

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017

GT-2 – Organização e Representação do Conhecimento

CRIAÇÃO E AMPLIAÇÃO DE ONTOLOGIAS POR FOLKSONOMIAS: UMA REVISÃO SISTEMÁTICA DA LITERATURA

Jeferson Gonçalves de Oliveira - Universidade Fumec (FUMEC)

Tissiane Torres Vieira - Universidade Fumec (FUMEC)

Viviâne de Almeida Torres - Universidade Fumec (FUMEC)

Débora Sarmiento - Universidade Fumec (FUMEC)

Luiz Cláudio Gomes Maia - Universidade Fumec (FUMEC)

Fernando Silva Parreiras - Universidade Fumec (FUMEC)

CREATE AND EXTEND ONTOLOGIES BY FOLKSONOMY: A SYSTEMATIC REVIEW OF LITERATURE

Modalidade da Apresentação: Comunicação Oral

Resumo: A segunda geração de comunidades e serviços da internet, conhecida como *Web 2.0*, proporcionou velocidade, facilidade de uso, interatividade, sistemas colaborativos e vários aplicativos baseados em redes, folksonomia e *blogs*. Com isso, a quantidade de conteúdo na internet cresce exponencialmente e, com ela, a complexidade de organizar e recuperar rapidamente a informação. A motivação deste trabalho foi contribuir para a comunidade acadêmica sobre os métodos de conversão de *tags* em ontologias, visto que o conteúdo da *Web* está em contínuo crescimento e se torna um trabalho moroso. Neste artigo, fez-se a revisão sistemática da literatura, buscando responder a esse problema com uma análise de artigos identificados que informem quais técnicas, métodos e algoritmos são utilizados para a criação e ampliação de ontologias a partir de folksonomias. Foram selecionados artigos indexados, em inglês, das bases *Web of Science*, *Scielo*, *Scopus* e *Ebsco*. Como resultado, observa-se que a maioria dos métodos de conversão apresentados nos estudos são automáticos e os pesquisadores desenvolveram algoritmos próprios. Métodos como a distância de *Levenshtein* e o algoritmo matemático baseado em vetores e semelhança de cosseno foram encontrados em vários trabalhos. O *Wordnet* foi citado em doze estudos como fonte semântica para a integração com outras técnicas.

Palavras-Chave: Ontologia; Folksonomia; *Web Semântica*.

Abstract: The second generation of Internet communities and services, known as Web 2.0, provided speed, ease of use, interactivity, collaborative systems and various applications based on networks, folksonomy and blogs. With this, the amount of content on the Internet grows exponentially and with it, the complexity of quickly organizing and retrieving information. The motivation of this work was to contribute to the academic community about the methods of converting tags into ontologies, since web content is in continuous growth and becomes a slow work. In this article, a systematic review of the literature was made, seeking to answer this problem with an analysis of identified articles that inform which techniques, methods and algorithms are used for the creation and expansion of ontologies from folksonomies. We selected articles indexed in English from the Web of Science, Scielo, Scopus and Ebsco databases. As a result, it is observed that most of the conversion methods presented in the studies are automatic and the researchers have developed their own algorithms. Methods such as Levenshtein's distance and mathematical algorithm based on vectors and cosine similarity were found in several works. Wordnet was cited in twelve studies as a semantic source for integration with other techniques.

Keywords: Ontology; Folksonomy; Semantic Web.

1 INTRODUÇÃO

As ontologias são conhecimentos estruturados em que conceitos, instâncias, atributos e relações são modelados (ECHARTE *et al.*, 2007) para fornecerem um vocabulário de domínio e definirem, em diferentes níveis de formalização, o vocabulário semântico e a estrutura que relaciona os seus termos (GRUBER, 1993). Por outro lado, as *tags* são etiquetas de identificação que qualquer usuário descreve como conteúdo em repositórios de documentos digitais. Trata-se da marcação social que resulta na folksonomia (LEZCANO; GARCÍA-BARRIOCANAL; SICILIA, 2012). O termo folksonomia – *folksonomy*, em inglês – é composto a partir da junção do termo *folk* (povo, pessoas) com *taxonomy* (CATARINO; BAPTISTA, 2007). A folksonomia é o resultado da marcação de informações e objetos para a própria recuperação, feita pelos consumidores do conteúdo, cujas *tags* podem ser compartilhadas e usadas por todos os membros de uma comunidade (CHRISTIAENS, 2006).

No entanto, existem estudos que tratam da junção desses dois conceitos, começando pela análise dos termos disponíveis na *Web* (*tags*) e a sua posterior transformação em ontologias. Dessa forma, a criação de um conhecimento estruturado nas ontologias é um processo demorado que pode ser apoiado pela facilidade de uso encontrada nas folksonomias. Buscando aprofundar o estudo sobre essa relação, este trabalho fez uma revisão sistemática da literatura para responder à pergunta: quais são os métodos utilizados para a criação e ampliação de ontologias a partir de folksonomias?

A partir da análise de trabalhos indexados que propõem métodos para a criação e ampliação de ontologias a partir de folksonomias, o presente estudo tem os seguintes objetivos: i) identificar os métodos utilizados; ii) identificar as etapas no processo; iii) identificar os pesquisadores e centros de pesquisa nessa temática; iv) verificar se as abordagens encontradas possuem aplicação empírica ou teórica.

A motivação deste trabalho foi contribuir para o enriquecimento das informações sobre a conversão das folksonomias em ontologias, apontando aspectos positivos e negativos, destacando as dificuldades no processo e apresentando as abordagens utilizadas. O crescimento contínuo de conteúdo da *Web* aponta para a necessidade de identificação de um método automático que possa colaborar nessa conversão, cujo trabalho é moroso e de alto custo.

Na sequência, apresentam-se as seções de referencial teórico, metodologia, resultados, trabalhos relacionados e as considerações.

2 REFERENCIAL TEÓRICO

2.1 Ontologias, Folksonomias e a *Web*

Em 2005 as citações do termo "*Web 2.0*" no Google chegavam a 9,5 milhões e, em 2007, alcançaram 213 milhões. O termo, que faz um trocadilho com a notação em informática e indica a versão de um software, foi popularizado, em 2004, pela O'Reilly Media e pela *MediaLive International*, como denominação de uma série de conferências.

Para Tim O'Reilly (2005), um dos responsáveis por cunhar e difundir o termo *Web 2.0*, não há como demarcar precisamente as suas fronteiras. Segundo ele, trata-se de um núcleo ao redor do qual gravitam princípios e práticas que aproximam os sites que os seguem. Os princípios que têm recebido ênfase são a colaboração e a interatividade, que podem ser consideradas uma "regulamentação" para a *Web* e levam o autor a enfatizar o desenvolvimento do que chama de "arquitetura de participação", na qual o sistema informático incorpora recursos de interconexão e compartilhamento.

Uma característica marcante e que traduz a diferença entre a *Web 1.0* e *2.0*, segundo O'Reilly (2005), é a passagem da utilização de diretórios e taxonomia para a aplicação dos princípios de etiquetagem (*tagging* e folksonomia). A utilização feita por e para pessoas de *tags*, etiquetas ou palavras-chave, ganhou o nome de folksonomia. A expressão criada por Thomas Vander Wal faz uma analogia à taxonomia, indicando que é uma maneira de indexar

informações, ou seja, folks (do inglês, pessoas, povo, gente) e nomia (do grego “nomos”, regra de conduta, lei, ordem) (WAL, 2007).

Para muitos, essa história de *Web 2.0* é uma jogada de marketing, permitida pela extensão e subjetividade do conceito. Porém, antes que ela pudesse se consolidar, surgiu a versão 3.0, que também foi muito criticada. O termo criado em 2006 pelo jornalista John Markoff, do *The New York Times*, buscava definir uma geração da *Web* baseada nos conceitos de semântica de redes, pesquisas em linguagem, *data mining*, inteligência artificial, *machine learning* e *recommendation agents*. De outra forma, designa uma tendência de se atribuir à informação um significado definido para integrar computadores e pessoas, deixando para trás a *Web* sintática – na qual a interpretação dos dados disponíveis ficava a cargo das pessoas. Desde 2001, o W3C¹ (*World Wide Web Consortium*) incluiu a preocupação com a questão semântica em suas recomendações.

A *Web Semântica*, sendo fundamentada nas ontologias - cujos modelos conceituais capturam e explicitam o vocabulário utilizado e servem como base para uma comunicação objetiva - retoma o princípio das categorias taxonômicas. Contudo, utiliza mecanismos robustos de inteligência artificial baseados em linguagem *Ontology Web Language* (OWL²), esquemas de metadados, que otimizam a recuperação de informações (BREITMAN, 2005), e linguagem RDF³ (CHRISTIAENS, 2006). Sendo assim, a *Web 3.0* ou *Web Semântica* busca reverter o caos semântico trazido pelo paradigma da produção coletiva e a ontologia é vista como a tecnologia para habilitar e gerenciar essas aplicações (CHRISTIAENS, 2006; LIU, GRUEN, 2008) por meio das linguagens RDF e OWL (CHRISTIAENS, 2006). Nesse sentido, na informática, uma ontologia formal é uma especificação explícita e compartilhada de uma conceituação usando uma estrutura específica e uma linguagem de representação (GRUBER, 1993).

Com a intenção de conciliar os benefícios advindos da *Web 2.0*, colaborativa, e da *Web 3.0*, semântica, surge a preocupação em criar ou ampliar ontologias a partir de folksonomias. O Quadro 1 mostra uma comparação sintética entre taxonomias formais ou ontologias e as folksonomias e etiquetagem livre.

¹ <http://www.w3c.br/Padroes/WebSemantica>

² <https://www.w3.org/2001/sw/wiki/OWL>

³ <https://www.w3.org/2001/sw/wiki/RDF>

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

Quadro 1: Comparação entre ontologia, taxonomia e folksonomia.

CRITÉRIO	TAXONOMIAS FORMAIS OU ONTOLOGIAS	FOLKSONOMIAS E ETIQUETAGEM LIVRE
Evolução	Categorizações ou modelos são visto como estáticos e podem ser criados com antecedência.	São criados e atualizados como parte de uma atividade em andamento.
Forma	As ontologias são, muitas vezes baseadas, em estruturas hierárquicas.	As folksonomias criam “ <i>namespaces</i> ” planos.
Suporte	Estruturas hierárquicas fornecem mais expressividade e suporte para o raciocínio de vários tipos.	Menos expressividade e suporte por várias razões.
Mudanças	Estruturas hierárquicas mais sensíveis a alterações.	Menos sensíveis a mudanças.
<i>Namespaces</i>	Os “ <i>namespaces</i> ” são fechados. Os usuários escolhem quaisquer <i>tags</i> , dentre as existentes, para descrever uma entidade.	Os <i>namespaces</i> são abertos. Os usuários podem escolher as <i>tags</i> que querem utilizar para descrever uma entidade.
Informações estruturadas	Fornecem um <i>framework</i> para lidar com informações estruturadas e para extrair conclusões a partir dessa estrutura de informação.	Não fornecem uma estrutura para lidar com informações estruturadas.
Manutenção	São de difícil manutenção.	São de fácil manutenção.
Representação	No espectro de sistemas de representação do conhecimento, o que exige mais recursos no que diz respeito a criação e manutenção são as ontologias.	São mais fáceis de criar, editar, usar e reusar.
Formalismo	Requerem acordo consensual sobre os seus conteúdos por parte dos membros de uma comunidade.	Não requerem acordo consensual.
Metadados	Os metadados são gerados por <i>experts</i> .	Podem ser gerados por <i>experts</i> e pelos consumidores da informação.
Vocabulário	Utilizam vocabulário controlado.	Palavras-chave são escolhidas ao invés de se utilizar vocabulário controlado.

Fonte: Adaptado de Binzabiah e Wade (2012).

PENG *et al.* (2010) considera que os sistemas de folksonomia têm como vantagem a flexibilidade, mas aponta que, para terem efetiva utilização, seria necessário considerar a semântica das *tags*, minimizando a polissemia, sinonímia e granularidade. Esses são os desafios enfrentados pelos estudiosos que se propõem a criar ou ampliar ontologias a partir

de folksonomias. Neste trabalho, faz-se uma análise sobre métodos e algoritmos utilizados para esse fim.

2.2 RELAÇÃO ENTRE FOLKSONOMIAS E ONTOLOGIAS

Ontologias são esquemas de categorização controlados e hierárquicos, em que conceitos, instâncias, atributos e relações são modelados. Desta forma, à medida que modela e fornece um vocabulário de um domínio, a ontologia é um conjunto de conceitos que podem ser usados por agentes para dialogar com uma linguagem (ECHARTE *et al.*, 2007). O desenvolvimento da ontologia é abordado de forma centralizada e formal (LIU, GRUEN, 2008).

Segundo Gruber (1993), uma ontologia é uma especificação explícita de uma conceituação. “O termo é emprestado da filosofia, em que uma ontologia é um relato sistemático da existência. Para sistemas baseados no conhecimento, o que ‘existe’ é exatamente o que pode ser representado” (GRUBER, 1993, p.199).

Borst (1997, p.12), por sua vez, define conceituação como “uma interpretação estruturada de uma parte do mundo que as pessoas usam para pensar e se comunicar sobre o mundo” e, a partir daí, faz uma modificação sutil, mas fundamental, na definição de Gruber: “ontologias são especificações formais de conceituações compartilhadas”. Por essa característica, as ontologias tornam-se necessárias para os sistemas que almejam buscar, combinar ou integrar informações de fontes variadas.

Para Echarte *et al.* (2007, p.3), explícito refere-se ao requisito de nomear todos os conceitos e elementos do domínio; formal significa que uma linguagem formal deve ser usada; e compartilhada significa que os pontos de vista de todos os usuários envolvidos no domínio estão incluídos na ontologia.

As ontologias melhoram a comunicação entre humanos ou máquinas, uma vez que fornecem uma compreensão compartilhada de um domínio. Um exemplo de aplicação é a possibilidade de melhorar a comunicação entre os sistemas de informação (USCHOLD; GRUNINGER, 1996; FENSEL, 2003 citados por VAN DAMME; COENEN; VANDIJCK, 2008).

No entanto, a construção de ontologias com as técnicas atuais de engenharia ontológica é um processo demorado que apresenta um descompasso conceitual entre os desenvolvedores e o vocabulário dos usuários (VAN DAMME; COENEN; VANDIJCK, 2008).

As folksonomias são o resultado de um processo de anotação colaborativa, são dados anotados pelos usuários. Elas são compostas de rótulos de texto livre - conjunto de etiquetas (*tags*) - e um conjunto de recursos associados a essas *tags*, que são atribuídas a objetos heterogêneos como recursos da *Web*, imagens, documentos, vídeos, etc. (ECHARTE *et al.*, 2007). Elas são utilizadas em ambientes *Web* como Flickr⁴ e Del.icio.us⁵, nos quais os usuários anotam imagens e *links*.

A *tag* é um recurso de organização da informação; existe para ajudar os usuários a armazenar e gerenciar seus recursos informacionais e fornecer uma plataforma para compartilhá-los. O usuário marca, de acordo com sua necessidade, os recursos que adicionaram a um vocabulário conhecido. É uma forma de classificação que não utiliza taxonomia ou vocabulário preestabelecido, na qual o espaço de cada usuário é compartilhado e, portanto, há influência mútua (WANG; JINDONG, 2013).

Segundo Echarte *et al.* (2007), nesse processo de anotação não há qualquer restrição, hierarquia predefinida ou restrição para definir as *tags* e, portanto, depende dos critérios dos usuários. Para o autor, a ausência de estrutura na classificação da informação é fator de sucesso das folksonomias.

Ao analisar as folksonomias e as ontologias, pode-se observar que o sucesso das folksonomias advém da facilidade de uso. Mas, o conhecimento estruturado em ontologias pode ser processado de forma eficiente, pois permite conclusões elaboradas devido ao uso de raciocinadores (ECHARTE *et al.*, 2007). As folksonomias são catálogos *bottom-up* de *tags* e ontologias são especificações *top-down* legíveis por máquina de conhecimento de domínio, usando URL's para todos os elementos de dados, propriedades e tipos de relacionamento (CHRISTIDIS; MENTZAS; APOSTOLOU, 2012). As folksonomias e ontologias são duas maneiras de organizar o conhecimento na *Web* atual (ECHARTE *et al.*, 2007).

3 METODOLOGIA

Para a realização da revisão sistemática da literatura foi adotado o processo proposto por Kitchenham (2004), que se divide em três etapas distintas: Planejamento da Revisão, Condução da Revisão e Análise dos Resultados.

⁴ <https://www.flickr.com/>

⁵ <https://del.icio.us/>

3.1 Planejamento da Revisão

Na fase de planejamento foi definido um protocolo para a execução da Revisão Sistemática da Literatura que seguiu as etapas descritas na sequência:

a) Descrição dos objetivos: identifica os estudos que contenham os métodos que criam ou ampliam ontologias a partir de folksonomias.

b) Elaboração das questões de pesquisa: o trabalho considerou as seguintes questões para subsidiar a busca:

01) Quais são os métodos utilizados para a criação e ampliação de ontologias por folksonomias?

02) Quais são as etapas para a criação ou ampliação de uma ontologia por folksonomias?

03) Quais são os principais pesquisadores e centros de pesquisa que tratam dessa temática?

04) Quais dessas abordagens têm aplicação empírica ou teórica?

c) A estratégia de busca utilizada neste trabalho baseia-se em dois critérios:

- Seleção das fontes de pesquisa: as bases que serviram como fontes para obtenção dos estudos foram: *Web of Science*, *Scielo*, *Scopus* e *Ebsco*.
- Elaboração da *string* de busca: foi elaborada uma *string* de busca que levou em consideração as palavras ontologia e folksonomia. Esses termos foram traduzidos para a língua inglesa e a *string* de busca padrão foi: “((ontology) AND (folksonomy))”.

d) Adoção de critérios para a inclusão e exclusão dos trabalhos selecionados:

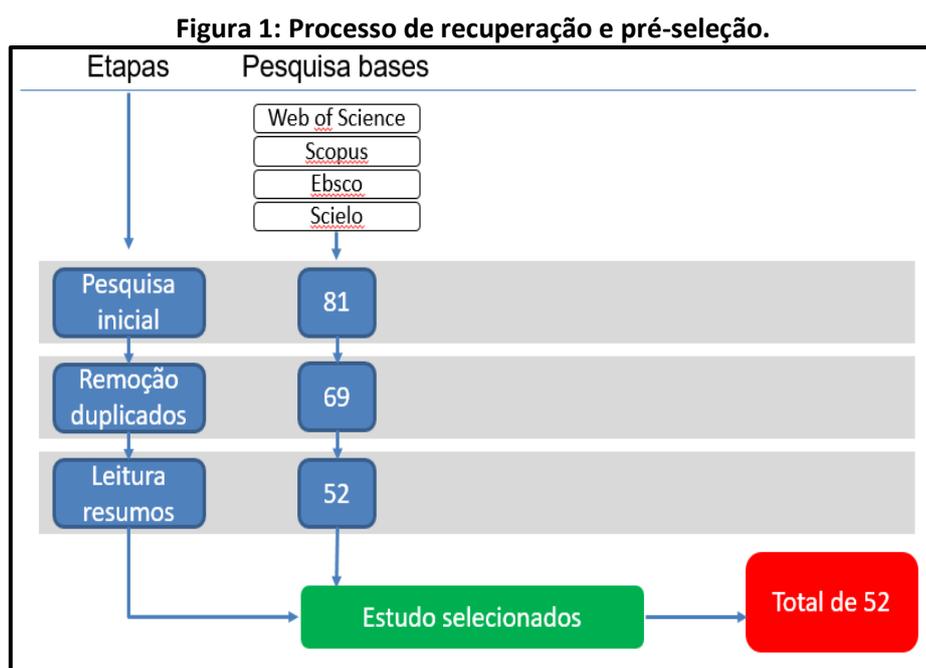
- Para a inclusão dos estudos: as publicações devem estar disponíveis na *Web*, nas bases citadas, na língua inglesa e devem representar metodologias ou métodos que tratem da criação ou ampliação de ontologias por folksonomias de forma analítica e que respondam a qualquer uma das questões de pesquisa.
- Critérios para a exclusão dos estudos: estudos duplicados e que não abordem metodologias para a ampliação ou criação de ontologias por folksonomias ou não respondam as questões de pesquisa.

3.2 Condução da Revisão Sistemática

A condução da revisão sistemática observou os passos: as *strings* de busca foram executadas nas fontes selecionadas; os estudos primários foram identificados; os documentos retornados foram selecionados de acordo com os critérios de inclusão/exclusão e avaliados segundo os critérios de qualidade estabelecidos durante o planejamento.

3.2.1 Processo para recuperação de pré-seleção dos estudos primários

Nesta etapa foram recuperados os trabalhos preliminares e realizado um processo de filtro nos estudos. Foram encontrados 81 estudos e, após a remoção dos duplicados, obteve-se um total de 69 estudos, dos quais resultaram 59, após a leitura dos títulos e dos resumos. Sete estudos estavam incompletos e essa fase terminou com 52 estudos selecionados conforme Figura 1.



Fonte: Elaborado pelos autores - 2017.

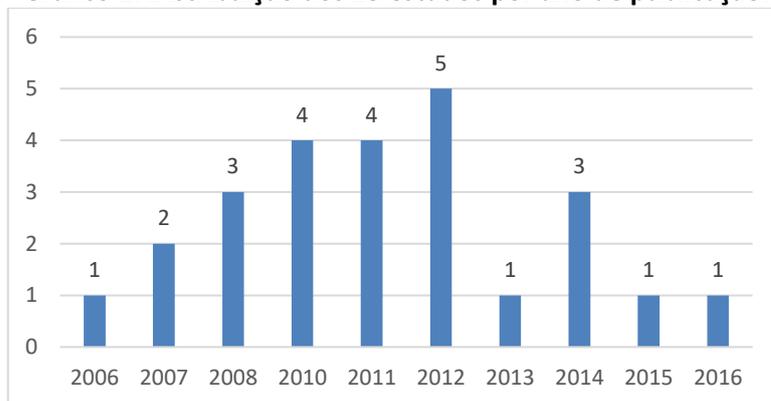
3.2.2 Processo de seleção dos estudos

Após a recuperação e pré-seleção, os artigos recuperados foram analisados pela leitura do seu conteúdo. Marconi e Lakatos (2003) citam que essa avaliação detalhada garante a qualidade dos estudos selecionados. Dessa forma, os 52 documentos recuperados foram analisados e, destes, foram selecionados 25 artigos.

4 RESULTADOS

Em relação ao ano de publicação, observa-se uma concentração nos anos de 2010, 2011 e 2012 conforme mostrado no Gráfico 1:

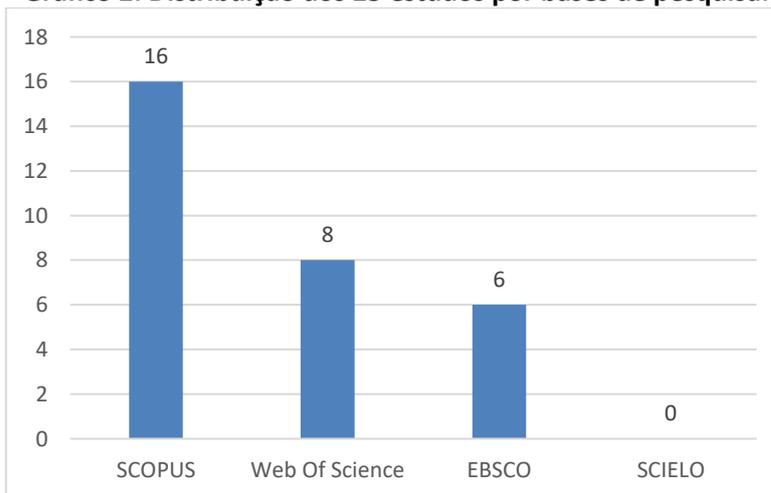
Gráfico 1: Distribuição dos 25 estudos por ano de publicação.



Fonte: Elaborado pelos Autores – 2017.

A distribuição dos artigos pelas bases pesquisadas evidencia que a *Scopus* possui o maior número de artigos encontrados para o tema pesquisado (Gráfico 2). A base *Scielo* não possui artigos selecionados. A soma resulta em um número maior que 25 porque houve artigos duplicados em mais de uma base.

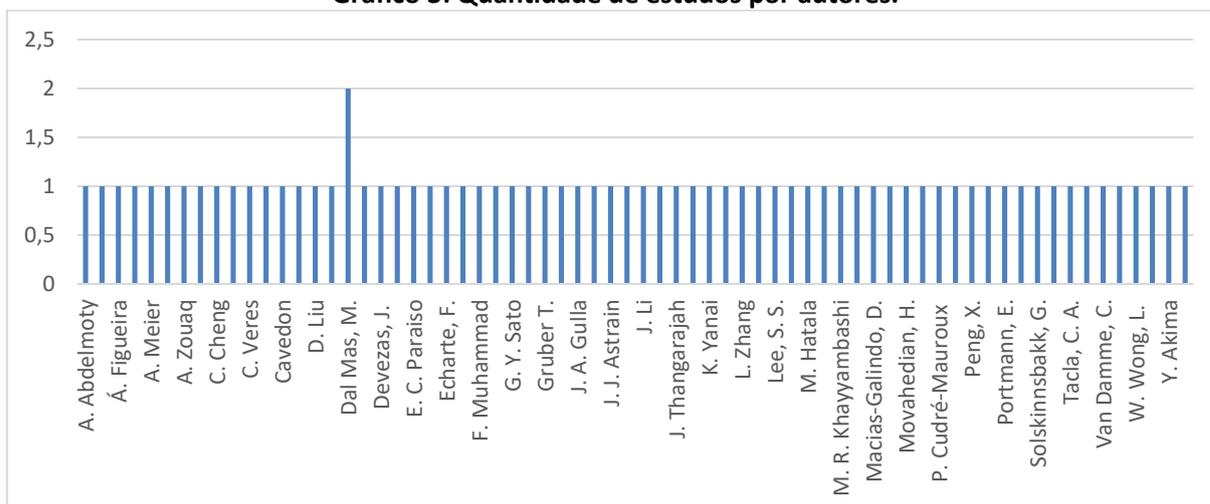
Gráfico 2: Distribuição dos 25 estudos por bases de pesquisa.



Fonte: Elaborado pelos autores – 2017.

Considerando os artigos selecionados, somente um autor possui mais de um estudo, ou seja, não é possível identificar um pesquisador como referência do tema analisado (Gráfico 3).

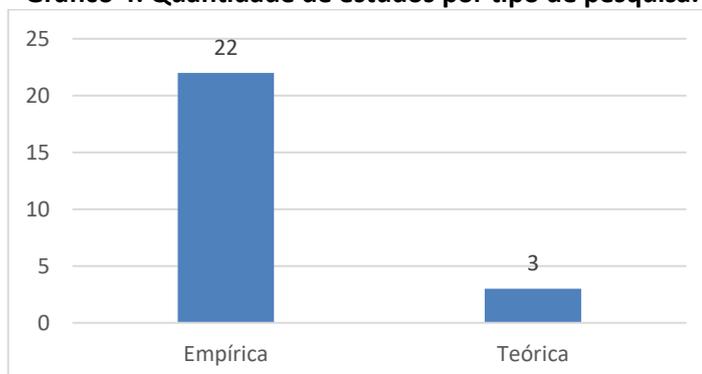
Gráfico 3: Quantidade de estudos por autores.



Fonte: Elaborado pelos Autores – 2017.

Nota-se uma dispersão de autores e de centros de pesquisa sobre o tema. Por fim, 22 dos 25 trabalhos selecionados são do tipo empírico e permitem uma aplicação prática para a solução de problemas relacionados à utilização das folksonomias para enriquecimento ou criação de ontologias. Inclusive, a maioria dos 22 trabalhos empíricos demonstra uma aplicação prática como resultado dos estudos (Gráfico 4).

Gráfico 4: Quantidade de estudos por tipo de pesquisa.



Fonte: Elaborado pelos autores – 2017.

Após a verificação dos 25 estudos selecionados (Quadro 2), foi realizada uma análise comparativa com o intuito de proporcionar uma visão sobre as abordagens utilizadas em

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP**

cada um dos documentos, facilitando a escolha da metodologia adequada ao processo de criação ou ampliação de ontologias por folksonomias.

Quadro 2: Os 25 estudos selecionados para a Revisão Sistemática da Literatura.

REFERÊNCIA	CARACTERÍSTICA
CHRISTIAENS, S. (2006)	Ampliação manual de uma ontologia base por folksonomia e taxonomias em que os especialistas ajudam com o contexto semântico.
ECHARTE, F., et al. (2007)	Os autores apresentaram um método que consiste em uma ontologia usada para representar qualquer tipo de folksonomia e um algoritmo para transformar as folksonomias na ontologia proposta. Ele atualiza a ontologia resultante à medida que a folksonomia evolui no tempo.
GRUBER, T. (2007)	Trabalho que é referenciado na área. O autor aborda uma forma de pensar em ontologias e folksonomias, aplicando as duas ideias de forma conjunta.
LEE, S. S. & YONG, H. S. (2008)	Os autores têm o objetivo de melhorar a forma de anotação para recuperação de imagens. Eles propõem uma ontologia genérica para ajudar os usuários com as <i>tags</i> (<i>Ontosonomy</i>). Propõem o uso da inteligência coletiva para enriquecer ontologias.
LIU, J. & GRUEN, D. M. (2008)	Propõem uma abordagem na qual os usuários ampliam a ontologia de forma manual e colaborativa através da ferramenta <i>Collaborative Reasoning and Analysis Framework</i> (CRAFT). As ontologias resultantes mostraram uma combinação de aspectos das folksonomias e ontologias.
VAN DAMME, C.; COENEN, T. & VANDIJCK, E. (2008)	Trabalham com uma abordagem em 6 passos que incluem várias técnicas. Na abordagem eles distinguiram dois para enriquecer folksonomias ou transformar folksonomias em ontologias: (1) encontrar <i>tags</i> e (2) encontrar conceitos e relações entre <i>tags</i> . Na etapa 1, foi aplicada a técnica de métrica de Levenshtein. Na etapa 2, foi aplicada as técnicas de co-ocorrência, função de probabilidade condicional, técnicas de redução transitiva e visualização.
DAUD, A., et al. (2010)	Os autores trabalham a modelagem ontologias por folksonomias e levam em consideração a latência semântica que considera a semântica, os atores, os conceitos e os recursos. Para isso, o trabalho desenvolveu algoritmos.
KAWAKUBO, H; AKIMA, Y; & YANAI, K (2010)	Propõem um método de 4 etapas para construir uma ontologia por folksonomias em uma base de dados de imagens: gerar vetores gráficos, vetoriais e combinados usando <i>Probabilistic Latent Semantic Analysis</i> (PLSA), remover imagens com ruídos para cada conceito, computar vetores de conceito e construir um grafo acíclico (DAG).
MØLLER, T. S; VERES, C; & NÆSS, B. (2010)	Apresentam uma abordagem manual para ampliação de ontologias por folksonomias. Usuários podem adicionar <i>tags</i> à instâncias existentes na ontologia.
PENG, X., et al. (2010)	Propõem uma solução baseada em ontologia que usa a folksonomia conjuntamente para oferecer uma estrutura flexível. O trabalho usa o <i>Resource Description Framework</i> (RDF) e a <i>Web Ontology Language</i> (OWL) para descrever o conhecimento gerado.
GAŠEVIĆ, D. et al. (2011)	O trabalho foca na evolução e manutenção de uma ontologia por folksonomias em um ambiente de <i>e-learning</i> . A evolução é feita por intervenção manual de alunos e professores e usa um algoritmo para auxiliar na manutenção da ontologia.
MACÍAS-GALINDO, D., et al. (2011)	Os autores utilizaram a ferramenta <i>Modular Knowledge Bases</i> (MKBUILD) para extrair conceitos para construção de ontologias. A ferramenta utiliza 3 estágios: definição do domínio, criação da camada MKB e extensão do MKB.
PETRUCCO, C. (2011)	O trabalho propôs um ambiente participativo " <i>EduOntoWiki</i> ", no qual especialistas acadêmicos no campo das Ciências da Educação desenvolvem ontologias que os membros de várias comunidades de prática (professores, formadores, etc.) podem modificar e integrar usando uma folksonomia e abordagem narrativa (<i>storytelling</i>).
TACLA, C. A., et al. (2011)	Cita formas de aquisição de conhecimento por ampliação de ontologias. Essa ampliação é feita por folksonomias criadas pela equipe e que são matéria prima para a aplicação de ontologia. A ferramenta utilizada é o KACSA (Agente de suporte,

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

	conceitualização e aquisição do conhecimento).
BINZABIAH, R. & WADE, S. (2012)	Construção automática de uma ontologia para recuperação de filmes baseada em folksonomias. Utiliza três fases: aquisição, conceitualização e avaliação.
DAL MAS, M. (2012)	Construção de uma ontologia a partir de relações entre folksonomias em que os usuários resolvem problemas inerentes a um vocabulário de <i>tags</i> .
DEVEZAS, J. & FIGUEIRA, A. (2012)	Construção automática de um modelo organizado por folksonomias retiradas de fragmentos de notícias. Utiliza o sistema “ <i>Breadcrumbs</i> ” que possui três fases: coleta, organização e agregação.
DOU, Y; HE, J. & LIU, D. (2012)	O trabalho utiliza o algoritmo de agrupamento espectral (SCA) que se baseia no conceito de semelhança entre ponto em vez de distância para construção automática da ontologia. Possui 3 fases que são: pré-processamento das <i>tags</i> , aplicação do SCA e construção da estrutura hierárquica.
SOLSKINNSBAKK, G., <i>et al.</i> (2012)	Apresenta uma abordagem para avaliar a qualidade das relações hierárquica em ontologias e estruturas baseadas em folksonomias. Utiliza um algoritmo matemático baseado em vetores e semelhança de cosseno.
ALVES, H. & SANTANCHÊ, A. (2013) ⁶	Os autores propõem uma abordagem de fusão (combinação de dados) em que a semântica viaja em ambas direções – da folksonomia à ontologia e vice-versa. O resultado é a Ontologia Folksonomizada (FO). A técnica tem 3 etapas descritas como: Extração, Enriquecimento e Evolução. O algoritmo utilizado para mapear <i>tags</i> em conceitos ontológicos foi desenvolvido no trabalho e é utilizado na fase de enriquecimento.
DAL MAS, M. (2014)	Construção de uma ontologia a partir de relações entre folksonomias em que os usuários resolvem problemas inerentes a um vocabulário de <i>tags</i> .
ELGINDY, E. & ABDELMOTY, A. (2014)	O trabalho tem como objetivo descobrir a relação semântica entre <i>tags</i> geolocalizadas para a criação de um perfil de usuário de forma automática. Utiliza 4 fases: limpeza das <i>tags</i> usando o método da distância de <i>Levenshtein</i> , identificação dos tipos, criação da ontologia usando o método cosseno de similaridade e criação do perfil por grafos.
MOVAHEDIAN, H. & KHAYYAMBASHI, M. R. (2014).	Parte do trabalho diz respeito à proposta de construção/ampliação de ontologias. O método é dividido em 3 partes: filtragem, mapeamento para conceitos ontológicos e unificação de sinônimos. A primeira fase utiliza o <i>Google Web Search Engine</i> e uma base de conhecimento semântico denominada YAGO2. A terceira fase usa um algoritmo de singularização personalizado.
PORTMANN, E., <i>et al.</i> (2015)	Os autores propõem um <i>framework</i> com arquitetura de 3 camadas: pesquisa de reputação, base de conhecimento e exibição. A criação automática da ontologia se dá na primeira camada e utiliza o algoritmo <i>metaphone</i> para normalizar as <i>tags</i> , o algoritmo <i>plotting points</i> para calcular a similaridade e o algoritmo <i>flake</i> para determinar a proximidade das <i>tags</i> .
FANG, Q., <i>et al.</i> (2016)	Os autores propõem um <i>framework</i> de 3 fases para converter folksonomias em ontologias automaticamente: descoberta de conceitos, extração de relacionamentos e construção da hierarquia. Utilizam o <i>Kernel Density Estimation</i> (KDE) na fase 1, algoritmos de similaridade na fase 2 e algoritmo de grafo acíclico dirigido (DAG) para construção do grafo na fase 3.

Fonte: Elaborado pelos autores – 2017.

No estudo, observou-se que a maioria dos trabalhos demonstra uma abordagem automática para o processo de conversão das folksonomias em ontologias. Dos 25 trabalhos selecionados, sete deles utilizaram técnicas que preconizam a intervenção dos usuários para

⁶ O artigo recuperado na *string* de busca (*Ontology AND Folksonomy*) foi: WANG, Shufeng *et al.* *An ontology evolution method based on folksonomy*; no entanto, trata-se de plágio do artigo: ALVES, Hugo; SANTANCHÊ, André. *Folksonomized ontology and the 3E steps technique to support ontology evolution*. Por esse motivo considerou-se adequado utilizar o segundo artigo.

enriquecimento semântico das *tags* que são utilizadas para ampliar ou criar ontologias. Os 17 estudos restantes abordam formas automáticas de fazer essa conversão.

Desses trabalhos, que demonstram técnicas automatizadas, observa-se que 10 deles utilizam três ou mais fases no processo. O número de fases depende da finalidade do trabalho. Os que lidam com extração de *tags* de bases de imagens utilizam um número maior de fases no processo. Pode-se concluir que as três fases mais comumente encontradas na literatura são:

- a) **Primeira fase:** utilizada para aquisição, extração e coleta das *tags*. Nessa fase o objetivo é fazer a coleta dos termos a serem trabalhados;
- b) **Segunda fase:** utilizada para agregar semântica, organizar, mapear conceitos e encontrar relações entre os termos. Nessa fase o objetivo é organizar as folksonomias;
- c) **Terceira fase:** utilizada para exibir os grafos da ontologia. Nessa fase é feita a construção da hierarquia de conceitos.

Na maioria dos trabalhos foram desenvolvidos algoritmos. Porém, nota-se que técnicas matemáticas se repetem nos estudos. O método denominado distância de *Levenshtein* foi descrito em três trabalhos para a limpeza e a organização dos termos. O algoritmo matemático baseado em vetores e semelhança de cosseno foi utilizado em quatro trabalhos para busca de similaridade entre as *tags*.

Algoritmos de mineração de dados, como o algoritmo de agrupamento espectral (SCA), foram utilizados em conjunto com bases de dados semânticas para a criação das relações hierárquicas. Dessas bases, o *Wordnet* foi citado em 12 trabalhos como a fonte semântica utilizada para integração com outras técnicas. Há autores que utilizam *frameworks* ou sistemas no processo, e, dessa forma, inviabilizam a análise da abordagem utilizada por cada um deles.

5 TRABALHOS RELACIONADOS

Este trabalho possui poucos aspectos comuns com outros estudos encontrados na literatura. No caso de revisões sistemáticas da literatura (RSL), não foram encontrados trabalhos que abordassem o tema, embora o estudo de Verdonck *et al.* (2015), que demonstra uma RSL sobre a modelagem conceitual orientada por ontologia, tenha pontos consonantes com esta pesquisa. Os autores relataram dificuldade em encontrar estudos

empíricos e descreveram a falta de métodos de avaliação experimentais. Apesar de trabalhar com outro tema, neste estudo, a maioria dos artigos selecionados demonstrou uma aplicação empírica para a criação ou ampliação de ontologias por folksonomias e parte deles resultam de experimentos realizados. Apesar dos estudos tratarem do tema “ontologias”, as RSL chegaram a conclusões distintas nesses quesitos.

6 CONSIDERAÇÕES

A partir da necessidade de organização e de recuperação do conteúdo disponível na *Web*, observa-se o desafio de minimizar a polissemia, sinonímia e granularidade dos termos utilizados nos sistemas colaborativos. Uma proposta é a criação e/ou ampliação de ontologias a partir das folksonomias e o objetivo deste trabalho foi realizar uma revisão sistemática da literatura sobre o tema para identificar métodos, etapas, pesquisadores e se essas abordagens teriam aplicações práticas ou teóricas.

Na revisão sistemática notou-se que os trabalhos são da década de 2006-2016 e que a maioria dos autores trabalha de forma isolada, sem vínculo aos estudos anteriores e sem validação de um algoritmo por pelo menos dois autores, ainda que haja uma repetição de técnicas matemáticas e estatísticas. O método denominado distância de *Levenshtein* foi descrito para a limpeza e organização de termos em três dos trabalhos analisados e o algoritmo baseado em vetores e semelhança de cosseno, para identificar similaridade entre as *tags*, foi utilizado em quatro dos estudos que compõem esta revisão sistemática de literatura. A utilização de uma base de dados semântica para o estabelecimento das relações hierárquicas foi recorrente e o *Wordnet* foi utilizado pela maioria dos autores como a fonte semântica para integração com outras técnicas.

Na análise dos trabalhos selecionados para esta revisão sistemática de literatura, verificou-se que há uma tendência para a conversão automática com, no mínimo, três fases: 1) aquisição ou coleta dos termos; 2) mapeamento de conceitos e relações e 3) hierarquia de conceitos.

As pesquisas analisadas sugerem que o processo automático é vantajoso devido ao conteúdo que cresce continuamente na *Web* e, para isso, demandam que os algoritmos sejam aperfeiçoados e os estudos sejam compartilhados para que as pessoas contribuam a fim de melhorar o processo.

REFERÊNCIAS

ALVES, H.; SANTANCHÈ, A. Folksomized ontology and the 3E steps technique to support ontology evolution. **Web Semantics: Science, Services and Agents on the World Wide Web**, pp. 19-30, 18 (2013).

BINZABIAH, R.; WADE, S. Proposed method to build an ontology based on Folksonomy. In: **Information Society (i-Society), 2012 International Conference on**. IEEE, p. 441-446, 2012.

BORST, Willem Nico. **Construction of engineering ontologies for knowledge sharing and reuse**. Universiteit Twente, 1997. Disponível em: <<http://doc.utwente.nl/17864/1/t0000004.pdf>>. Acesso em 30 de junho de 2017.

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

CATARINO, Maria Elisabete; BAPTISTA, Ana Alice. Folksonomia: um novo conceito para a organização dos recursos digitais na *Web*. **DataGramaZero-Revista de Ciência da Informação**, v. 8, n. 3, p. 1-20, 2007.

CHRISTIAENS, S. Metadata mechanisms: From ontology to folksonomy... and back. In: **On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops**. Springer Berlin/Heidelberg, p. 199-207, 2006.

CHRISTIDIS, Konstantinos; MENTZAS, Gregoris; APOSTOLOU, Dimitris. Using latent topics to enhance search and recommendation in Enterprise Social Software. **Expert Systems with Applications**, v. 39, n. 10, p. 9297-9307, 2012.

DAL MAS, M. Elastic adaptive ontology matching on evolving folksonomy driven environment. In: **Evolving and Adaptive Intelligent Systems (EAIS), 2012 IEEE Conference on**. IEEE, p. 35-40, 2012.

DAL MAS, M. Elastic adaptive dynamics methodology on ontology matching on evolving folksonomy driven environment. **Evolving Systems**, v. 1, n. 5, p. 33-48, 2014.

DAUD, A. *et al.* Modeling ontology of folksonomy with latent semantics of tags. In: **Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on**. IEEE, p.516-523, 2010.

DEVEZAS, J.; FIGUEIRA, A. Finding language-independent contextual supernodes on coreference networks. **IAENG International Journal of Computer Science**, v. 39, n. 2, p.200-207, 2012.

DOU, Y.; HE, J.; LIU, D. A method for ontology construction derived from folksonomy. **International journal of services technology and management**, v. 18, n. 1-2, p. 88-101, 2012.

ECHARTE, F. *et al.* Ontology of Folksonomy: A New Modelling Method. **SAAKM**, v. 289, p. 36, 2007.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

ELGINDY, E.; ABDELMOTY, A. Enriching user profiles using geo-social place semantics in geofolksonomies. **International Journal of Geographical Information Science**, v. 28, n. 7, p. 1439-1458, 2014.

FANG, Q. *et al.* Folksonomy-based visual ontology construction and its applications. **IEEE Transactions on Multimedia**, v. 18, n. 4, p. 702-713, 2016.

GAŠEVIĆ, D. *et al.* An approach to folksonomy-based ontology maintenance for learning environments. **IEEE Transactions on Learning Technologies**, v. 4, n. 4, p. 301-314, 2011.

GRUBER, T. A Translation approach to portable Ontology Specifications. **Knowledge Acquisition**, 5(2):199-220, 1993.

GRUBER, T. Ontology of Folksonomy: A Mash-up of Apples and Oranges. **International Journal on Semantic Web & Information Systems**, v.3, n.1, 2007.

KAWAKUBO, H.; AKIMA, Y.; YANAI, K. Automatic construction of a folksonomy-based visual ontology. In: **Multimedia (ISM), 2010 IEEE International Symposium on**. IEEE, p.330-335, 2010.

KITCHENHAM, B. **Procedures for Performing Systematic Reviews**, 2004. Joint Technical Report Software Engineering Group, Keele University, United Kingdom and Empirical Software Engineering, National ICT Australia Ltd, Australia, 2004.

LEE, S.; YONG, H. Ontosonomy: Ontology-based extension of folksonomy. In: **Semantic Computing and Applications, 2008. IWSCA'08. IEEE International Workshop on**. IEEE, p.27-32, 2008.

LEZCANO, Leonardo; GARCÍA-BARRIOCANAL, Elena; SICILIA, Miguel-Angel. Bridging informal tagging and formal semantics via hybrid navigation. **Journal of Information Science**, v. 38, n.2, p. 140-155, 2012.

LIU, J.; GRUEN, D. Between ontology and folksonomy: a study of collaborative and implicit ontology evolution. In: **Proceedings of the 13th international conference on Intelligent user interfaces**. ACM, p.361-364, 2008.

MACÍAS-GALINDO, D. *et al.* Using a lexical dictionary and a folksonomy to automatically construct domain ontologies. **AI 2011: Advances in Artificial Intelligence**, p. 638-647, 2011.

MARCONI, M.; LAKATOS, E. **Metodologia científica: ciência e conhecimento científico**. 5. ed. Atlas, São Paulo, 2003.

MØLLER, T. S.; VERES, C.; NÆSS, B. Position Paper on User Input in Ontology Alignment-Folksonomy Based Alteration of an Interactive City Walk. In: **Complex, Intelligent and Software Intensive Systems (CISIS), 2010 International Conference on**. IEEE, p. 998-1001, 2010.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

MOVAHEDIAN, H.; KHAYYAMBASHI, M.R. Folksonomy-based user interest and disinterest profiling for improved recommendations: An ontological approach. **Journal of Information Science**, v. 40, n. 5, p. 594-610, 2014.

O'REILLY, T. **What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software**, March 9, 2005. *O'Reilly* [site]. Disponível em: <<http://www.oreilly.com/pub/a/Web2/archive/what-is-Web-20.html>> Acesso em 30 de junho de 2017.

PENG, X. *et al.* A folksonomy-ontology-based digital gazetteer service. In: **Geoinformatics, 2010 18th International Conference on**. IEEE, p. 1-6, 2010.

PETRUCCO, C. Learning about evaluation and assessment: teacher's use of folksonomies and ontologies in an online narrative environment. **Journal of Computer Assisted Learning**, v. 27, n. 5, p.399-410, 2011.

PORTMANN, E. *et al.* FORA—A fuzzy set based framework for online reputation management. **Fuzzy sets and systems**, v. 269, p. 90-114, 2015.

SOLSKINNSBAKK, G. *et al.* Quality of hierarchies in ontologies and folksonomies. **Data & Knowledge Engineering**, v. 74, p. 13-25, 2012.

TACLA, C. A. *et al.* Supporting small teams in cooperatively building application domain models. **Expert Systems with Applications**, v. 38, n. 2, p. 1160-1170, 2011.

VAN DAMME, C.; COENEN, T.; VANDIJCK, E.. Turning a Corporate Folksonomy into a Lightweight Corporate Ontology. **BIS**, v. 7, p. 36-47, 2008.

VERDONCK, Michaël *et al.* Ontology-driven conceptual modeling: A systematic literature mapping and review. **Applied Ontology**, v. 10, n. 3-4, p. 197-227, 2015.

WAL, T. V. **Folksonomy** (2007). Disponível em: <<http://vanderwal.net/folksonomy.html>>. Acesso em 30 de junho de 2017.

WANG, X.; JINDONG, C. The Standardization of Folksonomy Based on Ontology Technology. In: **Advanced Materials Research**. Trans Tech Publications, Switzerland, V. 655-657, p. 1830-1833, 2013.

World Wide Web Consortium [W3C]. **Web Semântica**, 2011. Disponível em: <<http://www.w3c.br/Padroes/WebSemantica>> Acesso em 05 de julho de 2017.