por diversas transformações, tanto na forma de interação, quanto no surgimento de novas tecnologias.

Sua concepção original tinha como objetivo utilizar os recursos disponíveis para realizar consultas e acesso a documentos fazendo uso de estrutura de hipertextos (links), e valendo-se de uma arquitetura cliente-servidor² para acesso aos conteúdos requisitados (LAUFER, 2015, p.10).

A estrutura Web concebida por Tim Berners-Lee, em 1989, fez uso de três pontos básicos: a) um identificador de documentos ou de nós denominado de URL (*Uniform Resource Locator*); b) uma linguagem de marcação para descrição dos documentos, por meio do HTML (*Hypertext Markup Language*); c) por fim, um protocolo responsável para estabelecer comunicação entre o cliente/servidor sendo responsável por fornecer como resultado as requisições dos serviços e dos documentos ao cliente, tal é denominado de HTTP (*Hypertext Transfer Protocol*), onde utiliza como base uma interface de navegação (*browser*) para intermediação entre essas requisições (LAUFER, 2015, p.10).

Evidente que outros requisitos se fizeram necessários para a implementação da Web, onde apontam: a estrutura física de conexão, equipamentos e suporte técnico. Mas, de forma sucinta, os documentos da Web são acessados por meios de um identificador (URL) direcionado para uma página Web e, via hipertextos (links), podem interligar-se a outros documentos ou serviços, assim, criando uma grande rede de recursos disponibilizados por

² Servidor é computador que executa serviços ou programas a fim de compartilhar recursos com computadores clientes.

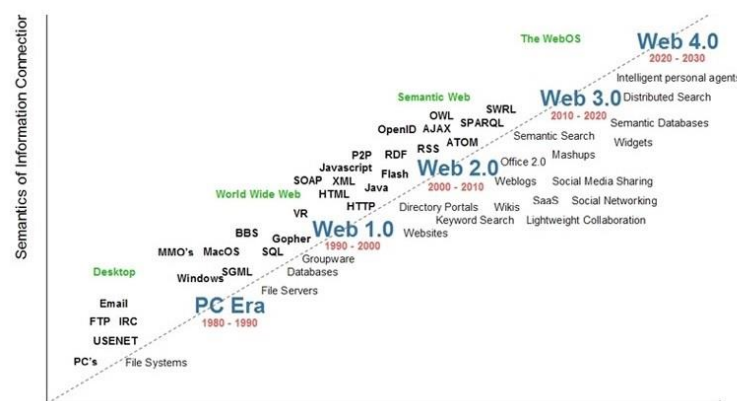
meios de protocolos de comunicação – o HTTP – e acessados pelos navegadores (*browser*) (LAUFER, 2015, p. 11).

Dentro dessa proposta, as estruturas das páginas Web construídas por meio da linguagem HTML fez-se uso de marcadores denominados de *tags*, os quais têm a função de formatar a estrutura dos documentos para exibição no navegador possibilitando ser visualizados pelo usuário (LAUFER, 2015, p.19).

Em análise ao processo evolutivo da Web, desde sua concepção até o momento atual, Manishkumar e Abhijit (2016, p. 75) destacam as fases da Internet e suas tecnologias e recursos disponibilizados para os usuários (Figura 1). Os autores apontam que:

- A Web 1.0 foi a primeira geração sendo denominada de “Web somente de leitura”, visto que, as páginas eram estáticas e com poucas mudanças em atualizações;
- Posteriormente, surgiu a Web 2.0, onde o foco era dado na colaboração e compartilhamento diante dos usuários, sendo denominada de Web Social;
- Em seguida, apontou a Web 3.0, onde o processo de decisão é deslocado do ser humano e passando mais para as máquinas em fornecer conteúdo de forma mais legível, neste caso, é adicionado semântica, sendo denominada de “Web Semântica”, que se encontra em andamento em sua consolidação;
- Por último, apresenta-se a 4.0, sendo uma Web do futuro (ou em andamento), sendo direcionada para ambientes inteligentes com o mundo interligado e gerando capacidade para o ser humano se auto atualizar e obter processamento de computação nas nuvens.

Figura 1 – Fases de desenvolvimento de tecnologias e construção da Web 1.0 até 4.0.



Fonte: Manishkumar e Abhijit (2016, p.77).

Observa-se que processo de evolução da Internet passou por várias fases e incorporação de diversas tecnologias, recursos e linguagens computacionais. Onde, inicialmente, apresentava páginas sem muitos recursos e de forma estática, em seguida, passando para páginas mais dinâmicas e uso dos recursos de consultas para os usuários, posteriormente, seguindo uma trajetória em busca de significado semântico e uso de agentes (softwares) que possibilitem interpretação do conteúdo informacional por máquinas.

3.2 LINGUAGEM DE MARCAÇÃO

O termo marcação é utilizado, de uma forma geral, para descrever anotações ou marcas em um texto, como exemplos: textos em sublinhado, textos com letras em itálico ou outras formas de destacar informações para o usuário. No entanto, no campo da computação, o ato de codificar ou marcar um texto para o processamento determinará como o conteúdo do texto deverá ser interpretado por máquinas, dessa forma, apresentam-se as Linguagem de Marcação (LM) (ALMEIDA, 2002, p. 6). A linguagem de marcação, segundo o autor, entende-se como: um conjunto de convenções que são utilizados para codificação de textos em uma forma específica com a utilização de tipos de marcas permitidas e seus significados, além de meios para distinguir entre as marcas e o texto do documento.

As linguagens de marcação têm oferecido destaque na apresentação de conteúdos para a Internet, ou seja, na construção de sites e páginas para a Web e serem interpretadas pelos *browsers* e *softwares* (ALMEIDA, 2002, p. 6).

Para tanto, uma boa linguagem de marcação deve atender a diferentes sistemas informatizados. Nesse sentido, um dos primeiros padrões utilizados como linguagem de marcação foi a SGML (*Standard Generalized Markup Language*), um padrão não proprietário e de código aberto, que proporcionou a manipulação de informações entre diferentes ambientes computacionais sem que houvesse perda de informações. Tais características fizeram com que a partir do SGML, derivassem outras linguagens de marcação (ALMEIDA, 2002, p. 6).

Entre os diversos tipos de linguagem de marcação, cada uma possui finalidades específicas. De forma geral, as LM utilizam-se de *tags* para delimitar textos para formatação de um documento, a exemplos as *tags* em HTML `<B>` `</B>` e `<I>` `</I>`, que representam as formatações em negrito e itálico, respectivamente, logo o texto `<B><I>` Web Semântica e

Linked Data </I>, será apresentado da seguinte forma em um navegador: **Web Semântica e Linked Data** (FURGERI, 2006, p. 226).

Ainda como exemplo, Laufer (2015, p, 11) relata que a *tag*: <h1> </h1> determina um estilo de formatação em maior destaque no documento em uma forma de título, tal sintaxe é representada da forma: <h1> A Web de Documentos </h1>, assim, resultaria um texto com maior realce apresentado pelo navegador ao usuário.

As *tags* apresentadas determinam a forma da construção das páginas, mas não permitem uma dedução do conteúdo informacional presente nelas; apenas à forma na qual o navegador deve disponibilizá-las de acordo com a sintática do HTML incorporado nos documentos, ou seja, há apenas o entendimento da relação de formatação dos documentos sem nenhuma dedução do seu conteúdo (LAUFER, 2015, p.11). Ressalta-se que essa forma de estrutura permite, apenas, a compreensão do conteúdo dos documentos pelos usuários, mas não cria a possibilidade de uma interpretação ou inferência³ por máquinas, nas quais poderiam mensurar o conteúdo apresentados na Web e satisfazer melhor aos resultados de consultas, portanto, atualmente, apresenta-se uma Web de documentos, assim, não sendo direcionada para a informação de conteúdo semântico, ou seja, a informação com significado computacional (DIAS; SANTOS, 2003, p. 79). Nesse sentido, de acordo com os autores, para obterem-se informações pertinentes nos documentos disponíveis na Web, decorrem maiores esforços de interpretação e extração das informações pelos usuários.

Com interesse em proporcionar uma maior facilidade interpretativa dos conteúdos informacionais, foram lançadas novas propostas de linguagem de marcação, a exemplo da linguagem de marcação XML (*Extensible Markup Language*), uma extensão da linguagem SGML, onde permitem ações complementares com a utilização de padrões de metadados, a fim de conferir um aspecto semântico⁴ às informações disponibilizadas na Internet (ALMEIDA, 2002, p. 6; FURGERI, 2006, p. 229).

3.3 METADADOS

Para que as informações contidas em documentos Web possam ser interpretadas por máquinas, fazem-se necessárias informações adicionais, além da estrutura de marcação

³ Consiste em deduzir um resultado utilizando meios lógicos, mas tendo com base a interpretação de outras informações.

⁴ Direcionada para o significado da informação, ou seja, o conteúdo informacional.

HTML. Nesse caso, ocorre a necessidade do uso descritivo das informações, ou seja, dados que descrevam os dados tendo-se, então, os metadados (ROCHA, 2004, p. 113).

Segundo Rocha (2004, p. 113), no ambiente digital, utiliza-se o termo objeto, que faz referência a todos os recursos descritos pelos metadados, como é o caso dos documentos, pessoas, conjunto de páginas na web, sistemas, aplicações, ou seja, o que possa ser referenciado por meio de URI (*Uniform Resource Identifier*). Desse modo, ao descrever os recursos na web, objetiva-se facilitar a descoberta desses objetos para que possam ser utilizados posteriormente.

Os metadados possibilitam aos desenvolvedores criarem aplicações com melhores mecanismos de interpretações para usuários e por máquina. Contudo, esses recursos devem preservar uma forma legível de interpretação das informações para que haja uma comunicação eficiente (LAUFER, 2015, p. 20). Assim sendo, a marcação de documentos é um mecanismo importante, não só para garantir mecanismos de busca mais eficientes, mas por proporciona uma melhor organização na representação do conhecimento (VITAL; CAFÉ, 2011, p. 125). Dentro dessa perspectiva, algumas linguagens de marcação têm sido utilizadas para fornecer significado ao texto (forma semântica) possibilitando com que os computadores possam manipular os documentos para uma melhor representação (FURGERI, 2006, p. 226).

Acrescenta-se que, os metadados também podem ser observados como dados que descrevem propriedades de recursos para algum propósito, a exemplo do contexto no qual o recurso é inserido e suas qualidades. Nessa visão, exemplifica-se o esquema de metadados MARC, utilizado para indicar as propriedades de um documento, tais como: seu título, assunto, autor, etc., o qual é muito utilizado na Biblioteconomia possibilitando uma identificação e descoberta de documentos de forma eficiente (ROCHA, 2004, p. 114).

Os metadados podem ser apresentados de diferentes tipos e classificados segundo certa estrutura taxonômica. Além da viabilidade de serem inclusos no código HTML possibilitam uma interpretação por máquinas das informações contidas nos documentos (ROCHA, 2004, p. 114; LAUFER, 2015, p. 20). Assim, representando um contexto de significados que proporcionam facilidades na recuperação de documentos por meios de mecanismos de busca e evita ambiguidades das linguagens naturais quanto à recuperação de informações não eficientes (VITAL; CAFÉ, 2011, p. 125).

Gilliland-Swetland (2000, p. 1 apud CAMPOS, 2007, p. 19) define metadados como: “a soma total do que pode ser dito sobre algum objeto informacional em algum nível de agregação”. O autor faz referência à classificação dos metadados digitais de acordo com os seguintes atributos:

- a) Podem ter sua fonte interna (quanto criada pelo gerador informacional), ou externa (gerado após a criação do objeto informacional por outro criado);
- b) Podem ser criados automaticamente por computadores ou humanos;
- c) Podem atender uma natureza leiga ou especializada, a depender se foram criados por especialistas;
- d) Podem ter um estado estático ou dinâmico, nesse caso, sendo alterado com a manipulação;
- e) Podem ser de longo prazo (para acesso e uso) ou de curto prazo (usado para transações);
- f) Podem ter uma semântica controlada, nesse caso, obedecendo a um vocabulário estatizado, ou não controlada (a exemplo uso de *tags* em HTML);
- g) Podem se referir a uma coleção ou a itens individuais.

Acrescenta-se que no ambiente Web os recursos que são disponibilizados, provenientes das diversas áreas do conhecimento, têm recebido especial atenção na descrição de seus esquemas de metadados, nesse sentido, visando alcançar uma padronização com a finalidade de facilitar a utilização dos metadados no ambiente da Web, fez-se uso do padrão Dublin Core (ROCHA, 2001, p. 115).

3.3.1 O padrão Dublin Core

A expansão da Internet na década de 90 e o compartilhamento de documentos demandaram pesquisas em busca de padrões mínimos que pudessem ser utilizados no compartilhamento de documentos no ambiente Web possibilitando sua identificação. Nesse sentido, o padrão de metadados *Dublin Core* (DC) surgiu como uma proposta em um encontro realizado pela *Online Computer Library Center* (OCLC), conjuntamente, com a *National Center for Supercomputing Applications* (NCSA), em 1995, nos Estados Unidos, tendo como objetivo: identificar, organizar e recuperar recursos digitais disponíveis em bibliotecas para uso em ambiente web (ARAKAKI *et al.*, 2015, p. 86).

O padrão *Dublin Core* apresenta-se como um esquema de metadados⁵ simples para disponibilizar documentos digitais na Internet, sendo inicialmente representado por 15 elementos (Título, Autor, Assunto/palavras-chave, Descrição, Editor, Contribuidor, Data,

⁵ Tem como função de descrever, representar, manter a consistência e a persistência do recurso digital no ambiente informacional (SANTAREM SEGUNDO; CONEGLIAM, 2016, p.107).

Tipo do recurso, Formato, Identificador do recurso, Fonte, Idioma, Relação, Cobertura, Direitos autorais) para descrição de arquivos para posterior recuperação no ambiente Web (PEREIRA *et al.*, 2005, p. 15). O conjunto desses 15 elementos simples permitiu uma compreensão semântica dos documentos e puderam ser utilizados por diversas sintaxes de marcação como o HTML, XML e outras linguagens possibilitando, assim, o intercâmbio e reuso de informações (FORMENTON; GRACIOSO, 2022, p. 12).

3.4 REPRESENTAÇÃO DO CONHECIMENTO

De acordo com Vital e Café (2011, p. 116), os sistemas de representação e organização do conhecimento são fundamentais, pois seus processos contribuem na representação, recuperação e intercâmbio de informações, em especial, ao crescente número de dados que são produzidos atualmente.

Santarem Segundo e Coneglian (2016, p. 107) mencionam que na representação do conhecimento diversos temas são importantes para o processo de organização da informação e sua representação, tais como: vocabulário controlado⁶, tesauros⁷, taxonomia⁸ e ontologias⁹. Fato corroborado por Vital e Café (2011, p. 116), que destacam a necessidade de conhecê-los para que haja a recuperação da informação de forma eficiente para os usuários, bem como a produção de sistemas de representação e organização de informação que possam garantir informações precisas.

Nesse contexto, a necessidade de uma compreensão do conteúdo dos documentos Web, que pudessem ser tratados por máquinas e agentes computacionais, fez com que os pesquisadores Tim Berners-Lee, Hendler James e Lassila Ora propusessem uma nova visão para a web: a estrutura da Web Semântica, descrita no artigo “*The Semantic Web*”, publicado pela *Scientific American*, em 2001. Tais ideias vislumbraram uma nova perspectiva de uma Web, onde os documentos representados na forma sintática, agora, passariam para uma representação semântica, por meio de novas estruturas a serem adicionadas para representação da informação (SANTAREM SEGUNDO; CONEGLIAN, 2016, p. 220).

⁶ Conjunto de termos organizados em estrutura relacional, cuja função é padronizar e facilitar o fluxo de dado em um sistema informacional (SANTAREM SEGUNDO; CONEGLIAN, 2016, p.107).

⁷ Uma lista de palavras com significados semelhantes, dentro de um domínio específico de conhecimento (SANTAREM SEGUNDO; CONEGLIAN, 2016, p. 107).

⁸ Ordenamento e rotulação de metadados, que permite organizar de forma sistemática a informação (SANTAREM SEGUNDO; CONEGLIAN, 2016, p. 107).

⁹ Modelo de representação conceitual do conhecimento que é utilizada para representar e recuperar as informações.

4 WEB SEMÂNTICA

Nessa seção são apresentados os conceitos basilares e tecnológicos para possibilitar a construção de uma Web Semântica. Inicia-se com a proposta de uma nova visão para a Web, proposta por Berners-Lee e colaboradores. Posteriormente, são mencionadas as tecnologias para a representação da informação, linguagens de marcação, recursos descritivos, esquemas de restrições. Por último, é apresentado o conceito de ontologia e seu uso na web Semântica, além de ferramentas computacionais para sua construção e as linguagens de consulta para o padrão da Web Semântica.

4.1 A PROPOSTA DA WEB SEMÂNTICA

No contexto atual, a forma como são tratadas as informações na Web, a maior parte delas, é direcionada para a interpretação por seres humanos e não para programas de computadores que possam ter um entendimento significativo das informações. Em virtude disso, o processamento computacional ainda é direcionado para a forma de apresentação do conteúdo ao usuário e a interligação de links entre páginas, portanto, não havendo um processamento confiável sobre informações semânticas (BERNERS-LEE *et al.*, 2001, p. 1).

A proposta em mudar de uma Web direcionada aos dados (atual) para uma Web de significado, ou seja, de uma Web Sintática para uma Web Semântica possibilitará “estrutura ao conteúdo significativo das páginas da Web, criam um ambiente onde os agentes de software que transitam de uma página para outra possam facilmente realizar tarefas sofisticadas para os usuários” (BERNERS-LEE *et al.*, 2001, p. 1). Nesse sentido, os autores apontam que:

A Web Semântica não é uma Web separada, mas uma extensão da atual, na qual a informação recebe um significado bem definido, permitindo que computadores e pessoas trabalhem em cooperação¹⁰ (BERNERS-LEE *et al.*, 2001, p. 1, tradução nossa).

Para que a Web Semântica funcione, os computadores devem ter acesso a coleções estruturadas de informações e conjuntos de regras de inferência que possam usar para conduzir o raciocínio automatizado [...] ¹¹ (BERNERS-LEE *et al.*, 2001, p. 21, tradução nossa).

¹⁰ The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation (BERNERS-LEE *et al.*, 2001, p. 1).

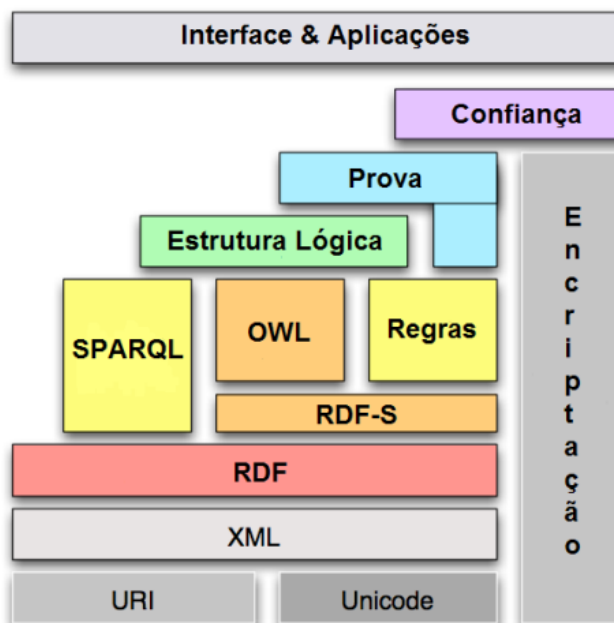
¹¹ For the semantic web to function, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning [...] (BERNERS-LEE *et al.*, 2001, p. 1).

Portanto, de forma sucinta, a concepção da Web Semântica é direcionada para que os dados da Web tenham significado e possam ser interpretados por máquinas por meios de agentes computacionais. Para tanto, faz-se necessário a utilização linguagem de marcação semântica para a construção de páginas Web, bem como modelos que possam representar os metadados, além do uso de ontologias para uma organização do conhecimento (SANTAREM SEGUNDO; CONEGLIAN, 2016, p. 220).

Acrescentam os autores que para possibilitar a construção da Web Semântica são propostos recursos computacionais como: XML (*Extensible Markup Language*), RDF (*Resource Description Framework* - Estrutura de descrição de recursos), RDF Schema (*Resource Description Framework Schema* - Esquema da Estrutura de Descrição de Recursos) e diversos tipos de ontologias (SANTAREM SEGUNDO; CONEGLIAN, 2016, p. 220).

De acordo com a proposta de Berners-Lee, a Web Semântica foi concebida em uma estrutura que agrega diversas camadas denominadas de Diagrama da Web Semântica (Figura 2), onde cada camada desempenha uma função específica. Segundo Santos e Carvalho (2007, p. 5), as camadas inferiores da arquitetura dispõem suas tecnologias para as camadas superiores.

Figura 2 – Diagrama da Web Semântica.



Fonte: Santos e Carvalho (2007, p. 5).

4.2 XML – EXTENSIBLE MARKUP LANGUAGE

O padrão XML tem como característica de ser uma linguagem de marcação mais flexível no processo de codificação, por meios de definição de *tags*, visto que permite aos desenvolvedores criarem suas próprias marcações (ALMEIDA, 2002, p. 6). Nas afirmações de Santarem Segundo e Coneglin (2016, p. 220), essa característica possibilita que seja agregada semântica aos documentos, assim, ficando a cargo das aplicações a interpretação de cada marcação que foi atribuída ao objeto informacional.

O padrão XML, assim como o HTML, foi desenvolvido para ser utilizado no ambiente Web. As diferenças entre o XML e o HTML são as seguintes: O XML possui a flexibilidade para a criação de marcação, assim criando a possibilidade de implementação de novas linguagens, portanto, é uma linguagem de marcação extensiva a outras; A linguagem XML é direcionada para o tratamento da estrutura para representar o significado da informação, ou seja, a semântica. Já a linguagem HTML trata da formação do documento em sua exibição no navegador, ou seja, apenas a forma sintática, Tanto o HTML como o XML são linguagens de marcação de padrões abertos (FURGERI, 2006, p. 229).

Lima e Carvalho (2005, p.1) relatam que a linguagem HTML é facilmente legível por seres humanos, contudo, não é fácil de ser manipulada por computadores, ao contrário, da linguagem XML que atende aos dois requisitos, ou seja, os documentos escritos em XML são de fácil entendimento por seres humanos e de tratamento por máquinas.

Devido à flexibilidade e extensibilidade da linguagem de marcação XML, essa característica permite aos desenvolvedores criarem outras linguagens que enquadram as propriedades do XML, a exemplo do OWL (*Ontology Web Language*), RDF (*Resource Description Framework*) e SPARQL (*SPARQL Protocol and RDF Query Language*). Tais formatos criam modelos semânticos que possibilitam a representação de classes, propriedades e axiomas de um objeto informacional (SOUSA *et al.*, 2018, p. 67). De acordo com os autores, o padrão XML, por seu formato simples, mas de alto nível de qualidade, tem contribuído na implementação de soluções que são direcionadas: Para a Web, banco de dados, aplicações em bibliotecas virtuais, além de garantir a interoperabilidade entre sistemas para a troca de informações.

O *World Wide Web Consortium* (W3C), é uma organização internacional responsável pela implementação de padrões e sua organização com a finalidade de facilitar a criação de

conteúdos que possam ser publicados no ambiente da Internet. Portanto uma de suas responsabilidades é definir as normas para o XML (SOUSA *et al.*, 2018, p. 71). Entre outros padrões em que o W3C tem em sua responsabilidade cita-se o RDF, uma linguagem criada a partir do XML que é utilizada para codificação, troca e reutilização de metadados na Web (FURGERI, 2006, p. 236).

A linhagem XML caracteriza-se com alguns elementos básicos que são apontados pelo World Wide Web Consortium (2015b, tradução nossa) (on-line)

A Extensible Markup Language (XML) é um formato simples baseado em texto para representar informações estruturadas: documentos, dados, configuração, livros, transações, faturas e muito mais¹².

O XML é um dos formatos mais usados para compartilhar informações estruturadas atualmente: entre programas, entre pessoas, entre computadores e pessoas, tanto localmente quanto em redes.¹³

São apontados as seguintes vantagens do XML: i) A marcação XML é muito detalhada; ii) A legibilidade do XML (por ser um formato baseado em texto) e a presença de nomes de elementos e atributos em XML significa que as pessoas que consultam um documento XML podem, muitas vezes, obter uma vantagem inicial na compreensão do formato; iii) Qualquer documento XML pode ser lido e processado por qualquer ferramenta XML.¹⁴

A forma da descrição das *tags* em um documento XML é realizada de maneira estruturada e hierárquica, sendo essa hierarquia semelhante ao conceito de árvore, nos quais os documentos possuem um único elemento raiz que irão derivar os outros elementos, denominados de elementos filhos (LIMA; CARVALHO, 2005. p. 4). Como exemplo, vejamos a representação a seguir:

```
<bibliografia>
  <livro>
    <título> Web Semântica: a internet do futuro</título>
    <autor> Karin Breitman </autor>
  </livro>
</bibliografia>
```

¹² The Extensible Markup Language (XML) is a simple text-based format for representing structured information: documents, data, configuration, books, transactions, invoices, and much more (WORLD WIDE WEB CONSORTIUM, 2015b).

¹³ XML is one of the most widely-used formats for sharing structured information today: between programs, between people, between computers and people, both locally and across networks (WORLD WIDE WEB CONSORTIUM, 2015b).

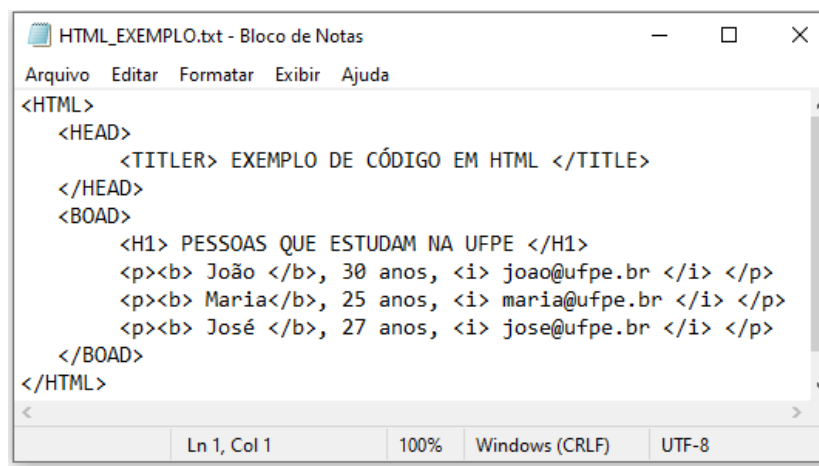
¹⁴ i) XML markup is very verbose; ii) The readability of XML (as it is a text-based format) and the presence of element names and attributes in XML means that people querying an XML document can often get a head start on understanding the format; iii) Any XML document can be read and processed by any XML tool (WORLD WIDE WEB CONSORTIUM, 2015b).

A forma simples de criação de *tags* em XML é exibida em um exemplo do World Wide Web Consortium (2015b) (on-line) para a representação de uma peça de automóvel (para-brisa) e suas características como: número da peça, nome, descrição da peça e item que contem a peça. Observa-se que cada *tag* (marcação) inicia-se com **<nome da tag>** e fecha com uma barra “/” acrescentada antes do nome da *tag* respectiva **</nome da tag>**.

```
<part number="1976">
  <name>Limpador de para-brisas</name>
  <description> O limpador de para-brisa remove automaticamente a
    chuva do seu para-brisa, se não limpar direito. Tem uma <ref
    part="1977"> lâmina </ref> de borracha que pode ser
    encomendado separadamente se você precisar substituí-la.
  </description>
</ part >
```

Almeida (2002, p. 10) descreve exemplo no qual relaciona o código escrito em HTML (figura 3) e seu resultado exposto em um browser apresentado na figura 4. Posteriormente, o mesmo exemplo é transcrito em XML para uma melhor representação do conteúdo, nesse caso, apresentando semânticas nas informações codificadas, assim, possibilitando a interpretação por aplicações e sistemas de busca semântica.

Figura 3 – Representação de informação em código HTML de uma página Web.



```
HTML_EXEMPLO.txt - Bloco de Notas
Arquivo  Editar  Formatar  Exibir  Ajuda
<HTML>
  <HEAD>
    <TITLE> EXEMPLO DE CÓDIGO EM HTML </TITLE>
  </HEAD>
  <BOAD>
    <H1> PESSOAS QUE ESTUDAM NA UFPE </H1>
    <p><b> João </b>, 30 anos, <i> joao@ufpe.br </i> </p>
    <p><b> Maria</b>, 25 anos, <i> maria@ufpe.br </i> </p>
    <p><b> José </b>, 27 anos, <i> jose@ufpe.br </i> </p>
  </BOAD>
</HTML>
```

Fonte: Adaptações de Almeida (2002, p. 11).

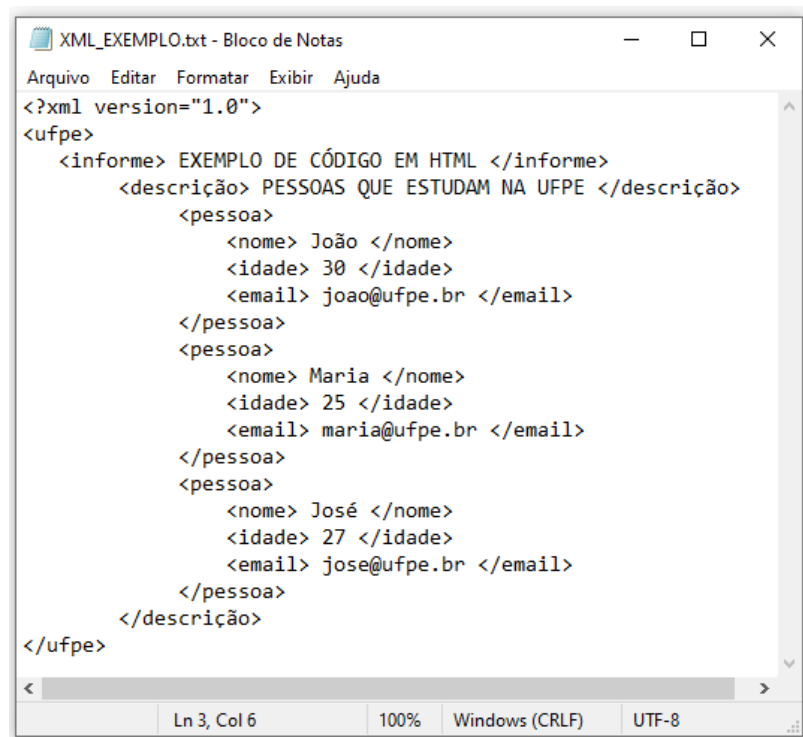
Figura 4 – Resultado do código HTML apresentado em um navegador.



Fonte: Adaptações de Almeida (2002, p. 11).

Na figura 5, representa-se um fragmento do exemplo na linguagem XML (ALMEIDA, 2002, p. 10). Acrescenta-se que o XML dispõe de outras características diferentes do HTML para representar as informações tais como: novas formas de marcação que podem ser definidas, as estruturas de dados podem ser agrupadas, um documento XML pode ter outro documento que define as descrições das marcações.

Figura 5 – Transcrição de código HTML para linguagem XML.



Fonte: Adaptações de Almeida (2002, p. 11).

Nesse contexto, Lima e Carvalho (2005, p. 2) ressaltam que os recursos para definição da gramática utilizada nos arquivos em XML deverão especificar os elementos, atributos e

regras que sejam bem definidas para a composição desses documentos. Para tanto, faz-se uso de documento e esquema que irão impor tais regras, sendo eles: DTD (*Document Type Definitions*) e XML schema.

4.3 DTD – DOCUMENT TYPE DEFINITIONS

Trata-se de um arquivo no qual dispõem os vocabulários descritivos das marcações utilizadas em um documento em XML indicando os tipos de marcações interna ao documento, como também, externamente por meio de uma URL, o que tende a facilitar o intercambio de dados entre aplicações. Como exemplo podemos mencionar um fragmento do arquivo DTD para o elemento < Pessoa > (ALMEIDA, 2002, p. 12). As declarações que são colocadas em um documento DTD irão especificar quais serão os elementos e atributos que farão parte de um documento em XML e a forma como serão os relacionamentos (LIMA; CARVALHO, 2005, p. 9).

```
<!DOCTYPE descrição [
    <!ELEMENT descrição (Pessoa*)>
    <!ELEMENT descrição (Pessoa (nome, idade, email)>
    <!ELEMENT nome (#PCDATA)>
    <!ELEMENT idade (#PCDATA)>
    <!ELEMENT email (#PCDATA)>
]>
```

Nesse sentido, o DTD apresenta-se como uma forma de garantir a integridade sintática e semântica, onde especifica a estrutura dos termos utilizados, os atributos de dados em que um documento XML utilizará. Dessa feita, os documentos DTD fornecem uma condição básica de validação de elementos em um documento XML (FEITOSA, 2006, p. 93).

Por outro lado, conforme mencionado por Feitosa (2006, p. 94), a utilização do DTD para a descrição dos elementos XML possui limitação em virtude de não apresentar a possibilidade de um maior refinamento sobre os formatos, os tipos dos elementos e seus atributos, pois o DTD restringe-se: apenas a 10 tipos de dados para a representação e uma sintaxe própria, diferentemente, da flexibilidade da linguagem XML. Além disso, um documento DTD não possibilita ser manipulado da mesma forma que um documento em XML, pois um documento DTD faz referência à estrutura do documento XML e não aos seus conteúdos (LIMA; CARVALHO, 2005, p.10). Assim sendo, a linguagem XML necessitou de outra forma de representação que pudesse suprir suas necessidades, o que foi implementado

com a utilização do XML schema (FEITOSA, 2006, p. 94). Conceitualmente, um esquema é definido como: “Uma descrição formal que compreende um documento válido, isto é, um XML schema é um documento contendo uma especificação formal do que é um documento XML válido” (FEITOSA, 2006, p. 96).

4.4 XML SCHEMA

É uma forma para a validação de um documento em XML, sendo representado por um documento também em XML (LIMA; CARVALHO, 2005, p.10). O World Wide Web Consortium (2015b, tradução nossa) (on-line) define XML schema como:

Uma linguagem que expressa as restrições de um documento produzido em linguagem XML, e fornece uma lista de elementos e atributos em um vocabulário; associa os tipos, como: integer, string, entre outros com valores encontrados no documento XML.¹⁵

Feitosa (2006, p. 94) enfatiza que diferentemente do DTD, o XML schema pode suportar 44 tipos ou mais de elementos, inclusive, tendo a usabilidade da criação de tipos personalizados. Essa extensibilidade permite declarar tipos simples, tipos de múltiplos valores ou complexos. Dessa forma, segundo o autor, um tipo ao ser definido é caracterizado de duas partes: um nome e um conjunto de valores atribuídos. Exemplifica-se com um elemento válido em um documento XML para o elemento “autor”, onde seria definido da seguinte maneira no XML schema:

[...] <xsd:element names = “autor” types = “xsd:string”>, assim, no arquivo em XML poder-se-ia atribuir o elemento <autor> Rui Barbosa </autor>, visto que, foi definido no XML schema o elemento “autor” com o tipo de valor aceito *string* (uma cadeia de caracteres) (FEITOSA, 2006, p. 95).

Uma característica do XML schema é a possibilidade de adotar *namespaces* por meio da seguinte sintaxe: <xsd:identificador>, onde xsd é utilizado como um prefixo em lugar de uma definição em uma URI, no qual aponta a localização de um vocabulário que é aceito no XML schema.

A parte *identificador* faz referência a um atributo identificador do metadado. Como exemplo, cita-se: <xsd:schema xmlns:xsd = http://www.w3.org/2001/xmlschema>, o prefixo *xsd* é uma abreviação de *XML Schema Definition*, mas poderia ser qualquer declaração, o que importa é o endereço da URI que aponta onde encontra-se o vocabulário válido para ser usado no documento XML (FEITOSA, 2006, p. 96).

¹⁵ An XML Schema is a language for expressing constraints about XML documents, provide a list of elements and attributes in a vocabular, associate types, such as integer, string, etc. with values found in documents (WORLD WIDE WEB CONSORTIUM, 2015b).

Como forma de esclarecimento aponta-se o código em XML schema (exemplo1) com uso do prefixo, o termo “xs” indicando a URL de localização do vocabulário, no qual descreve uma sequência de elementos (*to*, *from*, *heading* e *body*) como tipo de dado “strings”, que fazem parte do elemento complexos (*note*) que serão apresentados, posteriormente, em um documento XML (exemplo 2)

Exemplo1:

```
<?xml version="1.0"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="note">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="to" type="xs:string"/>
        <xs:element name="from" type="xs:string"/>
        <xs:element name="heading" type="xs:string"/>
        <xs:element name="body" type="xs:string"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Fonte: https://www.w3schools.com/xml/schema_intro.asp.

No código em XML exposto em seguida é referente ao código do XML schema apresentado no exemplo 1, onde foram definidos os novos tipos de marcadores (elementos): *to*, *from*, *Heading*, *body*.

Exemplo 2:

```
<?xml version="1.0" encoding="UTF-8"?>
<note>
  <to>Tove</to>
  <from>Jani</from>
  <heading>Reminder</heading>
  <body>Don't forget me this weekend!</body>
</note>
```

Fonte: <https://www.w3schools.com/xml/>

Diferentemente das marcações da linguagem HTML, em que as *tags* utilizadas indicam a formatação do documento a ser apresentado no navegador, no XML as marcações fazem referência ao conteúdo do documento, assim, necessitando de uma forma para indicar como o documento será apresentado ao ser visualizado no navegador ou impresso. Para tanto, faz-se

uso de estruturas denominadas de folhas de estilo que poderão ser de dois tipos: *Cascading Style Sheets - CSS* e a *Extensible Stylesheet Language – XSL* (LIMA e CARVALHO, 2005, p.14).

Os autores descrevem que a CSS, originalmente, foi criada para atender as necessidades da linguagem HTML, posteriormente, também passou a ser utilizada junto com XML para a representação do estilo das páginas referentes à forma de sua apresentação (tipo de fonte, espaçamento, margem, entre outros). Enquanto que o XSL é composto por um conjunto de três tipos de linguagem recomendadas pela W3C para transformação e apresentação dos documentos escritos em XML, são elas:

XSLT (*XSL Transformations*): linguagem para descrever como transformar um documento XML em outro documento;
 XPath (*XML Path Language*): linguagem usada pelo XSLT para acessar ou referenciar partes específicas de um documento XML;
 XSL-FO (*XSL Formatting Objects*): um vocabulário XML para especificação da formatação da semântica (LIMA; CARVALHO, 2005, p.14).

Aponta-se que os padrões DTD e XML schema são utilizados para representarem a gramática e o vocabulário no intuito de descrever as marcações na linguagem XML. Por outro lado, há diferenças entre eles: o DTD é um conjunto de marcações para definição da estrutura de um documento XML, enquanto o XML schema pode ser utilizado para definir a estrutura de marcação de um documento extensível em XML, bem como seu conteúdo e restrições.

Dentro dessa perspectiva, no processo de construção da Web Semântica utilizaram-se os padrões DTD e o XML schema devido à combinação de suas características e flexibilidade. Todavia, tais padrões não foram suficientes para oferecer estruturas de descrição dos documentos na web como um todo (LIMA, 2003 apud FEITOSA, 2006, p. 99). Fato mencionado por Ferreira e Costa (2013, p.13), visto que a Web é composta por vários dados heterogêneos em formatos diversos e de diferentes fontes de informação, além do que, para que haja troca de informações, fazem-se necessário o estabelecimento de padrões para compartilhamento das informações e recursos tanto para os seres humanos quanto para as máquinas.

Portanto, com a finalidade de buscar meios que pudessem fornecer estruturas nas informações para esse ambiente heterogêneo e, conseqüentemente, fosse possível descrever o conteúdo das páginas, recursos e entidades e outros que, normalmente, não são representadas no ambiente digital (a exemplo pessoas) e que dessa forma pudessem ser acessados no ambiente Web, deu-se à base motriz para se fundamentar os princípios da Web Semântica (FERREIRA; COSTA, 2013, p. 14).

Com base nesses princípios, surge o padrão RDF (*Resource Description Framework*) para estruturar os diversos tipos de informações no ambiente Web e, conseqüentemente, possibilitar a interoperabilidade entre aplicações (SOUSA *et al.*, 2018, p. 72).

4.5 RDF – RESOURCE DESCRIPTION FRAMEWORK

O desenvolvimento do padrão RDF se deu para situações em que as informações possam ser processadas por aplicações, ao invés de apenas ser visualizadas pelos usuários. Portanto, trata-se de uma arquitetura genérica de metadados, que permite apresentar informações sobre recursos na Web sendo uma tecnologia importante para a Web Semântica (LIMA; CARVALHO, 2005, p.14).

Acrescenta-se que o RDF pode usufruir da sintaxe da linguagem XML para o intercâmbio e processamento dos metadados, além do que, igualmente à linguagem XML que possui o XML schema para definir seu vocabulário, o RDF possui o RDF schema possibilitando a publicação de vocabulários que possam ser legíveis por pessoas e por máquinas (FERREIRA; COSTA, 2013, p, 15). Apresenta-se como definição de vocabulário, no cenário da Web Semântica, segundo o World Wide Web Consortium (2015a, tradução nossa) (on-line):

[...] os vocabulários definem os conceitos e relacionamentos (chamados de “termos”), os quais são utilizados para descrever e representar uma área de interesse. Os vocabulários são usados para classificar os termos que podem ser empregados em uma determinada aplicação, podendo, assim, caracterizar possíveis relacionamentos e definir possíveis restrições ao uso desses termos [...].¹⁶

As palavras vocabulário e ontologia, no enquadramento da Web Semântica, não são tão claras. De acordo com o World Wide Web Consortium (2015a) (on-line), normalmente, utiliza-se a palavra ontologia para descrever coleções de termos mais complexos e, possivelmente, bastante formais. Já a palavra vocabulário é utilizada quando não há certo formalismo estrito. O World Wide Web Consortium (2015) (on-line) ainda menciona que: “os vocabulários são usados como blocos de construções básicas para as técnicas de inferência utilizadas na Web Semântica, onde ajuda na integração dos dados sem que haja ambiguidade de termos usados por diferentes conjuntos de dados”.

¹⁶ On the Semantic Web, vocabularies define the concepts and relationships (also referred to as “terms”) used to describe and represent an area of concern. Vocabularies are used to classify the terms that can be used in a particular application, characterize possible relationships, and define possible constraints on using those terms. (WORLD WIDE WEB CONSORTIUM, 2015a).

Além disso, o RDF foi concebido para fornecer interoperabilidade e proporcionar semântica para os metadados objetivando facilitar a busca por recursos disponíveis na Web, que o torna diferente do modelo em que os recursos eram obtidos por meio de mecanismos de busca textual simples (BREITMAN, 2005, p. 50).

Outro ensinamento de Laufer (2005, p. 34), aponta sobre as aplicações diante dos modelos tradicionais, visto que:

[...] as aplicações encontram dificuldade para processar os dados que estão expostos nos documentos, sem que seja colocada uma informação específica e codificada em um formato adequado. A solução proposta é usar metadados para descrever os dados publicados na web, que ajudem a localizar e a processar informações, fornecendo descrições sobre a estrutura, o conteúdo ou outras informações relacionadas.

Isso tudo vem a corroborar com Feitosa (2006, p. 101), quando explica que o RDF foi desenvolvido para controlar a semântica, a sintaxe e a estrutura de metadados de um determinado domínio, além da possibilidade de ser mesclado com outros domínios mantendo a clareza e legibilidade originais. Para tanto, o RDF é baseado na identificação de objetos por meios de identificadores na Web denominados de URI (*Uniform Resource Identifier*), além da descrição de recursos, de suas propriedades e valor (LIMA; CARVALHO, 2005, p. 24).

Como referenda Breitman (2005, p. 50), o RDF possui três princípios: recursos, propriedades e frases. A autora explica que um recurso pode se referir a um objeto ou uma coisa que se deseja mencionar algo, mas sendo identificados por meios de URIs. Já as propriedades têm a função de descrever os relacionamentos entre os diversos recursos, a exemplos: “escrito por”, “cargo”, “escolaridade”, “idade”. Some-se a isso que da mesma forma que os recursos, as propriedades, também podem ser identificadas por URIs, assim, cria-se um modo de identificação global e única. Por fim, as frases são uma sentença ou literal para a formação da tripla em RDF, que será composta por sujeito (recursos), predicado (propriedade) e o objeto.

Laufer (2005, p. 37) ressalta que as URIs são a forma mais abrangente das URL (*Uniform Resource Locator*). Essas são usadas pelos navegadores para identificação de um endereço e recuperação das informações, enquanto aquelas não estão ligadas a localização de recursos da web, mas com o intuito de localizar as coisas, ou seja, o que faz parte da web ou que não podem ser recuperados diretamente nesse ambiente, a exemplo dos indivíduos. Outro ensinamento de Laufer (2005, p. 34) defende que os recursos caracterizam-se por quaisquer

coisas, assim, podendo ser concretos ou abstratos, como é o caso de uma pessoa, uma página na web, empresa, sentimento, cor, entre outras.

4.5.1 Modelo de representação dos dados em rdf

A sintaxe em RDF está baseada na estrutura da linguagem XML sendo processada por máquinas e utiliza como base as *URIs* para ligar as informações na Web (LIMA; CARVALHO, 2005. p.25).

Para tanto, a estrutura declarativa em RDF consiste de três elementos, ou seja, uma tripla composta por: *sujeito, predicado e objeto*, os quais interagem para expressar uma afirmação RDF entre recursos. Tal interação dá-se da seguinte forma: “*O sujeito e o objeto representam os dois recursos sendo relacionados, enquanto, o predicado representa a natureza desta relação, que é formulada de modo direcional (do sujeito para o objeto) e é chamada em RDF de propriedade*” (LAUFER, 2005, p. 34). As relações entre *sujeito predicado e objeto* tomaram como base as propriedades linguísticas utilizadas na formação de uma oração, dessa forma, foram usadas para uma representação semântica entre os conjuntos declarativos a serem interpretados por máquinas (SOUSA *et al.*, 2018, p. 72).

No dizer de Feitosa (2006, p. 102), um conjunto de triplas é representado por um grafo RDF. Cada tripla é uma declaração nos quais, também, podem ser denominadas pelos seguintes termos: *objetos, propriedades e valores* e que irão compor a informação como uma unidade de conhecimento.

Para representar a estrutura de tripla RDF, Laufer (2005, p. 36) demonstra grafos dirigidos por meio de seta (propriedade) formando uma relação entre o sujeito e valor, a exemplo informações de livros: onde cada livro é constituído por recurso, conforme apresentados nas figuras 6, 7, 8 e 9.

De acordo com o World Wide Web Consortium (2014, tradução nossa) (on-line), em suas recomendações sobre a sintaxe XML do RDF 1.1 “O gráfico RDF tem nós e rotulado direcionado por arcos que ligam pares de nós e isso é representado como um conjunto de triplas RDF, onde cada tripla contém um nó de assunto, predicado e nó de objeto”.¹⁷

¹⁷ The RDF graph has nodes and labeled directed arcs that link pairs of nodes and this is represented as a set of RDF triples where each triple contains a subject node, predicate and object node (WORLD WIDE WEB CONSORTIUM, 2014).

Figura 6 – Grafo de tripla de um livro descrevendo seu título.



Fonte: Laufer (2005, p. 36).

Representação gráfica de uma tripla RDF que poderá ser codificada em XML e ser referenciada por uma URI.

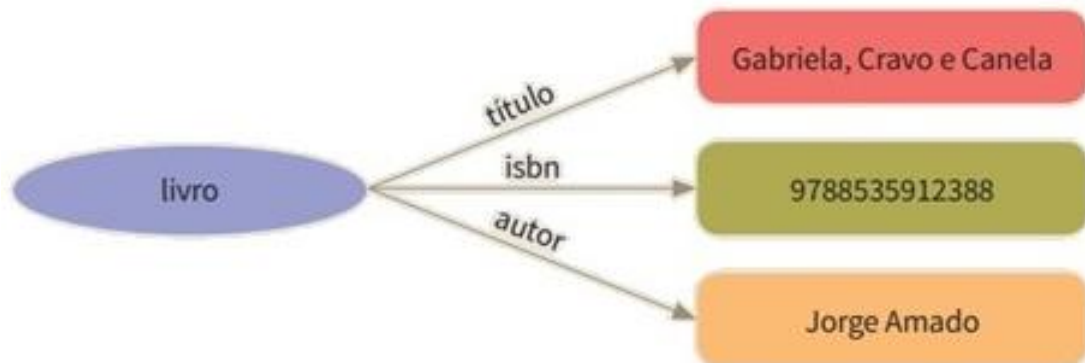
Figura 7 – Grafo genérico de uma tripla (Sujeito, predicado e valor).



Fonte: Laufer (2005, p. 36).

Apresentam-se em seguida três triplas com um único sujeito (livro) e diferentes predicados (título, isbn e autor) ligados a diferentes objetos (Gabriela, Cravo e Canela; 9788535912388; Jorge Amado).

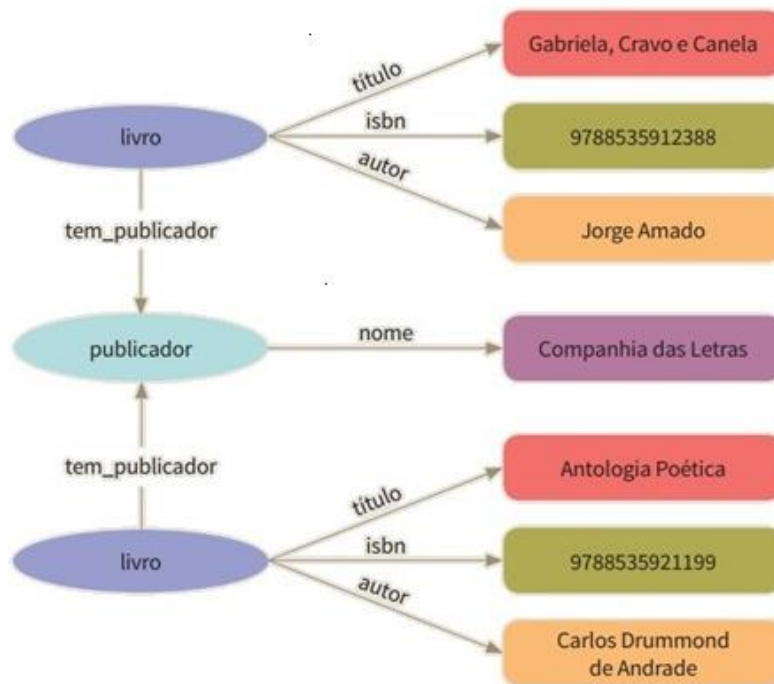
Figura 8 – Grafo RDF com três triplas de um mesmo recurso com propriedade diferente.



Fonte: Laufer (2005, p. 37).

As relações contidas em uma tripla também podem ser expandidas em outras relações com outro conjunto de triplas, ou seja, com outros recursos, dessa forma, criando uma rede de informações, conforme exemplificado no grafo em seguida que esquematiza dois livros de um único publicador (editora) (LAUFER, 2005, p. 37).

Figura 9 – Grafo RDF composto de relações entre as triplas com outros recursos referentes a um mesmo publicador para dois livros.

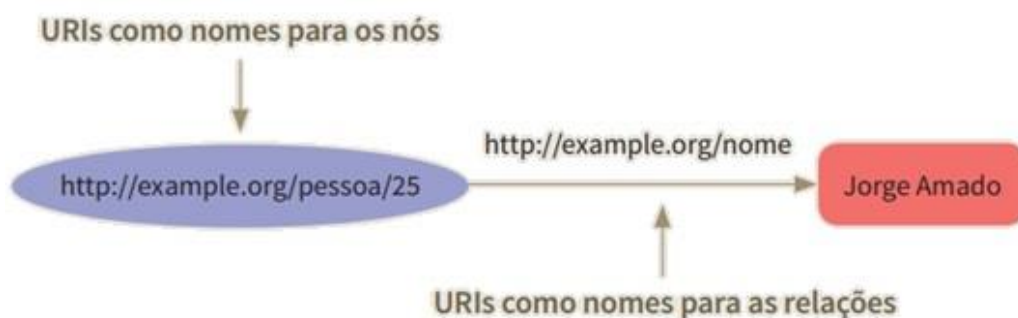


Fonte: Laufer (2005, p. 38).

Outra forma de interligação dos recursos dá-se por meios dos URIs, no qual identificam os recursos de forma única e universal. Assim sendo, um identificador que pode ser utilizado para referenciar, também, coisas que não fazem parte da web, como exemplo URIs de pessoas (figura 10) (LAUFER, 2005, p. 39).

Sousa *et al.* (2018, p. 72) esclarecem que o RDF fornece mecanismos para uma construção semântica na web, por meios de interligação de dados (recursos), assim esses poderão ser recuperados, posteriormente, por um Sistema de Recuperação de Informação (SRI).

Figura 10 – Representação de URI interligando um endereço a uma pessoa no formato RDF.



Fonte: Laufer (2005, p. 39).

Sousa *et al.* (2018, p. 72) ainda acrescentam que, em busca de uma semântica que possa ser amplamente utilizada, o RDF pode fazer uso de vocabulários conhecidos de um determinado domínio, a exemplo o FOAF (*friend of a friend* – *amigo de um amigo*) para uso na descrição de pessoas por meio da URI (<http://xmlns.com/foaf/0.1/name>) que nesse exemplo faz referência a propriedade nome de uma pessoa Figura 11.

Figura 11 – Utilização de vocabulário FOAF/name para descrever pessoas com uso de URI.



Fonte: Laufer (2005, p. 39).

Um grafo pode ser entendido como um conjunto de caminho composto por nó, arco e predicado. Contudo, para serem processadas por máquinas, essas informações deverão ser representadas com os termos em XML (WORLD WIDE WEB CONSORTIUM, 2014) (on-line).

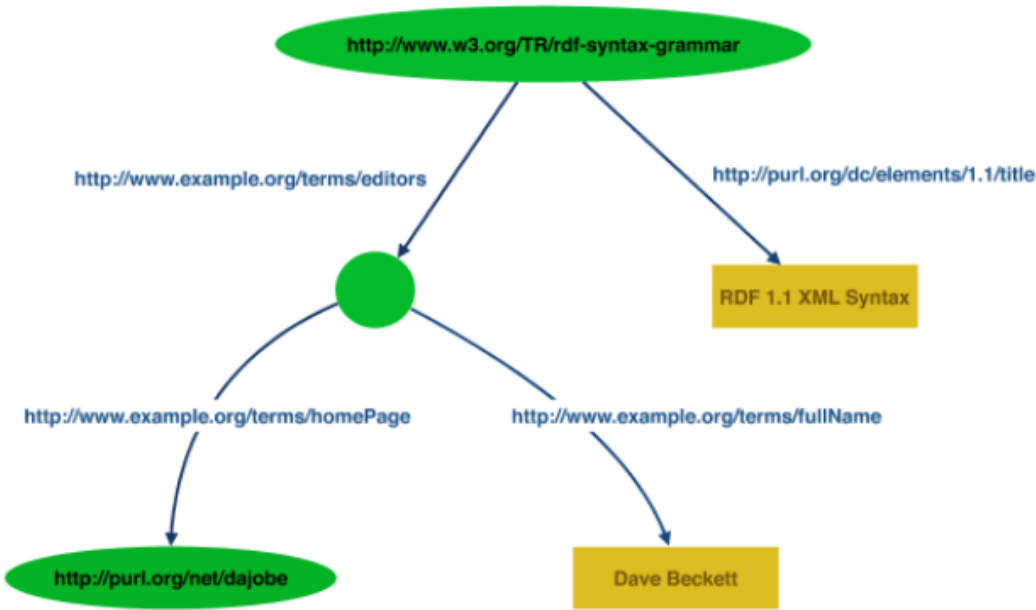
De certo que a estrutura apresentada no formato de grafos facilita o entendimento para o ser humano, mas não são apropriadas para computadores. Nesse sentido, deve ser transformado em um conjunto de cadeias de caracteres legíveis, o processo de transformação é denominado de serialização (FERREIRA; COSTA, 2013, p. 19). Existem algumas sintaxes que são utilizadas para esse processo como: Notation3 (N3), N-Triples, Turtle, JSON e RDF/XML (LAUFER, 2005, p. 41; FERREIRA; COSTA, 2013, p. 20). De acordo com o World Wide Web Consortium (2014, tradução nossa) (on-line), em suas recomendações sobre a sintaxe XML do RDF 1.1, o processo de serialização é definido como:

[...] uma sequência de elementos dentro de elementos que alternam entre elementos para nós e arcos de predicados. O nó no início da sequência se transforma no elemento mais externo, o próximo arco de predicado se transforma em um elemento filho e assim por diante. As faixas geralmente começam na parte superior de um documento RDF/XML e sempre começam com nó.¹⁸

O World Wide Web Consortium (2014) (on-line) exemplifica a transformação de um gráfico RDF (figura 12) em uma representação de termos em XML do RDF 1.1, de acordo com as informações apresentadas no quadro 1.

¹⁸ [...] the sequence turns into the outermost element, the next predicate arc turns into a child element, and so on. The stripes generally start at the top of an RDF/XML document and always begin with nodes (WORLD WIDE WEB CONSORTIUM, 2014).

Figura 12– Gráfico em RDF composto com 5 nós e URIs.



Fonte: World Wide Web Consortium (2014).

Quadro 1. Representação das informações de grafo RDF.

Lado esquerdo da figura 11	Lado direito da figura 11
Nó com URI: <code>http://www.w3.org/TR/rdf-syntax-grammar</code> ;	Arco Predicado marcado com URI: <code>http://purl.org/dc/elements/1.1/title</code> ;
Arco Predicado marcado com URI: <code>http://example.org/terms/editor</code> ;	Nó com <i>string</i> RDF 1.1 XML Syntax;
Nó sem URI;	
Arco Predicado marcado com URI: <code>http://example.org/terms/homePage</code> ;	
Nó com IRU: <code>http://purl.org/net/dajobe</code> ;	
Arco Predicado marcado com URI: <code>http://example.org/terms/fullname</code> ;	
Nó com <i>string</i> David Backer.	

Fonte: Adaptação de World Wide Web Consortium (2014).

Descrição completa do caminho do gráfico em RDF/XML (figura 11) apresentado em World Wide Web Consortium (2014, adaptação nossa) (on-line)

```
<rdf: Description rdf:about="http://www.w3.org/TR/rdf-syntax-grammar">
  <ex:editor>
    <rdf: Description >
```

```

        <ex:homepage>
            <rdf:Description rdf:about="http://purl.org/net/dajobe/">
                </rdf:Description>
            </ex:homepage>
        </rdf: Description >
    </ex:editor>
</rdf: Description >

<rdf: Description rdf:about="http://www.w3.org/TR/rdf-syntax-grammar">
    <ex:editor>
        <rdf: Description >
            <ex:fullName>Dave Beckett</ex:fullName>
        </rdf:Description>
    </ex:editor>
</rdf: Description >

<rdf: Description rdf:about="http://www.w3.org/TR/rdf-syntax-grammar">
    <dc:title>Sintaxe XML RDF 1.1</dc:title>
</rdf:Description>

```

Outro exemplo de serialização em XML é descrito por Ferreira e Costa (2013, p, 20), referente ao relacionamento entre livro, título do livro e autor.

```

<?xml version="1.0" encoding="utf-8"?>
<rdf: RDF xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
    xmlns:cod="http://padraoemetadados-y.org/uri#">

<rdf:description rdf?about=http://biblioteca.x.org/uri#livro-X>
    <cod:titulo>Cibercultura</cod:titulo>
    <cod:possuiAutor>
        <rdf:Description rdf:about=" http://biblioteca.x.org/uri#livro-Z">
            <cod:nomeDpAutor> Lévy, Pierr</cod:nomeDoAutor>
        </rdf:Description>
    </cod:possuiAutor>
</rdf:Description>

</rdf:RDF>

```

Uma das vantagens na utilização da sintaxe XML é a de possibilitar o intercâmbio e o processamento de metadados e, em conjunto com o RDF, torna-se uma ação positiva, visto que vários sistemas de informação poderão interagir para construção de sistemas de busca mais eficientes e, conseqüentemente, fornecerão aos usuários serviços mais especializados e criação de modelos que possam ser processáveis por máquinas (FERREIRA; COSTA, 2013, p. 21).

Apesar da flexibilidade e da simplicidade na representação dos metadados, o RDF não oferece recursos necessários para uma linguagem de ontologia, então, para suprir essa lacuna,

foi desenvolvido o RDF schema (BREITMAN, 2005, p. 50). Em outras palavras, o modelo RDF não faz suposição a respeito da semântica dos recursos identificados pelos URIs, fazendo-se necessário fornecê-lo por meio de vocabulários (LAUFER, 2015, p. 46).

4.6 RDFS – RESOURCE DESCRIPTION FRAMEWORK SCHEMA

O RDF schema é uma linguagem descritiva de vocabulários tendo como objetivo especificar propriedades e classes, além da possibilidade de hierarquização dessas com a finalidade de organização dos recursos RDF (BREITMAN, 2005, p. 50). Conforme afirma Feitosa (2006, p. 108), o RDF schema pode descrever propriedades e relações de um dado recurso, com as propriedades de outros recursos. O autor ressalta que: *“O RDF schema foi concebido para fornecer um rigor sintático e semântico a um documento e funciona como uma extensão do RDF, fornecendo a arquitetura para a descrição de classes e propriedade específica para uma aplicação”*.

Portanto, a utilização de classes no RDF schema tem papel importante, pois é por meio delas que são especificadas categorias que poderão ser utilizadas na classificação de recursos (LAUFER, 2015, p.46). Dentro dessa linha Breitman (2005, p. 52), aponta que as classes definidas no RDF schema possibilitam que os recursos que estão descritos dentro do documento RDF possam ser definidos como instâncias ou subclasses das classes que estão presentes no RDF schema. Para tal, é necessário criar a relação da instância com sua classe, onde é definida por meio da propriedade “*type*” de RDF (LAUFER, 2015, p. 46).

Uma das possibilidades do uso na Web Semântica é por meio da utilização de inferência. O termo inferência segundo o dicionário online de português¹⁹ pode ser definido como:

Processo intelectual segundo o qual é possível chegar a uma conclusão a partir de premissas, proposições, pressupostos: não há provas do assassinato, somente inferências;

Raciocínio através do qual uma proposição é considerada verdadeira pela sua ligação com outras já tidas como verdadeiras; a proposição que se assume como sendo verdadeira.

No campo da Web Semântica a inferência é utilizada para fazer declarações menos extensas a respeito de um conjunto de triplas, como exemplo: ao se definir uma tripla sobre pessoas, não há necessidade de definir uma tripla sobre um parente como exemplo: um tio, também, tido como pessoa, pois será feito uma inferência. Destarte, na inferência há uma

¹⁹ www.dicio.com/inferencia

dedução de triplas não declaradas de forma explícita, todavia, o processo de inferência no RDF schema é de forma limitada (LAUFER, 2015, p. 47).

Outra limitação apontada por Breitman (2005, p. 53) é a de que o RDFs possui primitivas simples para o processo de modelagem de ontologias tendo pouca expressividade na construção de modelos ontológicos de terminado domínio, haja vista, a impossibilidade de utilização de conectivos como: negação, disjunção, lógicos, entre outros que não existem no RDF do qual o RDF schema é derivado. Então, para ser suprida essa necessidade na Web Semântica, uma nova camada é adicionada ao modelo: a ontológica.

Diversas linguagens para a construção de ontologias foram implementadas, cada uma com suas características na representação da informação, sendo elas: SHOE (*Simple HTML Ontology Extension*), OIL (*Ontology Inference Layer*), DAML (*DARPA, Agent Markup Language*), DAML+OIL e OWL (*Ontologic Web Language*), que é uma revisão da DAML+OIL. A OWL é uma recomendação do W3C para atender as necessidades da Web Semântica (BREITMAN, 2005, p. 59).

4.7 ONTOLOGIA

A utilização do termo ontologia é de época remota, sendo utilizada por Aristóteles e relacionada aos estudos da existência. Foi no campo da Ciência da Computação, na área de inteligência artificial, que novas ideias foram incorporadas ao termo no sentido de reutilização de componentes como forma de facilitar a construção de sistemas de conhecimento (BRANDÃO; LUCENA, 2002, p. 3).

Uma definição clássica foi dada por Gluber (1993, apud BRANDÃO; LUCENA, 2002, p. 4, tradução nossa), sendo definida a ontologia como: “Uma ontologia é uma especificação explícita de uma conceituação”²⁰. Posteriormente, é redefinida por Borst (1997 apud BRANDÃO; LUCENA, 2002, p. 4, tradução nossa), sendo “Ontologias são definidas como uma especificação formal de uma conceituação compartilhada”²¹

Em Gluber (1993), tem-se a ideia de conceitualização, fazendo menção aos conceitos, objetos e entidades que existem em uma determinada área de interesse a ser representada e seus possíveis relacionamentos, enquanto que na concepção delineada por Borst, é

²⁰ “An ontology is an explicit specification of a conceptualization” (GLUBER, 1993 apud BRANDÃO; LUCENA, 2002, p. 4).

²¹ “Ontologies are defined as a formal specification of a shared conceptualization” (BORST, 1997 apud BRANDÃO; LUCENA, 2002, p. 4).

demonstrada a importância do compartilhamento da informação como base nas premissas de conceitualização e sedimentada em uma formalidade para a representação das informações na web e sistemas de softwares (BRANDÃO; LUCENA, 2002, p. 4).

Na visão de Breitman (2005, p.7), a ontologia no campo da Web Semântica é definida como: “especificações formais e explícitas de conceitualização compartilhadas”, ou seja, são modelos que têm como função explicar o vocabulário utilizado nas aplicações semânticas, assim, servindo como base para que haja uma comunicação livre de ambiguidades.

Nas explicações de Breitman o termo “*conceitualização*” é tido como um modelo abstrato que representa algum fenômeno que identifica os conceitos relevantes. Entendem-se como “*explícitas*”, os elementos e suas restrições que são definidas de forma clara na ontologia, enquanto, o termo “*formal*” atribui-se que as ontologias devem possibilitar um processamento automático e compartilhado dentro de um conhecimento consensual os quais são aceitos por pessoas sobre um determinado domínio (BREITMAN, 2005, p. 30).

Em Laufer (2015, p, 49), ontologia é definida como “[...] uma especificação de um conceito dentro de um determinado domínio de interesse”. Enquanto em Santarem Segundo (2015, p. 226 apud SANTAREM SEGUNDO; CONEGLIAN, 2016, p. 221) declara que a ontologia é um dos principais elementos para a construção da Web Semântica para utilização de informações em virtude de proporcionar:

[...] uma das maneiras de se construir uma relação organizada entre termos dentro de um domínio, favorecendo a possibilidade de contextualizar os dados, tornando mais eficiente e facilitando o processo de interpretação dos dados pelas ferramentas de recuperação da informação.

Para o uso como tecnologia da Web Semântica, entendem-se as ontologias como: artefatos computacionais que descrevem um domínio do conhecimento e forma estruturada, através de classes, propriedades, relações, restrições, axiomas e instâncias.

Nesse aspecto, apreende-se por classes a representação de um conjunto ou coleções que fazem referências a pessoas, objetos e coisas nas quais compartilham de mesmo grupo de características, mas que possam diferenciá-las dos demais grupos. Nesse sentido, elas irão representar um tipo de domínio específico que servirá para ser usada em uma taxonomia (BREITMAN, 2005, p. 61). As propriedades são utilizadas no sentido de descrever fatos referentes a membros (de uma forma geral ou individual) que façam parte de determinada classe. Acrescenta-se que as propriedades são utilizadas, também, para criar certas restrições a membros de uma determinada classe (BREITMAN, 2005, p. 63 - 64).

No caso das instâncias, elas representam as classes, ou seja, são os objetos do mundo; é aquilo que pertence a determinada classe e que fazem relação a outros indivíduos e outras classes por meios das propriedades. Portanto, o indivíduo é definido como os membros das classes (BREITMAN, 2005, p. 62). Em complemento, os axiomas, segundo Laufer (2015, p, 49), são representados pelos conjuntos de afirmações que fazem parte de uma ontologia.

4.7.1 Engenharia de ontologia

Foi concebida para tratar questões que envolviam a engenharia do conhecimento nos aspectos das necessidades de compreender seus conceitos e, assim, poder compartilhá-los, além da possibilidade em sua reutilização. Portanto, a utilização dos conceitos de ontologias possibilitou uma representação explícita e sem ambiguidade do conhecimento em um dado domínio, logo, veio a preencher uma lacuna em virtude da escassez de tecnologias que possibilitassem a utilização e implantação de ferramentas para o tratamento crescente das diversas formas de conhecimentos gerados (BONINI, 2019, p. 44).

Para o processo de criação de ontologias, diversas metodologias foram propostas tais como: Método Cyc, skeletal methodology (metodologia proposta pela Uschold), metodologia Toronto Virtual Enterprise (TOVE), Methontology. Cada uma com suas características e processo intrínsecos de construção de ontologia (BREITMAN, 2005, p. 68 - 72).

Além de metodologia empregada na construção de ontologias Breitman (2005, p. 171 - 175), aponta que foram desenvolvidas ferramentas computacionais²² para auxiliar na tarefa de modelagem e construção das ontologias, tais como: i) OildEd (um editor simples de ontologia com suporte para DAM+OIL e OWL; ii) OntoEdit (um editor onde a construção de ontologias é dividida em três fases: requisitos, refinamento e validação), iii) Protégé (um ambiente para construção de ontologia e base de conhecimento, sendo bastante utilizado por ter uma arquitetura aberta).

4.7.2 Ontologic web language – owl

O World Wide Web Consortium (2013, tradução nossa) (on-line) define OWL como uma linguagem que foi projetada para “representar conhecimentos complexos sobre coisas

²² Site com diversas ferramentas computacionais (DAML TOOL) para consulta: <http://www.daml.org/tools>.

(objetos, pessoas, etc.) e suas relações para serem exploradas em programas de computadores, sendo recomendada para atender as necessidades da Web Semântica”²³.

A OWL foi criada para os seguintes requisitos: construção de ontologia e explicação de seus conceitos e propriedades; explicar fatos sobre determinado domínio, por meios de informações sobre os indivíduos que fazem parte do mesmo; determinar as consequências do que foi construído e explicitado sobre as ontologias e seus fatos inerentes (BREITMAN, 2005, p. 59). Dentro dessa perspectiva Laufer (2015, p. 49), esclarece que a OWL é uma linguagem que estende as características do RDF e RDFs proporcionando melhores tipos de restrições para as triplas utilizadas na representação da informação. São por meios de sublinguagens (OWL *Lite*, OWL *DL*, OWL *Full*) que a OWL projeta o uso de implementadores específicos (SANTAREM SEGUNDO; CONEGLIAN, 2016, p. 222).

Explica Feitosa (2006, p. 113) que o OWL *Lite* fornece os recursos de suporte para os usuários que desejam uma hierarquia simples de classificação e restrições taxonômicas. Enquanto a OWL *DL* (*Description Logic*) é direcionada para os usuários que desejam o máximo de expressividade e completeza computacional, nessa ontologia são incluídos todos os construtores da OWL. Contudo, há restrições na utilização desses construtores quanto no processo de instâncias entre classes e subclasses. Já a OWL *Full* é direcionada para os usuários que desejam máxima expressividade e liberdade sintática que não é encontrada no modelo RDF.

Alguns elementos são necessários para construção de ontologia (classes, propriedades, relações, restrições, axiomas e instâncias), conforme definidos anteriormente. No caso da OWL, são acrescentadas outras estruturas sintáticas tais como: *namespaces* e *cabeçalhos*, os quais são definidos da seguinte forma Breitman (2005, p. 60):

Namespaces – São declarações que são colocadas dentro do código de um documento onde permite que as informações contidas em uma determinada ontologia sejam interpretadas sem que ocorram ambiguidades.

Como exemplo, a autora cita uma ontologia que foi criada referente a comidas (food). Assim, uma declaração do *namespaces* que seria adicionado em um documento na seguinte forma: `xmlns:food=http://w3.org/TR/2004/REC-owl-guide-20040210/food#`

²³ “represent rich and complex knowledge about things, groups of things, and relations between things. OWL is a computational logic-based language such that knowledge, being recommended to meet the needs of the Semantic Web” (WORLD WIDE WEB CONSORTIUM, 2013).

Outro ponto importante são os cabeçalhos definidos como: “[...] sentenças sobre a ontologia adicionada, por meio da etiqueta owl:ontology, referente a comentários informativos do uso da ontologia” (BREITMAN, 2005, p. 61), ou seja, é o local destinado para colocar os metadados que fazem referência ao documento

De acordo com Feitosa (2006, p. 114), no OWL as classes são utilizadas para organização de recursos:

[...] cada classe pode fornecer um mecanismo de abstração para agrupamento de recursos e está associada a um conjunto de objetos individuais, denominados de extensão da classe. Cada objeto individual em uma classe é denominado de instância da classe, todos os recursos em um documento OWL são descritos como: classes, subclasses e propriedades das classes.

No quadro 2 é exemplificado um arquivo escrito na linguagem OWL e expresso em XML com seus requisitos básicos.

Quadro 2 – Representação do código OWL em sintaxe XML.

Código expresso em XML	Descrição do código
<?xml version “1.1”>	Indica o início do código do arquivo XML.
<!DOCTYPE Ontology [<!ENTITY xsd http://www.w3.org./2001/XMLSchema#> >]	Indica que o arquivo é uma ontologia no qual utiliza um vocabulário que está expresso em XML schema no qual foi definido pela W3C.
<owlx: Ontology owlx:names=http://www.ci.df.gov.br/leia xmlns: owlx=http://www.w3.org/2003/05/owl-xml>	Local do elemento raiz, onde indica a marca de início das declarações referente às classes owlx:names=http://www.ci.df.gov.br/leia e faz referência a uma ontologia definida pelo autor, de onde serão validados os nomes dos elementos xmlns: owlx que faz referência aos elementos que foram definidos pela W3C para uso na linguagem OWL.
Cabeçalhos dos elementos de classes e indivíduos.	Local onde serão definidos os cabeçalhos dos elementos de classes e indivíduos e suas propriedades.
</owlx: Ontology>	Indica o final do código.

Fonte; Adaptação de (W3C apud FEITOSA, 2006, p. 114).

Alguns elementos são utilizados no OWL para estabelecer as relações entre as classes e subclasses feitos por meios das propriedades e restrições na ontologia que são importados do modelo RDF e RDFs, são eles: *Class*, *subClassOf*, *Proprrty*, *SubPropertyOf*, *domain*, *range*, entre outros (FEITOSA, 2006, p. 115).

Para o processo de consulta (*query*), no ambiente da Web Semântica, tanto em bases de dados de triplas quanto em dados conceituais como o RDF, RDFs e OWL é feito por meio do protocolo SPARQL (*SPARQL Protocol and Query Language*), onde as sentenças de

consultas são semelhantes às aquelas utilizadas em banco de dados relacionais utilizando o SQL²⁴; contudo, adaptando-se ao uso de triplas (LAUFER, 2015, p. 50).

4.8 SPARQL - SIMPLE PROTOCOL AND RDF QUERY LANGUAGE

De acordo com o World Wide Web Consortium (2013) (on-line), o SPARQL é um protocolo e uma linguagem de consulta para o RDF. Assim sendo, específica para a Web Semântica com uma sintaxe específica para a realização de consultas, visto que no RDF têm-se informações armazenadas no formato de triplas, onde irão constituir as bases de dados (*triple store*) ou em banco de dados relacionais que contenham um mapeamento RDF (LAUFER, 2015, p. 55).

O World Wide Web Consortium (2013) (on-line) define a forma como o SPARQL realiza as consultas, sendo constituída por duas operações HTTP (*Hypertext Transfer Protocol*): uma para a realização de consulta, por meio da SPARQL Query Language e outra parte para operações de atualização onde executar as solicitações. Assim, ao realizar uma consulta, o cliente do protocolo SPARQL envia as solicitações para o protocolo HTTP e para os serviços do protocolo SPARQL que terão a função de realizar o tratamento das solicitações e enviar as respostas ao cliente por meio do *Hypertext Transfer Protocol*. Para a submissão de consultas, o SPARQL utiliza-se de comandos semelhantes à linguagem SQL usados nos modelos relacionais como: Select, Where, Order By, entre outros. Contudo, algumas características da linguagem SPARQL são diferentes do SQL, entre elas, relacionam-se as variáveis de consulta nas quais devem começar com o sinal de interrogação (?), além de poderem ser definidas em qualquer posição de uma tripla (LAUFER, 2015, p. 55).

Como meios exemplificativos de consultas SPARQL, em bases de dados de triplas na Web (*Dataset*), pode-se utilizar acesso via *endpoint* como no caso da DBPEDIA²⁵ e o GEODATA²⁶. *Endpoint* é um ponto de acesso a um *Dataset* que possibilita realizar consultas no formato RDF e obter resultados via protocolo HTTP (LAUFER, 2015, p. 58). As consultas no DBPEDIA são realizadas por meio do Virtuoso SPARQL Query Editor, este definido como: “um motor de busca virtual capaz de atender aos diversos servidores apenas informando grafos (recursos da web) a serem utilizados” (DANTAS FILHO; RAMALHO, 2018, p. 8). Para a implementação e funcionamento da Web Semântica, as informações devem

²⁴ Standard Query Language. Linguagem de consulta utilizada nos principais Banco de Dados (Microsoft SQL Server, Oracle, MySQL).

²⁵ Disponível em: <http://dbpedia.org/sparql>.

²⁶ Disponível em: <http://linkedgeodata.org/sparql>.

ser apresentadas de formas estruturadas e terem um conjunto de regras de inferência que possibilitem um raciocínio de forma automatizada por computadores, portanto, um dos desafios da Web Semântica é adicionar lógica computacional nas informações que são disponibilizadas na Web (BERNERS-LEE *et al.*, 2001, p. 2). Acrescentam os autores que a lógica adicionada deve ser poderosa para que possa descrever propriedades complexas referentes aos objetos, contudo, que permitam aos agentes (softwares de busca) sua interpretação.

Conforme Berners-Lee *et al.* (2001, p. 2), são por meio das tecnologias XML, RDF, Ontologia a base para o desenvolvimento da Web Semântica. A Linguagem XML trouxe a flexibilidade para a criação de *tags* de marcação possibilitando aos usuários adicionarem estruturas diversas aos seus documentos. No entanto, as criações de tais estruturas carecem de significado informacionais. Para tanto, a tecnologia RDF apontou para desempenhar o papel de fornecer significado por meios dos conjuntos de triplas (sujeito, predicado, objeto) os quais formam um conjunto interligado de informações sobre as coisas relacionadas e serão expressas no formato de *URI*, assim, fornecendo uma identificação única na Web, e permitindo o processamento por máquinas, todavia sem criar a possibilidade de inferência sobre as informações, ou seja, deduções lógicas.

Para suprir as lacunas referentes a deduções lógicas, faz-se uso das taxonomias que possibilitam a criação de classes, subclasses, propriedades e estender tais propriedades como herança para outras subclasses. Dessa forma, permite criar regras de inferência por meios de ontologias e soluções de equivalência entre os termos utilizados na linguagem de descrição XML. Destarte, as ontologias permitem uma maior precisão e eficiência em pesquisa no ambiente Web (BERNERS-LEE *et al.*, 2001, p. 3).

Dessa feita, os autores apontam que as pesquisas na Web alcançarão maior eficiência quando agentes²⁷ puderem ter acesso a um maior número de documentos criados com base nas premissas da Web Semântica e que possam trocar informações com outros agentes, conseqüentemente, criando uma interligação de informações conectadas e interpretáveis por máquinas contribuindo para a criação de uma Web de Dados Conectados.

²⁷ Software de busca utilizado para coletar informações estruturadas.

5 LINKED DATA

Nessa seção, inicia-se com o conceito sobre Linked Data e seus princípios de implementação. Em seguida, é apresentado o ciclo de vida para o Linked Data e o conceito sobre dados abertos. Por último, é descrito sobre Linked Open Data e consultas no DBPEDIA por meios de Endpoints.

5.1 LINKED DATA (DADOS VINCULADOS)

O Linked Data relaciona-se ao conjunto de dados estruturados e publicados na Web e que sejam interligados a outros conjuntos de dados externos mantendo-se legíveis, com significados, e que possam ser interpretados por máquinas (MENDONÇA, 2013, p. 33). Nesse aspecto Bizer, Berners-Lee, Heath (2009, p. 2, tradução nossa) definem Linked Data com o objetivo:

[...] serem usados na Web para criar links digitais entre dados de diferentes fontes. Eles podem ser de tipos diversos quanto são originados de bancos de dados mantidos por duas organizações em diferentes localizações geográficas, ou simplesmente sistemas heterogêneos dentro de uma organização que, historicamente, não interoperaram facilmente no nível dos dados. Tecnicamente, Linked Data refere-se a dados publicados na Web de forma que sejam legíveis por máquina, e seu significado seja explicitamente definido e que estejam vinculados a outros conjuntos de dados externos e, por sua vez, possam ser vinculados a conjuntos de dados externos.²⁸

Nesse contexto, os dados que se encontram estruturados na Web seriam interligados e possibilitaria maior acesso aos conteúdos informacionais em virtude de atenderem aos requisitos propostos pela Web Semântica com a utilização do modelo padrão RDF – o qual permite relacionar os dados de várias origens diferentes (RAUTENBERG *et al.*, 2017, p. 124).

A fim de possibilitar interligação de dados na Web e sua publicação algumas regras deverão ser seguidas. Nesse contexto, foi proposto os “Princípios do Linked Data” por Berners-Lee (2006, p.1), sendo os seguintes:

- I) Usar URIs como nomes para as coisas (recursos);

²⁸ [...] using the Web to create typed links between data from different sources. These may be as diverse as databases maintained by two organisations in different geographical locations, or simply heterogeneous systems within one organisation that, historically, have not easily interoperated at the data level. Technically, Linked Data refers to data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets (BIZER, BERNERS-LEE, HEATH, 2009, p. 2).

- II) Usar *Hipertext Transfer Protocol* (HTTP) e URIs para que pessoas possam consultar esses recursos;
- III) Quando alguém consultar uma URIs, forneça informações úteis, usando o padrão RDF e SPARQL;
- IV) Incluir links para outras URIs, para que possam descobrir mais coisas (recursos).

Em análise dos princípios do Linked Data, Laufer (2015, p. 72) esclarece que:

O ponto inicial dá-se com a utilização dos *URIs* que possibilitam a identificação dos diversos recursos que podem ser disponibilizados na Internet, tanto de forma abstrata quanto concreta. Tais recursos podem ser representados e referenciados por meio do padrão RDF com a utilização das triplas (sujeito, predicado e objeto) além das propriedades, valores e relacionamentos com outros recursos. Em seguida é apontado o HTTP, um protocolo de comunicação, no qual desempenha a função de atender as requisições via Web de acordo com as *URIs* apresentadas. Posteriormente, ao realizar consultas sobre determinada *URIs*, os servidores Web deverão retornar as informações requisitadas pertinentes aos recursos apontados pela URIs inicial, sendo apresentadas na forma de SPARQL, por meio de triplas ou um arquivo RDF de acordo com as configurações do servidor Web. Por último, é apontada a ligação entre outros dados para que possibilite uma expansão de conhecimento e a obtenção de novos recursos na Web (LAUFER, 2015, p. 72).

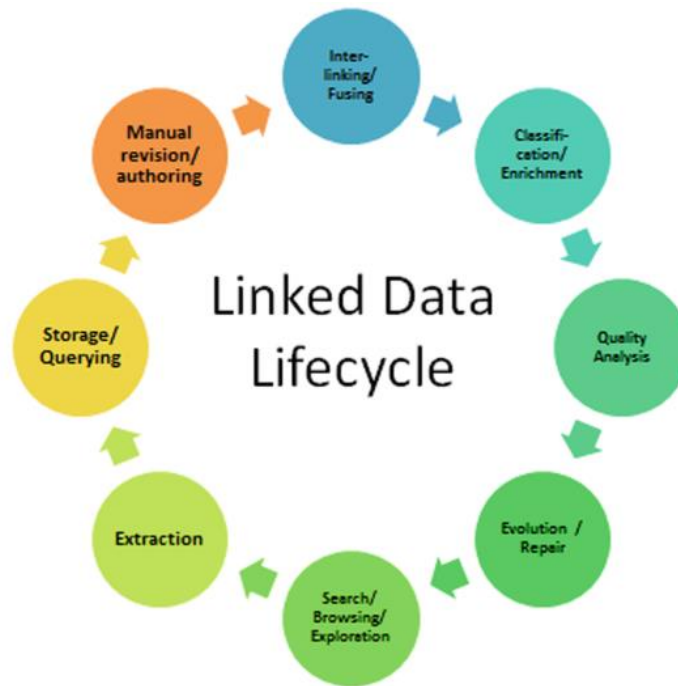
5.2 CICLO DE VIDA DO LINKED DATA

Com a finalidade de disponibilizar dados estruturados e conectados no ambiente Web o Instituto de pesquisa *Agile Knowledge and Semantic Web*²⁹ propôs uma metodologia “Linked Data Lifecycle” (figura 13), que orienta as diversas fases para estruturação de dados no contexto de Linked Data (RAUTENBERG *et al.*, 2017, p. 124; AUER, 2014, p. 4).

Segundo Auer (2014, p. 4), cada fase dessa metodologia não deve ser tratada individualmente, mas de maneira conjunta, que facilite o desenvolvimento e busca por solucionar os desafios na construção de dados conectados, sendo definidas da seguinte forma:

²⁹ Site do Instituto: <https://aksw.org/About.html>.

Figura 13 – Ciclo de Vida do Linked Data.



Fonte: Auer (2014, p. 4).

Partindo do ponto em que os dados que serão publicados encontram-se de forma não estruturada ou em diferentes formatos, segue as explicações de cada fase, de acordo com Rautenberg *et al.* (2017, p. 128),

Extraction (Extração) – é o mapeamento dos dados não estruturados ou estruturados e em diferentes formatos para o modelo RDF;

Storage/Querying (Armazenamento/Consulta) – é a disponibilização dos dados RDF com o uso de sistemas gerenciadores (*Triple Stores*) para potencializar as tarefas de publicação e de consumo de dados;

Manual Revision/Authoring (Revisão Manual/Autoria) – engloba as tarefas de criação ou editoração de novos dados em um grafo RDF;

Interlink/Fusion (Interligação/Fusão) – estabelece o relacionamento de dados de um grafo RDF para com os dados de outra fonte para explorar novas informações;

Classification/Enrichment (Classificação/Enriquecimento) – viabiliza a expressividade e a riqueza semântica de um conjunto de dados em relação a um contexto, representando os dados mediante o uso de ontologias ou vocabulários;

Quality Analysis (Análise de Qualidade) – pontualmente, tratam dos aspectos de integridade, precisão, consistência e validade de dados. E de forma geral, verifica os requisitos de compreensibilidade, disponibilidade e proveniência do modelo de dados;

Evolution/Repair (Evolução/Reparação) – uma vez encontradas inconsistências nos dados ou no modelo de representação, ações corretivas devem ser tomadas a fim de corrigir as não conformidades;

Search/Browsing/Exploration (Busca/Navegação/Exploração) – usa técnicas de busca, navegação, exploração para visualizar os dados RDF.

Uma vez os dados estruturados, eles poderão ser disponibilizados em diferentes formas de acesso para a sociedade a fim de usufruí-los. No sentido de direção à democratização de acesso à informação, iniciativas têm sido tomadas pela comunidade científica, como apontada para o *Open Data*. (Dados abertos). Sendo relacionado ao conjunto de dados que podem ser usados, distribuídos e reutilizados livremente, sem que haja restrições de propriedades, ou quando muito a exigência de serem atribuídas e compartilhadas com o mesmo tipo de licença (MENDONÇA, 2013, p. 34; RAUTENBERG *et al.*, 2017, p. 124). Desse modo, a proposta de *Open Data* foi adicionada ao Linked Data para o desenvolvimento de um projeto da W3C o Linked Open Data – LOD.

O LOD tem como objetivo a identificação de conjunto de dados abertos que possam ser convertidos nos moldes de Linked Data e serem publicados na Web (HEATH; BIZER, 2011, apud MENDONÇA, 2013, p. 34). Para que ocorram publicações das informações como Linked Open Data, desafios deverão ser superados, entre eles: a necessidade do compartilhamento de dados locais e sua publicação no ambiente Web; O mapeamento e a publicação de sistemas legados (sistemas antigos) que deverão utilizar vocabulários que são amplamente utilizados na Web de Dados, e por último, a verificação da qualidade dos dados e reparos necessários (RAUTENBERG *et al.*, 2017, p. 129),

Dentro desse conceito, Berners-Lee (2006, p.1) propôs critérios de maturidade de dados abertos (Linked Open Data - LOD) classificando por meios níveis de estrelas (Quadro 3):

Quadro 3 – Critérios de maturidade de dados abertos (Linked Open Data – LOD).

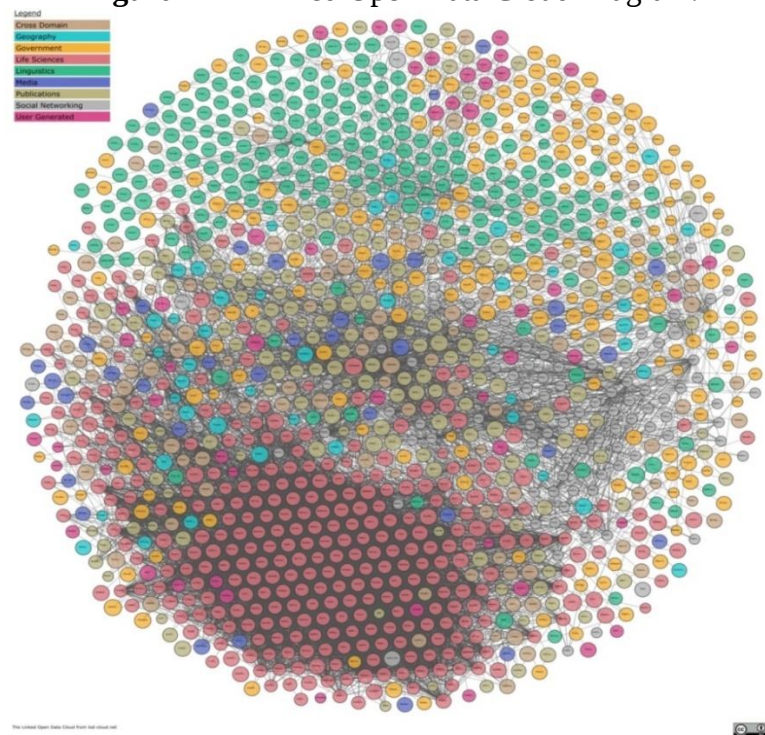
★	Disponibilizar na web (qualquer formato), mas com licença aberta, para ser <i>Open Data</i> .
★★	Disponibilizar como dados estruturados legíveis por máquina (por exemplo, excel em vez de digitalização de imagem de uma tabela).
★★★	Utilizar formato não proprietário (por exemplo, CSV em vez de Excel).
★★★★	Para todos os itens acima use padrões abertos do W3C (RDF e SPARQL) para identificar coisas, para que as pessoas possam apontar para suas coisas.
★★★★★	Use todos os itens acima, mas: vincule seus dados aos dados de outras pessoas para fornecer contexto.

Fonte: Adaptado de Berners-Lee, (2006, p.1).

O objetivo de apresentar tais critérios dá-se em virtude do alto grau de heterogeneidades dos dados no ambiente da Internet. Dessa forma, a descrição de um caminho evolutivo e sua representação facilitará a obtenção de formatos para uma Web de dados, ou seja, irá direcionar para obtenção de informações semânticas para que possam ser utilizadas como Linked Data (LAUFER, 2015, p. 73).

Apresenta-se como a principal representação de Linked Open Data a visualização da DBPEDIA, sendo seu diagrama esquematizado por grafo Linked Open Data Cloud Diagram³⁰ (Figura 14), que mapeiam diversos conjunto de triplas de *Datasets* de diferentes domínios (BIZER; HEATH; BERNERS-LEE, 2009, p. 5).

Figura 14 – Linked Open Data Cloud Diagram.



Fonte: lod-cloud.net (2020) (on-line).

A imagem representa o conjunto de dados que foi publicado no formato de Linked Open Data Cloud, sendo composta de 1.301 conjuntos de dados com 16.283 links publicados até maio de 2020 (LOD-CLOUD.NET, 2020).

Além das representações em forma de grafo, a DBPEDIA dispõe de formas de consultas utilizando a linguagem SPARQL a um SPARQL (*endpoint*) que pode ser acessado em <https://dbpedia.org/sparql/>. Esse *endpoint* utiliza um servidor de dados conectado denominado de Virtuoso (LAUFER, 2015, p. 83).

³⁰ Diagrama da DBPEDIA, disponível em: <https://lod-cloud.net>.

6 PROCEDIMENTOS METODOLÓGICOS

Para a construção de uma pesquisa bem consolidada, procedimentos científicos hão de ser seguidos com o objetivo de alcançar uma acurácia nos dados e oferecer subsídios para uma repetitividade do experimento quando possível. Dentro dessa visão, a pesquisa científica aponta como um conjunto de procedimentos realizados de forma racional e sistemática, objetivando responder as indagações sobre determinados problemas que são analisados e propostos (GIL, 2018, p.1). Dentro desse aspecto, uma das possibilidades de pesquisa dá-se no campo da bibliometria.

De acordo com Tavares e Celerino (2018, p 11), a bibliometria é definida como “[...] uma prática multidisciplinar utilizada para identificação do comportamento e da evolução da literatura em determinado período ou contexto”, fazendo uso de aplicações estatísticas e matemáticas com a finalidade de mensuração da produtividade, assim, possibilitando realizar análise quantitativa das informações em diversos documentos. Por conseguinte, possibilita reconhecer tendências e o conhecimento em determinado domínio do saber e prevê propensões de publicação, além de estimar a produtividades de autores, organizações e países, bem como avaliar o processo de citações, cocitações e o grau de crescimento de temas sobre determinada área do conhecimento (CAMARGO; BARBOSA, 2018, p. 114).

Para tanto, necessita-se dos indicadores bibliométricos, a fim de mensurar os dados. Diversas formas podem ser utilizadas para a construção de indicadores. Todavia, de modo geral, existem as seguintes etapas: recuperação e pré-processamento dos dados, definição das unidades de análise e métricas, definição do tipo de análise, apresentação dos resultados e conclusões (CAMARGO; BARBOSA, 2018, p. 117).

Camargo e Barbosa (2018, p. 118) mencionam que o ponto de partida nesse processo é definir a base de dados para recuperação das informações e, posteriormente, um pré-processamento para adequação dos dados obtidos (refinamento, limpeza dos dados, entre outros). Já as etapas subsequentes irão se enquadrar aos tipos de dados coletados. Por não ser uma tarefa trivial no processo de construção de indicadores, diversas ferramentas computacionais (softwares) encontram-se disponíveis, cuja finalidade é facilitar as análises bibliométricas. Contudo, seu uso dependerá dos dados coletados a serem analisados (CAMARGO; BARBOSA, 2018, p. 121).

Nesse aspecto, existe um conjunto de softwares direcionados para as análises bibliométricas, bem como para a coleta e análises de informações em bases de dados e sua visualização. Os softwares podem incluir diversas funcionalidades como: medidas de normalização, agrupamento, tipos de análises bibliométricas, entre outros (RUAS; PERREIRA, 2014, p. 59). Ferreira e Silva (2019, p. 452-453) apontam alguns tipos de softwares que podem ser utilizados para avaliação bibliométrica como: Citespace, VOSviewer, Ucinet, Pajek.

Entre os softwares apresentados, destaca-se o VOSviewer por ser uma ferramenta projetada para a construção e visualização de mapas bibliométricos em formas gráficas, além de permitir realizar agrupamentos das informações e apresentação destas de forma bidimensional, sendo visualizadas das seguintes formas: com rótulos, por densidade, por densidades de clusters e por dispersão (COBO *et al.*, 2011, p. 1390).

Essa pesquisa está direcionada para a análise bibliométrica sobre os temas Web Semântica e Linked Data por meio de artigos científicos completos disponíveis nas Bases de Dados da *Scielo e Web of Science*, com objetivo de averiguar as convergências sobre esses temas, portanto, caracterizando como uma pesquisa bibliográfica. Conforme Gil (2018, p. 28), uma pesquisa bibliográfica é elaborada a partir das análises de artigos científicos, teses, dissertações, livros, jornais, revistas e anais de eventos científicos, que tenham sido publicados de forma impressa ou que estejam disponíveis em formato digital.

Uma ressalva faz-se necessária a fim de elucidar quanto ao meio de pesquisa bibliográfica e documental:

[...] a pesquisa documental apresenta pontos semelhantes com a pesquisa bibliográfica. A principal diferença está na natureza das fontes. A pesquisa bibliográfica fundamenta-se em material elaborado por autores com o proposto específico a ser lido por público específico. Já a pesquisa documental vale-se de toda sorte dos documentos elaborados com finalidades diversas (GIL, 2018, p. 4).

Em síntese, para o processo de desenvolvimento dessa pesquisa foram instituídas etapas, algumas de forma isolada, outras, concomitantemente, ao desenvolvimento do trabalho, sendo em linhas gerais as seguintes:

- a) Levantamento do referencial teórico para um entendimento do tema proposto;
- b) Escolha das Bases de Dados para coleta dos dados;

- c) Estruturação das consultas (*query*) e uso de filtros de refinamento de pesquisa em bases de dados *Scielo* e *Web of Science*;
- d) Leitura dos tópicos (título, resumo e palavras-chave) dos artigos recuperados e posterior *downloads* para armazenamento e análise com maior profundidade ao conteúdo;
- e) Definição de indicadores métricos; Classificação e organização dos artigos;
- f) Extração dos dados de acordo com os indicadores referentes; Tratamento dos dados e armazenamento das informações em planilhas de excel;
- g) Análise dos dados por meios de softwares específicos;
- h) Análise e discussão das informações obtidas.

Para a classificação dos métodos de pesquisa adotaram-se as orientações de Gil (2018, p. 26) sendo classificada a pesquisa: quanto ao objetivo (fins), em geral, enquadra-se como uma pesquisa exploratória, visto há necessidade de obter-se conhecimento do tema no qual foi realizada na fase do referencial teórico; quanto aos meios a pesquisa apresenta-se como uma pesquisa bibliográfica; quanto à natureza dos dados obtidos a pesquisa classifica-se como quali-quantitativa. No aspecto qualitativo é referenciado com as análises dos artigos completos recuperados nas bases de dados, enquanto, quantitativa dá-se por meios das análises bibliométricas.

Esclarece-se que nas pesquisas quantitativas os resultados obtidos são expostos em formatos numéricos, enquanto, nas pesquisas qualitativas são direcionados em uma forma mais textual, ou seja, no campo da narrativa. Portanto, uma pesquisa quali-quantitativa, também, conhecida como mista é a junção de ambas nas descrições dos resultados (GIL, 2018, p. 40).

Para verificação dos dados quantitativos coletados foram utilizados elementos básicos da estatística (somatório, agrupamento, percentual e média) e análise por meio de métodos estatísticos descritivos que possibilitou: a organização dos dados e a formação de resumos visando a construção de tabelas e gráficos, permitindo realizar inferências sobre as informações coletadas.

Quanto à verificação dos dados qualitativos foi utilizada a visitação a conceitos básicos do método de análise de conteúdo (pré-análises, exploração do material, tratamento e

inferência dos resultados) propostos por Bardin (2016, p. 125). Para a realização da fase qualitativa foram realizadas análises de conteúdos dos artigos verificando os títulos, resumos e palavras-chave com a finalidade de criar agrupamentos temáticos. Dessa forma, foram concebidas 13 categorias que possibilitaram classificar o *corpus* oriundo das consultas realizadas nas bases de dados *Scielo* e *Web of Science*.

Quanto à coleta de dados diversas bases de armazenamento encontram-se disponíveis para realizar pesquisas científicas sobre os temas propostos, seja em âmbito nacional, seja à nível internacional. Entre elas temos: *Scielo*, *BRAPCI*, *Scopus*, *Web of Science*, entre outras.

Como o foco da pesquisa é analisar a produção científica dos artigos nacionais e internacionais, nesse sentido, a escolha entre a *Scielo* e a *BRAPCI* para os artigos nacionais deu-se a favor da *Scielo*, em virtude: a) de ser uma base multidisciplinar onde há publicações de diversas áreas, logo, sendo mais abrangentes em seu processo de indexação de periódicos, enquanto a *BRAPCI* congrega mais artigos direcionados a Ciência da Informação; b) de ambas as bases possibilitarem realizar pesquisas com diversos operadores booleanos e buscas por autores, títulos, palavras-chave, resumo; entretanto, a *Scielo* possibilita um maior refinamento dos dados de pesquisa por meios de maior quantidade de filtros; c) de a *Scielo* possibilitar exportar as informações sobre a pesquisa em textos com arquivos sem formatação ou com tabulação, além de outros formatos como o RIS; dessa forma, possibilitando utilizar os dados exportados em softwares como o VOSviewer, enquanto a *BRAPCI* não tem essa possibilidade; d) em face da possibilidade de realizar pesquisa no portal de periódico da CAPES, via plataforma CAFE, o que viabilizou um maior acesso a diversos artigos, sem restrições de leitura e downloads.

Em relação à escolha da base de dados internacional, tanto a *Scopus*, quanto a *Web of Science* possibilitam buscas refinadas, por meios de diversos filtros, e serem acessadas via o portal de periódicos da CAPES. Seus sistemas de buscas podem utilizar os operadores booleanos e filtros avançados, além de arquivos de exportação das consultas para serem utilizadas no VOSviewer. A *Scopus* e a *Web of Science* são bases de dados multidisciplinares, ou seja, possuem uma grande quantidade de publicações científicas indexadas por vários periódicos, nas avaliações dos índices de indicadores de impacto (índice h) existem poucas diferenças entre ambas (SILVA; GRÁCIO, 2017, p.197). Entretanto, segundo Silva e Grácio (2017, p.198), a *Web of Science* tem maior rigor para indexação de artigos em sua base de dados, assim, garantindo alto padrão de qualidade. Outro ponto refere-se às atualizações que

são semanalmente, enquanto na *Scopus* é diariamente; dessa forma, os dados coletados de uma pesquisa podem ser diferentes em outros dias de coleta. Acrescenta-se em virtude da escolha pela base de dados *Scielo* para realizar as pesquisas dos arquivos nacionais, e visando uma padronização na construção das consultas, na estrutura e no formato dos dados coletados para a exportação. Nesse contexto, aponta-se que tanto a *Scielo* quanto a *Web of Science* são acessadas por meio da Clarivate Analytics, desse modo, há uma padronização na forma de consultas das informações, enquanto a *Scopus* é feita pela plataforma Elsevier com padrões diferentes. Com base nesses pontos, a escolha de consultas em bases internacionais deu-se pela *Web of Science*.

Para a obtenção dos dados foram realizadas pesquisas nas Bases de Dados *Scielo* e *Web of Science*, ambas acessadas no Portal de Periódico da CAPES³¹, via acesso CAFE, disponível aos discentes e docentes da UFPE. Os dados coletados nas bases de dados foram artigos completos *Open Access*. A coleta dos dados deu-se no período da segunda quinzena de abril de 2022.

Para o critério de busca e seleção foram utilizadas palavras-chave que caracterizam o objeto de estudo, tais como: Web Semântica, Linked Data, Semantic Web, Web de significado, Dados conectados, Dados lincados. As três primeiras sendo utilizadas como termos principais e as demais como consultas acessórias para teste de consultas.

Foram elaboradas consultas (*queries*), por meios de operadores booleanos (AND, OR, NOT) com a finalidade de buscas avançadas nas bases de dados. Devido às bases terem sistemas de busca e uso de filtros diferenciados, nesse caso, buscamos homogeneizar as consultas com filtros básicos para o refinamento de artigos completos, todos os tópicos (Título, Resumo e Palavras-chave), todos os idiomas e ano de publicação para todos, posteriormente, alguns ajustes foram realizados.

Entretanto, ocorreram algumas diferenças quanto à realização das buscas nas bases de dados da *Scielo* e *Web of Science* para as coletadas de registros: na base de dados da *Scielo*, os artigos recuperados foram referentes à produção nacional, sendo excluído o ano de 2022 na consulta, em virtude, de não ter todas as informações anuais de produção científica ainda computados. Enquanto que na *Web of Science* não houve a filtragem de nacionalidade, entretanto, devido à grande quantidade de registros recuperados fez-se necessário incluir cortes por: a) área de pesquisa (*Information Science Library Science*) para ter-se uma

³¹ <https://www-periodicos-capes-gov-br.ez16.periodicos.capes.gov.br/>.

quantidade de artigos para a viabilidade de analisá-los; b) por exclusão de artigos de acesso antecipado; c) exclusão do ano vigente (2022) para não entrar na contagem; d) em ambas as consultas foram excluídos os registros referentes a Linked Open Data, visto que não era o propósito na avaliação. As *queries* de consulta para cada base de dados estão representadas em seguida.

1 – Query na Base de Dados da *Scielo*:

("Web semantic" OR "Web semântica" OR "Web semantica" OR "Semantic Web") (Tópico) and ("Linked Data") (Tópico) not ("Linked Open Data") (Tópico) and 2022 (Excluir – Anos da publicação) and Research Article (Tipos de documento) and BRAZIL (Países/Regiões)

2 – Query na Base de Dados da *Web of Science*:

("Web semantic" OR "Web semântica" OR "Web semantica" OR "Semantic Web") (Tópico) and ("Linked Data") (Tópico) not ("Linked Open Data") (Tópico) and 2022 (Excluir – Anos da publicação) and **Acesso antecipado** (Excluir – Tipos de documento) and Acesso Aberto and Artigos (Tipos de documento) and Information Science Library Science (Áreas de pesquisa)

Foram recuperados 5 registros na Base de Dados *Scielo* e na da Base de Dados da *Web of Science* 40 registros, de forma inicial. Tais registros serviram para a construção dos dados bibliométricos. Seguidamente, foram realizados os *downloads* dos registros das bases de dados em formato de arquivo TXT, os quais possibilitaram ser avaliados no software VOSViewer (versão 1.6.18) para a realização das inferências bibliométricas e, no software Excel para a construção das planilhas e gerar as tabelas e gráficos. Em um segundo momento, foram selecionados: os artigos da Base de Dado *Scielo* e os artigos mais relevantes da *Web of Science*, de acordo com seu critério de classificação.

Quanto às análises bibliométricas quantitativas foram divididas em duas etapas para cada uma das bases de dados, mas seguindo os mesmos métodos sistemáticos para ambas. Ao realizar as *queries* de consultas e análises dos resultados foram exportados arquivos da consulta no formato TXT, sem tabulação, e no formato excel tabulado.

Os arquivos possibilitaram criar os mapas de coautoria e análises dos registros, filtragens e aplicação dos métodos estatísticos para quantificação, média e percentagem. Para tanto, os indicadores de análises escolhidos para a pesquisa foram: a produtividade de publicações, os locais de publicações dos artigos, relações de autoria e coautorias, mapa de

interações de coautoria, o grau de colaboração, citações dos artigos e vínculos institucionais dos autores.

Tais indicadores possibilitaram mensurar os níveis de colaboração e tendências dos pesquisadores quanto aos temas nos quais estão pesquisando, além, de saber quais os principais centros de pesquisas acadêmicas, regiões e países que estão focando nos temas da Web Semântica e Linked Data. Possibilitou, também, inferir sobre os grupos de trabalhos formados e em que níveis estão as ligações: fortes ou fracas entre os clusters, e os números mínimos e máximos de documentos produzidos por cada autor.

Quanto às análises bibliométricas qualitativas, foi realizada a análise de conteúdo dos artigos. Inicialmente, com a etapa de pré-análise dos resultados obtidos com as *queries* nas bases de dados. Em seguida foram explorados, com maior acurácia, os artigos por meio de leitura de tópicos para gerar os agrupamentos de classificação, assim, possibilitando direcionar para criação das categorias temáticas. A partir desses conjuntos de descritores houve um ordenamento dos termos que pudessem identificar e classificar temas específicos de pesquisa. De posse dos descritores classificados, os termos mais importantes foram extraídos e transcritos para a criação dos temas para cada categoria temática. Cada categoria criada recebeu um código, com a finalidade de uma melhor identificação em quadro de classificação, além do título da categoria.

Em seguida, realizou-se análise textual das categorias com a finalidade de averiguar redundância ou possíveis reestruturações. Com as categorias temáticas definidas, iniciou-se a seleção dos artigos para uma descrição do seu conteúdo, análises e inferências e correlação entre as categorias para apresentação dos resultados na análise bibliométrica qualitativa. Dessa forma, possibilitando verificar quais as categorias estão sendo direcionadas às publicações de artigos científicos e os seus contextos.

Quanto ao recorte temporal para a realização da pesquisa, em virtude de não ter sido definido *a priori*, só ocorreu a restrição para exclusão do ano de 2022, pois os dados não se encontram totalmente completos para refletir a produção desse ano. A flexibilidade de período possibilitou uma maior abrangência dos resultados da pesquisa, visto que, o maior intervalo temporal ocorreu na base de dados da *Web of Science* de 2010 a 2021, enquanto na base de dados da *Scielo* deu-se entre 2016 a 2020, portanto, o período temporal da *Web of Science* inclui o da *Scielo*.

7 ANÁLISE BIBLIOMÉTRICA DAS PUBLICAÇÕES SOBRE A CONVERGÊNCIA DA WEB SEMÂNTICA E LINKED DATA

Como forma de esclarecimento, essa seção inicia-se com abordagem quantitativa das publicações recuperadas na base de dado da *Scielo* e, posteriormente, a abordagem quantitativa das publicações recuperadas na base de dado da *Web of Science*. Para ambas as bases de dados foram analisadas as informações quanto: à produtividade de publicações dos autores; aos locais de publicações dos artigos científicos, ou seja, em quais periódicos; às relações de autoria e coautorias nas publicações dos artigos; ao mapa de interações sobre autoria e coautoria; ao grau de colaboração (GC) entre os autores e coautores por ano de publicação; ao quantitativo de citações dos artigos e os vínculos intituionais dos autores.

Em segundo momento, são apresentados os aspectos qualitativos com base nos artigos recuperados e selecionados em ambas as bases de dados. O *corpus* foi composto por 37 artigos, nos quais foram categorizados por áreas temáticas, por meio das análises dos conteúdos. Também foi gerado uma quantificação dos artigos de acordo com as categorias temáticas e período de publicação.

As tabulações deram-se por meios de planilha Excel onde foram aplicados métodos estatísticos para gerar valores de média, total e percentuais para a criação de gráficos, quadros e tabelas, além, de aplicações de filtros de seleção. Com base nos dados filtrados e formatados realizaram-se as análises dos resultados quantitativos e as tabulações qualitativas.

7.1 ASPECTOS QUANTITATIVOS

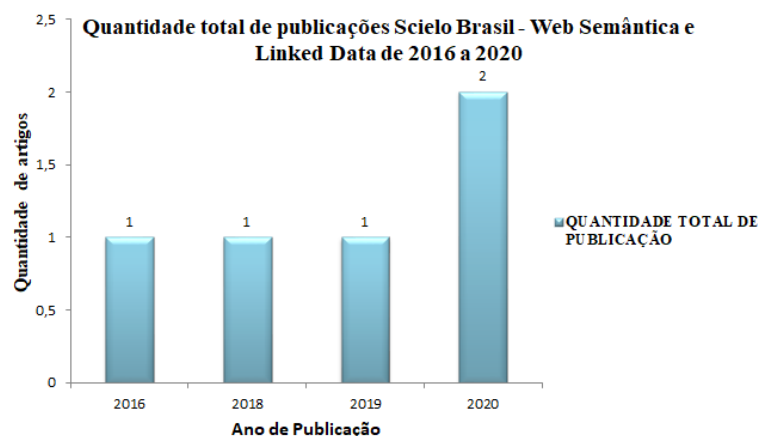
7.1.1 Das publicações recuperadas na base de dados da *Scielo*

A partir dos cinco registros recuperados na consulta foi possível averiguar a evolução de produção científica, entre 2016 e 2020, sobre os temas Web Semântica e Linked Data (Gráfico 1). Além de identificar a autoria dos artigos, vínculo institucional dos autores, rede de coautoria, periódicos utilizados para as publicações, ano de publicação dos artigos e número de citações de cada autor.

Em relação à produtividade de publicações dos autores por ano, verificou-se que a maior produtividade nacional ocorreu em 2020, com dois artigos publicados, seguindo com um único artigo nos anos de 2016, 2018 e 2019. Com base nos dados apresentados, pode-se inferir que, de modo geral, houve pouca publicação sobre os temas. Acrescenta-se uma

tendência de níveis de estabilidade produtiva em anos consecutivos e, posteriormente, uma leve alta de publicações em ano seguinte. A média geral de produtividade no período foi de 1.25 artigos publicados/ano. Logo, podendo ser considerada como baixa.

Gráfico 1 – Quantidade de artigos publicados na Base de Dados da *Scielo* Brasil sobre os temas: Web Semântica e Linked Data entre 2016 e 2020.



Fonte: Dados de Pesquisa (2022).

Portanto, há necessidade de estimular o interesse por esse tema nas diversas áreas da ciência, em especial na Ciência da Informação, visto que se trata de temas emergentes com grandes possibilidades de aplicabilidade em processos de recuperação de informações e interação homem/máquina, bem como no uso dos computadores para realizar inferências significativas das informações disponíveis na Internet.

Novas áreas têm proporcionado o uso das tecnologias voltadas para inteligência artificial e uso semântico das informações, a exemplo a telemedicina que unida com a Ciência da Informação tem proporcionado desde consultas remotas, acesso a prontuários e possibilidades de estudos de doenças com dados e diagnósticos correlacionados em bases de dados, a exemplo a Pubmed³², com conteúdos direcionados à área da saúde.

Quanto aos locais de publicações dos artigos científicos, as escolhas ficaram restritas a três periódicos nos períodos de 2016 a 2020. Apenas um periódico teve maior número de publicações: Transinformação, com 3 publicações, equivalendo a 60,00%, do total. Os demais periódicos ficaram com 20% cada um (Tabela 1).

De acordo com a plataforma Sucupira³³, o “qualis” para os periódicos Transinformação e Perspectivas em Ciência da Informação é classificado como A1 para

³² <https://pubmed.ncbi.nlm.nih.gov>.

³³ Plataforma Sucupira : <https://sucupira.capes.gov.br/sucupira/>.

ambos, referentes à área de Comunicação e Informação. Devido à classificação “qualis” A1 ser o mais alto nível, esses periódicos impõem maior rigor para publicações de artigos científicos.

Tabela 1 – Lista de periódicos para as publicações sobre Web Semântica e Linked Data - Base de Dados *Scielo* Brasil no período de 2016 a 2020.

PERIÓDICO	2016	2018	2019	2020	TOTAL	%
JISTEM *	0	1	0	0	1	20,00
Perspectivas em Ciência da Informação	1	0	0	0	1	20,00
Transinformação	0	0	1	2	3	60,00
TOTAL	1	1	1	2	5	100,00

* (*Journal of Information Systems and Technology Management*).

Fonte: Dados de Pesquisa (2022).

As preferências para publicações nos periódicos Transinformação e Perspectivas em Ciência da Informação refletem a busca por maiores projeções dos trabalhos científicos, além do que há um maior rigor em análises quanto aos pares dos artigos. Fato em que são indexados pelas grandes bases de dados como *Web of Science* e *Scopus*. O quadro 4 indica os artigos, os periódicos e o ano de publicação.

Quadro 4 – Periódicos que publicaram os artigos entre 2016 e 2020 – *Scielo* Brasil.

Título do artigo	Periódico	Ano
Um protótipo Linked data para catalogação semântica de publicações.	Perspectivas em Ciência da Informação.	2016
Recursos audiovisuais na Web de Dados: a construção do Portal Semântico Audiovisual.	Transinformação	2019
Um método para publicação semântica Linked Data de bases de dados convencionais e um estudo de caso real de artigos acadêmicos.	Transinformação	2020
Um método para criação de tesouros semânticos de requisitos usando a ontologia SKOS.	Transinformação	2020
Methodology for The Development of an Ontology Network on The Brazilian National System for the Evaluation of higher Education (OntoSINAES).	JISTEM - Journal of Information Systems and Technology Management	2018

Fonte: Dados de Pesquisa (2022).

Sobre o aspecto da relação entre autores e coautores, baseado nos cinco registros de publicações, percebe-se que não ocorreu nenhum artigo com publicação única, logo, havendo uma forte colaboração entre os pesquisadores (Quadro 5). Os dados apontam que o pesquisador Mark Douglas de Azevedo Jacyntho destaca-se na participação de três artigos, o

que equivale a 60% do total dos cinco artigos, os demais aparecem uma única vez equivalendo a 20% para cada um em relação aos cinco artigos publicados.

Quadro 5 – Relação de autores e coautores em publicação dos artigos sobre Web Semântica e Linked Data – *Scielo* Brasil.

AUTORES e COAUTORES	ARTIGO
Freitas Junior, Nilton; JACYNTHO, Mark Douglas de Azevedo	Um protótipo Linked data para catalogação semântica de publicações
CONEGLIAN, Caio Saraiva; ARAKAKI, Ana Carolina Simionato; GONÇALEZ, Paula Regina Ventura Amorim; SANTAREM SEGUNDO, José Eduardo	Recursos audiovisuais na Web de Dados: a construção do Portal Semântico Audiovisual
GONÇALVES, Adriano de Oliveira; JACYNTHO, Mark Douglas de Azevedo	Um método para publicação semântica Linked Data de bases de dados convencionais e um estudo de caso real de artigos acadêmicos
ZACARIAS, Rodrigo Oliveira; JACYNTHO, Mark Douglas de Azevedo; VASCONCELOS, Aline Pires Vieira de	Um método para criação de tesauros semânticos de requisitos usando a ontologia SKOS
SILA, Cleiton; BELO, Olando; BARROS, Victor	Methodology for The Development of an Ontology Network on The Brazilian National System for the Evaluation of higher Education (OntoSINAES)

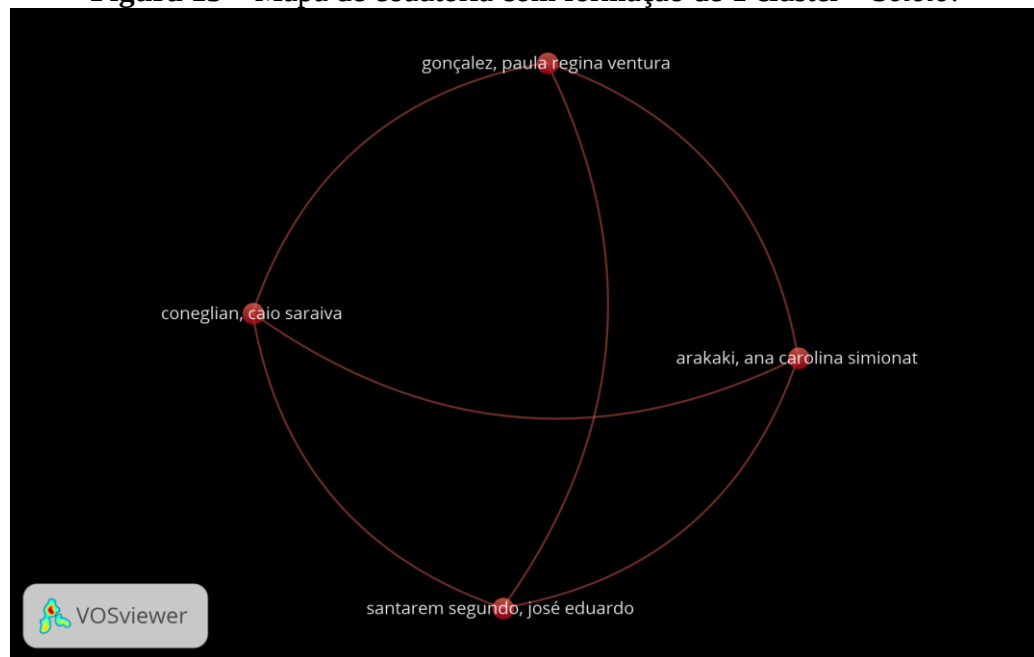
Fonte: Dados de Pesquisa (2022).

Os dados refletem baixa produtividade de forma individualizada e uma forte relação entre autores/coautores. Nesse aspecto, presume-se a formação de grupos de produção fortemente coesos para produção científica sobre Web Semântica e Linked Data. De um ponto de vista, a formação dos grupos fortalece o grau de confiança entre os pesquisadores de determinado grupo, contudo, pode configurar isolamento entre outros grupos, assim, prejudicando a troca de informações e avanços nos diversos campos de pesquisa científica.

Por meio do mapa de rede de interações pode-se visualizar o grau de interações entre autores e coautores. Apresenta-se na figura 15 o mapa de rede desenvolvido no VOSviewer com seus agrupamentos (Clusters). Como fonte de informação, com os mesmos dados, foi elaborado outro mapa sem que houvesse a rede de interações totalmente interligada (Figura 16). Na visualização da rede de coautoria, os nós (representados por círculos) indicam os autores, enquanto as arestas (linhas) indicar as relações com os coautores. Quanto maior o nó, maior será sua rede de interações com os coautores, assim, indicando uma frequência relativa na estrutura de rede.

No mapa em rede de interações totalmente interligada, apresentado na figura 15, percebe-se a formação de um único agrupamento (cluster) composto por 4 itens, sendo a relação dos coautores. Esse cluster único é formado por: ‘Arakaki, Ana Carolina Simiont; Coneglian, Caio Saraiva; Gonzalez, Paula Regia Ventura; Santarem Segundo, José Eduardo’. Tal agrupamento mostra a forte relação entre esses pesquisadores. Esclarece-se que para a formação da relação em redes de coautoria, as análises deram-se com relação de no mínimo um documento, visto a baixa quantidade de artigos produzidos por cada autor.

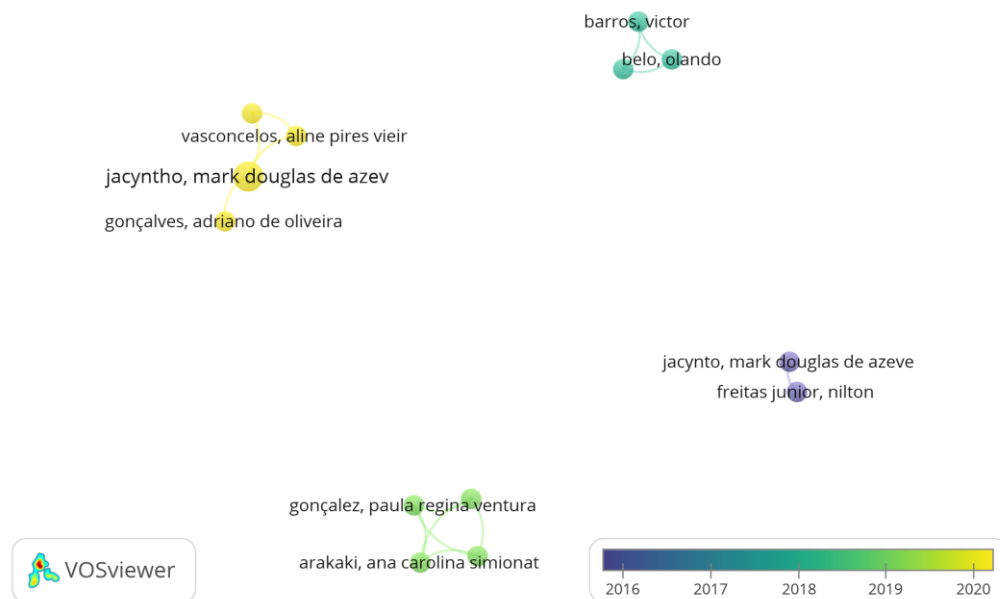
Figura 15 – Mapa de coautoria com formação de 1 cluster – *Scielo*.



Fonte: Dados de Pesquisa (2022). Uso do software VOSviewer.

A figura 16 mostra o mapa dos autores sem compor uma rede de interações totalmente interligada, ou seja, alguns dos itens do conjunto de dados não estavam conectados, desse modo, foram formados quatro clusters. Os clusters indicam quais os coautores que contribuíram na produção em um determinado artigo.

Figura 16 – Mapa de coautores sem a rede de interligações totalmente interligada de quatros clusters.



Fonte: Dados de Pesquisa (2022). Uso do software VOSviewer.

O nó ‘Jacynto, Mark Douglas de Azevedo; Freitas Junior, Nilton’ consta da produção de um único artigo com um autor e com um coautor. Os demais nós indicam que cada artigo produzido teve um autor e dois ou mais coautores por artigo, sem, contudo, haver interligações de rede de coautoria com outros clusters. Nesse mapa não contam nenhum autor com produção isolada de publicação, portanto, infere-se que pelo menos em cada publicação há um coautor. Verifica-se que os dados da relação de produção dos artigos ficaram na margem temporal de 2016 a 2020. Da mesma forma que a análise anterior, a formação da relação sem redes de coautoria totalmente interligadas, as análises deram-se com relação de no mínimo um documento.

Para verificar o grau de colaboração entre os coautores, com a finalidade de mensurar a relação de colaboração entre os autores por ano foi utilizada a fórmula proposta por Pandey e Sahoo (2020, p. 378), que permite calcular o grau de colaboração (GC), onde valores igual a 1,0 indica uma colaboração total entre os autores nas produções de artigos científicos levando em consideração os autores únicos e os vários autores que tiveram colaboração.

Grau de Colaboração (GC) = $N_m / (N_m + N_s)$, onde: N_m = Número de artigos científicos com vários coautores (múltiplos); N_s = Número de artigos científicos com um único autor (simples).

A tabela 2 representa os dados coletados para o *corpus* de 5 artigos sobre Web Semântica e Linked Data no período entre 2016 e 2020. Verifica-se que todos os anos obtiveram para GC = 1.000, assim, indicando que houve 100% do total de colaboração entre

autores e coautores, visto que não ocorreu nenhuma produção de artigo de forma isolada (único autor), o que vem a reforçar o mapa de rede já demonstrado.

Tabela 2 – Grau de colaboração de autores de produção científica entre 2016 e 2020 sobre os temas Web Semântica e Linked Data - *Scielo* Brasil.

ANO	NS	NM	NS + NM	GRAU DE COLABORAÇÃO (GC)
2016	0	1	1	1,0000
2018	0	1	1	1,0000
2019	0	1	1	1,0000
2020	0	2	2	1,0000
MÉDIA DE GC				1.0000

Fonte: Dados de Pesquisa (2022). Adaptado de Pandey e Sahoo (2020, p. 378).

Apesar da média de GC ser 1.000, ou seja, 100% de colaboração (cada artigo teve pelo menos um autor e um coautor) a quantidade de artigos publicados é baixa. Tendo em vista, os números denotarem alta colaboração entre os pesquisadores, frequentemente, há produção de artigos de uma única autoria, logo, presume-se que os valores obtidos de 100% deram-se em virtude do quantitativo de artigos recuperados no critério de consulta (*query*) ter sido baixo.

Quanto aos aspectos de números de citações dos artigos, as informações levantadas (Quadro 6), indicam um baixo número dentro do período avaliado de 2016 a 2021 entre os cinco artigos. Fato que pode demonstrar baixa publicidade na produção nacional sobre os temas da pesquisa. Apesar de haver publicação em revista *qualis* A1, mesmo assim, não houve uma projeção quanto ao nível de citação. Acrescenta-se que o ano em houve a citação de um dos artigos foi de 2020, entretanto, a publicação deste deu-se em 2016 havendo um grande intervalo de tempo até a citação ocorrer.

Quadro 6 – Quantidade de citações por artigo – *Scielo* Brasil.

Título do artigo	Periódico	Ano de publicação	Quantidade de Citações
Um protótipo Linked data para catalogação semântica de publicações.	Perspectivas em Ciência da Informação.	2016	1

Fonte: Dados de Pesquisa (2022).

Quanto aos locais de pesquisa apresentam-se as instituições cujos autores estão vinculados e receberam algum tipo de apoio à pesquisa e publicação dos artigos, de acordo com os registros recuperados na base de dados da *Scielo* Brasil.

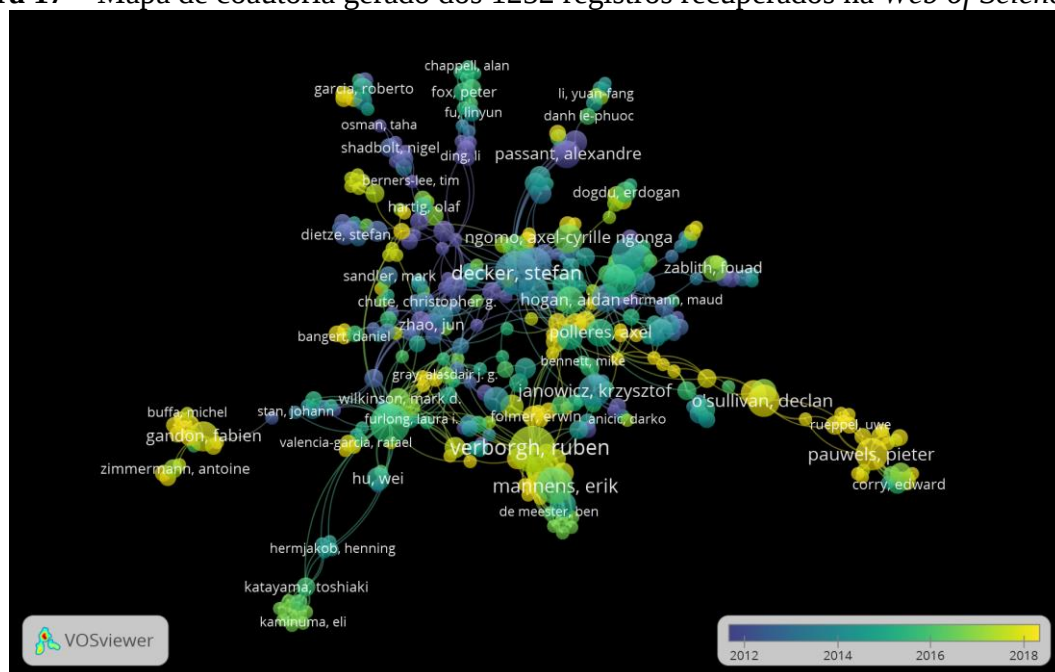
1. **Artigo:** Um protótipo Linked data para catalogação semântica de publicações; **Autor principal:** Freitas Junior, Nilton; **Instituição de Vínculo:** Faculdade de Minas, Brasil.
2. **Artigo:** Recursos audiovisuais na Web de Dados: a construção do Portal Semântico Audiovisual; **Autor principal:** Coneglian, Caio Saraiva **Instituição de Vínculo:** Universidade Estadual Paulista, Brasil.
3. **Artigo:** Um método para publicação semântica Linked Data de bases de dados convencionais e um estudo de caso real de artigos acadêmicos; **Autor principal:** Gonçalves, Adriano de Oliveira **Instituição de Vínculo:** Universidade Federal do Rio de Janeiro, Brasil.
4. **Artigo:** Um método para criação de tesouros semânticos de requisitos usando a ontologia SKOS; **Autor principal:** Zacarias, Rodrigo Oliveira; **Instituição de Vínculo:** Instituto Federal de Educação, Ciência e Tecnologia Fluminense, Brasil.
5. **Artigo:** Methodology for The Development of an Ontology Network on The Brazilian National System for the Evaluation of higher Education (OntoSINAES); **Autor principal:** Sila, Cleiton; **principal:** **Instituição de Vínculo:** Instituto Federal Goiano, Brasil.

Verifica-se que do total de instituições vínculos dos autores uma foi Universidades Federal; uma Faculdade; uma Universidade Estadual; dois Institutos Federais. Dessa forma, sendo diversificadas as instituições de pesquisa em termos de vínculos dos autores principais dos artigos.

7.1.2 Das publicações recuperadas na base de dados da *Web of Science*

Para a coleta de dados na *Web of Science*, inicialmente, realizou-se uma *query* mais abrangente, tendo o retorno de 1.252 registros. Diante do resultado, como modo informativo, foi criado um mapa de coautoria para demonstrar a complexidade de interligações dos nós com 1.252 registros. Os nós foram gerados no software VOSviewer com no mínimo de dois documentos por autores (Figura 17).

Figura 17 – Mapa de coautoria gerado dos 1252 registros recuperados na *Web of Science*.



Fonte: Dados de Pesquisa (2022). Uso do software VOSviewer.

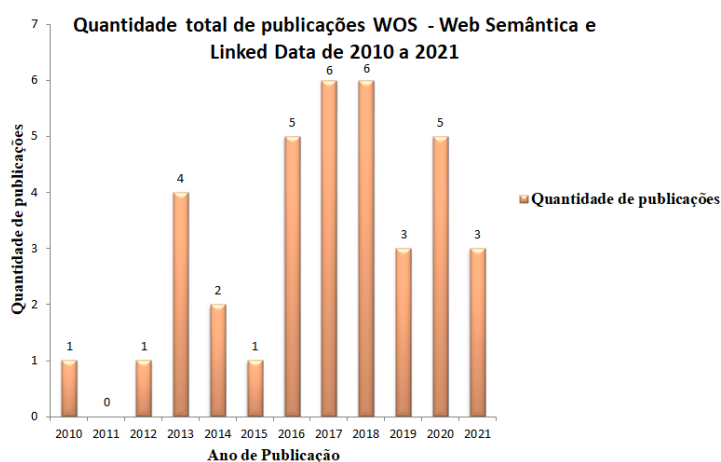
O quantitativo de registros recuperados tornaria inviável a coleta e leitura dos artigos completos. Portanto, refinamentos foram realizados. Primeiramente, foi excluído o ano de 2022, pois no ano vigente os registros na base de dados não estão completos. Em seguida, foram excluídos os documentos de acesso antecipado. Posteriormente, foram selecionados os documentos que fossem *Open Access* e artigos científicos completos. Por último, a concentração de área de pesquisa *Information Science Library Science*. A consulta deu-se por tópicos, ou seja, título do artigo, resumo (*abstract*) e palavras-chave do autor.

Esse processo resultou, ao final, em 40 registros recuperados, nos quais foram feitas leituras como seleção para os mais indicados e de maior relevância, três artigos foram eliminados, restando 37 artigos para formar o *corpus* da pesquisa. Tais eliminações deram-se em virtude de um artigo ser mais direcionado para projetos de dados abertos, e dois artigos com conteúdos vagamente abordados para o contexto desse trabalho.

Assim, foi possível averiguar a evolução da produção científica, entre 2010 e 2021 referentes aos temas propostos. De acordo com as informações coletadas na base de dados da *Web of Science*, verificaram-se quais os periódicos escolhidos para publicações, além de identificar a autoria dos artigos, quais os vínculos institucionais, a rede de coautoria, bem como os anos de publicações e o número de citações de cada publicação.

Em análise sobre a produtividade de publicações dos autores por ano, de acordo com os dados no gráfico 2, observa-se que há oscilações em quantitativo de produção científica sobre os temas. Apenas nos períodos de 2012 a 2013, 2016 a 2017 e 2019 a 2020 houve crescimentos contínuos. Em outros anos a produção aumentou, mas houve queda no ano seguinte. Os anos de maiores publicações foram 2017 e 2018, enquanto o ano de 2011 não ocorreu produção de artigo. No ano de 2021, a produção foi menor que os anos com maiores produções.

Gráfico 2 – Quantidade de artigos publicados na Base de Dados da *Web of Science* sobre os temas: Web Semântica e Linked Data entre 2010 e 2021.



Fonte: Dados de Pesquisa (2022).

A média geral de produtividade no período foi de 3,08 artigos publicados/ano, considerada uma média baixa de publicações na área de pesquisa analisada. Logo, há a necessidade de uma exploração maior sobre o tema, em especial, sobre o prisma da Ciência da Informação, em virtude de que as formas de informação modificam e, na Internet, tudo ocorre de modo muito dinâmico. A Web Semântica e Linked Data são uma realidade, de certo que em fase de consolidação e amadurecimento. Contudo, os pesquisadores não devem deixá-la de lado em suas publicações.

Quanto aos locais de publicações dos artigos científicos, as escolhas ficaram restritas entre 20 periódicos nos períodos de 2010 a 2021 (Tabela 3 e Quadro 7). O periódico que teve maior número de publicações (6 artigos), foi P10 - *Information Technology and Libraries* com 16,22% do total dos 37 artigos. Em segunda colocação está o periódico Transinformação com 5 artigos que equivale a 13,51%, os demais periódicos obtiveram percentuais entre 2,70% e 8,11% (de 2 a 3 artigos respectivamente) (Gráfico 3).

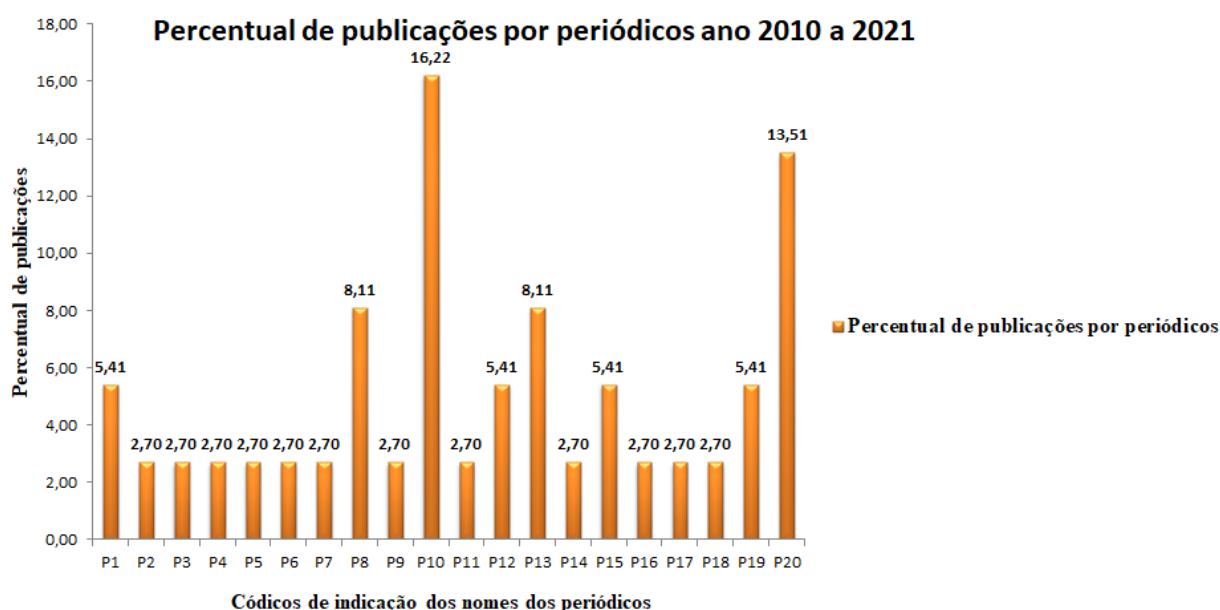
Tabela 3 – Quantidade de publicações nos periódicos entre 2010 e 2021 referente a Web Semântica e Linked Data – *Web of Scince*.

INDICADOR	PERIÓDICOS	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	TOTAL	%
P1	ANALES DE DOCUMENTACION	0	0	0	0	0	0	0	0	1	1	0	0	2	5,41
P2	ASLIB JOURNAL OF INFORMATION MANAGEMENT	0	0	0	0	1	0	0	0	0	0	0	0	1	2,70
P3	ASLIB PROCEEDINGS	1	0	0	0	0	0	0	0	0	0	0	0	1	2,70
P4	ATOZ	0	0	0	0	0	0	0	0	0	0	1	0	1	2,70
P5	BIBLIOTECAS REVISTA DE LA ESCUELA DE BIBLIOTECOLOGIA DOCUMENTACION E INFORMACION	0	0	0	0	0	0	0	0	0	1	0	0	1	2,70
P6	DOCUMENTATION ET BIBLIOTHEQUES	0	0	0	0	0	0	0	1	0	0	0	0	1	2,70
P7	E CIENCIAS DE LA INFORMACION	0	0	0	0	0	0	0	1	0	0	0	0	1	2,70
P8	EM QUESTAO	0	0	0	0	0	0	0	1	1	0	0	1	3	8,11
P9	ENCONTROS BIBLI REVISTA ELETRONICA DE BIBLIOTECONOMIA E CIENCIA DA INFORMACAO	0	0	0	0	0	0	0	0	1	0	0	0	1	2,70
P10	INFORMATION TECHNOLOGY AND LIBRARIES	0	0	0	0	0	1	2	1	1	0	0	1	6	16,22
P11	INVESTIGACION BIBLIOTECOLOGICA	0	0	0	0	0	0	0	0	0	0	1	0	1	2,70
P12	JOURNAL OF DATA AND INFORMATION SCIENCE	0	0	0	0	0	0	0	0	0	0	1	1	2	5,41
P13	JOURNAL OF INFORMATION SCIENCE	0	0	0	0	0	0	0	1	2	0	0	0	3	8,11
P14	JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY	0	0	0	1	0	0	0	0	0	0	0	0	1	2,70
P15	LIBRARY HI TECH	0	0	2	0	0	0	0	0	0	0	0	0	2	5,41
P16	LIBRARY RESOURCES TECHNICAL SERVICES	0	0	0	0	0	0	0	0	1	0	0	0	1	2,70
P17	MALAYSIAN JOURNAL OF LIBRARY INFORMATION SCIENCE	0	0	0	0	0	0	0	0	1	0	0	0	1	2,70
P18	PROFESIONAL DE LA INFORMACION	0	0	0	1	0	0	0	0	0	0	0	0	1	2,70
P19	REVISTA ESPANOLA DE DOCUMENTACION CIENTIFICA	0	0	0	0	1	0	1	0	0	0	0	0	2	5,41
P20	TRANSINFORMACAO	0	0	1	0	0	0	1	0	0	1	2	0	5	13,51
TOTAL		1	0	3	2	2	1	5	6	6	3	5	3	37	100,00

Fonte: Dados de Pesquisa (2022).

Destaque deve ser dado ao periódico Transinformação que, novamente, aponta entre as melhores nas escolhas para a publicação de artigos devido ao seu alto padrão e *qualis* A1.

Gráfico 3 – Percentual das publicações nos periódicos entre 2010 e 2021 referente a Web Semântica e Linked Data – *Web of Scince*.



Fonte: Dados de Pesquisa (2022).

Verifica-se que a maioria dos periódicos (no total de 12) teve apenas duas publicações nesse período, fato que aponta baixa produtividade. Um dos fatores que pode ser apontado é quanto ao custo, visto que, para cobrir as diversas etapas de publicações, além de taxas adicionais, em periódicos internacionais torna-se bastante oneroso. De acordo com Nassi-Caiò (2021), a publicação de um artigo em periódico internacional pode variar seu custo entre US\$ 194.00 a 720.00. Fato que, muitas vezes, torna-se inviável a submissão do artigo. Destarte,

fazem-se necessárias novas políticas para democratização do conhecimento. Nesse sentido, movimentos de acesso aberto às publicações têm direcionado como uma possível solução.

Quadro 7– Periódicos utilizados para publicações - 2010 a 2021 – *Web of Science*.

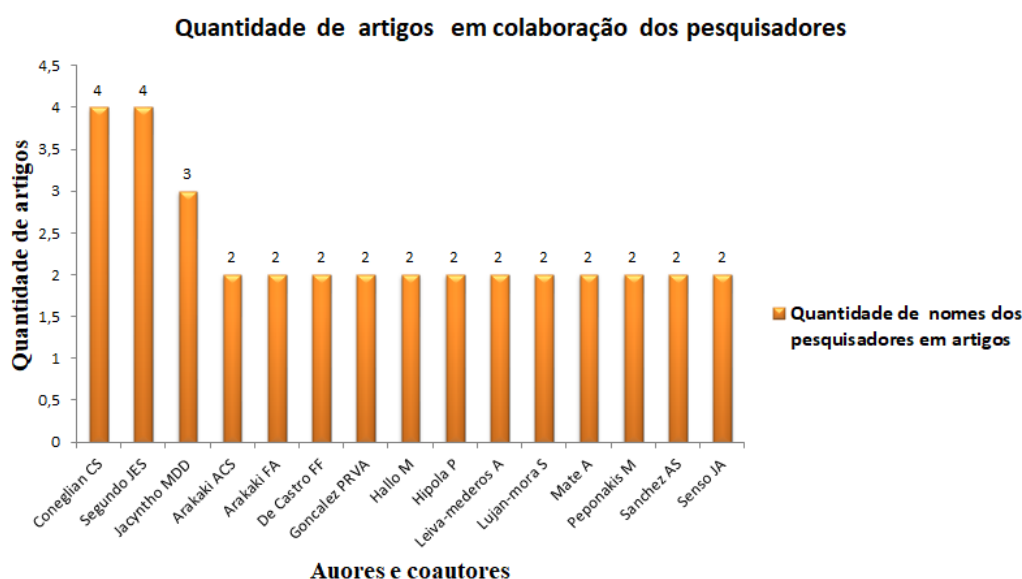
Título do artigo	Periódico
Availability of metadata in linked data for digital repositories; Personal authorities of the librarian community of practice revisited for semantic web.	Anales de documentacion.
User-centered design and evaluation of overview components for semantic data exploration.	Aslib journal of information management.
User-centered design and evaluation of overview components for semantic data exploration.	Aslib proceedings.
COVID-19 Information Panel: SPARQL query at Wikidata.	Atoz novas práticas em informação e conhecimento.
The Linked Data and the Discovery of Links in the Bibliographic Universe.	Bibliotecas revista de la escuela de bibliotecologia documentacion e informacion.
The Stages in the Development of an Open Data Project Linked to the Library.	Documentation et bibliotheques.
Knowledge Organization Systems: Definition and Historical Development.	E-ciencias de la informacion.
Audiovisual and Linked Data: a study of the DBpedia and LMDb databases; The library catalogue and the linked data; The impact of the IFLA Library Reference Model on cataloging practice: use cases, advantages and disadvantages.	Em questão.
Audiovisual and Linked Data: a study of the DBpedia and LMDb databases.	Encontros bibli revista eletrônica de biblioteconomia e ciência da informação.
Transitioning from XML to RDF: Considerations for an Effective Move Towards Linked Data and the Semantic Web; Bridging the Gap Using Linked Data to Improve Discoverability and Diversity in Digital Collections; Linked Data in Libraries: A Case Study of Harvesting and Sharing Bibliographic Metadata with BIBFRAME; A Case Study on the Path to Resource Discovery; Efficiently Processing and Storing Library Linked Data using Apache Spark and Parquet; In the Name of the Name: RDF Literals, ER Attributes, and the Potential to Rethink the Structures and Visualizations of Catalogs.	Information technology and libraries.
Transitioning from XML to RDF: Considerations for an Effective Move Towards Linked Data and the Semantic Web	Investigacion bibliotecologica.
Improving Archival Records and Service of Traditional Korean Performing Arts in a Semantic Web Environment; RDFAdaptor: Efficient ETL Plugins for RDF Data Process.	Journal of data and information Science.
Current state of Linked Data in digital libraries; Evaluating open access journals using Semantic Web technologies and scorecards; SKOS concepts and natural language concepts: An analysis of latent relationships in KOSs.	Journal of information Science.
Current state of Linked Data in digital libraries.	Journal of the american society for information science and technology.
AUTHORIS: a tool for authority control in the semantic web;	Library hi tech.

datos.bne.es and MARiMbA: an insight into library linked data.	
Continuing Education in New Standards and Technologies for the Organization of Data and Information A Report on the Cataloging and Metadata Professional Development Survey.	Library resources technical services.
Adoption of linked data technologies among university librarians in Pakistan: Challenges and prospects.	Malaysian journal of library information Science.
Automatic semantic markup in content management systems: integration and quantification.	Profesional de la informacion.
Challenges and opportunities for knowledge organization at the intersection with information Technologies; Implementation of the RDF data model in digital collections of Spanish libraries, archives and museums.	Revista espanola de documentacion cientifica.
A method for Linked Data semantic publishing of conventional databases and a real case study of academic papers; Use of Linked Data principles for semantic management of scanned documents; Audiovisual resources in the Data Web: The construction of the Semantic Audiovisual Portal; Descriptive representation in the semantic web contexto; A method for creating semantic thesauri of requirements using the SKOS ontology.	Transinformação.

Fonte: Dados de Pesquisa (2022).

Sobre o aspecto da relação entre autores e coautores, com base nos registros de publicações, contabilizaram-se 93 nomes entre autores e coautores nos artigos. Os pesquisadores em que seus nomes constam em maior quantidade de artigos foram: Coneglian C. S. e Segundo J. E. S. com quatro artigos cada um, em seguida Jacyntho M. D. D., em três artigos. Doze nomes de pesquisadores foram contabilizados com a participação em dois artigos cada um. Os demais 78 pesquisadores apareceram uma única vez dentro de 37 artigos. No gráfico 4 são mostrados os nomes dos autores que participaram de dois a quatro artigos em colaboração.

Gráfico 4 – Quantidade de autores com maior colaboração na produção de artigos sobre Web Semântica e Linked Data – *Web of Science*.



Fonte: Dados de Pesquisa (2022).

Percebe-se que os autores Coneglian C. S., Segundo J. E. S. e Jacyntho M. D. D. participam de uma rede mais expressiva de colaboração com os demais pesquisadores, enquanto a maioria dos pesquisadores só participou em colaboração em um só artigo, e cinco realizaram publicações individuais (Quadro 8).

Quadro 8 – Autores com publicação individualizada.

AUTORES	TÍTULO DA PUBLICAÇÃO
ST-GERMAIN, Marielle	The Stages in the Development of an Open Data Project Linked to the Library.
FRIPP, Dominic	Using linked data to classify web documents.
THARANI, Karim	Linked Data in Libraries: A Case Study of Harvesting and Sharing Bibliographic Metadata with BIBFRAME.
GUAY, Beth	A Case Study on the Path to Resource Discovery.
PEPONAKIS, Manolis	In the Name of the Name: RDF Literals, ER Attributes, and the Potential to Rethink the Structures and Visualizations of Catalogs.

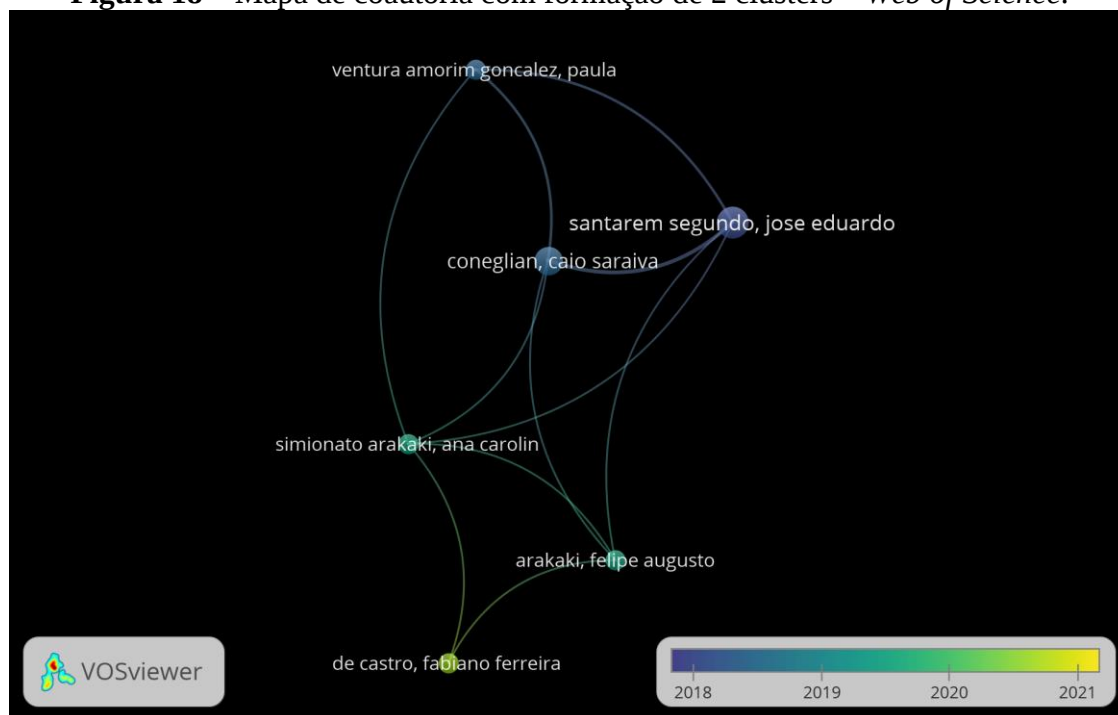
Fonte: Dados de Pesquisa (2022).

Portanto, existe uma forte relação de colaboração no aspecto de produção de artigos, mas essa relação está direcionada em formação de grupos de pesquisa.

Por meio do mapa de rede de interações entre autores e coautores, apresenta-se na figura 18 o mapa de rede de colaboração totalmente interligada. Sendo gerado no VOSviewer,

esclarece-se que para a formação da relação em redes de coautoria as análises deram-se com relação de nó mínimo com dois documentos.

Figura 18 – Mapa de coautoria com formação de 2 clusters – *Web of Science*.



Fonte: Dados de Pesquisa (2022). Uso do software VOSviewer.

No mapa de rede de interações totalmente interligadas, percebe-se a formação de dois agrupamentos (clusters). O primeiro cluster é formado por Caio Saraiva Coneglian, José Eduardo Santarem Segundo e por Paula Ventura Amorim Gonçalves.

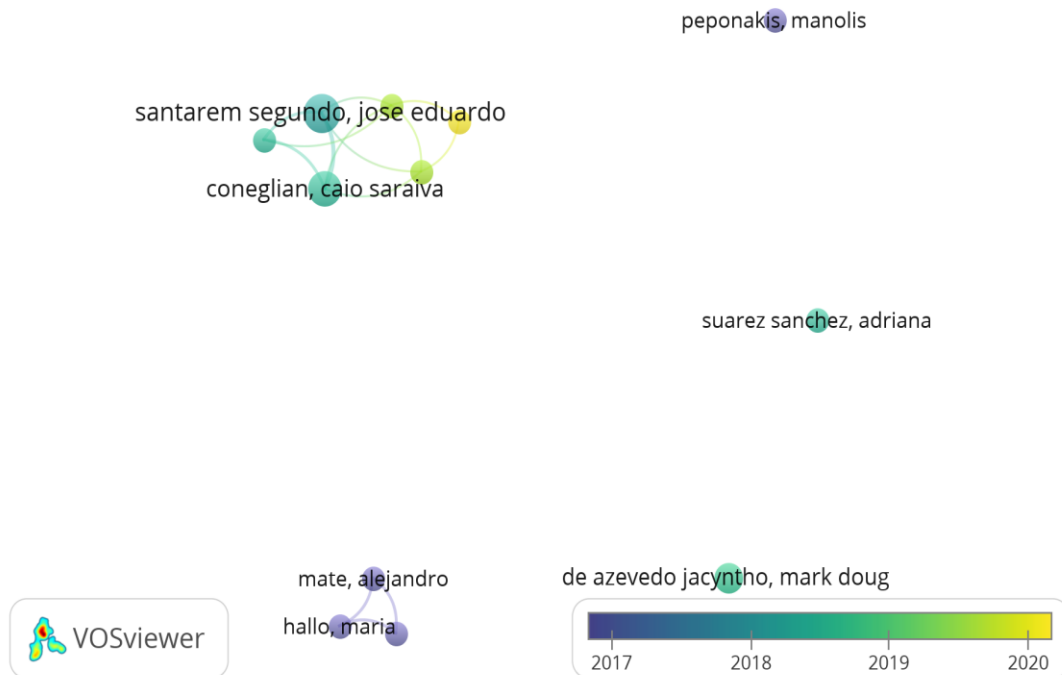
O segundo cluster foi formado por Ana Carolin Simionato Arakaki, Felipe Augusto Arakaki e Fabiano Ferreira de Castro. Como apenas dois agrupamentos foram formados, isso indica uma forte relação entre esses pesquisadores dentro do cluster, mas uma fraca relação de colaboração em rede de coautores com os demais pesquisadores.

Verifica-se que Caio Saraiva Coneglian e José Eduardo Santarem Segundo destacaram-se dentro da rede de colaboração. Outro ponto é que esses clusters iniciaram a colaboração entre si a partir do período de 2018, sendo recentes as publicações dos artigos.

A figura 19 mostra o mapa dos autores sem compor uma rede de interações totalmente interligada. Este indica a formação de seis clusters. O primeiro cluster é composto por Jose Eduardo Santarem Segundo; Caio Saraiva Coneglin e Paula Ventura Amorin Gonçalves. O segundo cluster, próximo ao primeiro, é composto por Ana Carolina Simionato Arakaki; Fabiano Ferreira de Castro e Felipe Augusto Arakaki. O terceiro cluster é composto por Maria

Hallo; Alejandro Mate e Sérgio Lujan-Mora. Nesses três primeiros clusters, os pesquisadores realizam colaboração na produção de artigos. Os clusters 4, 5 e 6 são formados por pesquisadores isolados. Foi levada em consideração a relação de, no mínimo, dois documentos por autor para a geração do mapa.

Figura 19 – Mapa de coautores sem a rede de interligações totalmente interligada de seis clusters.



Fonte: Dados de Pesquisa (2022). Uso do software VOSviewer

Diferente do primeiro mapa, o período temporal de publicação dos artigos inicia-se a partir de 2017. Como não foi levada em conta a rede totalmente interligada, os autores com publicação isolada são representados, dessa forma, o autor Manolis Peponakis teve o artigo publicado em 2017, de acordo com a representação da faixa de cores dos nós em intervalo temporal localizada na parte inferior da figura.

Conforme realizado nas análises do grau de colaboração da produção nacional, apresenta-se o grau de colaboração (GC) dos dados referentes à *Web of Science*. A tabela 4 representa os dados coletados para o *corpus* de 37 artigos sobre Web Semântica e Linked Data no período entre 2010 e 2021.

Tabela 4 – Grau de colaboração de autores de produção científica entre 2010 e 2021 sobre os temas Web Semântica e Linked Data – *Web of Science*.

ANO	NS	NM	NS + NM	GRAU DE COLABORAÇÃO (GC)
2010	1	0	1	0,0000
2012	0	1	1	1,0000
2013	0	4	4	1,0000
2014	0	2	2	1,0000
2015	1	0	1	0,0000
2016	2	3	5	0,6000
2017	3	3	6	0,5000
2018	0	6	6	1,0000
2019	1	2	3	0,6667
2020	1	4	5	0,8000
2021	0	3	3	1,0000
MÉDIA DE GC				0,8667

Fonte: Dados de Pesquisa (2022). Adaptado de Pandey e Sahoo (2020, p. 378).

Nos anos de 2010 e 2015, o GC = 0.0000 indica que nestes não houve a produção de artigos de forma conjunta (autor e coautor), apenas, produção individual de cada pesquisador. No ano de 2017 foram produzidos seis artigos, sendo três artigos de forma individualizada e três artigos compostos por colaboradores, o que resultou no GC = 0.5000, ou seja, 50%.

Nos anos de 2016, 2019 e 2020 os GC estiveram entre 0.6000 a 0.8000 indicando oscilações entre publicação de artigos de forma individual e produção de artigos com colaboradores. Esclarece-se que no ano de 2020 houve um total de cinco artigos publicados e, apenas, um desse de forma individualizada; o restante se deu em colaboração. Logo o GC = 0.8000 tendeu para 1.0000, isso ocorre quando há 100% da colaboração nas publicações de artigos. Fato também percebido nos anos de 2012, 2013, 2014, 2018 e 2021.

Outro ponto a considerar é que não houve produção de artigos no ano de 2011, pois não ocorreram registros de artigos na coleta de dados. A média de GC nesse período de 2010 a 2021 foi de 0,8667, o que indica uma tendência para maior número de coautores, consequentemente, uma menor produção de artigos de forma individual. Portanto, o GC vem a confirmar os dados apontados anteriormente, com oscilações de produção de artigos, formação de clusters e algumas publicações de forma isolada.

Isto posto, podemos inferir que a produção científica com coautores direciona-se para uma maior colaboração na produção de artigos científicos no âmbito internacional. Em nossa visão, apesar dessa tendência colaborativa, pode ocorrer a formação de clusters setorializados dentro dos centros de pesquisas e instituições de ensino, sem que haja uma interligação em

rede de compartilhamento para fora desses ambientes. Nesse sentido, o fluxo da informação científica será mais propagado dentro desses clusters e restrito a outros.

Quanto aos aspectos de números de citações dos artigos, os autores que receberam maiores números de citações (consideradas a partir de três) estão representado no Quadro 9, enquanto os títulos dos artigos descritos na ordem dos respectivos autores abaixo do quadro.

Quadro 9 – Autores mais citados sobre os temas Web Semântica e Linked Data entre os anos de 2010 a 2021 na *Web of Science*.

Autores mais citados (primeiro autor)	Total de citações
Hallo, Maria;	35
Van Hooland, Seth;	16
Vila-Suero, Daniel;	12
Leiva-Mederos, Amed;	10
Tharani, Karim.	9
Catarino, Maria Elisabete;	8
Mastora, Anna;	6
Maria Brunetti, Josep;	6
Lisena, Pasquale;	5
Tosaka, Yuji;	3
Warraich, Nosheen Fatima;	3
Hallo, Maria; Lujan-Mora, Sergio;	3
Peponakis, Manolis.	3
Martinez-Avila, Daniel;	3

Fonte: Dados de Pesquisa (2022).

Artigos mais citados:

1. Current state of Linked Data in digital libraries;
2. Evaluating the success of vocabulary reconciliation for cultural heritage collections;
3. datos.bne.es and MARiMbA: an insight into library linked data;
4. AUTHORIS: a tool for authority control in the semantic web;
5. Linked Data in Libraries: A Case Study of Harvesting and Sharing Bibliographic Metadata with BIBFRAME;
6. Descriptive representation in the semantic web contexto;
7. SKOS concepts and natural language concepts: An analysis of latent relationships in KOSs;
8. User-centered design and evaluation of overview components for semantic data exploration;

9. Improving (Re-) Usability of Musical Datasets: An Overview of the DOREMUS Project;
10. Continuing Education in New Standards and Technologies for the Organization of Data and Information A Report on the Cataloging and Metadata Professional Development Survey;
11. Adoption of linked data technologies among university librarians in Pakistan: Challenges and prospects;
12. Evaluating open access journals using Semantic Web technologies and scorecards;
13. In the Name of the Name: RDF Literals, ER Attributes, and the Potential to Rethink the Structures and Visualizations of Catalogs;
14. Challenges and opportunities for knowledge organization at the intersection with information Technologies;

Verifica-se que a autora Maria Hallo, obteve o maior número de citações com seu artigo “*Current state of Linked Data in digital libraries*”, publicado em 2016, no periódico “*Journal of Information Science*”; sendo indicado como um ponto de referência para aprofundamento e estudos.

Quanto aos locais de pesquisa apresentam-se as instituições cujos autores estão vinculados e receberam algum tipo de apoio à pesquisa para publicações dos artigos, de acordo com os 37 registros recuperados na base de dados da *Web of Science*.

1. **Artigo:** Audiovisual and Linked Data: a study of the DBpedia and LMDb databases; **Autor principal:** Simionato, Ana Carolina **Instituição de Vínculo:** Univ Federal São Carlos, São Carlos, SP, Brazil.
2. **Artigo:** The library catalogue and the linked data; **Autor principal:** Serra, Liliana Giusti; **Instituição de Vínculo:** Univ Estadual Paulista, Sao Paulo, SP, Brazil.
3. **Artigo:** Transitioning from XML to RDF: Considerations for an Effective Move Towards Linked Data and the Semantic Web; **Autor principal:** Hardesty, Juliet L. **Instituição de Vínculo:** Indiana Univ Lib, Bloomington, IN USA.
4. **Artigo:** A method for Linked Data semantic publishing of conventional databases and a real case study of academic paper; **Autor principal:** Gonçalves, Adriano de Oliveira **Instituição de Vínculo:** Univ Fed Rio de Janeiro, Superintendencia Tecnol Informacao & Comunicacao, Macae, RJ, Brazil.
5. **Artigo:** The Linked Data and the Discovery of Links in the Bibliographic Universe **Autor principal:** Avila Barrientos, Eder; **Instituição de Vínculo:** Univ Nacl Autonoma Mexico, Posgrad Bibliotecol & Estudios Informac, Mexico City, DF, Mexico.
6. **Artigo:** Current state of Linked Data in digital libraries; **Autor principal:** Hallo, Maria; **Instituição de Vínculo:** Natl Polytech Sch, Isabel Catolica E11-253, Ecuador, Quito 01, Ecuador.

7. **Artigo:** The role of semantic web in the big data process; **Autor principal:** Coneglian, Caio Saraiva; **Instituição de Vínculo:** Univ Estadual Paulista Unesp, Ciencia Informacao, Ilha Solteira, Brazil.
8. **Artigo:** Evaluating open access journals using Semantic Web technologies and scorecards; **Autor principal:** Hallo, Maria; **Instituição de Vínculo:** Natl Polytech Sch, Isabel Catolica E11-253,, Quito, Ecuador.
9. **Artigo:** Challenges and opportunities for knowledge organization at the intersection with information Technologies; **Autor principal:** Martinez-Avila, Daniel ; **Instituição de Vínculo:** Sao Paulo State Univ UNESP, Informat Org Res Grp FAPOI, Sao Paulo, Brazil.
10. **Artigo:** Use of Linked Data principles for semantic management of scanned documents; **Autor principal:** Pessanha Monteiro, Luciane Lena; **Instituição de Vínculo:** Univ Candido Mendes, Dept Comp, Coordenacao Comp, Campos Dos Goytacazes, RJ, Brazil.
11. **Artigo:** Availability of metadata in linked data for digital repositories **Autor principal:** Arakaki, Felipe Augusto **Instituição de Vínculo:** Univ Brasilia, Campus Univ Darcy Ribeiro, Fac Ciencia Informacao, Brasilia, DF, Brazil.
12. **Artigo:** Audiovisual resources in the Data Web: The construction of the Semantic Audiovisual Portal; **Autor principal:** Coneglian, Caio Saraiva; **Instituição de Vínculo:** Univ Fed Sao Carlos, Ctr Educ & Ciencias Humanas, Programa Posgrad Ciencia Informacao.
13. **Artigo:** The Stages in the Development of an Open Data Project Linked to the Library; **Autor principal:** St-Germain, Marielle **Instituição de Vínculo:** Univ Montreal, EBSI, Montreal, PQ, Canada.
14. **Artigo:** Descriptive representation in the semantic web contexto; **Autor principal:** Catarino, Maria Elisabete; **Instituição de Vínculo:** Univ Estadual Londrina, Dept Ciencia Informacao, Londrina, PR, Brazil.
15. **Artigo:** Improving Archival Records and Service of Traditional Korean Performing Arts in a Semantic Web Environment; **Autor principal:** Park, Ziyong; **Instituição de Vínculo:** Hansung Univ, Dept Lib & Informat Sci, Seoul, South Korea.
16. **Artigo:** Personal authorities of the librarian community of practice revisited for semantic web **Autor principal:** Garzon-Farinos, Fernanda **Instituição de Vínculo:** Univ Catolica Valencia San Vicente Martir, Valencia, Spain.
17. **Artigo:** AUTHORIS: a tool for authority control in the semantic web; **Autor principal:** Leiva-Mederos, Amed **Instituição de Vínculo:** Univ Granada, Dept Informac & Comunicac, Granada, Spain.
18. **Artigo:** Using linked data to classify web documents; **Autor principal:** Fripp, Dominic; **Instituição de Vínculo:** Univ W England, Dept Informat Sci & Digital Media, Bristol, Avon, England.
19. **Artigo:** Continuing Education in New Standards and Technologies for the rganization of Data and Information A Report on the Cataloging and Metadata Professional Development Survey; **Autor principal:** Tosaka, Yuji; **Instituição de Vínculo:** Coll New Jersey, Ewing, USA.
20. **Artigo:** Adoption of linked data technologies among university librarians in Pakistan: Challenges and prospects; **Autor principal:** Warraich, Nosheen Fatima; **Instituição de Vínculo:** Univ Punjab, Dept Informat Management, Lahore, Pakista.

21. **Artigo:** datos.bne.es and MARiMbA: an insight into library linked data **Autor principal:** Vila-Suero, Daniel; **Instituição de Vínculo:** Univ Politecn Madrid, Ontol Engn Grp, Madrid, Spain.
22. **Artigo:** Bridging the Gap Using Linked Data to Improve Discoverability and Diversity in Digital Collections; **Autor principal:** Boczar, Jason; **Instituição de Vínculo:** Univ S Florida, Tampa, USA.
23. **Artigo:** Linked Data in Libraries: A Case Study of Harvesting and Sharing Bibliographic Metadata with BIBFRAME **Autor principal:** Tharani, Karim **Instituição de Vínculo:** Univ Saskatchewan, Saskatoon, SK, Canada.
24. **Artigo:** Management of archival materials with Linked Data and federated queries; **Autor principal:** Hidalgo-Delgado, Yusniel; **Instituição de Vínculo:** Univ Ciencias Informat, Dept Programac, Havana, Cuba.
25. **Artigo:** SKOS concepts and natural language concepts: An analysis of latent relationships in KOSs; **Autor principal:** Mastora, Anna; **Instituição de Vínculo:** Ionian Univ, Fac Informat Sci & Informat, Corfu, Greece.
26. **Artigo:** Implementation of the RDF data model in digital collections of Spanish libraries, archives and museums ; **Autor principal:** Sule, Andreu; **Instituição de Vínculo:** Univ Barcelona, Dept Bibliotecon & Documentac, Barcelona, Spain.
27. **Artigo:** A Case Study on the Path to Resource Discovery **Autor principal:** Guay, Beth; **Instituição de Vínculo:** Univ Maryland, Univ Maryland Lib, College Pk, MD USA.
28. **Artigo:** The impact of the IFLA Library Reference Model on cataloging practice: use cases, advantages and disadvantages **Autor principal:** Alves de Oliveira, Rhuan Henrique **Instituição de Vínculo:** Univ Fed Sao Carlos, Sao Carlos, SP, Brazil.
29. **Artigo:** User-centered design and evaluation of overview components for semantic data exploration; **Autor principal:** Maria Brunetti, Josep **Instituição de Vínculo:** Univ Lleida, Dept Comp Sci & Engn, Lleida, Spain.
30. **Artigo:** Ontologies: analysis about their implementations in library science **Autor principal:** Suarez Sanchez, Adriana; **Instituição de Vínculo:** Univ Nacl Autonoma Mexico, Inst Invest Bibliotecol & Informac, Mexico City, DF, Mexico.
31. **Artigo:** Efficiently Processing and Storing Library Linked Data using Apache Spark and Parquet; **Autor principal:** Sharma, Kumar; **Instituição de Vínculo:** Univ Kalyani, Dept Comp Sci & Engn, Kalyani, W Bengal, India.
32. **Artigo:** Automatic semantic markup in content management systems: integration and quantification; **Autor principal:** Pastor-Sanchez, Juan-Antonio; **Instituição de Vínculo:** Univ Murcia, Fac Comunicac & Documentac, Murcia, Spain.
33. **Artigo:** Knowledge Organization Systems: Definition and Historical Development; **Autor principal:** Suarez Sanchez, Adriana; **Instituição de Vínculo:** Univ Nacl Autonoma Mexico, Posgrad Bibliotecol & Estudios Informac, Mexico City, DF, Mexico.
34. **Artigo:** In the Name of the Name: RDF Literals, ER Attributes, and the Potential to Rethink the Structures and Visualizations of Catalogs; **Autor principal:** Peponakis, Manolis. **Instituição de Vínculo:** Natl Hellen Res Fdn, Natl Documentat Ctr, Athens, Greece.
35. **Artigo:** COVID-19 Information Panel: SPARQL query at Wikidata; **Autor principal:** Simionato Arakaki, Ana Carolina; **Instituição de Vínculo:** Univ Fed Sao Carlos UFSCar, Sao Paulo, Brazil.

36. **Artigo:** A method for creating semantic thesauri of requirements using the SKOS; ontology **Autor principal:** Zacarias, Rodrigo Oliveira **Instituição de Vínculo:** Inst Fed Educ Ciencia & Tecnol Fluminense, Campos Dos Goytacazes, RJ, Brazil.
37. **Artigo:** RDFAdaptor: Efficient ETL Plugins for RDF Data Process; **Autor principal:** Li, Jiao; **Instituição de Vínculo:** CAAS, Agr Informat Inst, Beijing Peoples R China.

Nota-se que a maioria das instituições de vínculos dos autores está no ambiente acadêmico, sendo que 31 são universidades, correspondendo a 83,78% do total de instituições. Portanto, nas academias estão as fontes de pesquisa, produção científica e de apoio para os pesquisadores. Dessa forma, iniciativas de apoio financeiro às instituições universitárias são de fundamental importância para aprofundar estudos e disponibilizar novos recursos de acesso à informação para a sociedade.

7.2 ASPECTOS QUALITATIVOS

Para as análises qualitativas, foram tomadas como aportes as informações do *corpus* coletado nas bases de dados *Scielo* e *Web of Science*, bem como as informações já projetadas nas análises quantitativas. Logo, para verificação dos pontos referentes aos dados qualitativos, inicialmente, realizaram-se as análises de conteúdos dos artigos com a finalidade de criar agrupamentos por categorias temáticas. Nesse sentido, foram identificados os principais aspectos abordados em cada artigo, pela leitura do resumo, título e palavras-chave, possibilitando, assim, gerar uma listagem com pontos-chave.

Posteriormente, essa listagem serviu como base para averiguar os conteúdos semelhantes com outros artigos. Dessa maneira, permitiram ter-se 13 categorias temáticas (Quadro 10), tais categorias, para facilitar a identificação, foram representados em códigos (CT01 a CT13), onde as letras “C” da inicial de categoria, “T” da inicial de temática e 01 a 13 a classificação ordinal.

Quadro 10 – Categorias de classificação temática.

Código	Classificação temática
CT01	Prototipação para uso nos princípios da Web Semântica e Linked Data.
CT02	Conversão e compartilhamento de Bases de Dados para padrões da Web Semântica e Linked Data.
CT03	Reuso de informações e engenharia de requisitos nos padrões da Web Semântica e Linked Data.
CT04	Métodos de rede de ontologias semânticas.
CT05	Estudos de Linked Data para aplicação em recursos audiovisuais.
CT06	Uso dos princípios da Web Semântica e Linked Data em catalogação bibliográfica.
CT07	Organização da Informação nos padrões da Web Semântica e Linked Data.
CT08	Publicação de dados para uso dentro dos padrões SPARQL nos Sistemas de Consulta.
CT09	Métodos para enriquecimento de dados semânticos.
CT10	Classificação facetada e anotações de documentos nos padrões da Web Semântica.
CT11	Interconexão de matadados nos padrões da Web Semântica.
CT12	Tecnologias para incorporação de marcação semântica.
CT13	Estudos de Caso sobre Web Semântica e Linked Data.

Fonte: Dados de Pesquisa (2022).

Em seguida, os registros das duas bases de dados foram concatenados em uma única planilha do excel, eliminado os registros redundantes. Foram utilizados os títulos dos artigos no idioma inglês como forma de padronização dos registros coletados na *Web of Science*. Também ocorreram adições dos campos de código de classificação temático e de atribuição do nome das bases de dados (*Scielo* ou *WOS*) na planilha do excel. Para cada artigo do *corpus* foi adicionado o respectivo código de classificação temática após as apreciações textuais. Facilitando, dessa forma, as análises qualitativas e inferências. Como ponto inicial, no quadro 11, são apontadas as classificações temáticas dos artigos, com seus títulos e o quantitativo de cada área temática.

Quadro 11 – Classificação dos títulos de artigos de acordo com as classes temáticas.

Código Temático	Título dos artigos	Quantidade
CT01	A Linked Data prototype for semantic cataloguing of publications; Audiovisual resources in the Data Web: The construction of the Semantic Audiovisual Portal; Use of Linked Data principles for semantic management of scanned documents; Efficiently Processing and Storing Library Linked Data using Apache Spark and Parquet.	04
CT02	A method for Linked Data semantic publishing of conventional databases and a real case study of academic papers; Transitioning from XML to RDF: Considerations for an Effective Move Towards Linked Data and the Semantic Web; The role of semantic web in the big data process; Evaluating open access journals using Semantic Web technologies and scorecards; Availability of metadata in linked data for digital repositories;	08

	AUTHORIS: a tool for authority control in the semantic web; Adoption of linked data technologies among university librarians in Pakistan: Challenges and prospects; datos.bne.es and MARiMbA: an insight into library linked data.	
CT03	A method for creating semantic thesauri of requirements using the SKOS ontology; The impact of the IFLA Library Reference Model on cataloging practice: use cases, advantages and disadvantages.	02
CT04	Methodology for The Development of an Ontology Network on The Brazilian National System for the Evaluation of higher Education (OntoSINAES); Current state of Linked Data in digital libraries; Ontologies: analysis about their implementations in library Science.	03
CT05	Audiovisual and Linked Data: a study of the DBpedia and LMDb databases.	01
CT06	The library catalogue and the linked data; Personal authorities of the librarian community of practice revisited for semantic web; In the Name of the Name: RDF Literals, ER Attributes, and the Potential to Rethink the Structures and Visualizations of Catalogs.	03
CT07	The Linked Data and the Discovery of Links in the Bibliographic Universe; Challenges and opportunities for knowledge organization at the intersection with information Technologies; Improving Archival Records and Service of Traditional Korean Performing Arts in a Semantic Web Environment; Continuing Education in New Standards and Technologies for the Organization of Data and Information A Report on the Cataloging and Metadata Professional Development Survey; SKOS concepts and natural language concepts: An analysis of latent relationships in KOSs; Knowledge Organization Systems: Definition and Historical Development.	06
CT08	COVID-19 Information Panel: SPARQL query at Wikidata; RDFAdaptor: Efficient ETL Plugins for RDF Data Process.	02
CT09	Descriptive representation in the semantic web context; Implementation of the RDF data model in digital collections of Spanish libraries, archives and museums.	02
CT10	Using linked data to classify web documents.	01
CT11	User-centered design and evaluation of overview components for semantic data exploration.	01
CT12	Automatic semantic markup in content management systems: integration and quantification.	01
CT13	Bridging the Gap Using Linked Data to Improve Discoverability and Diversity in Digital Collections; Evaluating the success of vocabulary reconciliation for cultural heritage collections; Implementation of the RDF data model in digital collections of Spanish libraries, archives and museums; A Case Study on the Path to Resource Discovery.	04

Fonte: Dados de Pesquisa (2022).

Verifica-se uma tendência dos pesquisadores em publicar artigos sob a temática CT02 - Conversão e compartilhamento de Bases de Dados para padrões da Web Semântica e Linked Data, categoria que totalizou oito artigos. Assim, apontando há necessidade dos estudos referentes à migração de conteúdo informacionais das bases de dados legadas, onde foram projetados no modelo de tabelas relacionais, a exemplo o SQL.

Nesse aspecto, o artigo intitulado *A method for Linked Data semantic publishing of conventional databases and a real case study of academic papers* menciona como realizar o

mapeamento de uma base de dados relacional para o acesso à informação de forma legível, assim, podendo ser acessada por máquinas dentro do padrão proposto para Linked Data. O artigo aponta a necessidade de criação dos links para compartilhamento na LOD. De fato, por meios dos padrões propostos por Berners-Lee, no contexto da Web Semântica, um dos pontos-chave recai sobre o modelo RDF que possibilita colocar semântica nas informações utilizando o conceito de Triplas. Como efeito, as informações poderão ser interligadas e acessadas com uso de URI e conectadas na DBpedia como forma de ampliação de acesso da informação na Internet.

Outro artigo referente à temática é *Transitioning from XML to RDF: Considerations for an Effective Move Towards Linked Data and the Semantic Web*, no qual é direcionado aos aspectos sobre os metadados, em especial, encontrados nas bibliotecas com grande volume de informações que necessitam ser disponibilizadas na Web. Esses metadados estão disponíveis no formato XML, o que vem a facilitar a conversão para o padrão RDF, visto que o XML é uma das bases das camadas da Web Semântica, em conjunto, com RDF e OWL. O artigo destaca os meios de transição dos metadados em XML para os padrões RDF, com uso de ferramentas computacionais para interligação dos dados bibliográficos e disponibilizá-los na Internet. Fato apontado de grande importância que, também, são referenciados nos estudos em outras temáticas nesta pesquisa. Ressalta-se que as pesquisas sobre as estruturas e meios de conversões dos metadados são de extrema importância.

Nesse aspecto, o artigo *Availability of metadata in linked data for digital repositories* faz considerações sobre a necessidade de enriquecimento semântico aos metadados em repositórios digitais, a fim de possibilitar a criação de modelos de conversão para o padrão RDF. Importante ressaltar que para tal ação um dos preceitos é a utilização de ontologias semânticas que possam oferecer significados às informações. Assim, são por meios dessas ontologias que podem ser adicionadas regras e axiomas que irão possibilitar inferências e deduções de dados. Tais ontologias serão disponibilizadas e intercambiadas em forma de URI nos documentos, possibilitando uma estrutura de leitura dos metadados enriquecidos semanticamente.

Dentro desses estudos, as referências adotadas pela W3C descrevem um conjunto de ontologias e destaca a OWL como base para Web Semântica e Linked Data. A estruturação dos metadados com o uso de Dublin Core tornou-se limitada para a representação das informações no ambiente da Internet. As novas tendências tecnológicas que emergem no

campo de acesso à informação, em especial, as de grande volume de dados, em que tudo estará conectado possibilitará novas dimensões de conhecimento.

O artigo *The role of semantic web in the big data process* faz referência ao contexto do *Big Data* e seu potencial como tecnologias convergentes entre a Web Semântica e Linked Data. Dentro dessa visão, pontos primordiais são apontados e deverão ser analisados pelos pesquisadores e profissionais da Ciência da Informação, visando os moldes das novas tendências, como o uso das ontologias e a interoperabilidade proposta pela Web Semântica dentro do ambiente do *Big Data*, além da utilização de aprendizagem de máquinas juntamente com a Inteligência artificial que possibilitará a extração de dados e conversão para os modelos da Web Semântica. Portanto, em análise à temática CT02, pode-se inferir um grande campo de estudo e atuação dos profissionais da informação nos novos contextos que apontam para o uso das tecnologias da Web Semântica e Linked Data.

A segunda categoria que se destaca com maior número de publicações é a CT07 - Organização da Informação nos padrões da Web Semântica e Linked Data, com seis publicações. Esta se relaciona com a temática CT02, haja vista ser um ponto primordial para a estruturação e organização da informação nos princípios da Web Semântica e Linked Data.

O artigo *The Linked Data and the Discovery of Links in the Bibliographic Universe* destaca a necessidade do entendimento dos objetos informacionais dentro do universo bibliográfico, no aspecto de analisar os padrões existentes, atualmente, e como poderão ser organizados para sua utilização na Internet nos princípios da Web Semântica e Linked Data. Nesse sentido, as bibliotecas devem avançar nos estudos da estruturação e da organização da informação.

Fato apontado no artigo *Challenges and opportunities for knowledge organization at the intersection with information Technologies*, onde relata a necessidade dos profissionais que atuam nas bibliotecas em aprofundar estudos referentes aos metadados e a catalogação, haja vista, que esses temas encontram-se em estágios iniciais, ainda, na forma exploratória. Tal observação corrobora com o levantamento bibliográfico feito por essa pesquisa, em que aponta a categoria CT01 - Prototipação para uso nos princípios da Web Semântica e Linked Data, com apenas quatro artigos publicados, ou seja, há pouca produção de ferramentas e de prototipação de recursos e produtos no campo informacional direcionado para as bibliotecas e convergindo para a Web Semântica. Além disso, há necessidade de formação de profissionais qualificados dentro das bibliotecas, demais unidades de informação e instituições como um

todo que possam, efetivamente, realizar trabalhos para as novas realidades propostas para dados vinculados.

Na etapa de análise quantitativa dessa pesquisa foi mencionado o artigo *Current state of Linked Data in digital libraries*, de 2016, do periódico *Journal of Information Science*, cuja autoria é de Maria Hallo, sendo destacado devido à quantidade de 35 citações recebidas. Tal artigo está classificado na categoria CT04 - Métodos de rede de ontologias semânticas. O referido artigo aponta para a importância dos estudos e pesquisas direcionadas aos vocabulários e ontologias, além dos benefícios e as principais dificuldades direcionadas a aplicabilidade do Linked Data em bibliotecas digitais e os desafios nessa implementação.

Esse artigo destacou-se em virtude da utilização de uma metodologia de revisão bibliográfica no âmbito do Linked Data buscando mapear as lacunas das aplicações sobre ontologias e vocabulários no ambiente das bibliotecas digitais. Além de apontar os problemas relacionados ao Linked Data em bibliotecas digitais e quais as tendências que são direcionadas para o futuro do uso de desse padrão nesse contexto. Fato, que o artigo faz uma imersão nos principais conceitos da Web Semântica, como RDF, a reutilização dos dados nas bibliotecas para seu compartilhamento e uso de novos recursos. Acrescentam-se as análises dos principais vocabulários e ontologias utilizadas nas bibliotecas para os recursos direcionados ao Linked Data. Sendo, também, relatados os principais problemas encontrados dentro das bibliotecas digitais para a implementação do Linked Data.

Apesar das análises destacadas no artigo, em nossa visão, há necessidade de uma ampliação dos estudos, não só em bibliotecas digitais, mas em instituições governamentais que necessitam dispor seus dados de acordo com as diretrizes para transparência das informações. De certo, que nas bibliotecas digitais os estudos são mais focados, assim, facilitando a coleta de informações e suas análises.

Na tabela 5 é apresentado o quantitativo de publicações por ano de cada categoria temática, além de sua totalização e percentual relativo ao total de artigos analisados. Conforme apontam os dados e como já mencionado, as categorias CT02 - Conversão e compartilhamento de Bases de Dados para padrões da Web Semântica e Linked Data e a CT07 - Organização da Informação nos padrões da Web Semântica e Linked Data obtiveram maiores destaques com 25,05% e 15,79%, respectivamente.

Tabela 5 – Quantificação das categorias por ano de 2010 a 2021 das bases de dados *Scielo* e *Web of Science*.

CATEGORIA	2010	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	TOTAL	%
CT01	0	0	0	0	0	2	0	1	1	0	0	4	10,5263
CT02	0	0	2	0	0	1	1	2	1	1	0	8	21,0526
CT03	0	0	0	0	0	0	0	0	0	1	1	2	5,2632
CT04	0	0	0	0	0	1	0	1	0	1	0	3	7,8947
CT05	0	0	0	0	0	0	0	1	0	0	0	1	2,6316
CT06	0	0	0	0	0	1	1	1	0	0	0	3	7,8947
CT07	0	0	0	1	0	0	2	1	1	1	0	6	15,7895
CT08	0	0	0	0	0	0	0	0	0	1	1	2	5,2632
CT09	0	1	0	0	1	0	0	0	0	0	0	2	5,2632
CT10	1	0	0	0	0	0	0	0	0	0	0	1	2,6316
CT11	0	0	0	1	0	0	0	0	0	0	0	1	2,6316
CT12	0	0	1	0	0	0	0	0	0	0	0	1	2,6316
CT13	0	0	1	0	0	1	1	0	0	0	1	4	10,5263
TOTAL	1	1	4	2	1	6	5	7	3	5	3	38	100

Fonte: Dados de Pesquisa (2022).

Verifica-se que o ano com maior produção conjunta foi em 2018, com sete registros e em segunda posição o ano de 2016 com seis artigos, onde ocorreu coincidência na produção de artigos em quatro categorias temáticas: CT01, CT02, CT04 e CT06. Assim, apontando convergências para essas e divergências entre as categorias CT07 e CT13.

8 DISCUSSÃO

Para a verificação dos resultados desta pesquisa são apresentados pontos comparativos e dedutivos das informações coletadas e processadas das bases de dados *Scielo e Web of Science*. Devido à quantidade de registros recuperados, em ambas as bases, serem diferentes adotaram-se os parâmetros estatísticos da média e percentual, pois, desse modo, proporcionam meios de equivalência nas medições das análises quantitativas.

Tomando como ponto inicial a produtividade de publicações, a média de produtividade dos autores nacionais foi de 1,25 artigos publicados/ano, enquanto as publicações internacionais foram de 3,08 artigos. Apesar das diferenças obtidas, não há divergência significativa. Assim, de forma geral, há pouca produção de artigos sobre esses temas de forma conjunta. Na produção nacional houve uma estabilidade de publicações durante anos consecutivos, seguindo de uma pequena alta. Já nas publicações internacionais ocorreram mais oscilações de produção em cada ano. Como pontos convergentes, aponta-se o baixo nível de publicações, mas diferenciando nas oscilações de produtividade durante os períodos analisados para produção nacional e internacional. Verifica-se que em ambas as bases de dados, a quantidade de publicações de artigos por ano ficou abaixo de três artigos.

Tendo em vista os altos custos para publicações de artigos em periódicos de alto padrão, pode ocorrer uma retenção de exposição de pesquisas, o que vem a refletir no quantitativo de artigos, assim, indicando uma baixa produtividade sobre os temas. Outro fator que pode influenciar é em relação ao campo de pesquisa, haja vista que os temas, apesar de sua importância para área da Ciência da Informação e áreas correlatas, podem não despertar interesse no campo de pesquisa de alto impacto e com alto índice de financiamento, como existe com as pesquisas direcionadas para a produção farmacológica e industrial.

Quanto aos periódicos para publicação, é notório que haveriam quantitativos diferenciados devidos às características de análise de cada consulta. Entretanto, os pontos convergentes recaem sobre a escolha de periódicos nacionais de *qualis* A1, visto que o periódico Transinformação é mencionado em ambas as bases, com maior quantidade de publicações nacionais e, em segunda posição, nas análises de publicações indexadas na *Web of Science*. Fato que aponta a qualidade de produção científica brasileira, apesar de baixa produtividade nacional. Os periódicos com alto nível de qualidade (A1) são bastante criteriosos na verificação aos pares dos artigos científicos, exigindo dos pesquisadores maior rigor científico.

No aspecto de autores e coautores os dados apontam que há uma forte relação de colaboração na produção dos artigos, havendo poucas produções de artigos com um único autor. Destaques são dados aos autores Caio Saraiva Coneglian, José Eduardo Santarem Segundo e Mark Douglas de Azevedo Jacyntho que aparecem nas consultas em ambas as bases de dados, onde os dois primeiros autores obtiveram maior índice de publicações na base de dados da *Web of Science* (1ª e 2ª posição, respectivamente), enquanto Mark Douglas de Azevedo Jacyntho foi o primeiro colocado na base dados da *Scielo* e o terceiro na *Web of Science*. Portanto, indicando um ponto de convergência em termos de autores de maiores destaques entre as bases de dados analisadas para os temas propostos. Apesar dessa relação de colaboração entre autores e coautores fortalecer o grau de confiança, por outro lado, pode ocorrer a formação de grupos que ficam restritos aos seus pares, criando um isolamento que vem a prejudicar a troca de informações e avanços nos diversos campos de pesquisa científica.

Verificando as relações colaborativas, por meio do mapa de rede e interações desenvolvidos com o software VOSviewer, analisamos as formações dos clusters. No caso das informações da *Scielo*, houve a formação de um único cluster, configurando uma forte relação de colaboração para publicação. Mas, ao verificar o mapa sem compor a rede totalmente interligada, obtiveram-se quatro clusters, sendo que apenas um único artigo apresenta-se com um autor e um coautor (Mark Douglas de Azevedo Jacyntho; Nilton Freitas Junior), os demais nós produzidos tiveram um autor e dois ou mais coautores por artigo, sem manter a interligação com outros clusters.

Quando da visualização do mapa de rede referente às informações coletadas da *Web of Science*, houve a formação de dois clusters indicando uma forte relação de colaboração entre autores e coautores dentro e entre esses dois clusters. Entretanto, ao verificar o mapa sem compor a rede totalmente interligada verificou-se a formação de um maior número de clusters isolados (6 no total), logo, configurando a formação de grupos que não mantem uma interligação com outros clusters, além de três clusters serem formados por autores isolados (Manolis Perponakis, Adeiana Suares Sanchs, Mark Doug de Azevedo).

Portanto, verificam-se que há convergências na formação de clusters de colaboração de coautores interligados entre as duas bases analisadas. Mas, também, existem clusters sem que sejam totalmente interligados, nesse caso, formado cluster de um único pesquisador ou clusters com dois ou mais pesquisadores por artigo. Dito isto, infere-se que apesar de uma

tendência de colaboração na produção de artigos sobre os temas levantados, dificilmente haverá 100% de colaboração, pois há formação de grupos de pesquisa que não interligam-se com outros grupos para uma rede de colaboração.

Em análise ao grau de colaboração (GC), que indica a participação do nível de colaboradores na produção de artigos, foi apontado para os registros da *Scielo* o valor médio $GC = 1.000$, logo com 100% dos artigos produzidos tiveram um autor e pelo menos um coautor. Enquanto na *Web of Science*, o valor médio foi $GC = 0,8667$, indicando que em alguns artigos produzidos não houve coautores. Isso vem a ratificar as análises do mapa de rede de coautores. Ressalta-se que apesar de $GC = 0,8667$ a diferença entre os dois pontos médios apresentados é de 0,1333 para alcançar o valor de 1,000 (100% de grau de colaboração), assim, não sendo muito grande, o que aponta uma tendência para haver maior número de colaboração na produção de artigos científicos entre as análises realizadas.

Quanto ao número de citações dos artigos, as informações apontaram baixo nível para os artigos nacionais na *Scielo* e alto nível de citações dos artigos na *Web of Science*; configurando que os pesquisadores, de forma geral, preferem realizar suas pesquisas em base de maior abrangência como a *Web of Science*. Destaque se dá à pesquisadora Maria Hallo e colaboradores com 35 citações, referente ao artigo “*Current state of Linked Data in digital libraries*”, de 2016, publicado no periódico *Journal of Information Science*. Portanto, não houve convergência entre a quantidade de citações entre as duas bases de dados, seus artigos relacionados, nem entre nomes de pesquisadores nas duas bases de dados.

Na perspectiva de quais são os vínculos institucionais dos pesquisadores, as informações apontaram que a grande maioria faz parte do ambiente acadêmico, entre universidades ou faculdades ou institutos de pesquisa educacional. Ponto convergente entre os dados analisados. O que só vem a reforçar o papel de destaque que o ambiente acadêmico tem tanto a nível nacional, quanto internacional. Todavia, os quantitativos de universidades nacionais que apoiaram os pesquisadores ficaram restritos ao Sul e Sudeste do Brasil, não havendo artigos de publicações da região Nordeste e Norte. Há, assim, há necessidade de estimular a produção científica sobre os temas Web Semântica e Linked Data nas instituições dessas últimas regiões.

Em respeito às análises dos dados qualitativos, verificam-se que para todas as categorias temáticas existem, de alguma forma, uma interrelação. Haja vista, a amplitude dos temas Web Semântica e Linked Data nos quais possibilitam diversas pesquisas nessas áreas, conforme

apresentados nesse trabalho. Alguns campos são mais explorados, outros ainda estão na fase de conceituação e estudos de casos. De certo, que há necessidade de amadurecimento e aprofundamento dos temas que possam, efetivamente, serem aplicados. Há uma carência de estudos que possam orientar as questões práticas de implantação da Web Semântica e Linked Data, visto que pelo próprio conceito da Web Semântica ser um padrão a almejar, dentro da Internet que já existe com toda sua heterogeneidade de informações.

Nas análises qualitativas dos artigos, todos merecem destaque de uma forma ou de outra, visto que possibilitam um amadurecimento em pesquisas e direcionamento para outros pesquisadores. A tendência de produção de artigos sobre a categoria temática CT02 - Conversão e compartilhamento de Bases de Dados para padrões da Web Semântica e Linked Data, só apontam um direcionamento, não excluindo ou invalidando as outras categorias.

De certo, as produções de artigos necessitam serem ampliadas, uma vez que, de modo geral, há necessidade de mais informações, em especial, no âmbito nacional. Fato comprovado em que apenas três artigos convergiram entre as duas bases de dados analisadas.

Com as novas tecnologias emergentes criam-se diversas possibilidades de pesquisas, como apontado nas categorias temáticas. Uma das possibilidades como ponto de partida é a conversão e compartilhamento das bases de dados atuais e reuso das informações por meios de engenharia de requisitos direcionados aos padrões da Web Semântica e Linked Data, conforme os métodos propostos pela Web Semântica como: uso do XML, RDF e OWL que irão possibilitar diversas instituições que trabalham com o objeto informacional, em destaque, as bibliotecas para disponibilizar um grande volume de informações na Web, assim, criando uma rede interligações por meios dos princípios de Linked Data. Dentro desse ambiente, os estudos sobre metadados e suas conversões toma papel essencial, além dos recursos humanos que estejam capacitados para realizar a interoperabilidade das informações. Destarte, criando a possibilidade de serem consultadas mais eficientemente e com uso de inferências por máquinas e por meios de linguagem de consultas, como SPARQL nos *endpoints* disponíveis na DBPEDIA e em outros locais.

Tais fatos irão proporcionar uma democratização das informações, além do que tudo poderá ser representado dentro da Internet e ser acessado de forma significativa e interpretável por máquinas, conseqüentemente, criando uma Web estruturada permitindo aos usuários recuperarem informações com maior acurácia.

9 CONCLUSÕES

A informação é apontada como primordial para a sociedade. Nesse contexto, a Ciência da Informação destaca-se como uma facilitadora para o acesso ao conteúdo informacional utilizando-se, também, da área da Gestão da Informação. Desse modo, os pontos abordados nesta pesquisa destacaram as tecnologias emergentes para o compartilhamento e acesso ao conteúdo informacional, em especial, no ambiente da Internet tendo como objeto de estudo a Web Semântica e Linked Data. O foco principal deram-se nas análises bibliométricas direcionadas à produção de artigos científicos que convergem a Web semântica e Linked Data na organização da informação, e estão disponibilizados em bases de dados nacionais e internacionais.

A pesquisa destaca-se como relevante, em virtude de proporcionar uma visão como estão sendo tratados tais temas, visto que, diante do grande volume de informações disponibilizadas, atualmente, a Web Semântica e Linked Data serão fundamentais para dispor meios em que os usuários e máquinas obtenham informações com sentido semântico. Acrescenta-se a importância da interligação desses conjuntos de dados, de diversos formatos, que poderão ser acessados adotando os princípios de Linked Data, onde tudo estará conectado na Web. As análises da produção de artigos científicos realizadas nesta pesquisa, por meios de análises quantitativa e qualitativa, referendou a importância em conhecer como estão sendo direcionados os saberes científicos e, assim, poderão possibilitar novas pesquisas no campo acadêmico.

Os diversos aspectos abordados destacam que a Web semântica e Linked Data estão sendo amadurecidos para que haja uma maior aplicabilidade dos recursos que poderão facilitar o acesso mais rápido e com interação e inferência por máquinas, dessa forma, proporcionando uma maior acurácia das informações recuperadas. Os estudos sobre os temas apresentados são um ponto inicial para novas pesquisas interligando com outras temáticas como: Inteligência Artificial, Internet das Coisas, Metaverso, entre outras. São por meios das publicações dos artigos científicos que os pesquisadores proporcionam a divulgação do conhecimento, logo, os estudos bibliométricos apresentados sobre os temas possibilitaram uma visão da produção acadêmica e sua importância.

Como ponto de verificação das propostas desse trabalho, destaca-se que o primeiro objetivo específico: caracterizar a Web Semântica e o Linked Data nos processos de gerenciamento da informação foi alcançado por meio do referencial teórico. No qual foram

destacados os conceitos da Ciência da Informação sobre a ótica dos diversos pesquisadores como: Le Coadic, Borko, Saracevic entre outros, o que proporcionou uma visão epistemológica da informação e seu uso. Quanto ao referencial teórico sobre os conceitos da Gestão da Informação, permitiu analisar a informação como fonte e geração de conhecimento e sua aplicabilidade na sociedade. Essa visão possibilitou a criação de links sobre a informação no contexto atual e as possibilidades de uso dentro dos princípios da Web Semântica e Linked Data, fato que foi alcançado com o referencial teórico sobre esses temas.

O segundo objetivo específico foi alcançado com o levantamento das publicações sobre o tema em bases de dados nacional e internacional, onde foi realizado com a formação dos critérios de pesquisa na formação de *query*, com testes de consultas e recuperação do *corpus* que da pesquisa, além das informações correlacionadas.

O terceiro objetivo específico, voltado à análise bibliométrica das publicações foi alcançado com as análises dos dados obtidos nas bases de dados e sistematização desses em tabelas, gráficos e quadros. Também, a partir das informações obtidas nas bases de dados e arquivos exportados, foi possível a construção de mapas de interações dos coautores com o software VOSviewer. Destarte, os objetivos específicos e, como efeito, o objetivo geral foram atendidos.

Recuperando a questão que norteou essa pesquisa: como está ocorrendo a produção científica acerca das convergências entre Web Semântica e Linked Data? Podemos concluir que sobre o aspecto quantitativo, há convergência na produção de artigos científicos de forma colaborativa entre os autores e coautores em ambas as bases de dados analisadas, contudo, existem baixos índices de produtividade sobre os temas relacionados conjuntamente. No aspecto qualitativo, os artigos produzidos foram classificados em 13 categorias temáticas, tendo maior destaque a categoria CT02 referente à conversão e compartilhamento de Bases de Dados para padrões da Web Semântica e Linked Data, assim indicando uma maior tendência e convergência de produção científica nessa temática.

Todo o processo de investigação tem seus desafios. Nesse caso, apontamos como fatores limitantes desta pesquisa o acesso às bases de dados que demandaram instabilidade de conexão via internet para consultas e verificação dos conteúdos e downloads das informações. Outro fator limitante deu-se em virude do pouco espaço de tempo para realizar a pesquisa, haja vista, a redução do semestre acadêmico devido a adequações impostas pela pandemia da

COVID-19. Isso também restringiu o acesso ao ambiente presencial para pesquisas e informações colaborativas no desenvolvimento da pesquisa com outros docentes e discentes.

Como ponto de sugestões de futuro aprofundamento do tema, sugere-se uma comparação com outras bases de dados, como a *Scopus*, a *BRAPCI* (em âmbito nacional) e correlação com outras áreas de pesquisas, além das possíveis aplicabilidades de ferramentas e recursos computacionais, como exemplo a criação de ontologias semânticas e aplicações de conversão de bases de dados relacionais para padrões da Web Semântica.

REFERÊNCIAS

- AGUIAR, F. C.; MENDES, V. L. P. S. Comunicação organizacional e Tecnologias da Informação e Comunicação (TIC) na gestão hospitalar. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 21, n. 4, p. 138-155, out./dez. 2016. Disponível em: <https://brapci.inf.br/index.php/res/v/37183>. Acesso em: 15 dez. 2021.
- ALMEIDA, M. B. Uma introdução ao XML, sua utilização na Internet e alguns conceitos complementares. **Ci. Inf.**, Brasília, v. 31, n. 2, p. 5-13, maio/ago. 2002. Disponível em: <https://www.scielo.br/j/ci/a/QQ3HX7JLBfTG3gpvPdfGK9j/?format=pdf&lang=pt>. Acesso em: 10 fev. 2022.
- ANNA, J. S. Aspectos epistemológicos da Ciência da Informação e o comportamento informacional: diálogos com Borko, Le Coadic e Saracevic. **RDBCI: Rev. Digit. Bibliotecon. Cienc. Inf., Campinas**, SP, v. 16, n. 2, p. 344-364, maio/ago. 2018. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8649807/pdf>. Acesso em: 20 dez. 2021.
- ARAKAKI, F. A.; SANTOS, P. L. V. A. C.; ALVES, R. C. V. Panorama das pesquisas sobre o padrão de metadados Dublin Core no Brasil. **Revista ACB: Biblioteconomia em Santa Catarina**, Florianópolis, v. 20, n. 1, p. 86-97, jan./abr., 2015. Disponível em: <https://revista.acbsc.org.br/racb/article/view/983>. Acesso em: 15 fev. 2022.
- ARAÚJO, C. A. A. Ciência da Informação, biblioteconomia, arquivologia e museologia: relações institucionais e teóricas. **Enc. Bibli. R. Eletr. Bibliotecon. Ci. Inf., Florianópolis**, v. 16, n. 31, p. 110-130, 2011. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2011v16n31p110/17765>. Acesso em: 20 dez. 2021.
- AUER, S. **Linked Open Data – Creating Knowledge Out of Interlinked Data**. Eds. AUER, S.; BRYL, V.; TRAMP, C. Springer: New York. 2014, 218 p. Disponível em: <https://link.springer.com/content/pdf/10.1007%2F978-3-319-09846-3.pdf>. Acesso em: 10 fev. 2022.
- BARDIN, L. **Análise do Conteúdo**. São Paulo: Edições 70, 1ª ed., 2016, 282 p. Disponível em: <https://madmunifacs.files.wordpress.com/2016/08/anc3a1lise-de-contec3bado-laurence-bardin.pdf>. Acesso em: 19 abr. 2022.
- BARRETO, A. A. A condição da informação. **São Paulo em Perspectiva**, São Paulo, v. 16, n. 3, p. 67-74, 2002. Disponível em: <https://www.scielo.br/j/spp/a/5Q85NCzRFvJ8BLjJd54jLMv/?format=pdf&lang=pt>. Acesso em: 15 dez. 2021.
- BERNERS-LEE T.; LASSILA, O.; HENDLER, J. The semantic web. **Scientific American**, New York, v. 5, maio, 2001. Disponível em: <http://www2.ic.uff.br/~bazilio/cursos/sistweb/material/Barners-Lee-Scientific-American-May-2001.pdf>. Aceso em: 15 fev. 2022.
- BERNERS-LEE. Linked Data - Design Issues. Retrieved, jul. 2006 on-line. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 10 fev. 2022.

BICALHO, L.; OLIVEIRA, M. A teoria e a prática da interdisciplinaridade em Ciência da Informação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 16, n. 13 p. 47-74, jul./set. 2011. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/1245>. Acesso em: 15 dez. 2021.

BIZER, C.; BERNERS-LEE, T.; HEATH, T. Linked Data: The Story so Far. **International Journal on Semantic Web and Information Systems**. v. 5, n. 3, p. 1-22, jun. 2009. Disponível em: https://www.researchgate.net/publication/225070216_Linked_Data_The_Story_so_Far. Acesso em: 10 fev. 2022.

BONINI, J. A. **Engenharia ontológica aplicada para o projeto e especificação da ontocancro**. 2019. 113 f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Santa Maria, Centro de Tecnologia, Programa de Pós-graduação em Ciência da Computação, Rio Grande do Sul, 2019, p. 113. Disponível em: https://repositorio.ufsm.br/bitstream/handle/1/19434/DIS_PPGCC_2019_BONINI_JESSICA.pdf?sequence=1&isAllowed=y. Acesso em 10 fev. 2022.

BORKO, H. Information Science: What is it? **American Documentation**, v.19, n.1, p.1-5, jan. 1968. Disponível em: https://edisciplinas.usp.br/pluginfile.php/1992827/mod_resource/content/1/Borko.pdf. Acesso em: 10 dez. 2021.

BRANDÃO, A. A. F.; LUCENA, C. J. P. Uma Introdução à Engenharia de Ontologias no contexto da Web Semântica. **PUC-RioInf**, MCC29/02, Rio de Janeiro, p. 1-16, nov., 2002. Disponível em: https://www.dbd.puc-rio.br/depto_informatica/02_29_brandao.pdf. Acesso em 10 mar. 2022.

BREITMAN, K, K, **Web semântica: a internet do futuro**. Rio de Janeiro: LTC, 2005. 190 p.

BUCKLAND, M. K. Information as thing. **Journal of the American Society for Information Science (JASIS)**, Berkeley, v.45, n.5, p.351-360, 1991. Disponível em: <https://ppggoc.eci.ufmg.br/downloads/bibliografia/Buckland1991.pdf>. Acesso em: 12 mar. 2022.

CAMARGO, L. S.; BARBOSA, R. R. Bibliometria, cienciometria e um possível caminho para a construção de indicadores e mapas da produção científica. **PontodeAcesso**, Salvador, v.12, n.3, p.109-125, dez. 2018. Disponível em: <https://periodicos.ufba.br/index.php/revistaici/article/view/28408/18030>. Acesso em: 12 mar. 2022.

CAMPOS, L. F. B. METADADOS DIGITAIS: revisão bibliográfica da evolução e tendências por meio de categorias funcionais. **Enc. Bibli. R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, n.23, 1º sem., p. 1-31, 2007. Disponível em: <https://brapci.inf.br/index.php/res/download/49386>. Acesso em: 20 mar. 2022.

CARVALHO, L. A.; CRIPPA, G. Ciência da Informação: histórico, delimitação do campo e a sua perspectiva sobre a área da Comunicação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 18, n. 4, p. 241-251, out./dez. 2013. Disponível em: <https://brapci.inf.br/index.php/res/v/35073>. Acesso em: 12 dez. 2021.

COBO, M. J.; LÓPEZ-HERRERA, A. G.; HERRERA-VIEDMA, E.; HERRERA, F. Science mapping software tools: Review, analysis, and cooperative study among tools. **Journal of the American Society for Information Science and Technology**, New York, USA, v. 62 , n. 7 , p.1382-1402, 2011. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/asi.21525>. Acesso em: 15 fev. 2022.

CORREIA, M. C. S.; ZANDONADE, T. O conceito de informação como conhecimento registrado. **RICI: R.Ibero-amer. Ci. Inf.**, Brasília, v. 11, n. 1, p. 83-102, jan./abril. 2018. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/8432/7026>. Acesso em: 12 dez. 2021.

DANTAS FILHO, A. J.; RAMALHO, R. A. S. Novas possibilidades de representação e recuperação de informações usando o SPARQL. *in*: Educação e Tecnologias inovação em cenários de transição; Congresso Internacional de Educação e Tecnologia (CIET) e Encontro de pesquisadores de educação a distância EnPED), 2018, São Carlos - SP, **Anais [...]**. São Carlos - SP; Universidade Federal de São Carlos (UFSCar), jun. / jul., 2018, , P.1-13 . Disponível em: <https://cietenped.ufscar.br/submissao/index.php/2018/article/view/398/434>. Acesso em: 20 mar. 2022.

DAVENPORT, T. H. **Ecologia da informação**: por que só a tecnologia não basta para o sucesso na era da informação. São Paulo: Futura. 1998, 316 p. Disponível em: <https://ppgic.files.wordpress.com/2018/07/davenport-t-h-2002.pdf>. Acesso em: 12 dez. 2021.

DIAS, T. D.; SANTOS, N. Web Semântica: Conceitos Básicos e Tecnologias Associadas. **Cadernos do IME: Série Informática**, Rio de Janeiro, v. 14, p. 79-92, jun., 2003 . Disponível em: <https://www.e-publicacoes.uerj.br/index.php/cadinf/article/viewFile/6619/4734>. Acesso em: 10 fev. 2022.

FEITOSA, A. **A organização da informação na Web**: das tags à web semântica. Brasília: Thesaurus, 2006, 132 p.

FERREIRA, J. A.; COSTA, P. L. V. A. S. O modelo de dados Resource Description Framework (RDF) e o seu papel na descrição de recursos. **Inf. & Soc.:Est.**, João Pessoa, v. 23, n. 2, p. 13-23, maio/ago. 2013. Disponível em: <https://brapci.inf.br/index.php/res/v/92565>. Acesso em 10 fev. 2022.

FERREIRA, J. B.; SILVA, L. A. M. O uso da bibliometria e sociometria como diferencial em pesquisas de revisão. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 15, n. 2, p. 448-464, maio/ago. 2019. Disponível em: <https://brapci.inf.br/index.php/res/v/113880>. Acesso em: 12 mar. 2022.

FORMENTON, D.; GRACIOSO, L. S. Padrões de metadados no arquivamento da web: recursos tecnológicos para a garantia da preservação digital de websites arquivado. **RDBCI: Rev. Dig. Bibliotec e Ci. Info. / RDBCI: Dig. J. of Lib. and Info. Sci.**, Campinas, SP, v. 20, e022001, p. 1- 29, 2022. Disponível em: <https://brapci.inf.br/index.php/res/v/170063>. Acesso em 10 fev. 2022.

FURGERI, S. O papel das linguagens de marcação para a Ciência da Informação. **TransInformação**, Campinas, v. 3, n. 18, p. 225-239, set./dez., 2006. Disponível em: <https://www.redalyc.org/pdf/3843/384334744006.pdf>. Acesso e, 10 fev. 2022.

GIL, A. C. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo: Atlas, 2018, 173 p.

GUEDES, V. L. S. A bibliometria e a gestão da informação e do conhecimento científico e tecnológico: uma revisão da literatura. **PontodeAcesso**, Salvador, v.6, n.2, p. 74-109 ago. 2012. Disponível em: <https://brapci.inf.br/index.php/res/v/81547>. Acesso em 10 fev. 2022

GRÁCIO, M. C. C. Acoplamento bibliográfico e análise de cocitação: revisão teórico-conceitual. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Santa Catarina, v. 21, n. 47, p. 82-99, set./dez., 2016. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2016v21n47p82/32343>. Acesso em: 20 dez. 2021.

HEDLER, H. C.; FERNEDA, E.; DUARTE, B. S.; PRADO, H. A.; GUTIERREZ, C. E. C. Aplicação do modelo de aceitação de tecnologia à computação em nuvem. **Perspectivas em Gestão & Conhecimento**, João Pessoa, v. 6, n. 2, p. 188-207, jul./dez. 2016. Disponível em: <https://periodicos.ufpb.br/index.php/pgc/article/view/26392/16526>. Acesso em: 12 fev, 2022.

HJØRLAND, B. Domain analysis in information science: eleven approaches - traditional as well as innovative. **Journal of Documentation**, United Kingdom, v. 58, n. 4, p. 422-462, 2002. Disponível em: https://www.researchgate.net/publication/249366184_Domain_analysis_in_information_science_Eleven_approaches_-_Traditional_as_well_as_innovative. Acesso em: 10 fev. 2022.

LAUFER, C. **Guia da web semântica**. Comitê Gestor da Internet no Brasil, São Paulo, 2015. 132 p. Disponível em: <https://www.cgi.br/publicacao/guia-de-web-semantica>. Acesso em: 12 fev. 2022.

LE COADIC, Y. **A ciência da informação**. Brasília: Briquet de Lemos, 1996, p. 1-115. Disponível em: <https://bibliotextos.files.wordpress.com/2012/07/a-cic3aancia-da-informac3a7c3a3o-le-coadic.pdf> . Acesso em: 14 fev. 2022.

LE COADIC, Y. Princípios científicos que direcionam a ciência e a tecnologia da informação digital. **Transinformação**, Campinas, v. 16, n. 3, p. 205-213, set./dez. 2004. Disponível em: <https://www.scielo.br/j/tinf/a/QJ6H6JRRcB9h6kftcTdkQGm/?format=pdf&lang=pt>. Acesso em: 12 dez. 2021.

LIMA, J. C; CARVALHO, C. L. **Extensible Markup Language (XML)**. Relatório Técnico, Instituto de Informática da Universidade Federal de Goiás, Goiás, 2005. 34 p. Disponível em: www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_002-05.pdf . Acesso em 10 fev. 2022.

LOD-CLOUD.NET. Linked Open Data Cloud. 2020 on-line. Disponível em: <https://lod-cloud.net/#diagram>. Acesso em: 12 fev. 2022.

MACHADO, K. C.; ENSSLIN, S. R.; VIANNA, W. B. V.; ENSSLIN, L. Avaliação de desempenho na gestão da informação digital: contribuições da literatura científica. **Inf. & Soc.:Est.**, João Pessoa, v. 26, n. 2, p. 155-172, maio/ago. 2016. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/28508/16210>. Acesso em: 22 fev. 2022.

MAIMONE, G. D.; SILVEIRA, N. C. Cognição humana e os Paradigmas da Ciência da Informação. **Revista Eletrônica Informação e Cognição**, São Paulo, v. 6, n. 1, p. 55-67,

2007. Disponível em: <https://revistas.marilia.unesp.br/index.php/reic/article/view/748>. Acesso em: 12 fev. 2022.

MARQUES, M. B. Gestão da informação em sistemas de informação complexos. **Pesq. Bras. em Ci. da Inf. e Bib.**, João Pessoa, v. 12, n. 2, p. 60-76, 2017. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/pbcib/article/view/35505/18544>. Acesso em: 12 fev. 2022.

MANISHKUMAR, R. S.; ABHIJIT, D. A Journey of Human Comfort: Web 1.0 to Web 4.0. **International Journal of Research and Scientific Innovation (IJRSI)**, Mandsaur, v. 3, n. 9, p. 75-79, set. 2016. Disponível em: <https://www.rsisinternational.org/IJRSI/Issue31/75-78.pdf>. Acesso em: 25 mar. 2022.

MENDONÇA, R. R. **Uma abordagem para coleta e publicação de dados de proveniência no contexto de Linked Data**. 2013. 143 f. Dissertação (Mestrado em Informática) – Universidade Federal do Rio de Janeiro, Instituto de Matemática, Instituto Tércio Pacitti, Programa de Pós-Graduação em Informática, Rio de Janeiro, 2013. Disponível em: <http://objdig.ufrj.br/15/teses/823607.pdf>. Acesso em: 12 fev. 2022.

NASSI-CAIÒ, L. Quanto custa um artigo? Serviços de publicação acadêmica e seus valores de mercado. **SciELO em Perspectiva**, São Paulo, n.10, nov. 2021. (on line). Disponível em: <https://blog.scielo.org/blog/2021/11/10/quanto-custa-um-artigo/#.YmLF6VDMLIU>. Acesso em: 20 abr. 2022.

NASSIF, M. E.; RESENDE, W. C. Gestão da informação e do conhecimento e suas relações com segurança da informação, tecnologias da informação e compartilhamento. **Ci.Inf.**, Brasília, DF, v. 45, n. 3, p. 110-118, set./dez. 2016. Disponível em: <https://brapci.inf.br/index.php/res/v/17503>. Acesso em: 5 fev. 2022.

PANDEY, S.; SAHOO, S. Research Collaboration and Authorship Pattern in the field of Semantic Digital Libraries. **DESIDOC Journal of Library & Information Technology**, v. 40, n 6, p, 375-381, nov. 2020. Disponível em: <https://publications.drdo.gov.in/ojs/index.php/djlit/article/view/15680>. Acesso: 12 abr. 2022.

PEREIRA, A. M; RIBEIRO JÚNIOR, D.; I; NEVES, G. L. Metadados para a descrição de recursos da Internet: As novas tecnologias desenvolvidas para o padrão Dublin Core e sua utilização. **Rev. ACB: Biblioteconomia em Santa Catarina**, Santa Catarina, v.10, n.1, p. 241-249, jan./dez., 2005. Disponível em: https://www.brapci.inf.br/_repositorio/2010/06/pdf_2589fe6ce8_0011157.pdf. Acesso em 14 fev. 2022.

RAUTENBERG, S.; MOTYL, S. K.; BURDA, A. C.; SILVÉRIO, A.; MOURA, F. M. Dados abertos conectados e gestão do conhecimento: estudos de caso cientométricos em uma universidade brasileira. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.22, n.3, p.116-142, jul./set. 2017. Disponível em: <https://www.scielo.br/j/pci/a/KykXkxTPkz369RZjZCfmjnR/?format=pdf&lang=pt>. Aceso em: 5 fev. 2022.

ROCHA, R. P. Metadados, Web Semântica, Categorização Automática: combinando esforços humanos e computacionais para a descoberta e uso dos recursos da web. **Em Questão**, Porto Alegre, v. 10, n.1, p. 109- 121, jan. / jun. 2004. Disponível em: <https://brapci.inf.br/index.php/res/v/9968>. Acesso em: 5 fev. 2022.

ROZA, R. H. Ciência da informação, tecnologia e sociedade. **Biblos: Revista do Instituto de Ciências Humanas e da Informação**, Rio Grande, v. 32, n. 2, p. 177-190, jul./dez. 2018. Disponível em: <https://periodicos.furg.br/biblos/article/view/7546/5861>. Acesso em: 10 dez. 2021.

RUAS, T. L.; PEREIRA, L. Como construir indicadores de Ciência, Tecnologia e Inovação usando Web of Science, Derwent World Patent Index, Bibexcel e Pajek? **Perspectivas em Ciência da Informação**, Belo Horizonte, v.19, n.3, p.52-81, jul./set. 2014. Disponível em: <https://www.scielo.br/j/pci/a/B8tPGKxB4nh7YjF5xCsMn/?format=pdf&lang=pt>. Acesso em: 10 fev. 2022.

SANTAREM SEGUNDO, J. E.; CONEGLIAN, C. S. Web Semântica e Ontologias: um estudo sobre construção de axiomas e uso de inferências. **Informação & Informação**, Londrina, Londrina, v. 21, n.2, p. 217-244, 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/26417>. Acesso em: 10 fev. 2022.

SANTOS, F. C; CARVALHO, C. L. Aplicações de Suporte à Web Semântica. **Instituto de Informática Universidade Federal de Goiás**, RT-INF_004-07 - Relatório Técnico, Goiás, dez, p. 1-25, 2007. Disponível em: https://ww2.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_004-07.pdf. Acesso em: 20 mar. 2022.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspec. Ci. Inf.**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235>. Acesso em: 20 dez. 2021.

SILVA, D. D.; GRÁCIO, M. C. C. **Índice h de Hirsch**: análise comparativa entre as bases de dados Scopus, Web of Science e Google Acadêmico. Em *Questão*, Porto Alegre, v. 23, p. 196-212, Edição Especial 5 EBBC, 2017. Disponível em: <https://brapci.inf.br/index.php/res/download/47141>. Acesso em: 22 abr. 2022.

SILVA, J. L. C.; GOMES, H. F. Conceitos de informação na Ciência da Informação: percepções analíticas, proposições e categorizações. **Inf. & Soc.:Est.**, João Pessoa, v. 25, n. 1, p. 145-157, jan./abr. 2015. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/145/13200>. Acesso em: 12 dez. 2021.

SOUSA, J. L.; MARTINS, P. G. M.; RAMALHO, R. A. S. Análise dos padrões XML e RDF para a representação na web sob a perspectiva da Ciência da Informação: um estudo preliminar. **Informação & Tecnologia (ITEC)**, João Pessoa, v.5, n.1, jan./jun. 2018. Disponível em: <https://periodicos.ufpb.br/index.php/itec/article/view/38458/22109>. Acesso em: 11 fev. 2022.

TAVARES, W. Q.; CELERINO, V. G. A Importância da Bibliometria para a Indexação Automática. **Folha de Rosto: Revista de biblioteconomia e Ciência da Informação**, Juazeiro do Norte, v. 4, n. 1, p. 7-14, jul./dez., 2018. Disponível em: <https://periodicos.ufca.edu.br/ojs/index.php/folhaderosto/article/view/273/287>. Acesso em: 20 fev. 2022.

VITAL, L. P.; CAFÉ, L. M. A. Ontologias e taxonomias: diferenças. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.16, n.2, p.115-130, abr./jun. 2011. Disponível em: <https://brapci.inf.br/index.php/res/v/35752>. Acesso em: 10 fev. 2022.

WORLD WIDE WEB CONSORTIUM. Web Ontology Language (OWL). W3C Recommendation, 11 dec., 2013. on-line. Disponível em: <https://www.w3.org/OWL/>. Acesso em: 25 fev. 2022.

_____. RDF 1.1 XML Syntax. W3C Recommendation, 25 fev., 2014. on-line. Disponível em: <https://www.w3.org/TR/2014/REC-rdf-syntax-grammar-20140225>. Acesso em: 10 fev. 2022.

_____. Vocabularies. 2015a. on-line. Disponível em: <https://www.w3.org/standards/semanticweb/ontology>. Acesso em: 8 fev. 2022.

_____. XML Essentials. 2015b. on-line. Disponível em: <https://www.w3.org/standards/xml/core>. Acesso em: 10 fev. 2022.