

ANÁLISE BIBLIOMÉTRICA DA PRODUÇÃO CIENTÍFICA SOBRE LINKED DATA

ANÁLISIS BIBLIOMÉTRICO DE LA PRODUCCIÓN CIENTÍFICA ACERCA LINKED DATA

Leandro Dal Pizzol – leandrodalpizzol@gmail.com
Mestre em Engenharia e Gestão do Conhecimento pela
Universidade Federal de Santa Catarina (UFSC).

Rafael de Moura Speroni – speroni@egc.ufsc.br
Doutorando no programa de Pós-Graduação em Engenharia e
Gestão do Conhecimento da Universidade Federal de Santa
Catarina (UFSC). Professor do Instituto Federal Catarinense

Airton Zancanaro – airtonz@egc.ufsc.br
Doutor em Engenharia e Gestão do Conhecimento pela
Universidade Federal de Santa Catarina (UFSC). Pós-doutorando
pela Universidade Regional de Blumenau.

Fernando Ostuni Gauthier – fernando.gauthier@gmail.com
Professor do Programa de Pós-Graduação em Engenharia e
Gestão do Conhecimento da Universidade Federal de Santa
Catarina (UFSC).

José Leomar Todesco – tite@lec.ufsc.br
Professor do Programa de Pós-Graduação em Engenharia e
Gestão do Conhecimento da Universidade Federal de Santa
Catarina (UFSC).

RESUMO

Introdução: Desde que Tim Berners-Lee cunhou o termo *Linked Data* em 2006, muitos estudos surgiram baseados nessa iniciativa. Seja na forma de movimentos de acesso à informação, seja na de aplicações baseadas nesse tipo de dados, o interesse por essa nova maneira de publicar e consumir informação na *web* está cada dia mais em foco.

Objetivo: Neste trabalho, apresenta-se uma revisão bibliométrica com o propósito de contribuir para o melhor entendimento e a clarificação de conceitos e termos associados à área de *Linked Data*.

Metodologia: Utilizou-se as bases de dados científicas *Web of Science (WoS)*, *Scopus*, EBSCO e IEEE, de modo a localizar estudos existentes sobre a temática até outubro de 2012.

Resultados: Para isso, foram mapeados 148 trabalhos científicos que abordam o referido tema, escritos por 409 autores pertencentes a 160 instituições de 30 países diferentes. Pôde-se, com isso, constatar que as pesquisas sobre a temática *Linked Data* estão concentradas num pequeno grupo de autores, o que sugere oportunidades e espaços para pesquisas futuras.

Conclusões: Ao sintetizar as informações referentes às publicações científicas de uma determinada área, se possibilita a construção de um arcabouço teórico e que novas pesquisas sejam realizadas.

Palavras-chave: *Linked data*. Bibliometria. Pesquisa bibliográfica. Bases de dados.

1 INTRODUÇÃO

O termo *Linked Data* tem despertado interesse acadêmico e organizacional desde a segunda metade da década de 2000 (BERNERS-LEE, 2006; BIZER, 2009; BRADLEY, 2009; HEATH, 2011). Proposto por Sir. Tim Bernes-Lee, em 2006, o termo *Linked Data* refere-se a um estilo de publicar e interligar dados estruturados de diferentes fontes na *Web*. Por utilizar a rede mundial de computadores, *Linked Data* altera significativamente a maneira como é publicado e consumido o conhecimento. Características como a diminuição de barreiras e a natureza genérica da *Web* em conjunto as técnicas de *Linked Data* e iniciativas de publicação de dados abertos resultaram em um repositório global de informação que possibilita estabelecer construções coletivas de conhecimento.

Costumeiramente as iniciativas de publicação de dados não passam da simples disponibilização de dados heterogêneos em formatos como CSV, XML ou páginas HTML sem nenhum tratamento integrado, sacrificando muito de sua estrutura e semântica. Para que esses dados realmente façam parte de um espaço único de dados global, Berners-Lee (2006) delineou um conjunto de quatro “regras” para a publicação de dados na *web*, a saber: 1) Deve-se usar URIs (*Uniform Resource Identifiers*) como nomes para as coisas; 2) Utilizar HTTP URIs

(*HyperText Transfer Protocol*) de modo que as pessoas possam procurar por esses nomes; 3) Quando alguém procurar um URI, fornecer informações úteis usando os padrões RDF (*Resource Description Framework*) e SPARQL; e 4) Incluir *links* para outros URIs, a fim de que se possa descobrir mais coisas.

Estas regras ficaram conhecidas como os princípios da *Linked Data* e fornecem uma receita básica para a publicação e conexão de dados usando a infraestrutura da *web* em aderência à sua arquitetura e padrões. Assim como na *web* tradicional, formada por *links hypertext*, a rede de dados em *Linked Data* é resultante da ligação entre documentos presentes na *web*. Porém, as ligações são feitas usando *links hyperdata* com informação expressa em RDF, em que, não somente os documentos podem ser ligados, mas também a informação presente nestes (BERNERS-LEE, 2006).

Ao analisar-se as tecnologias presentes na abordagem de *Linked Data* depara-se com aquelas que também são fundamentais para a *web*: *Uniform Resource Identifiers* (URIs) (BERNERS-LEE; FIELDING; MASINTER, 2005) e o HTTP (FIELDING *et al.*, 1999). Enquanto URLs (*Uniform Resource Locator*) tornaram-se familiares como endereços de documentos e de outras entidades que podem ser localizados na *web*, URIs fornecem um meio mais genéricos para identificar qualquer entidade que existente, seja esse, um documento ou descrição deste documentos (BIZER; HEATH; BERNERS-LEE, 2009). Sendo assim, ao identificarem entidades com URIs que utilizam o padrão “http://”, estas podem ser consultadas simplesmente dereferenciando a URI por meio do protocolo HTTP, o que proporciona uma maneira simples e universal para a recuperação de recursos.

Além das URIs e HTTP outra tecnologia fundamental para a *web* de dados é o RDF (LASSILA; SWICK, 1998). Ele fornece grafos genéricos baseados no modelo de dados com os quais é possível descrever a estrutura de dados e *links* usada para representar as coisas no mundo (MANOLA; MILLER, 2004). O modelo RDF expressa os

dados na forma de triplas, isto é, sujeito, predicado, objeto, nas quais o sujeito e o objeto são recursos identificados por URIs. Por sua vez, o predicado especifica como o sujeito e o objeto estão relacionados, sendo também representado por uma URI (BRICKLEY; GUHA, 2004).

A união das tecnologias de *Linked Data* permite que pessoas compartilhem dados de forma tão fácil quanto documentos são atualmente compartilhados na *web*. Essa facilidade de publicação, em conjunto com iniciativas de abertura de dados, fez surgir uma vasta rede de dados (BIZER; CYGANIAK; GAUß, 2007).

Valendo-se de técnicas bibliométricas, este artigo objetiva mapear o perfil e as características dos trabalhos científicos que abordam o tema *Linked Data*. A realização desse levantamento da literatura contribui para sua melhor compreensão, aponta as iniciativas que estão sendo desenvolvidas, novas linhas de pesquisas, além de apresentar os mais representativos autores e grupos de pesquisas que têm desenvolvido estudos associados à temática (PIZZANI *et al.*, 2012).

O artigo está dividido da seguinte forma: Na seção 2, é apresentado o detalhamento dos procedimentos metodológicos utilizados para o desenvolvimento da pesquisa bibliométrica. Na seção 3, são apresentados e discutidos os principais resultados. E, por fim, na seção 4, são descritas as considerações finais do artigo, seguidas pelo referencial teórico.

2 PROCEDIMENTOS METODOLÓGICOS PARA A REALIZAÇÃO DA PESQUISA BIBLIOMÉTRICA

Esta pesquisa foi desenvolvida em duas etapas combinando diferentes técnicas bibliométricas e bibliográficas, seguindo algumas recomendações e procedimentos metodológicos propostas por Macias-Chapula (1998), Guedes e Borschiver (2006) e Kobashi e Santos (2006). Na primeira etapa foi realizada uma busca sistêmica da literatura, constituída da coleta, filtragem e padronização dos dados. Já na

segunda, foi feita uma análise descritiva composta pela análise e seleção dos dados coletados, culminando na elaboração do documento final. Cada fase da pesquisa é composta por um conjunto de etapas que foram seguidas para a confecção do artigo. Nas subseções seguintes serão descritos os procedimentos adotados em cada uma das etapas do processo de desenvolvimento.

Etapas 1 - Definição dos termos de busca

Para a realização das buscas bibliográficas nas bases científicas foi utilizado o termo em inglês “*Linked* Data”. O uso do símbolo “*” possibilita a busca de outras palavras antes ou entre a expressão, contemplando assim expressões similares, como “Open Linked Data”, “Linked Open Data” ou “Linked Government Data”.

Etapas 2 - Consulta nas bases de dados

Os dados usados nessa bibliometria são provenientes de quatro bases de dados científicos internacionais: *Web of Science* (WoS), *Scopus*, EBSCO e IEEE. A escolha dessas bases de dados justifica-se pelo fato de serem bases de caráter multidisciplinar (ALMEIDA, 2006), de terem retornado registros consistentes sobre os termos pesquisados e de permitirem a exportação dos dados para *softwares* gerenciadores de referências bibliográficas. A primeira base de dados indexa mais de 12.700 periódicos de diferentes áreas do conhecimento (COSTA *et al.*, 2012) e é considerada uma das bases mais relevantes e utilizadas para os estudos bibliométricos (BRAMBILLA; STUMPF, 2012). A segunda é uma base científica lançada em 2004, mantida pela editora Elsevier e indexa artigos desde 1960 com mais de 18.500 títulos com atualizações diárias (OLIVEIRA; GRACIO, 2011). Abrange as áreas do conhecimento como a engenharia, as ciências sociais, a economia e as ciências em geral (MESQUITA *et al.*, 2006). A terceira, permite o acesso ilimitado ao texto completo das publicações de diversas áreas do conhecimento (SABURIT; GUERRERO; MANTECÓN, 2008). A última base indexa

tanto trabalhos científicos como técnicos publicados pela IEEE (*Institute of Electrical and Electronics Engineers*) e por instituições parceiras. Para tal, indexa mais de 3 milhões de documentos com textos completo publicados em mais de 160 periódicos científicos, 1.200 anais de eventos e mais de 3.800 padrões técnicos (IEEE, 2012).

Cada uma das bases pesquisadas apresenta características próprias de estruturação e indexação dos termos de pesquisa, o que requer estratégias de busca diferentes para cada uma das bases. A pesquisa foi realizada em outubro de 2012, e, além da restrição resultante da escolha do campo de busca, os artigos foram restritos aos idiomas inglês, português e espanhol, pertencentes a revistas e congressos, e classificados nas áreas de engenharia, computação, administração e ciências sociais.

Etapa 3 – Exportação dos registros

O resultado das buscas bibliográficas nas bases científicas foi exportado e carregado na ferramenta de gerenciamento de referências bibliométricas *EndNote*, gerando um conjunto único com todos os artigos. Criada pela Thompson Reuters, essa ferramenta permite a exportação de informações, como título, autores, ano de publicação, local de publicação e instituição, usadas neste trabalho.

Etapa 4 – Aplicação de critérios de seleção

Nesta etapa do desenvolvimento foram aplicados os seguintes critérios de seleção: 1) exclusão de artigos sem autoria; 2) exclusão de artigos duplicados, ou seja, artigos que estavam indexados em mais de uma base de dados; 3) retirada dos artigos que estavam fora do contexto deste estudo; e 4) retirada de artigos no qual os textos completos não estão disponíveis gratuitamente para *download*.

Etapa 5 – Padronização dos dados

Para garantir a consistência das informações provenientes das diferentes bases de dados, foi necessária a padronização dos

resultados. Como exemplo, a base *Scopus* apresenta o cadastro do nome do autor apenas com a primeira letra do nome. Já na WoS o nome do autor costuma ser cadastrado por completo. Nesta fase, além da padronização dos dados, foram incluídas informações não disponíveis nas bases de dados, como, por exemplo, a latitude e longitude das instituições às quais os autores estão vinculados, que foram preenchidas manualmente por meio das informações presentes nos documentos. Para auxiliar no cadastro dessa informação, foi criada uma base de dados usando o *Microsoft Access*, no qual foi feito o preenchimento manual de informações faltantes, como informações dos autores (sobrenome, nome, afiliação e localização da instituição), dos artigos (ano de publicação, idioma, tipo de documento, *journal* de publicado, resumo e principais temáticas), das referências citadas (autor, ano, título do trabalho) e as palavras-chave. Esse procedimento, apesar de trabalhoso, garante a padronização e consistência das informações presentes nas publicações anteriormente selecionadas para a realização das análises.

Etapa 6 – Análise dos dados

Após a padronização de todos os dados, foi possível realizar operações de análise com rapidez e precisão. Alguns dos dados bibliométricos retirados das publicações foram: quantidade de publicações por ano, tipos das fontes de publicação (eventos ou periódicos) e dos autores a quantidade, a instituições de vínculo e os países de origem. Os dados coletados possibilitam, por exemplo, a identificação dos principais autores e grupos de pesquisas que desenvolvem estudos sobre *Linked Data*. Outra importante fonte de informação analisada são as referências citadas nos artigos, as quais possibilitam identificar referencial teórico nas bases teóricas em que se fundamentam os trabalhos.

Por fim, outra atividade desenvolvida nesta fase foi a classificação dos artigos em grupos conforme sua temática principal.

Para isso, os pesquisadores realizaram a leitura dos resumos dos artigos ou do texto completo quando o primeiro não fosse suficiente para definir a temática do artigo. Como resultado, foram identificadas as principais abordagens seguidas pelos autores e os macrotemas e temas de interesse dentro do universo de *Linked Data*.

Nesta etapa, ainda, os pesquisadores utilizaram o *software* SCI 2¹ que permite, através dos dados exportados da base de dados do Microsoft Access, no formato ISI, gerar a rede de autoria e coautoria.

Etapa 7 – Elaboração do documento final

De posse das informações coletadas, foi elaborado o documento que contém todas as constatações, descrições e os resultados das análises compiladas e publicadas na forma de um artigo científico. Nas próximas seções deste artigo são apresentados os resultados obtidos por meio do desenvolvimento de cada uma das etapas descritas anteriormente.

3 RESULTADOS

Nesta seção, apresentam-se os resultados das análises e sínteses das informações, selecionados a partir das buscas bibliográficas realizadas nas bases de dados EBSCO, *Scopus*, WoS e IEEE.

Dados bibliométricos gerais sobre as pesquisas de Linked Data

A busca bibliométrica nas bases de dados WoS, *Scopus*, EBSCO e IEEE apresentou inicialmente um total de 631 registros. Após a

¹ Science of Science (SCI2) foi desenvolvido pela Universidade de Indiana com o objetivo de ser um conjunto de ferramentas modulares projetado especificamente para o estudo da ciência (<https://sci2.cns.iu.edu>).

aplicação dos diversos filtros, restou um total de 148 trabalhos que serão analisados no decorrer desta pesquisa. A Tabela 1 ilustra o processo de seleção dos artigos para a análise final.

Tabela 1 - Número de publicações selecionadas nas bases de dados

Bases de Dados	Publicações Localizadas	Exclusão das Publicações Repetidas	Exclusão das Publicações sem Autoria Disponível	Exclusão das Publicações sem Texto Completo Disponível ou Fora de Contexto	Total de Publicações Selecionadas
WoS	184	89	4	78	13
Scopus	205	1	0	139	65
Ebsco	100	33	7	50	10
IEEE	142	0	10	72	60
Total	631	123	21	339	148

Fonte: Os autores.

Entre os 148 trabalhos selecionados, 62 são artigos de congressos e 86, indexados em periódicos científicos. Esses trabalhos foram escritos por 459 autores, vinculados a 160 instituições de 30 países diferentes. Os autores utilizaram 286 palavras-chave para indexar seus trabalhos e 2.182 referências (média de 15 referências citadas por publicação). A Tabela 2 sintetiza os dados da pesquisa.

Tabela 2 - Dados bibliométricos gerais das publicações selecionadas

Dados bibliométricos	Frequência
Publicações	148
Fontes de publicações	109
Autores	459
Instituições	160
Países	30
Palavras-chave	286
Referências citadas	2.182

Fonte: Os autores.

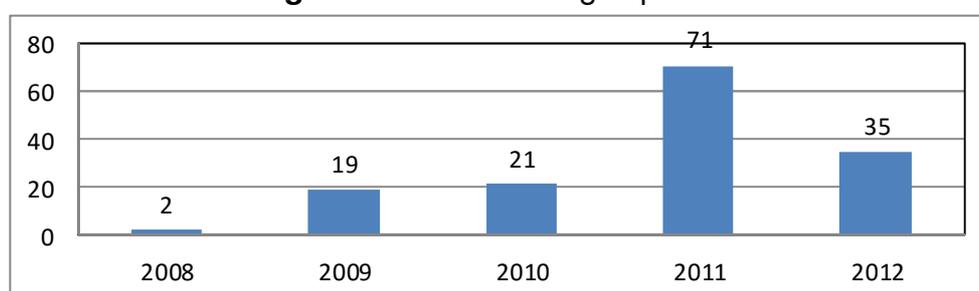
Na sequência serão descritas as análises resultantes de cada um dos dados bibliométricos apresentados na Tabela 2.

Publicações por ano

A série histórica aponta que o maior número de artigos, conforme demonstra o gráfico da Figura 1, foi no ano de 2011. Nesse ano, 23 artigos foram publicados em conferências ou *Workshops* internacionais, como, por exemplo, *International Conference on Intelligent Networking*

and Collaborative Systems, *International Enterprise Distributed Object Computing Conference Workshops* e *International Workshop on Database and Expert Systems Applications (DEXA)*, com dois artigos publicados em cada evento, totalizando 71 artigos publicados. Em se tratando de revistas, o destaque é dado para o *Journal of Web Semantics* com oito publicações e para a *IEEE Internet Computing* com quatro artigos publicados.

Figura 1 - Total de artigos por ano



Fonte: Os autores.

Observa-se que, pelo fato do *Linked Data* ser um tema relativamente novo, as primeiras publicações identificadas nas bases de dados pesquisadas, que tratam do assunto, são de 2008. Nesse ano, dois trabalhos foram identificados: *Revyu: Linking reviews and ratings into the Web of Data* (Heath and Motta, 2008) e *Sindice.com: A document-oriented lookup index for open linked data* (Oren et al., 2008). Ambos tratam de ferramentas que utilizam recursos da *Web Semântica* para buscar conteúdos de fontes externas.

Principais fontes de publicações

Os artigos foram publicados em 109 fontes de publicações diferentes, e as quatro mais utilizadas pelos autores estão apresentadas na Tabela 3.

Tabela 3 - Principais fontes de publicações

Fonte de publicação	Qtd	ISSN	Tipo	Fator de Impacto(JRC)	Editora
Journal of Web Semantics	14	1570-8268	Journal	1,302	Elsevier
IEEE Internet Computing	6	1089-7801	Journal	2,000	IEEE Computer Society
IEEE Intelligent Systems	5	1541-1672	Journal	2,154	IEEE
International Journal on Semantic Web and Information Systems	4	1552-6283	Journal	2,308	University of Leipzig, IBM TJ Watson, University of Aberdeen e ISTI-CNR

Fonte: Os autores.

Destaca-se o periódico científico interdisciplinar *Journal of Web Semantics* que indexa pesquisas e aplicações de várias áreas do conhecimento cujo objetivo principal é a indexação do conhecimento intensivo e serviços inteligentes da *web*. Essa revista foi a mais utilizada pelos autores (14 artigos) para publicação; entretanto, possui o menor fator de impacto dentre as mais citadas. Já a *IEEE Internet Computing* publica artigos relacionados com as tecnologias da *Internet* e aplicações. A *IEEE Intelligent Systems* aborda pesquisas de bases teóricas sobre metodologias e tecnologias da computação social e as suas principais áreas de aplicação. Por último, a *International Journal on Semantic Web and Information Systems* é a revista que possui maior fator de impacto dentre o grupo selecionado, pois disponibiliza pesquisas que abrangem todos os aspectos da *Web Semântica*.

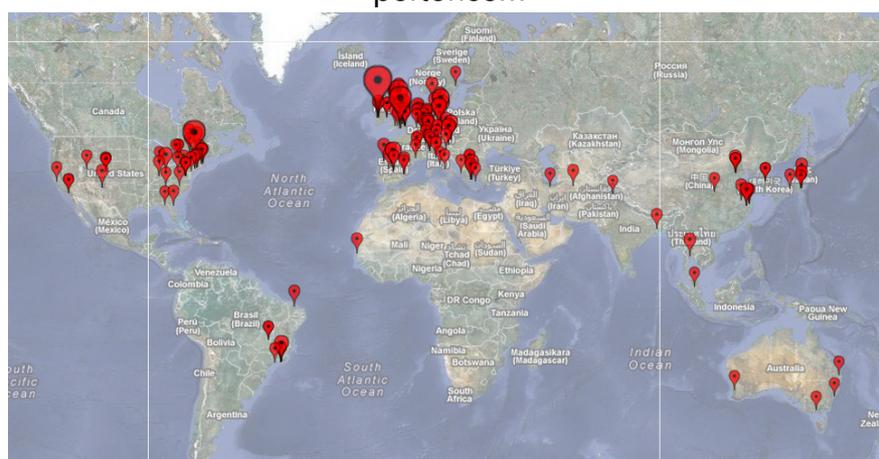
Além dos periódicos científicos, 41,89% dos artigos foram publicados em congressos, como, por exemplo, a *International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, *International Conference e-Science (e-Science)* e *International Workshop on Database and Expert Systems Applications (DEXA)*, todos com o apoio da IEEE.

Principais autores, instituições e países

As publicações analisadas neste trabalho foram escritas por 459 autores que estão afiliados a 160 instituições. O mapa da Figura 2 ilustra a localização dessas instituições, e percebe-se ali que a Europa (Inglaterra com 109, Alemanha com 73, Irlanda com 52 e Espanha com

36 pesquisadores) e os Estados Unidos (com 83 pesquisadores) são os locais com maior número de publicações. Vale destacar que a China, com 29, e o Brasil, com 24 pesquisadores, também compõem a lista dos países que estudam o tema *Linked Data*.

Figura 2 - Mapa com a distribuição das instituições às quais os autores pertencem



Fonte: Os autores usando GPS Visualizer²

Por meio da análise das autorias dos trabalhos selecionados, foi possível identificar a existência de 85 *clusters* de grupos de pesquisa sobre o tema *Linked Data*. Dessa forma, os autores com maior número de publicações são apresentados na Tabela 4 e descritos individualmente na sequência.

Tabela 4 - Autores com maior número de publicações

Autor	Publicações		Vínculo Institucional	Cidade	País
	Total	1º Autor			
Hendler, J	5	2	Rensselaer Polytechnic Institute	Troy	EUA
Auer, S	4	-	Universitdt Leipzig	Leipzig	Alemanha
Bizer, C	4	3	Freie Universitat	Berlin	Alemanha
Haslhofer, B	4	2	University of Vienna	Viena	ustria
Passant, A	4	-	National University of Ireland	Galway	Irlanda
Sheth, A	4	-	Wright State University	Fairborn	EUA
Zhao, J	4	1	University of Oxford	Oxford	Inglaterra

Fonte: Os autores.

² http://www.gpsvisualizer.com/map_input?form=google

O grupo (rede de pesquisa) mais expressivo em termos de publicações é formado por James Hendler, que é autor de dois trabalhos e coautor de três (HENDLER, 2009; DING et al., 2011; DOWNIE et al., 2011; XIAN; JIE; HENDLER, 2011; HENDLER et al., 2012). Seu grupo de pesquisa é composto por 23 pesquisadores de quatro instituições. Hendler é professor de Computação e Ciência Cognitiva, chefe do Departamento de Ciência da Computação do Instituto Politécnico de Rensselaer (RPI) e também professor visitante da Universidade DeMontfort, em Leicester, no Reino Unido. Tendo sido um dos criadores da *Web Semântica*, é autor de mais de 200 artigos técnicos nas áreas de *Web Semântica*, Inteligência Artificial, computação baseada em agentes e processamento de alto desempenho, e tem atuações em órgãos como Associação Americana de Inteligência Artificial, Sociedade Britânica de Computação, Associação Americana para o Avanço da Ciência, DARPA e IEEE.

Sören Auer publicou quatro trabalhos (BIZER et al., 2009; HELLMANN; LEHMANN; AUER, 2009; MORSEY et al., 2012; SALAS et al., 2012) como coautor, e a sua rede de pesquisa é composta por 16 integrantes de duas instituições. Auer é cientista sênior e chefe do grupo de pesquisa *Agile Engineering and Semantic Web*, no Instituto de Informática Aplicada e Sistemas de Informações para Negócios, da Universidade de Leipzig. É também diretor de estudos acadêmicos na Escola de Mídia de Leipzig para estudos de mestrado.

Christian Bizer publicou quatro trabalhos, dos quais é autor em três (BECKER; BIZER, 2009; BIZER; HEATH; BERNERS-LEE, 2009; BIZER, 2009; BIZER et al., 2009). Já a sua rede de pesquisa é composta por nove integrantes de cinco instituições. Ele obteve o título de doutor pesquisando o tema qualidade da informação no contexto de Sistemas Baseados na Web. Desde julho de 2012, Bizer é professor da Universidade de Mannheim e o foco das suas pesquisas é a evolução da *web*, passando de um meio de publicação de documentos para um espaço de dados globais. Como resultados de seu trabalho, incluem o

modelo de dados *Named Graphs*, adotado na recomendação de SPARQL da W3C; a linguagem de mapeamento D2RQ; o *Silk – Linking Framework*; e o Berlin SPARQL Benchmark. Foi um dos fundadores da comunidade W3C Linking Open Data, e co-fundador do projeto DBPedia.

Bernhard Haslhofer publicou quatro trabalhos (HASLHOFER; SCHANDL, 2010; HASLHOFER *et al.*, 2011; POPITSCH; HASLHOFER; ROOCHI, 2010; POPITSCH; HASLHOFER, 2011), é autor em dois deles, e a sua rede de colaboradores é composta por seis pesquisadores. Haslhofer é professor no Departamento Ciência da Informação na Universidade de Cornell e pesquisador nas áreas de redes de dados globais e descentralizadas e seus contextos sociais, culturais e técnicos, além de trabalhar em soluções para avaliação e manutenção da qualidade dos dados em ambientes abertos. Suas pesquisas envolvem, ainda, a investigação de como as redes de dados e os algoritmos construídos sob os princípios da abertura e descentralização podem ser aplicados no âmbito acadêmico e vice-versa.

Alexandre Passant é coautor em quatro trabalhos (KIM *et al.*, 2011; MENDES *et al.*, 2010; ORLANDI; PASSANT, 2011; PFISTERER *et al.*, 2011) e a sua rede de colaboração é composta diretamente por 20 pesquisadores. Passant é professor da Universidade Nacional da Irlanda, pesquisador do Digital Enterprise Research Institute (DERI) e membro dos grupos de trabalho da W3C nas áreas de Social Web, SPARQL e *Semantic Web*. É ainda fundador de uma *start-up* chamada seevl.net que tem o objetivo de facilitar a descoberta de músicas do youtube.

Amit P. Sheth possui quatro trabalhos publicados (JUN *et al.*, 2011; LAKSHMANAN *et al.*, 2011; MENDES *et al.*, 2010; PATNI; HENSON; SHETH, 2010) em conjunto com outros 11 pesquisadores. Professor de Ciência da Computação e Engenharia da Wright State University, atualmente é diretor do Centro de Excelência em

Computação baseada em Conhecimento de Ohio. Em 2012, Sheth foca as suas pesquisas na *Web 3.0*, incluindo áreas da *Web Semântica*, da *Web Social*, de Sensores, da Computação móvel e da Computação em nuvem, além de ter trabalhos na área médica, de gerenciamento de água e gestão de desastres e de movimentos sociais.

Por último, Jun Zhao também publicou quatro trabalhos (KEI-HOI et al., 2009; ROURE et al., 2010; JUN et al., 2011; MARSHALL et al., 2012), dos quais é autora em um. É afiliada ao departamento de Zoologia da Universidade de Oxford. Atua em projetos como o de utilização das tecnologias da *Web Semântica* para fornecer métodos e ferramentas para garantir a preservação a longo prazo dos *workflows* científicos e o de apoio mais eficiente a biomedicina usando as tecnologias da *Web*.

A Tabela 5 ilustra a quantidade de autores que estão vinculados às instituições que pesquisam assuntos relacionados com o *Linked Data*.

Tabela 5 - Quantidade de autores por instituição

Instituição	Quantidade de Autores
National University of Ireland	31
Rensselaer Polytechnic Institute	22
University of Southampton	22
University of Manchester	17
Hasso-Plattner-Institut	13

Fonte: Os autores.

A instituição mais atuante, ou seja, com maior número de pesquisadores sobre o tema *Linked Data*, é a *National University of Ireland* através do *Digital Enterprise Research Institute* (DERI). Este é um Centro de Ciência, Engenharia e Tecnologia estabelecido em 2003 pela Science Foundation Ireland, reconhecido internacionalmente em pesquisas sobre *Web Semântica*, educação e transferência de tecnologia. Atualmente, o DERI conta com aproximadamente 130 membros, tendo mais de 550 publicações de pesquisa e participando de 12 grupos de padronização.

O objetivo do DERI é tornar-se reconhecido como o principal instituto de pesquisa sobre a *web*, interligando tecnologias, informações e pessoas para o avanço dos negócios e o benefício da sociedade. Dentre os projetos desenvolvidos pelo DERI, o *Cloud4SOA* trata da interoperabilidade semântica e combina computação em nuvens, arquiteturas orientadas a serviço (SOA) e semântica. O projeto *digital.me*, por sua vez, visa à integração e à gestão de dados pessoais por meio de um ponto único de acesso de usuário e compartilhamento com redes e serviços sociais. Já o projeto *Linked2Media* visa facilitar o acesso das pequenas e médias empresas da Europa aos dados de usuários de sistemas e redes sociais para identificar e melhorar estratégias de *marketing* e retorno de investimento. O projeto *Linked2Safety* visa à criação de um *framework* de interoperabilidade semântica, além de infraestrutura técnica e plataformas eficientes para companhias farmacêuticas, profissionais de saúde e pacientes acessarem as informações médicas contidas nos registros eletrônicos de saúde (EHRs), na Europa. Já o projeto *LOD Around the Clock* (LATC) tem por objetivo auxiliar pessoas e instituições a publicar e consumir *Linked Data*.

Já o Rensselaer Polytechnic Institute (RPI), localizado em Nova York, é a mais antiga universidade tecnológica dos Estados Unidos e conta com o *Tetherless World Constellation*, um grupo de pesquisa que visa a uma *web* cada vez mais acessível, em que as informações e comunicações não estejam presas a uma localização ou a um dispositivo. Desse grupo fazem parte, por exemplo, o professor James Hendler, um dos inventores da *Semantic Web*, e a professora Deborah McGuinness, uma das criadoras da *Web Ontology Language* (OWL).

Dentre os projetos de pesquisa desenvolvidos pelo grupo o RPI, o *Linking Open Government Data* (LOGD) trata da abertura e ligação de dados governamentais utilizando tecnologias de *Web Semântica*, provendo demos e tutoriais de consumo e visualização. O *Tetherless World Constellation* desenvolve, ainda, pesquisas como *Inference Web*:

A Knowledge Provenance Infrastructure, que trata da confiabilidade das informações recebidas na *web*, e *Using Knowledge to Filter Text-Based Data*, que filtra as informações vindas da *web* para determinar quais são novas e quais são repetições daquilo que já havia sido obtido.

Por último, a Southampton University, em seu departamento de Eletrônica e Ciência da Computação, conta com um grupo de pesquisa pioneiro em *web science*, uma nova disciplina que tem por objetivo prover um entendimento completo da *web* como fenômeno técnico e social.

O Web and Internet Science Research Group realiza pesquisas nos campos da Ciência da *web* e *Internet*, e seus principais focos de pesquisa são: examinar a *web* e entender seu impacto na sociedade contemporânea; explorar as mudanças propiciadas pela *web* na forma como se vive; e desenvolver serviços *web* inovadores e melhorias na forma como ela deverá funcionar no futuro. Dentre as iniciativas desenvolvidas na Southampton University, relacionadas a *Linked Data*, cita-se o *Open Data Service*, através do qual é disponibilizado acesso aos dados administrativos da universidade.

A University of Manchester conta com o *Information Management Group* (IMG), um grupo que conduz pesquisas sobre projeto, desenvolvimento e uso de dados e de sistemas de gestão do conhecimento. Suas atividades incluem pesquisas básicas em modelos e linguagens que dão suporte a algoritmos, tecnologias e arquiteturas, abordando, em particular as áreas de *Web Semântica* e *e-Science*.

No Hasso Plattner Institut, dois grupos de pesquisa desenvolvem trabalhos voltados para a área de *Linked Data*: o *Internet technologies and systems* e o *Information systems*. Dentre os projetos relacionados, citam-se o *GovWILD*, que faz a integração de dados governamentais de diversas fontes, o *Yovisto*, um motor de busca específico para vídeos de conteúdo acadêmico, e o *Mediaglobe*, que visa publicar conteúdos multimídia seguindo os princípios do *Linked Data*.

Principais palavras-chave e temáticas abordadas pelos trabalhos

Os trabalhos analisados utilizaram 286 palavras-chave. As principais foram: *Linked Data*, utilizada em 61 trabalhos; *Web Semântica*, com 35 ocorrências; *Linked Open Data*, com 17 ocorrências; RDF, citado por 11 artigos; Metadados, com 10 ocorrências; e ontologias e *web* de dados, com oito ocorrências.

Conforme pode ser observado na Figura 3 (nuvem de termos), o termo *data* (dados) foi utilizado pelos autores juntamente com outras palavras, como: *archives*, *casting*, *documentation*, *exploration*, *federation*, *handling*, *management*, *mappings*, *portability*, *preservation*, *quality* e *systems*. Outro termo em destaque é *semantic* (semântica), que vem acompanhado de termos como: *analysis*, *annotation*, *information*, *distribution*, *integration*, *mapping*, *mashup*, *navigation*, *sensor web*, *techiques*. Com isso, observou-se que os trabalhos que exploram o tema *linked data* tratam a informação como gerenciamento, rede, organização, espaço, sistema e visualização.

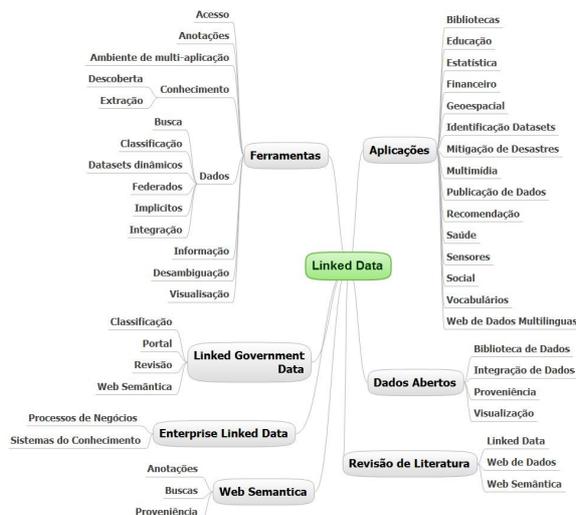
Figura 4 - Representação das palavras-chave



Fonte: Os autores.

Com as análises das palavras-chave e a leitura dos resumos dos trabalhos, foi possível identificar os macrotemas (principais temáticas) e microtemas abordados neste estudo (Figura 5).

Figura 5 - Macrotemas e microtemas abordados pelos trabalhos



Fonte: Os autores.

Exemplos de artigos que tratam da temática “Aplicações” discutem o uso de *Linked Data* nos microtemas são apresentados na Tabela 6.

Tabela 6 - Microtemas e trabalhos relacionados com a temática “Aplicações”

Microtemas	Exemplos de trabalhos
Bibliotecas	Méndez e Greenberg (2012).
Educação	Yu e colaboradores (2012).
Estatística	Salas e colaboradores (2012).
Financeiro	Xian, Jie e Hendler (2011).
Geoespacial	Becker e Bizer (2009).
Identificação de <i>Datasets</i>	Popitsch e Haslhofer (2011).
Mitigação de Desastres	Silva, Wuwongse e Sharma (2011).
Multimídia	Kamura e colaboradores (2011).
Publicação de Dados	Roure e colaboradore (2010).
Recomendação	Dolog e colaboradores (2011).
Saúde	Marshall e colaboradores (2012).
Sensores	Patni, Henson e Sheth(2010) e Pfisterer e colaboradores (2011).
Social	Mendes e colaboradores (2010).
Vocabulários	Méndez e Greenberg (2012).
<i>Web of Data</i> (WoD)	
Multilínguas	Gracia e colaboradores (2012).

Fonte: Os autores.

No macrotema “Dados abertos” os artigos tratam das seguintes temáticas apresentadas na Tabela 7.

Tabela 7 - Microtemas e trabalhos relacionados com a temática “Dados abertos”

Microtemas	Exemplos de trabalhos
Biblioteca de Dados	Parsons e colaboradores (2011).
Integração de Dados	Sall, Thiam e Lo (2011).
Proveniência	Marjit, Sarkar e Biswas (2012).
Visualização	Papantoniou e outros (2011).

Fonte: Os autores.

Os trabalhos classificados com o macrotema “Revisão de literatura” abordam os seguintes temas apresentados na Tabela 8.

Tabela 8 - Microtemas e trabalhos relacionados a temática “Revisão de literatura”

Microtemas	Exemplos de trabalhos
<i>Linked Data</i>	Bradley (2009); Bizer (2009); Bizer, Heath e Berners-Lee (2009)
<i>Web de Dados</i>	Rodriguez (2009) e Bizer e colaboradores (2009).
<i>Web Semântica</i>	Ferrara, Nikolov e Sharffe (2011).

Fonte: Os autores.

Na temática “Ferramentas” os autores abordam os temas apresentados na Tabela 9.

Tabela 9 - Microtemas e trabalhos relacionados com a temática “Ferramentas”

Microtemas	Exemplos de trabalhos
Acesso	Kopecky, Pedrinaci e Duke (2011).
Anotações	Haslhofer e outros (2011).
Descoberta de conhecimento	Chen e outros (2012).
Extração do conhecimento	Hellmann, Lehmann e Auer (2009) e Morsey e outros (2012).
Buscas de dados	Hollink, Tsirikia e De Vries (2011).
Classificação de dados	Zhang e outros(2010).
Dados de <i>datasets</i> dinâmicos	Sanderson e Van De Sompel (2012).
Dados federados	Walshe, Brennan e O'Sullivan (2012).
Dados implícitos	Farouk e Ishizuka(2012).
Integração de dados	Blanke e outros (2012).
Desambiguação	Hogan e outros (2012).
Estruturação da informação	Kim e outros (2011).
<i>Multi-application Enviroment</i>	Na e outros (2011).
Visualização	Downie e colaboradores (2011).

Fonte: Os autores.

Os artigos que tratam do macrotema “*Linked Government Data*” abordam as seguintes temáticas conforme são apresentadas na Tabela 10.

Tabela 10 - Microtemas e trabalhos relacionados com a temática “*Linked Government Data*”

Microtemas	Exemplos de trabalhos
Classificação	Kalampokis, Tambouris e Tarabanis (2011).
Portal	Hendler e colaboradores (2012)
Revisão de literatura	Casanova e outros (2012);
Web Semântica	Hoxha e Brahaj (2011).

Fonte: Os autores.

O macrotema “*Enterprise Linked Data*” aborda as temáticas conforme descritas na Tabela 11.

Tabela 11 - Microtemas e trabalhos relacionados a temática “*Enterprise Linked Data*”

Microtemas	Exemplos de trabalhos
Processos de negócios	Feng e outros (2011).
Sistemas de conhecimento	Tudhope e colaboradores (2011).

Fonte: Os autores.

Por fim, o macrotema “*Web Semântica*” aborda as seguintes temáticas que estão relatadas na Tabela 12.

Tabela 12 - Microtemas e trabalhos relacionados com a temática “*Web Semântica*”

Microtemas	Exemplos de trabalhos
Anotações	Heath e Motta (2008) e; Bonatti e outros (2011).
Buscas	Kei-Hoi e outros (2009) e Hogan e outros (2011).
Proveniência	Jun e colaboradores (2011) e Orlandi e Passant (2011).

Fonte: Os autores.

Principais referências citadas

Como embasamento teórico, os artigos analisados utilizaram 2.183 referências, o que representa uma média de aproximadamente 15 referências por trabalho. A Tabela 13 resume as principais referências utilizadas pelos autores, as quais, em seguida, serão descritas individualmente.

Tabela 13 - As 10 referências mais utilizadas pelos trabalhos analisados

	Autores	Ano	Título	Fonte de publicação	Qtd
1	Berners-Lee, T.	2006	Linked data – Design issues	Publicação online W3C Consortium Recommendations	53
2	Bizer, C., Heath, T. e Berners-Lee, T.	2009	Linked Data - The Story So Far	International Journal on Semantic Web and Information Systems (IJSWIS)	49
3	Berners-Lee, T., Hendles, J. e Lassila, O.	2001	The Semantic Web	Scientific American	22
4	Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R. e Ives, Z	2007	DBpedia: A Nucleus for a Web of Open Data	In 6th Int'l Semantic Web Conference, Busan, Korea	20
5	Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R. e Hellmann, S.	2009	DBpedia - A Crystallization Point for the Web of Data	Web Semantics: Science, Services and Agents on the World Wide Web	16
6	Bizer, C., Cyganiak, R. e Heath, T.	2007	How to Publish Linked Data on the Web	t ISWC2008 27th October 2008, Karlsruhe, Germany	15
7	Alexander, K., Cyganiak, R., Hausenblas, M, e Zhao, J.	2009	Describing Linked Datasets - On the Design and Usage of void	Linked Data on the Web Workshop (LDOW 09)	15
8	Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., Lerer, A. e Sheets, D.	2006	Tabulator: Exploring and analyzing linked data on the semantic web	<i>Proceedings of the 3rd International Semantic Web User Interaction (2006).</i>	13
9	Auer, S., Dietzold, S., Lehmann, J., Hellmann, S. e Aumueller, D.	2009	Triplify - light-weight linked data publication from relational databases	<i>WWW '09 Proceedings of the 18th international conference on World wide web Pages 621-630</i>	13
10	Klyne, G. e Carroll, J.J.	2004	Resource Description Framework (RDF): Concepts and Abstract Syntax	Publicação online W3C Consortium Recommendations	11
11	Bizer, C., Heath, T., Idehen, K. e Berners-Lee, T.	2008	Linked data on the Web	The LDOW2008 Proceedings	11

Fonte: Os autores.

O artigo mais citado, com 53 ocorrências, foi o trabalho seminal sobre o *Linked Data* de Sir. Tim Berners-Lee denominado *Linked Data – Designs Issues*. Publicado em 2006, o artigo cunhou o termo *Linked Data* e defendeu que a *Web Semântica* não é somente a publicação de dados na *web*, mas sim promover *links* para que pessoas e maquinas possam explorar a *Web* de dados. Segundo o autor, assim como a *web* de hipertexto, a *web* de dados é construída por documentos presentes na *web*. Contudo, diferentemente daquela, esses *links* são ligações entre documentos hipertexto escritos em HTML. Para promover essa ligação, o artigo apresenta quatro recomendações que ficaram conhecidas como as quatro regras para a publicação de *Linked Data*, já apresentadas na introdução deste artigo.

O segundo artigo com o maior número de citações foi *Linked Data - The Story So Far*, publicado em 2009 pelos autores Bizer, Heath

e Berners-Lee, com 49 aparições. Nesse trabalho os autores apresentam conceitos e princípios técnicos sobre *Linked Data*, além de situá-la dentro de um universo mais amplo, constituído pelos desenvolvimentos tecnológicos relacionados. São descritos, também, o processo e a publicação de aplicações *Linked Data* na *web*; é feita uma revisão sobre o que foi desenvolvido para explorar a *web* de dados; e por fim traçadas diretrizes de pesquisa para o avanço da comunidade de dados ligados.

O terceiro trabalho, com 22 citações, foi *The Semantic Web*, de Berners-Lee, Hendles e Lassila. Escrito em 2001, esse artigo apresenta o conceito da *Web Semântica* que, na época, prometia ser uma forma revolucionária de publicação de conteúdo na *web*, em que são atribuídos aos dados significados para que computadores possam compreender o que está sendo disponibilizado, o que desencadeou aquilo que hoje conhecemos por *Web 2.0*.

O quarto e o quinto artigos mais citados possuem como tema a *DBpedia*, uma iniciativa para extrair informações estruturadas da *Wikipedia*, tornando-a disponível na *web*. Escrito por Auer, Bizer, Kobilarov, Lehmann, Cyganiak e Ives, *DBpedia: A Nucleus for a Web of Open Data* descreve a extração dos conjuntos de dados *DBpedia* e como a informação resultante é publicada na *web*, para ser consumida por humanos e máquinas. Os autores descrevem algumas aplicações resultantes da *DBpedia*, como disponibilizar conteúdo, o estado atual, sua interligação com outros conjuntos de dados abertos, além de descrever como a *DBpedia* pode representar um núcleo para uma emergente rede de dados abertos. Já *DBpedia - A Crystallization Point for the Web of Data*, de Bizer, Lehmann, Kobilarov, Auer, Becker, Cyganiak e Hellmann, tem como foco principal a forma como ocorre a extração da base de conhecimento presente na *DBpedia*. Igualmente ao outro artigo, traz ainda um panorama do estado atual da *DBpedia*, sua articulação com outras fontes de dados na *web* e uma visão geral de aplicações que constituem rede de dados ao redor *DBpedia*.

O sexto artigo mais citado, com 15 ocorrências, *How to Publish Linked Data on the Web*, de Bizer, Cyganiak e Heath, fornece um tutorial sobre como publicar *Linked Data* na *web* por meio da visão geral do seu conceito, da descrição de conceitos relacionados como URIs e RDF e da descrição de diversas receitas práticas para a publicação de informações na forma de *Linked Data* na *Internet*, como, por exemplo, que vocabulários devem ser usados para representar a informação, o que deve retornar como descrição RDF para uma URI e como definir *links* RDF para outras fontes de dados. Recentemente, esse tutorial foi substituído pelo livro *Linked Data: Evolving the Web into a Global Data Space*, escrito por Heath e Bizer e publicado em 2011, já que este representa uma visão mais detalhada e atual sobre *Linked Data*.

O sétimo artigo mais citado foi *Describing Linked Datasets - On the Design and Usage of void*, de Alexander, Cyganiak, Hausenblas e Zhao, com 15 ocorrências. Nesse artigo os autores discutem a concepção e implementação do void, *Vocabulary Of Interlinked Datasets*, um vocabulário que permite descrever formalmente conjuntos de dados ligados RDF. O artigo traz casos de uso do vocabulário void, seu estado atual e suas potenciais aplicações no contexto de conjuntos de *Linked Data*.

Oitavo artigo, *Tabulator: Exploring and Analyzing linked data on the Semantic Web*, de Berners-Lee, Chen, Chilton, Connolly, Dhanaraj, Hollenbach, Lerer, e Sheets, propõe um navegador semântico capaz de navegar por dados ligados RDF que pode ser usado tanto por novos usuários, com o intuito de provocar interesse na *web*, quanto por desenvolvedores de conteúdo RDF.

O nono artigo mais referenciado, com 13 citações, foi *Triplify - light-weight linked data publication from relational databases*, de Auer, Dietzold, Lehmann, Hellmann e Aumueller. Nesse artigo os autores apresentam uma ferramenta chamada *Triplify*, uma abordagem simples mas eficiente de publicar *Linked Data* a partir de bancos de dados relacionais. Baseada no mapeamento HTTP URI das requisições de

consultas de dados relacionais, a ferramenta transforma os resultados dessas buscas em declarações RDF publicando-os na *web* em serializações RDF, em especial na forma de *Linked Data*. Segundo os autores, o desenvolvimento dessa ferramenta se justifica já que a maior parte da informação na *web* já está armazenada de forma estruturada, em bancos de dados relacionais.

Por fim, na décima posição, com 11 citações, estão os artigos *Resource Description Framework (RDF): Concepts and Abstract Syntax*, de Klyne e Carroll, e *Linked data on the Web*, de Bizer, Heath, Idehen e Berners-Lee. O primeiro apresenta a sintaxe do RDF, uma estrutura para representar as informações na *web* que é fundamental para publicação *Linked Data*. O documento define a sintaxe abstrata em que o RFD se baseia e que liga esta com a sua semântica formal. Os autores também apresentam as diferenças entre as sintaxes RDF e XML, os objetivos de desenvolvimento, conceitos-chave, a sua normalização e como é feita a manipulação de referências de URI. O segundo artigo, por sua vez, apresenta os resultados do *Linked Data on the Web Workshop* de 2008 que aconteceu em Pequim, na China. O objetivo do *workshop* foi promover um fórum em que a comunidade *Linked Data* pudesse apresentar e discutir suas abordagens para a publicação de dados na *web*.

Ao analisar-se conjuntamente o conteúdo dos artigos mais citados, nota-se que estes são artigos seminais ou introdutórios sobre *Linked Data* ou temas correlatos. Entre eles estão o artigo seminal de *Web Semântica*, do qual *Linked Data* deriva, o artigo que cunhou a expressão *Linked Data* e artigos de caráter introdutório sobre o tema. Ainda estão presentes trabalhos sobre a *web* de dados, em especial a *DBpedia*, que é a principal representação do tema, e sobre a sintaxe RDF. A predominância de artigos com esse enfoque é esperada, já que esses trabalhos constituem a base teórica para outros estudos, sejam estes avanços teóricos, sejam aplicações dos conceitos iniciais de *Linked Data*.

4 CONSIDERAÇÕES FINAIS

Este artigo apresentou os resultados de uma pesquisa bibliométrica usando o termo *Linked Data* com o objetivo de mostrar as características das pesquisas que abordam o tema. Para isso foram utilizadas técnicas bibliométricas de recuperação de informação que possibilitaram traçar um perfil abrangente sobre o tema, identificar padrões e definir algumas linhas promissoras de pesquisa na área.

A principal dificuldade encontrada no desenvolvimento deste trabalho foi a junção das informações contidas em diferentes bases científicas (EBSCO, IEEE, *Scopus*, WoS). Utilizar mais de uma base de dados representou maior abrangência, contudo, resultou em um acréscimo significativo na quantidade de trabalho para uniformizar as informações. Como as bases possuem padrões de estruturação de informação diferentes entre si, foi necessária a criação de um banco de dados auxiliar para a padronização dos dados antes de iniciar as análises. Além disso, informações como a localização dos autores, que não estavam presentes nos artigos, foram identificadas e complementadas manualmente pelos pesquisadores.

Um resultado interessante foi a grande concentração de citações de trabalhos pertencentes a um pequeno grupo de autores. Dentre os dez artigos mais referenciados (238 citações), nove contam com a participação de pelo menos um dos seguintes autores: Tim Berners-Lee, Richard Cyganiak e Christian Bizer. Isso se justifica por serem estes os autores pioneiros na área de *Linked Data* e que continuam desenvolvendo pesquisas sobre o tema no ano de 2012.

Outros resultados são o ganho na compreensão sobre as iniciativas já desenvolvidas, além de apontar o estado atual de *Linked Data* no nível internacional.

Em relação as limitações deste trabalho temos a grande quantidade de trabalhos fora de contexto que associam o termo *Linked Data* a interligação de dados sem qualquer relação com os conceitos de

Linked Data. Depois, ao analisar as referências utilizadas pelos autores, observou-se a falta de informações principalmente sobre o ano e o título dos artigos, o que impediu que algumas referências fossem utilizadas nas análises. Outra restrição foi a falta de palavras-chave informadas pelos autores dos trabalhos analisados, o que dificultou uma análise mais apurada desse tema. Além disso, muitos dos artigos selecionados nas bases de dados que poderiam ser utilizados nesta pesquisa foram rejeitados por não disponibilizarem o documento completo para uma análise mais apurada do assunto tratado. Por último, pelo fato de a pesquisa nas bases de dados ter sido realizada no mês de outubro de 2012, certamente não abarcou todos os artigos publicados neste ano, ficando a sugestão para os próximos trabalhos.

Para trabalhos futuros, recomenda-se a pesquisa em outras bases de dados incluindo as nacionais, como a Scielo. Sugere-se, também, a análise bibliométrica de cada macrotema levantado nesta pesquisa levando em consideração outros trabalhos disponibilizados nas diferentes bases de dados científicas.

Portanto, ao apresentar uma síntese das informações das principais fontes de publicações que abordam o termo *Linked Data*, construiu-se um sólido arcabouço teórico que abre portas para que novas pesquisas possam aprofundar seus estudos.

REFERÊNCIAS

ALMEIDA, Elenara Chaves Edler de. **O Portal de periódicos da Capes: estudo sobre a sua evolução e utilização**. Brasília: Universidade de Brasília, 2006.

BECKER, Christian; BIZER, Christian. Exploring the Geospatial Semantic Web with DBpedia Mobile. **Journal of Web Semantics**, Amsterdam, v.7, n.4, p.278-286, 2009.

BERNERS-LEE, Tim. **Linked Data: Design Issues**. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 5 jul. 2012.

BERNERS-LEE, Tim; FIELDING, Roy; MASINTER, L. **Uniform Resource Identifier (URI): Generic Syntax**, 2005. Disponível em:<<http://tools.ietf.org/html/rfc3986>>. Acesso em: 5 jul. 2012.

BIZER, Christian. The emerging web of linked data. **IEEE Intelligent Systems**, Los Alamitos, v. 24, n. 5, p. 87-92, 2009.

BIZER, Christian *et al.* DBpedia - A crystallization point for the Web of Data. **Journal of Web Semantics**, Amsterdam, v. 7, n. 3, p. 154-165, 2009.

BIZER, Christian *et al.* Linked data - The story so far. **International journal on Semantic Web and information systems**, Hershey, v. 5, n. 3, p. 1 - 22, 2009.

BIZER, Christian; CYGANIAK, Richard; GAUß, Tobias. The RDF book Mashup: From Web APIs to a Web of data. In: WORKSHOP ON SCRIPTING FOR THE SEMANTIC WEB (SFSW2007), 3., 2007, Austria. **Proceedings...** Austria, 2007.

BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. Linked data - The story so far. **International Journal on Semantic Web and Information Systems**, Hershey, v. 5, n. 3, p. 1-22, 2009.

BLANKE, Tobias *et al.* Linked data for humanities research — The SPQR experiment. In: IEEE INTERNATIONAL CONFERENCE: Digital Ecosystems Technologies (DEST), 6., 2012. **Anais...**Campione d'Italia: IEEE, 2012.

BONATTI, Piero A. *et al.* Robust and scalable Linked Data reasoning incorporating provenance and trust annotations. **Journal of Web Semantics**, Amsterdam, v. 9, n. 2, p. 165-201, 2011.

BRADLEY, Fiona. Discovering Linked Data. **Library Journal**, New York, v. 134, n. 7, p. 48-50, 2009.

BRAMBILLA, Sônia Domingues Santos; STUMPF, Ida Regina Chittó. Produção científica da UFRGS representada na Web of Science Scientific production of UFRGS INDEXED at Web of Science (2000-2009). **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 17, n. 3, p. 34-50, 2012.

BRICKLEY, Dan; GUHA, R. V. **RDF Vocabulary Description Language 1.0: RDF Schema**. Disponível em:<<http://www.w3.org/TR/rdf-schema>>. Acesso: 2 ago. 2012.

CASANOVA, Marco A. *et al.* Open Government Data in Brazil. **IEEE Intelligent Systems**, Los Alamitos, n. May-June, 2012, p. 45-49, 2012.

CHEN, Hua-jun *et al.* A multi-agent framework for mining semantic relations from Linked Data. **Journal of Zhejiang University: Science A**, Hangzhou, v. 13, n. 4, p. 295-307, 2012.

COSTA, Maria Teresa *et al.* A Bibliometria e a Avaliação da Produção Científica: indicadores e ferramentas. **Actas dos Congressos Nacionais de Bibliotecários, Arquivistas e Documentalistas...** Lisboa, 2012.

DING, Li *et al.* TWC LOGD: A portal for linked open government data ecosystems. **Journal of Web Semantics**, Amsterdam, v. 9, n. 3, p. 325-333, 2011.

DOLOG, Peter *et al.* Recommending open linked data in creativity sessions using web portals with collaborative real time environment. **Journal of Universal Computer Science**, Graz, v. 17, n. 12, p. 1690-1709, 2011.

DOWNIE, Marc *et al.* Evolving a rapid prototyping environment for visually and analytically exploring large-scale Linked Open Data. In: SYMPOSIUM ON LARGE DATA ANALYSIS AND VISUALIZATION, 2011. **Anais...**Seattle: IEEE, 2011.

FAROUK, Mamdouh; ISHIZUKA, Mitsuru. An inference based query engine for RDF data. In: CONFERENCE : INFORMATION RETRIEVAL & KNOWLEDGE MANAGEMENT (CAMP), 2012. **Anais...**Kuala Lumpur: IEEE, 2012.

FENG, Gao *et al.* Extending BPMN 2.0 with Sensor and Smart Device Business Functions. Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE),. **Anais...**Paris: IEEE, 2011

FERRARA, Alfio; NIKOLOV, Andriy; SCHARFFE, François. Data linking for the semantic web. **International Journal on Semantic Web and Information Systems**, Hershey, v. 7, n. 3, p. 46-76, 2011.

FIELDING, Roy *et al.* **Hypertext Transfer Protocol -- HTTP/1.1**. Disponível em: <<http://www.w3.org/Protocols/rfc2616/rfc2616.html>>. Acesso em: 3 dez. 2012.

GRACIA, Jorge *et al.* Challenges for the multilingual Web of Data. **Journal of Web Semantics**, Amsterdam, v. 11, p. 63-71, 2012.

GUEDES, VÂNIA. L. S.; BORSCHIVER, Suzana. Bibliometria: uma ferramenta estatística para a gestão da informação e do conhecimento, em sistemas de informação, de comunicação e de avaliação científica e tecnológica. In: ENCONTRO NACIONAL DE CIÊNCIAS DA INFORMAÇÃO, 7., 2006, Marília. **Anais...** Marília: Unesp, 2006

HASLHOFER, B.; SCHANDL, B. Interweaving OAI-PMH data sources with the Linked Data cloud. *International Journal of Metadata, Semantics and Ontologies*, v. 5, n. 1, p. 17-31, 2010.

HASLHOFER, Bernhard *et al.* The Open Annotation Collaboration (OAC) Model. In: WORKSHOP ON MULTIMEDIA ON THE WEB: Multimedia on the Web (MMWeb), 2011. **Anais...** Graz, Austria: MMWeb, 2011

HEATH, Tom. Linked Data - Welcome to the Data Network. **IEEE Internet Computing**, v. 15, n. 6, p. 70-73, 2011.

HEATH, Tom; MOTTA, Enrico. Revyu: Linking reviews and ratings into the Web of Data. **Web Semantics**, Amsterdam, v. 6, n. 4, p. 266-273, 2008.

HELLMANN, Sebastian; LEHMANN, Jens; AUER, Sören. Learning of OWL class descriptions on very large knowledge bases. **International Journal on Semantic Web and Information Systems**, Olney, v. 5, n. 2, p. 25-48, 2009.

HENDLER, Jim. Web 3.0 emerging. **Computer**, Long Beach, v.42, n.1, p. 111-113, 2009.

HENDLER, Jim *et al.* US government linked open data: Semantic.data.gov. **IEEE Intelligent Systems**, Los Alamitos, v. 27, n. 3, p. 25-31, 2012.

HOGAN, Aidan *et al.* Searching and browsing Linked Data with SWSE: The Semantic Web Search Engine. **Journal of Web Semantics**, Amsterdam, v. 9, n. 4, p. 365-401, 2011.

HOGAN, Aidan *et al.* Scalable and distributed methods for entity matching, consolidation and disambiguation over linked data corpora. **Journal of Web Semantics**, Amsterdam, v. 10, p. 76-110, 2012.

HOLLINK, Vera; TSIKRIKA, Theodora; VRIES, Arjen P. de. Semantic search log analysis: A method and a study on professional image search. **Journal of the American Society for Information Science and Technology**, New York, v. 62, n. 4, p. 691-713, 2011.

HOXHA, Julia; BRAHAJ, Armand. Open Government Data on the Web: A Semantic Approach. In: INTERNATIONAL CONFERENCE ON EMERGING INTELLIGENT DATA AND WEB TECHNOLOGIES, 2011. **Anais...**Tirana: EIDWT, 2011.

IEEE. **About IEEE Xplore® Digital Library**. Disponível em: <<http://ieeexplore.ieee.org/xpl/aboutUs.jsp>>. Acesso em: 29 nov. 2012.

JUN, Zhao *et al.* Extending Semantic Provenance into the Web of Data. **IEEE Internet Computing**, v. 15, n. 1, p. 40-48, 2011.

KALAMPOKIS, Evangelos; TAMBOURIS, Efthimios; TARABANIS, Konstantinos. A classification scheme for open government data: Towards linking decentralised data. **Intern. Journal of Web Engineering and Technology**, v. 6, n. 3, p. 266-285, 2011.

KAMURA, Tetsuro *et al.* Study Support and Integration of Cultural Information Resources with Linked Data. In: INTERNATIONAL CONFERENCE ON CULTURE AND COMPUTING, 2., 2011, Kyoto. **Anais...**Kyoto: IEEE, out. 2011.

KEI-HOI, Cheung *et al.* A journey to Semantic Web query federation in the life sciences. **BMC Bioinformatics**, London, v. 10, p. 1-16, 2009.

KIM, Hak-Lae *et al.* Integrating tagging into the web of data: Overview and combination of existing tag ontologies. **Journal of Internet Technology**, Taipei, v. 12, n. 4, p. 561-572, 2011.

KOBASHI, Nair Yumiko; SANTOS, Raimundo Nonato M. dos. Institucionalização da pesquisa científica no Brasil : cartografia temática e de redes sociais. **TransInformação**, Campinas, v. 18, n. 1, p. 27-36, 2006.

KOPECKY, Jacek; PEDRINACI, Carlos; DUKE, Alistair. RESTful Write-oriented API for Hyperdata in Custom RDF Knowledge Bases. In: INTERNATIONAL CONFERENCE ON NEXT GENERATION WEB SERVICES PRACTICES, 7., 2011. **Anais...**Salamanca: IEEE, 2011.

LAKSHMANAN, Geetika T. *et al.* Guest Editors' Introduction: Provenance in Web Applications. **IEEE Internet Computing**, v. 15, n. 1, p. 17-21, jan. 2011.

LASSILA, Ora; SWICK, Ralph R. **Resource Description Framework (RDF) Model and Syntax Specification**. Disponível em: <<http://www.w3.org/1998/10/WD-rdf-syntax-19981008>>. Acesso em: 28 ago. 2012.

MACIAS-CHAPULA, Cesar A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 134-140, 1998.

MANOLA, Frank; MILLER, Eric. **RDF Primer W3C Recommendation**. Disponível em: <<http://www.w3.org/TR/rdf-primer/>>. Acesso em: 2 ago. 2012.

MARJIT, Ujjal; SARKAR, Arup; BISWAS, Utpal. A novel approach to develop linked data with provenance. In: CONFERENCE COMPUTING, COMMUNICATION AND APPLICATIONS, 2012. **Anais...** Dindigul: ICCCA, 2012

MARSHALL, M. Scott *et al.* Emerging practices for mapping and linking life sciences data using RDF. **Journal of Web Semantics**, Amsterdam, v. 14, p. 2-13, 2012.

MENDES, PABLO N. *et al.* Linked Open Social Signals. In: INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE AND INTELLIGENT AGENT TECHNOLOGY, 2010. **Anais...**Toronto: IEEE, 2010.

MESQUITA, Rosa *et al.* Elaboração e aplicação de instrumentos para avaliação da base de dados Scopus. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 2, p. 187-205, ago. 2006.

MORSEY, Mohamed *et al.* DBpedia and the live extraction of structured data from Wikipedia. **Program**, v. 46, n. 2, p. 157-181, 2012.

MÉNDEZ, Eva María; GREENBERG, Jane. Datos enlazados para vocabularios abiertos: marco global de HIVE. **El profesional de la información**, Barcelona, v. 21, n. 3, p. 236-244, 2012.

NA, Ye *et al.* A linked data-driven solution to user modeling in a Multi-Application Environment. In: INTERNATIONAL CONFERENCE: Software Engineering and Service Science, 2., 2011. **Anais...Beijing**, IEEE, 2011.

OLIVEIRA, Ely Francina Tannuri de; GRACIO, Maria Cláudia Cabrini. Indicadores bibliométricos em ciência da informação: análise dos pesquisadores mais produtivos no tema estudos métricos na base Scopus. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 16, n. 4, p. 16-28, 2011.

OREN, Eyal *et al.* Sindice.com: A document-oriented lookup index for open linked data. **International Journal of Metadata, Semantics and Ontologies**, Olney, v. 3, n. 1, p. 37-52, 2008.

ORLANDI, Fabrizio; PASSANT, Alexandre. Modelling provenance of DBpedia resources using Wikipedia contributions. **Web Semantics: Science, Services and Agents on the World Wide Web**, Amsterdam, v. 9, n. 2, p. 149-164, jul. 2011.

PAPANTONIOU, Agis *et al.* A Framework for Visualizing the Web of Data: Combining DBpedia and Open APIs. In: PANHELLENIC CONFERENCE ON INFORMATICS, 16., 2011. **Anais... Kastoria**, Greece, 2011.

PARSONS, Mark A. *et al.* A conceptual framework for managing very diverse data for complex, interdisciplinary science. **Journal of Information Science**, Cambridge, v. 37, n. 6, p. 555-569, 2011.

PATNI, Harshal; HENSON, Cory; SHETH, Amit. Linked sensor data. In: INTERNATIONAL SYMPOSIUM ON COLLABORATIVE TECHNOLOGIES AND SYSTEMS. **Anais... Chicago**: IEEE, 2010.

PFISTERER, Dennis *et al.* SPITFIRE: toward a semantic web of things. **IEEE Communications Magazine**, New York, v. 49, n. 11, p. 40-48, 2011.

PIZZANI, Luciana *et al.* A arte da pesquisa bibliográfica na busca do conhecimento. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 10, n. 1, p. 53-66, 2012.

POPITSCH, N.; HASLHOFER, B.; ROOCHI, E. M. An evaluation approach for dynamics-aware applications using linked data. *International Workshop on Web Semantics (WebS 10)*. Anais...José Luis del Val: WEBS 10, 2010.

POPITSCH, Niko; HASLHOFER, Bernhard. DSNotify - A solution for event detection and link maintenance in dynamic datasets. **Journal of Web Semantics**, Amsterdam, v. 9, n. 3, p. 266-283, 2011.

RODRIGUEZ, Marko A. A Reflection on the Structure and Process of the Web of Data. **Bulletin of the American Society for Information Science and Technology**, Silver Spring, v. 35, n. 6, p. 38-43, 2009.

ROURE, David D. *et al.* The Evolution of myExperiment. In: INTERNATIONAL CONFERENCE ON E-SCIENCE. **Anais...**Brisbane, Queensland: IEEE, 2010.

SABURIT, A. Borrell; GUERRERO, M. Prendes; MANTECÓN, M. Lopez. Utilización de Ebsco, Hinari, SeCiMed y Cumed en el Sistema Nacional de Información en Salud. **ACIMED: Centro Nacional de Información de Ciencias Médicas**, La Habana, v. 17, n. 1, 2008.

SALAS, Percy E. Rivera. *et al.* Publishing Statistical Data on the Web. In: International Conference on Semantic Computing. **Anais...**Palermo: IEEE, 2012.

SALL, Ousmane; THIAM, Mouhamadou; LO, Moussa. A formal model of data integration approach based on semantic data. In: INTERNATIONAL CONFERENCE. NETWORKED: Computing and Advanced Information Management, 7., 2011. **Anais...**Gyeongju, South Korea: IEEE, 2011.

SANDERSON, Robert; VAN DE SOMPEL, Herbert. Cool URIs and dynamic data. **IEEE Internet Computing**, v. 16, n. 4, p. 76-79, 2012.

SILVA, Thushari; WUWONGSE, Vilas; SHARMA, Hitesh Nidhi. Linked Data in Disaster Mitigation and Preparedness. In: INTERNATIONAL CONFERENCE ON INTELLIGENT NETWORKING AND COLLABORATIVE SYSTEMS, 3., 2011. **Anais...**Fukuoka, Japan: IEEE, 2011.

TUDHOPE, Douglas *et al.* A STELLAR Role for Knowledge Organization Systems in Digital Archaeology. **Bulletin of the American Society for Information Science & Technology**, Silver Spring, v. 37, n. 4, p. 15-18, 2011.

WALSHE, Brian; BRENNAN, Rob. O'SULLIVAN, Declan. Correspondence pattern attribute selection for consumption of federated data sources. In: NETWORK OPERATIONS AND MANAGEMENT SYMPOSIUM (NOMS), 2012. **Anais...**Maui, Hawaii: IEEE, 2012.

XIAN, Li; JIE, Bao; HENDLER, James A. Fundamental analysis powered by Semantic Web. In: IEEE SYMPOSIUM COMPUTATIONAL INTELLIGENCE FOR FINANCIAL ENGINEERING AND ECONOMICS, 2011. **Anais...**Paris: IEEE, 2011.

YU, Hong Qing *et al.* Using Linked Data to Annotate and Search Educational Video Resources for Supporting Distance Learning. **IEEE Transactions on Learning Technologies**, v. 5, n. 2, p. 130-142, abr. 2012.

ZHANG, Jing *et al.* A novel ranking framework for linked data from relational databases. **Tsinghua Science and Technology**, v. 15, n. 6, p. 642-649, 2010.

Title

Bibliometric analysis of scientific production about Linked Data

Abstract

Introduction: Since Tim Berners-Lee coined the term Linked Data in 2006, many studies have emerged based on this initiative. Whether in the form of information access initiatives, whether in applications based on this data type, the interest in this new way to publish and consume information on the Web is increasingly in focus.

Objective: In this paper, we present a bibliometric review with the aim of contributing to a better understanding and clarification of concepts and terms associated with the area of Linked Data.

Methodology: We have used the Web of Science (WoS), Scopus, EBSCO and IEEE scientific databases, in order to find the existing studies about Linked Data until October 2012.

Results: We have mapped 148 scientific papers that address the above issue, written by 409 authors from 160 institutions from 30 different countries. Through the analysis of the articles, it was found that researches about Linked Data are

concentrated in a small authors group, suggesting future research opportunities.

Conclusions: Synthesizing the information related to the scientific publications in a given area allows building a theoretical background that supports the development of future researches.

Keywords: Linked data. Bibliometrics. Literature search. Databases.

Titulo

Análisis bibliométrico de la literatura científica acerca Linked Data

Resumen

Introtucción: Desde que Tim Berners-Lee acuñó el término Linked Data en 2006, muchos estudios han surgido acerca la base de esta iniciativa. Sea en la forma de iniciativas de acceso a la información, o sea en aplicaciones basadas en este tipo de datos, el interés por esta nueva forma de publicar y consumir información en la Web es cada vez más en el foco.

Objetivo: En este trabajo se presenta una revisión bibliométrica con el objetivo de contribuir a una mejor comprensión y aclaración de conceptos y términos asociados con el área de Linked Data.

Metodología: Se utilizó las bases de datos científicas Web of Science (WoS), Scopus, EBSCO y IEEE para localizar los estudios existentes al respecto de el tema hasta octubre de 2012.

Resultados: Fueron asignados 148 artículos científicos, escritos por 409 autores de 160 instituciones de 30 países diferentes. Se constato que la investigación sobre el tema Linked Data se concentra en un grupo pequeño de autores, lo que sugiere oportuidades y áreas para futuras investigaciones.

Conclusiones: Hacer una síntesis de la información sobre las publicaciones científicas de una determinada área permite la construcción de um marco teórico para que se desarrollen futuras investigaciones.

Palabra clave: Linked data. Bibliometría. Búsqueda bibliográfica. Bases de datos.

Recebido em: 03.05.2014

Aceito em: 17.06.2015