

A INTEGRAÇÃO ENTRE FOLKSONOMIAS E ONTOLOGIAS:
uma análise qualiquantitativa

*THE INTEGRATION OF FOLKSONOMIES AND ONTOLOGIES: a
qualiquantitative analysis*

 Juliana Assis¹
 Vanessa da Costa²
 Isaura Nogueira³

¹ Doutora em Ciência da Informação pela Universidade Federal de Minas Gerais (UFMG). Professora adjunta na Universidade Federal do Rio de Janeiro (UFRJ).

E-mail: juliana.assis@facc.ufrj.br

² Graduanda em Biblioteconomia e Gestão de Unidades de Informação Universidade Federal do Rio de Janeiro (UFRJ). Bolsista Bolsista de Iniciação Científica pela UFRJ.

E-mail: vanessacostabiblio@gmail.com

³ Graduanda em Biblioteconomia e Gestão de Unidades de Informação Universidade Federal do Rio de Janeiro (UFRJ). Bolsista de Iniciação Científica pelo CNPq.

E-mail: nogueiraisaas@gmail.com

 ACESSO ABERTO

Copyright: Esta obra está licenciada com uma Licença Creative Commons Atribuição 4.0 Internacional. 

Conflito de interesses: Os autores declaram que não há conflito de interesses.

Financiamento: Não há.

Declaração de Disponibilidade dos dados: Todos os dados relevantes estão disponíveis neste artigo.

Recebido em: 06 maio 2024.

Aceito em: 31 jul. 2024.

Publicado em: 04 nov. 2024.

Como citar este artigo:

ASSIS, J.; COSTA, V.; NOGUEIRA, I. A integração entre folksonomias e ontologias: uma análise qualiquantitativa. **Informação em Pauta**, Fortaleza, v. 9, p. 1-17, 2024.

RESUMO

Sistemas de Organização do Conhecimento são esquemas de cunho conceitual e terminológico utilizados para representar e organizar o conhecimento e a informação. Este estudo aborda dois dos Sistemas de Organização do Conhecimento que se destacam ao longo do desenvolvimento da web: as folksonomias e as ontologias. Objetiva identificar as principais perspectivas que integram folksonomias e ontologias; descrever modelos e métodos empregados nessa integração, bem como apontar suas potencialidades e limites, a fim de contribuir para a estruturação informacional da web. Trata-se de uma pesquisa de natureza exploratória e qualiquantitativa. O campo empírico é composto por trabalhos científicos indexados pela base de dados Scopus no recorte temporal: 2004-2023. No tratamento e interpretação dos dados foram aplicadas duas abordagens metodológicas: a pesquisa bibliográfica e a terminográfica. Os resultados revelam que a integração entre folksonomias e ontologias apresenta tanto benefícios, quanto desafios significativos. É capaz de promover

uma combinação única de flexibilidade e estruturação semântica, resultando em sistemas informacionais mais eficazes e adaptáveis às necessidades dos seus utilizadores. Porém, isso inclui o desenvolvimento de algoritmos complexos para mapear as tags das folksonomias para os conceitos da ontologia de forma precisa, a necessidade de manter essa integração atualizada ao longo do tempo diante da dinâmica das folksonomias e da evolução dos domínios de aplicação, e a importância de garantir a qualidade e consistência dos dados devido à natureza idiossincrática da linguagem natural.

Palavras-chave: sistemas de organização do conhecimento; folksonomias; ontologias; hibridização de sistemas de organização do conhecimento.

ABSTRACT

Knowledge Organization Systems are conceptual and terminological schemes used to represent and organize knowledge and information. This study addresses two of the Knowledge Organization Systems that stand out throughout the development of the web: folksonomies and ontologies. It aims to identify the main perspectives that integrate folksonomies and ontologies; describe models and methods employed in this integration, as well as highlight their

potentialities and limitations, in order to contribute to the informational structuring of the web. It is an exploratory and qualitative research. The empirical field consists of scientific works indexed by the Scopus database in the time frame: 2004-2023. In the treatment and interpretation of the data, two methodological approaches were applied: bibliographic research and terminographic research. The results reveal that the integration between folksonomies and ontologies presents both benefits and significant challenges. It is capable of promoting a unique combination of flexibility and semantic structuring, resulting in more effective and adaptable information systems to the needs of its users. However, this includes the development of complex algorithms to map folksonomy tags to ontology concepts accurately, the need to keep this integration updated over time given the dynamics of folksonomies and the evolution of application domains, and the importance of ensuring data quality and consistency due to the idiosyncratic nature of natural language.

Keywords: knowledge organization systems; folksonomies; ontologies; hybridization of knowledge organization systems.

1 INTRODUÇÃO

A Organização do Conhecimento (OC) é um campo do saber que se dedica a estruturar sistemas conceituais -e consequentemente, terminológicos- a fim de categorizar conhecimentos e informações de maneira lógica e acessível.

Os produtos da OC, denominados Sistemas de Organização do Conhecimento (SOC), possibilitam a representação e recuperação da informação nos mais diversos contextos. São ferramentas essenciais para a gestão eficaz do conhecimento e da informação em todos os ramos da atividade humana.

O presente estudo aborda a integração entre dois SOC que são comumente estudados no contexto de duas distintas fases da Web: folksonomias e ontologias. Nesse sentido, é possível afirmar que as folksonomias estão para a Web 2.0 assim como as ontologias estão para a Web Semântica.

A Web 2.0 é pautada por um modelo de negócios que faz com que seja caracterizada por uma maior interatividade e participação entre os seus utilizadores e pela introdução de uma variedade de ferramentas de colaboração que permitem a esses atores contribuir com a produção e circulação de conteúdos, além de organizá-los de forma coletiva. Tal abordagem inclui wikis, blogs, fóruns e redes sociais digitais. Seus utilizadores podem criar, editar e categorizar o

conteúdo, tornando a organização da informação e do conhecimento descentralizada, a partir de uma perspectiva bottom-up (que parte dos utilizadores para os gestores do sistema).

A Web 2.0 adota a folksonomia, enquanto principal abordagem, para organização da informação digital em seus sistemas colaborativos. Seus utilizadores atribuem livremente palavras-chave (tags) aos mais diversificados tipos de conteúdo, tornando a representação e a organização da informação mais flexível e adaptável, sob a perspectiva desses sujeitos. Nos ambientes digitais que adotam a folksonomia, as tags refletem a linguagem e os interesses dos próprios utilizadores.

Sendo assim, a organização da informação na Web 2.0 se apresenta como fluida, personalizável e orientada pela comunidade de utilizadores desta plataforma, que nela desempenham um papel fundamental na categorização, organização e distribuição de conteúdos. Essa abordagem descentralizada e colaborativa, aparentemente, contribui para tornar a Web uma plataforma mais dinâmica e adaptável às mudanças nas preferências e necessidades dos seus utilizadores.

Já a Web semântica tem por objetivo tornar as informações disponíveis de forma mais compreensível tanto para os seres humanos quanto para os sistemas computacionais. Foi proposta pelo cientista da computação Tim Berners-Lee, o mesmo criador da Web na década de 80, e representa uma visão em que os dados são estruturados de modo que os sistemas computacionais possam gerar inferências, permitindo uma melhor integração e interpretação dos conteúdos.

A Web semântica visa instaurar uma fase de desenvolvimento tecnológico em que as máquinas possam processar o significado dos dados, permitindo a automação de tarefas complexas, como pesquisa inteligente, integração de informações de várias fontes e o desenvolvimento de assistentes virtuais mais avançados. Isso se mostra especialmente útil em áreas como: Recuperação da Informação, E-commerce, pesquisa científica, E-learning e Ciência de Dados, por exemplo.

A organização do conhecimento e da informação na Web semântica se baseia em princípios fundamentais e padrões que sustentam a visão de uma plataforma na qual os dados são estruturados e sistematizados de maneira semanticamente explícita, tornando-os compreensíveis tanto para os humanos quanto para os sistemas computacionais.

Nesse cenário, as ontologias informacionais são tipos de SOC que constituem modelos formais que descrevem conceitos e relações a fim de propiciar a representação do conhecimento de uma área de aplicação. Fornecem um vocabulário comum e uma estrutura para representar esse conhecimento de maneira precisa. Em razão disso, são essenciais para a interoperabilidade de dados e informações (Campos, 2017).

Ressalta-se que a Web semântica está em desenvolvimento e não atingiu seu pleno potencial. Observa-se que muitas das tecnologias e conceitos da Web semântica estão sendo gradualmente adotados e experimentados em aplicativos e serviços da Web 2.0 à medida que a infraestrutura e as capacidades da Web continuam a se desenvolver. Desse modo, abordar os termos Web 2.0 e Web semântica é se referir a uma mesma plataforma, porém com diferentes conceitos e tecnologias estruturantes que denotam fases ou estágios de desenvolvimento em curso.

Tal panorama é a base para se analisar o movimento de integração entre os modelos de representação do conhecimento e da informação, preconizados pela Web 2.0 e pela Web semântica. Quais são as características e limites dessa integração?

Objetiva-se especificamente: a) identificar as principais abordagens que conciliam folksonomias e ontologias; b) descrever os modelos e métodos empregados e c) apontar potencialidades e limitações desse movimento de integração entre

dois SOC considerados opostos, no que se refere ao nível de formalismo da linguagem que os constitui, mas com possibilidades de contribuírem para o aprimoramento mútuo.

Folksonomias são sistemas de classificação colaborativa, baseados nas palavras-chave atribuídas pelos utilizadores da Web. Sob a perspectiva da Ciência da Informação (CI), são consideradas menos estruturadas que as ontologias, que são modelos de dados formais e explícitos. A integração entre esses dois SOC envolve a criação de estratégias que permitam que os dados sejam compartilhados e enriquecidos por meio das práticas colaborativas e de estruturação semântica.

Toma-se como pressuposto a tendência histórica e inescapável de hibridização dos SOC, a partir do entendimento de que tanto a linguagem natural, na qual se baseia a folksonomia, quanto a linguagem padronizada, na qual se baseia a ontologia, enquanto um artefato de software, possuem vantagens e desvantagens que podem ser conciliadas a fim de alcançarem benefícios mútuos.

2 METODOLOGIA

O estudo foi desenvolvido a partir da compreensão inicial de que se trata de uma pesquisa de natureza exploratória e de cunho tanto quantitativo, quanto qualitativo.

O campo empírico é composto pela literatura científica, no idioma inglês, indexada pela base de dados Scopus, no período de 2004 a 2023. O período inicial de 2004 foi escolhido por se tratar do ano em que o termo 'folksonomia' foi cunhado por Thomas Vander Wal. Sendo assim possível verificar a partir de que momento se deu o início dos estudos que abordam a integração entre os SOC supracitados.

O levantamento da literatura aplicou como expressão de busca: "folksonomies AND ontologies". Foram coletados em planilhas os dados de: título, autoria, fonte, pertencimento institucional, ano, resumos e palavras-chave.

Para as análises quantitativas foram considerados os 319 artigos recuperados a partir da expressão de busca supracitada. Já para a análise qualitativa e aplicação da abordagem terminográfica foi feito um refinamento desse corpus textual com base na identificação dos artigos mais relevantes em cada ano do período temporal delimitado, o que totalizou 61 artigos publicados entre 2005 e 2020, nos quais foram analisados: títulos, resumos e palavras-chave. A partir desse recorte, foram elencados 90 termos, suas definições e relações.

No tratamento e interpretação dos dados obtidos foram aplicadas duas abordagens metodológicas integradas: a pesquisa bibliográfica e terminográfica.

Com base em Bevilacqua e Finatto (2006) compreende-se que a Terminografia se concentra na análise e documentação da terminologia em uma área específica. Envolve a coleta, organização e descrição de termos técnicos e científicos. Sua principal finalidade é criar glossários ou dicionários especializados que ajudem a padronizar a terminologia e a facilitar a comunicação dentro da área. Tal perspectiva se mostrou pertinente à pesquisa, pois possibilitou documentar os termos, suas definições e suas relações com outros termos. A articulação entre as duas perspectivas metodológicas supracitadas possibilitou mapear e entender as conexões entre os conceitos obtidos na fase empírica da pesquisa.

3 FOLKSONOMIA: CONCEITO, HISTÓRICO E PROBLEMAS

Ao iniciar o estudo sobre folksonomia, destaca-se que se trata de um tipo de SOC, e como tal, atua como uma ferramenta para representar, agrupar e recuperar conteúdos e objetos informacionais em contextos digitais.

A palavra tem origem no inglês folk: agrupamento de pessoas, e foi cunhada por Thomas Vander Wal em 2004 para indicar um tipo de classificação popular observada em websites como Furl, Flickr e Del.icio.us. Embora alguns desses sites tenham deixado de existir ao longo do tempo, a abordagem folksonômica foi incorporada às diversas plataformas de conteúdos e redes sociais digitais, como Youtube, Instagram, Tik Tok e X. A ideia surge em meio a um debate sobre o termo no então chamado Asylomar Institute for Information Architecture (AIFIA). Na procura por uma nomenclatura adequada para a abordagem colaborativa de classificação observada na Web naquela época, Eric Scheid, um dos presentes, aponta como sendo um fenômeno de ‘classificação popular’ conceito que de imediato Thomas Vander Wal adaptou para ‘folksonomia’ (Vander Wal, 2007). Para o autor, a folksonomia:

[...] é o resultado da livre marcação pessoal de informações e objetos (qualquer coisa com uma URL) para recuperação própria. A marcação é feita em ambiente social (geralmente compartilhado e aberto a outras pessoas). A folksonomia é criada a partir do ato de marcação por parte de quem consome a informação. (Vander Wal, 2007, tradução nossa).

O conceito surge no âmbito do modelo de negócios da Web 2.0, que instaura uma expansão do potencial colaborativo e interativo entre os seus utilizadores, os sistemas e os serviços dessa plataforma multimídia a fim de propiciar o aproveitamento da inteligência coletiva. Segundo O’Reilly (2005) a Web 2.0 se caracteriza como uma plataforma digital interativa que promove a participação dos sujeitos mediante uma arquitetura mais amigável, o que permite que eles saiam do papel de receptores passivos de uma publicação para o de atores que geram valor agregado aos sistemas, constituindo um modelo de negócios que explora o potencial colaborativo e coletivo.

No campo da CI, isso significa que é um espaço hipermidiático que possibilita a criação e a estruturação de documentos, bem como de suas representações, por parte dos sujeitos informacionais. Salienta-se que esses atores classificadores geralmente não possuem nenhuma formação, ou conhecimento especializado, sobre a área de classificação. Dessa maneira, se revela um mecanismo que pode ser utilizado por leigos, devido ao seu caráter intuitivo. Por esse motivo, a folksonomia é considerada um tipo de classificação informal e multifacetada:

Hoje em dia, os sistemas colaborativos de marcação [...] permitem aos usuários adicionar conteúdo à web, anotá-los usando tags e compartilhá-los. Isso cria redes complexas de usuários, recursos e tags [...]. As tags simplificam a recuperação e navegação de recursos. Além disso, a marcação permite aos usuários anotar os mesmos recursos com vários termos, o que permite uma organização multifacetada (Mousselly-Sergieh et al. 2014, p. 3, tradução nossa).

A folksonomia se apresenta nos ambientes digitais a partir de um sistema classificatório flexível gerado pelas tags (palavras-chave que também são chamadas de etiquetas) ou ainda em uma das formas de classificação de conteúdos mais utilizadas nas plataformas de redes sociais digitais, as hashtags. Sendo assim, as folksonomias resultam da junção de todas as palavras-chave utilizadas para etiquetar conteúdos digitais, sejam eles fotos, vídeos, textos, trechos, áudios ou ainda documentos em quaisquer outros formatos.

Quando se faz o uso de uma tag no ambiente digital, essa serve para identificar, localizar e, também, disseminar um objeto informacional digital. É a partir dessa atribuição de tags que esse objeto é marcado como detentor de um determinado potencial informativo.

A folksonomia se mostra uma ferramenta altamente personalizável e flexível. Segundo Specia e Motta (2007), as tags são utilizadas pelos indivíduos de forma

personalizada, de acordo com suas próprias necessidades, sem depender de um vocabulário controlado ou estrutura definida previamente. Ou seja, a sua estrutura se baseia em uma linguagem natural, não em uma linguagem formal e controlada.

O uso de tags para representar a informação, demonstra os interesses e a interpretação do sujeito que as atribui, assim como indica seu contexto social, cultural, experiências e percepções de mundo. Araújo (2006) observa que antes de ser uma atividade específica do campo da CI e da Biblioteconomia, a classificação deve ser percebida enquanto um fenômeno social, pois o ato de classificar é algo inerente ao ser humano. Dessa forma, as folksonomias também se apresentam como classificações sociais, pois refletem a visão de mundo e o contexto sociocultural daquele que as utiliza:

Uma das suas características distintivas é que são sistemas abertos e não controlados, onde os utilizadores podem anotar recursos com tags diferentes, dependendo das suas origens sociais ou culturais, experiência e percepção do mundo. Por exemplo, um zoólogo pode marcar uma fotografia de um leão com {felidae, pantherinae, mamífero}, enquanto um não-especialista pode usar {leão, rei, animal, selva} para o mesmo propósito. (Angeletou, Sabou, Specia, Motta, 2007, p. 1, tradução nossa).

As tags acentuam os efeitos da subjetividade na representação da informação e do conhecimento. Dessa forma, em consonância com Spyns et al. (2006) um sujeito tem total liberdade para criar, renomear, agrupar, dividir, mesclar e excluir palavras-chave, utilizando desse recurso para representar e classificar seu próprio universo de conteúdos e percepções.

Apesar de toda essa dinâmica personalizável, colaborativa e flexível, há problemáticas no uso da folksonomia na representação do conhecimento e da formação justamente em razão da flexibilidade e da falta de uso de um vocabulário controlado, como aponta Dotsika (2009, p. 409, tradução nossa):

Os entusiastas da folksonomia, no entanto, não conseguem fornecer provas sérias de que a etiquetagem de bottom-up pode lidar com questões de interoperabilidade entre repositórios de conhecimento distribuídos, pesquisa automatizada e qualidade da recuperação de informação. Os problemas inerentes aos vocabulários não controlados, isto é ambiguidade, granularidade inconsistente, duplicações e sinônimos, levam a uma recuperação de conteúdo comprometida (por exemplo, a procura de objetos marcados como 'animal de estimação' não irá necessariamente recuperar aqueles marcados com 'gato' ou 'hamster') e problemas de qualidade que são difíceis de resolver (Dotsika, 2009, p. 409, tradução nossa).

É corriqueiro observar a atribuição de diversas tags com o mesmo significado para um só conteúdo digital. Esses fatores geram um retorno volumoso de conteúdos quando uma palavra-chave é utilizada como recurso de pesquisa, e isso resulta na falta de precisão na recuperação da informação. Ainda segundo Dotsika (2009), a falta de interoperabilidade pode levar a infortúnios como a inconsistência semântica, a ambiguidade lógica e a redundância de informações.

Como ferramenta que possibilita a disseminação rápida de informações personalizadas e carregadas de discursos, que refletem o cenário social, político e cultural de quem atribui as tags – e esses agentes classificadores podem ser tanto humanos quanto inteligências artificiais – a folksonomia também se mostra um mecanismo capaz de contribuir para a formação de bolhas informacionais e fomentar a desinformação (Assis, 2021).

Diante disso, uma das principais preocupações no uso da folksonomia está em

desenvolver uma forma de administrar os conteúdos de forma a suprir o descontrole de vocabulário, a inconsistência semântica, a imprecisão e a ambiguidade. E, por meio disso, construir um ambiente digital com uma representação do conhecimento e da informação capaz de contribuir no enfrentamento de problemas sociais, como a desinformação.

As abordagens que tentam conter os problemas gerados pela linguagem natural no âmbito das folksonomias sugerem a contribuição de SOC formais, ou seja, aqueles que apresentam uma estruturação lógica e explícita dos seus termos, a fim de contornar as idiosincrasias oriundas de uma abordagem informal e flexível de representação do conhecimento e da informação. Dentre os SOC abordados na literatura, destacamos as ontologias.

4 ONTOLOGIAS: ORIGENS, CONTEXTO E POSSIBILIDADES

O termo ontologia tem sido adotado pelas áreas da Filosofia, Ciência da Computação e Ciência da Informação (CI) com diferentes propostas. Na primeira, se integra a um ramo da Metafísica que estuda os entes. Na segunda, se emprega na modelagem do conhecimento, e na última, se relaciona à representação do conhecimento e da informação (Almeida, 2014). Neste artigo, as ontologias são abordadas a partir da perspectiva da CI.

Assim como a Web 2.0 proporcionou um ambiente favorável ao surgimento e disseminação das folksonomias, a Web semântica se mostra o espaço ideal para a evolução e disseminação do uso das ontologias como ferramentas de representação e organização do conhecimento e da informação.

Pickler (2007, p. 65) descreve a Web semântica como “uma extensão da web que acrescenta semântica ao atual formato de representação de dados”, ou seja, por trás do conteúdo visível dessa Web, existe uma estrutura interligada de significados atrelados àqueles dados e informações. Por isso, a Web semântica enseja um ambiente digital muito mais otimizado, já que se baseia em estruturação lógica e conseqüentemente, inferencial.

Da mesma forma que a Web 2.0 surgiu como um avanço tecnológico da Web 1.0 (também chamada de Web sintática), que revolucionou o seu modelo de negócios, a Web semântica está em processo de construção e atualmente ambas coexistem no ambiente virtual e hipermidiático, popularmente conhecido como Web.

Seu desenvolvimento busca melhorar a entrega de informação relevante por meio dos motores de busca, dos agentes inteligentes e de outros mecanismos que possibilitam a recuperação precisa e contextual de conteúdos informacionais digitais. Para Breitman (2005, p.7) a Web semântica tem como objetivo “permitir que as máquinas façam o processamento que atualmente, na Web Sintática, tem de ser realizado por seres humanos”. Nas fases tidas como anteriores da Web (1.0 e 2.0), dados e informações são oferecidos pelo computador e a ação interpretativa é feita totalmente pelos seres humanos. A respeito disso, Pickler (2007, p. 70) observa que, partindo da já existente infraestrutura da Web, se faz necessária a criação de padrões para a descrição de dados assim como de “uma linguagem que permita a construção e a codificação desses significados compartilhados”, o que geraria uma rede semântica global interpretável tanto por sistemas computacionais, quanto por seres humanos.

No atual modelo, as ferramentas de busca se guiam principalmente por algoritmos e palavras-chave para realizar suas pesquisas (Franco, 2018). Como já mencionado, o uso desses recursos para a recuperação de objetos informacionais digitais resulta em um grande volume de material, geralmente dotado de sinonímia e ambiguidade, tornando a busca um processo lento e que muitas vezes traz resultados pouco relevantes (Pickler, 2007), (Dotsika, 2009).

Esse processo está relacionado com a dificuldade em se determinar os contextos informacionais desses objetos e que tem como conseqüência a impossibilidade por parte

dos agentes de identificar de forma precisa a atinência dos documentos (Souza; Alvarenga, 2004). Logo, um dos empecilhos na recuperação relevante desses itens é justamente a ausência de uma estrutura inferencial estabelecida dos dados e informações que são armazenados na Web, parte da resolução desse problema se dá no uso das ontologias para criar uma conexão entre o que é registrado e os objetos informacionais já existentes.

Diante disso, as ontologias surgem como ferramentas que visam formalizar para as máquinas os modelos de conhecimento gerados pelo ser humano. De acordo com Sharif (2009), as ontologias são capazes de tornar um conhecimento que se apresenta de forma implícita em uma forma explicitada, compreensível e processável por máquinas e agentes virtuais. Sob a perspectiva informacional, a proposta da Web semântica é estruturar os dados de tal forma que o próprio sistema de busca seja capaz de identificar conteúdos e até mesmo de inferir significados, e o que se espera é que essa estruturação seja feita por meio das ontologias. Sendo assim, no contexto da CI, a ontologia é um esquema conceitual dotado de um alto grau de formalismo semântico e que vai detalhar as relações entre os termos e conceitos atinentes a um determinado objeto ou conteúdo digital.

A literatura da CI atesta que as ontologias mantêm algumas correspondências com os tesauros, por se tratarem de ferramentas utilizadas para definir vocabulários padronizados, em que se determinam relações lógico-semânticas entre termos. Dessa forma, as ontologias estão para os agentes virtuais assim como os tesauros estão para os profissionais da informação. A respeito disso, Dotsika (2009) aponta que as ontologias:

São metadados estruturados que representam conjuntos de conceitos e seus relacionamentos dentro de um domínio. Elas modelam entidades, restrições lógicas e relacionamentos na forma de gráficos direcionados. Assim como as taxonomias, são esquemas de cima para baixo em que o conhecimento é modelado em classes, propriedades e relacionamentos, e onde as associações taxonômicas é-um são estendidas para incluir tipos adicionais que permitem uma modelagem semântica mais refinada (Dotsika, 2009, p.408, tradução nossa).

Assim, as ontologias farão a ponte entre o conhecimento registrado, os motores de busca, os agentes computacionais que buscam esse conteúdo e os sujeitos informacionais, tudo isso por meio de uma rede semântica bem definida. No exemplo proposto por Dotsika (2009), ao realizar uma busca sobre 'animais de estimação' em ambientes digitais da Web 2.0, que adotam a folksonomia, os resultados recuperados apontariam apenas para os documentos marcados com a tag 'animaisdeestimação'. Já nos ambientes da Web semântica, as conexões feitas por meio de uma rede formal de significados, permitiriam às ferramentas de busca recuperar também os documentos marcados com as tags 'hamster' e 'gato', pois os agentes seriam capazes de inferir a relação é-um através da teia de significados que constitui sua infraestrutura. Nesse âmbito, destaca-se que:

podemos esperar que a Web tenha grande melhoria dos índices de revocação e precisão no atendimento das necessidades de informação, porque a semântica embutida nos documentos permitirá aos dispositivos de recuperação evitar os problemas comuns de polissemia e sinonímia, além de considerar as informações em seus contextos de significado (Souza e Alvarenga, 2004, p. 139).

Percebe-se que a ontologia fornece soluções para os problemas gerados pelo uso da linguagem natural. A princípio, um dos maiores empecilhos na efetivação do emprego e do uso das ontologias

está na rigidez de sua estrutura e no desconhecimento desse recurso pelos usuários. Dessa maneira, as ontologias se tornam ferramentas inacessíveis para as massas e isso seria um inconveniente em meio a crescente participação desses sujeitos informacionais nos meios de comunicação descentralizados e na disseminação da informação. Seria um retrocesso desvincular completamente esses sujeitos do processo de classificação e representação dos conteúdos informacionais digitais, visto que a tendência trazida pela Web 2.0 é de que a cooperação entre utilizadores e sistemas computacionais seja cada vez mais intensa.

Sendo assim, uma das questões a se resolver é desenvolver um processo que possibilite a integração dos utilizadores da Web na dinâmica da engenharia de ontologias enquanto ferramentas de representação do conhecimento e da informação.

Diante desse panorama, são necessários estudos e projetos que almejam a integração entre essas duas diferentes tipologias de SOC, capazes de flexibilizar a estrutura ontológica e permitir uma maior participação dos sujeitos informacionais a partir de plataformas com interfaces amigáveis e caracterizadas pela hibridização de SOC.

5 FOLKSONOMIAS E ONTOLOGIAS: UMA INTEGRAÇÃO POSSÍVEL

Historicamente desde a década de 60, estudos como os testes de Cranfield sugerem a integração entre a linguagem natural e vocabulários controlados a fim de prover melhorias na recuperação da informação (Cleverdon, 1967). No cenário analisado, com o propósito de aumentar a precisão da recuperação dos objetos digitais, os estudos convergem para o uso das folksonomias aliadas ao apoio estrutural de mecanismos de controle terminológico. Essa integração se dá na utilização conjunta de folksonomias e esquemas conceituais formais como, por exemplo, tesouros, taxonomias e ontologias.

A junção estrutural de dois ou mais sistemas conceituais é o que aqui se denomina hibridização de SOC (Assis, 2015). Ou seja, a hibridização é a integração de características de dois ou mais SOC a fim de gerar um esquema terminológico integrado, capaz de solucionar os problemas que esses sistemas apresentam quando utilizados separadamente.

No que tange à integração entre folksonomias e ontologias, os resultados da pesquisa apontam a existência de 319 trabalhos, publicados desde 2005 até 2021. Após um segundo recorte de dados que considerou os trabalhos mais relevantes, a partir dos critérios de citação e atinência, foram identificados 61 trabalhos. A partir desse corpus foram obtidos e analisados 90 termos.

Verificou-se que a primeira proposta que abordou a unificação entre folksonomias e ontologias observada na literatura surgiu no ano de 2005, trata-se do artigo oriundo de conferência, intitulado "A proposal of community-based folksonomy with RDF metadata". Nele Ohmukai, Hamasaki, Takeda (2005) apresentam um modelo de social bookmarking (um tipo de serviço para armazenamento e compartilhamento de links baseado em tags para descrevê-los) estruturado em RDF (Resource Description Framework), uma tecnologia semântica utilizada para a construção de ontologias para a Web semântica.

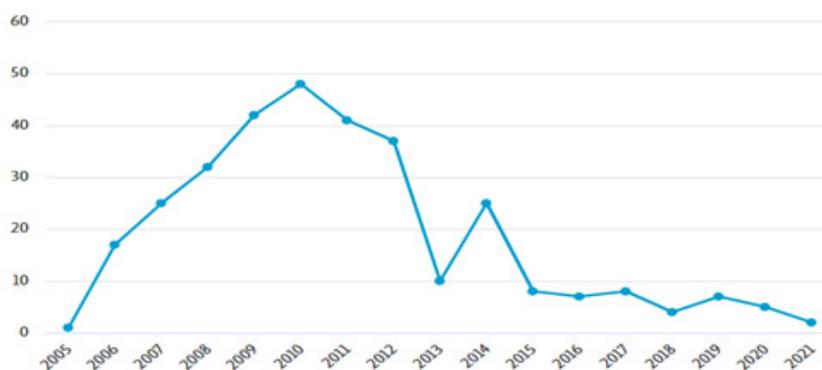
Já a primeira proposta de ferramenta híbrida foi identificada em 2006 em mais um artigo oriundo de conferência, intitulado "FolksAnnotation: a semantic metadata tool for annotating learning resources using folksonomies and domain ontologies" de Al-Khalifa e Davis (2006). A ideia foi desenvolvida para fornecer suporte à anotação semântica de recursos de aprendizagem a partir da combinação de folksonomias e ontologias de domínio.

Com relação aos artigos de periódicos pioneiros nessa abordagem, destaca-se o trabalho de Gruber (2007) "Ontology of folksonomy: A mash-up of apples and oranges" publicado no *International Journal on Semantic Web and Information Systems* que recebeu 207 citações. O autor abordou a relação entre folksonomias e ontologias,

destacando suas diferenças fundamentais e discutindo como elas podem ser reconciliadas ou combinadas. Sugere que as folksonomias e as ontologias têm vantagens e desvantagens distintas e que uma abordagem híbrida, que combina elementos de ambas, pode ser mais eficaz para lidar com a complexidade e a diversidade das informações na Web. O artigo também destaca os desafios e oportunidades associados à integração dessas duas abordagens e propõe direções para pesquisas futuras nessa área.

Sobre o volume de publicações, observou-se o ano de 2010 como o ápice das publicações sobre esse assunto, com 48 registros, conforme a figura abaixo:

Figura 1 - Trabalhos publicados por ano



Fonte: dados da pesquisa.

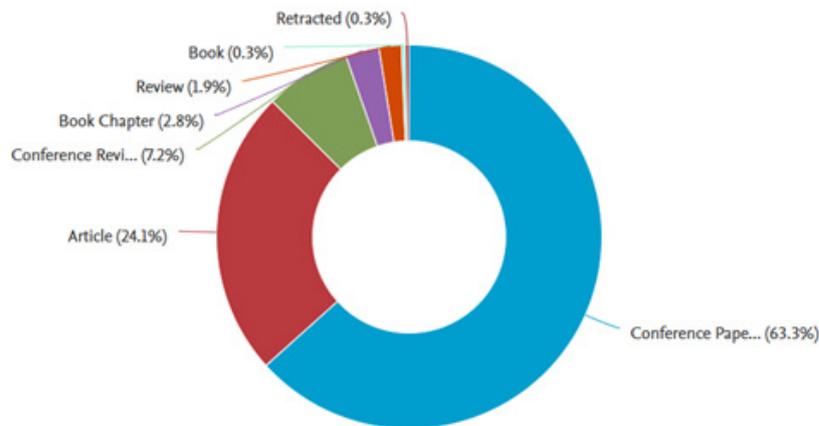
Os dados da pesquisa também mostraram que os países que mais produziram literatura sobre o assunto são Alemanha, Espanha e Estados Unidos. Destaca-se em quarto lugar os esforços de pesquisa oriundos do Reino Unido.

Figura 2 - Trabalhos publicados por regiões geográficas



Fonte: dados da pesquisa.

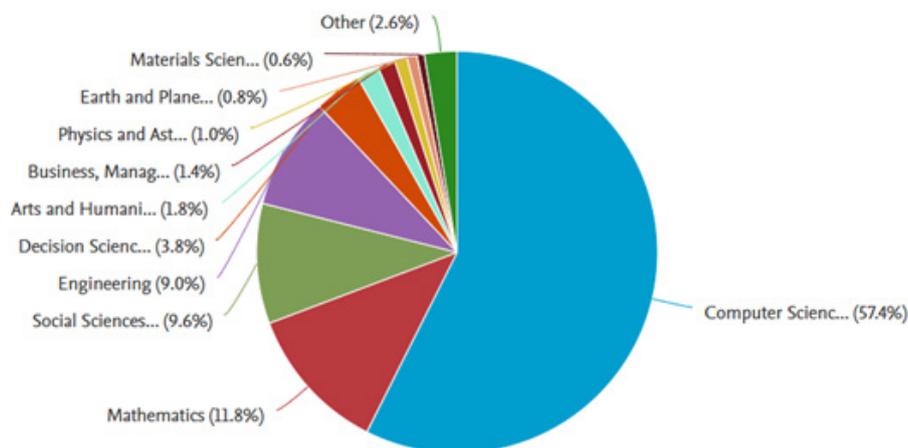
Quanto à natureza das publicações sobre a temática, constatou-se que 63,3% são artigos oriundos de eventos científicos e 24,1% são artigos de periódicos, o que revela a busca majoritária por soluções conceituais e tecnológicas, bem como consensos sobre o tema, a partir dos fóruns de apresentação de novas pesquisas. Outros tipos de publicações como livros, capítulos de livros e revisões não apresentaram percentuais relevantes na preferência de publicização dos trabalhos pelos pesquisadores dessa temática.

Figura 3 - Natureza das publicações sobre o assunto

Fonte: dados da pesquisa.

Uma síntese dos registros das discussões e reflexões promovidas nesses eventos revela que apesar de ambas serem ferramentas de organização e representação do conhecimento e da informação, folksonomias e ontologias diferem em vários pontos. A primeira se dá de forma colaborativa e informal, é uma abordagem classificatória em que as ações emergem dos utilizadores para os gestores das plataformas (bottom-up) e que possui um baixo grau de precisão na recuperação de material relevante. A segunda reúne conjuntos de conceitos e termos formalmente estruturados, é projetada dos gestores para os utilizadores (top-down). Seus criadores são indivíduos especializados e propicia aos agentes computacionais a capacidade de inferir significados a partir de dados semanticamente estruturados, o que resulta em uma recuperação da informação mais otimizada, porém a rigidez de sua estrutura se mostra um problema na massificação de seu uso (Dotsika, 2009), (Sharif, 2009).

Quando o foco da análise dos dados se decaiu sobre as áreas de origem dos estudos sobre essa possibilidade de integração, foi possível constatar que a Ciência da Computação é a área de maior interesse e investimento, respondendo por 57,4% das pesquisas. Destaca-se em terceiro lugar o campo das Ciências Sociais com 9,6%, aqui se inclui a contribuição da CI.

Figura 4 - Principais áreas do conhecimento

Fonte: dados da pesquisa.

Independentemente do campo de origem desses estudos, percebeu-se que as perspectivas de integração entre esses dois esquemas se pautam em formas de resolver os problemas de ambiguidade, imprecisão e descontrole vocabular apresentado pela folksonomia, intentam ainda tornar as ontologias mais versáteis, ao enriquecê-las semanticamente.

Apesar de não haver na literatura analisada um consenso que nomeie um SOC híbrido, compreendemos que o ápice semântico dos estudos sobre a integração entre folksonomias e ontologias se deu com o termo “ontologias folksonomizadas”, proposto por Alves e Santanché (2011) pois trata-se do único termo que reaparece nos anos posteriores em publicações relevantes e de outros autores. De acordo com Assis (2021, p. 9):

O conceito de ontologia folksonomizada revela um SOC que hibridiza a linguagem natural e a linguagem controlada agregando os aspectos positivos de ambas, a fim de promover avanços na estruturalógica que embasa a representação e a recuperação dos conteúdos e objetos informacionais.

No contexto dos sistemas informacionais, a abordagem de uma ontologia folksonomizada pode contribuir para o aprimoramento das recomendações de tags semanticamente relacionadas, como apontam Yadav, Kaur e Duhan (2016). Já em contextos mais específicos pode propiciar precisão e assertividade na recuperação da informação em domínios do conhecimento, como a área de finanças, explorada na pesquisa de Salgado (2019).

No período temporal analisado, foram muitos os autores que trabalharam a perspectiva de integração entre esses dois esquemas, enquanto alguns não chegaram a atribuir nenhuma palavra, sigla ou termo para nomeá-la, outros propuseram termos específicos para designá-la, como exemplificado no Quadro 1.

Quadro 1 - Termos que nomeiam a busca por integração entre folksonomias e ontologias

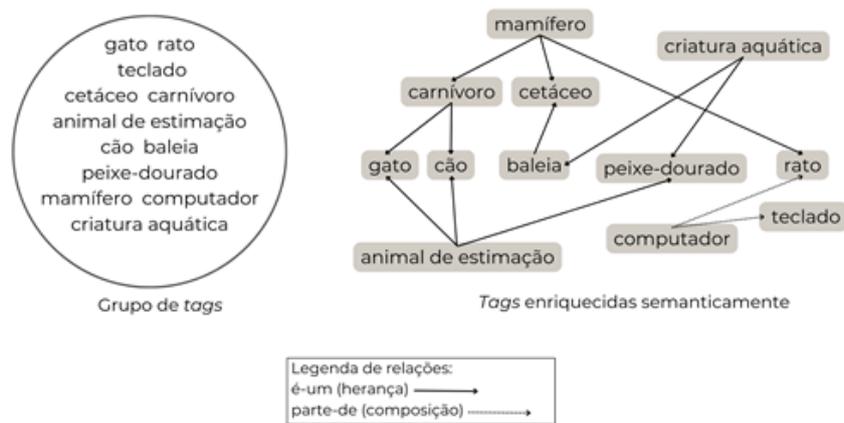
Modelo	Autores	Ano
TagOntologia	Gruber	2005
Folksonomias semanticamente enriquecidas	Angeletou, Sabou, Specia e Motta	2007
Folksontologia	Van Damme, Hepp e Siorpaes	2007
Flexonomias	Kapetanos e Schaal	2007
Folksonomias integradas com a Web 3.0	Specia e Motta	2007
Ontologias folksonomizadas	Alves e Santanché	2011
Ontologia baseada em folksonomia	Djuana	2018
Ontologias colaborativas	Allam, <i>et al</i>	2020

Fonte: dados da pesquisa.

Dentre os modelos que abordam a integração entre folksonomias e ontologias, destacamos: o de Sharif (2009) que propõe um esquema em que a experiência de usuário é integrada a um repositório onde ontologias e folksonomias se desenvolvem de forma dinâmica; o de Dotsika (2009) que ilustra com detalhes como pode se dar o enriquecimento semântico de folksonomias; o SocioOntoProcess de Guergour e Boufaida (2016) que propõe uma engenharia de ontologias centrada no usuário a partir da contribuição da folksonomia e da filosofia da Web 2.0 e o de Alruqimi e Aknin (2017) que sugere um modelo de algoritmo capaz de extrair uma ontologia específica de domínio a partir da folksonomia.

Observou-se no modelo de Dotsika (2009) uma solução para a explicitação das relações semânticas entre os conceitos representados pelas tags, possibilitando a formalização de uma rede semântica que contribua tanto para melhorar a experiência dos utilizadores dos sistemas, quanto para a realização de inferências e processos computacionais em sistemas ditos inteligentes.

Figura 5 - Enriquecimento semântico de folksonomias



Fonte: Adaptado de Dotsika (2009, p. 408, tradução nossa).

Na Figura 5 é possível observar um agrupamento de tags, tal qual ocorre nas nuvens de tags, que atuam como interfaces para a recuperação da informação mediante navegação (browsing) e a explicitação das relações lógico-semânticas entre as tags contribuindo para sanar ambiguidades e outros problemas decorrentes da linguagem natural.

Para além da explicitação das relações semânticas, as pesquisas sobre a construção de sistemas caracterizados pela hibridização entre folksonomia e ontologia podem gerar protótipos informacionais caracterizados por cinco aspectos: 1) tags controladas e livres: o sistema pode permitir que os usuários atribuam tags livremente (como ocorre na folksonomia), ao mesmo tempo em que oferece um conjunto controlado de tags baseadas em uma ontologia. Isso proporciona aos seus utilizadores flexibilidade para expressar seus interesses, enquanto também promove a consistência e a estruturação dos dados; 2) mapeamento automático de tags: o sistema pode usar algoritmos de correspondência semântica para interligar automaticamente as tags fornecidas pelos utilizadores aos conceitos da ontologia. Isso ajuda a melhorar a precisão e a relevância dos resultados de busca, aproveitando a estrutura semântica da ontologia; 3) feedback do usuário para refinar a ontologia: o sistema pode coletar feedback dos usuários a partir de tags e conceitos usados, e utilizar essas informações para atualizar e refinar a ontologia ao longo do tempo. Isso ajuda a manter a ontologia alinhada com os interesses e necessidades da sua comunidade

de utilizadores; 4) personalização da experiência do usuário: com base nas tags fornecidas pelos usuários e na estrutura da ontologia, o sistema pode personalizar a experiência do seu utilizador, oferecendo recomendações e conteúdo relevantes que correspondam aos seus interesses específicos e 5) utilização de metadados semânticos: o sistema pode enriquecer os metadados associados aos itens (como fotos, vídeos, etc.) com informações semânticas da ontologia, além das tags fornecidas pelos utilizadores da Web. Isso melhora a descrição e a compreensão dos conteúdos, facilitando sua organização e recuperação.

6 CONCLUSÃO

A integração entre folksonomias e ontologias é uma temática de pesquisas em desenvolvimento, especialmente em campos que visam: a organização de conteúdo em contextos digitais; a melhoria de sistemas de recomendação e a recuperação precisa da informação. Ao combinar o poder da colaboração e da abordagem bottom-up das folksonomias com a estruturação, precisão e abordagem top-down das ontologias, é possível criar sistemas aptos a organizar e acessar grandes quantidades de dados digitais interligados.

Em termos de esquematização conceitual, as ontologias tendem a ser mais estáveis e estruturadas, enquanto as folksonomias são mais flexíveis e dinâmicas. Integrar essas duas abordagens se mostra algo desafiador devido às diferenças em estrutura, escala e complexidade.

Observa-se ainda a existência de desafios técnicos a serem superados, como a interoperabilidade entre sistemas, o processamento de grandes volumes de dados e o gerenciamento de recursos computacionais. Mapear as tags das folksonomias para os conceitos de uma ontologia de forma precisa e eficaz, por exemplo, requer algoritmos sofisticados e técnicas de correspondência semântica.

Manter a integração entre folksonomias e ontologias atualizada e relevante com o decorrer do tempo se mostra um outro aspecto problemático, especialmente devido à natureza dinâmica das folksonomias e à evolução dos domínios de aplicação.

Por fim, garantir a qualidade e consistência dos dados constitui um fator crucial para que os sistemas baseados nessa integração possam ser bem-sucedidos uma vez que os dados gerados pelos utilizadores da Web a partir das folksonomias podem ser inconsistentes ou ambíguos dada a natureza idiossincrática da linguagem natural.

São recomendações para estudos futuros: a) desenvolvimento de algoritmos de correspondência semântica que atuem no mapeamento entre os SOC mediante a aplicação de técnicas de aprendizagem de máquina e processamento de linguagem natural, a fim de melhorar a precisão e a eficácia das correspondências; b) proposição de métodos para garantir a qualidade e consistência dos dados em sistemas integrados com o propósito de incluir a criação de métricas para avaliar a confiabilidade das tags e estratégias para mitigar a ambiguidade e inconsistência; c) realização de estudos de caso em diferentes domínios para avaliar a eficácia da integração em contextos reais, o que poderia ajudar a identificar desafios específicos e benefícios práticos dessa abordagem e d) investigar como as interações e contribuições dos utilizadores em ambiente folksonômicos podem influenciar a evolução e adaptação de ontologias ao longo do tempo gerando instrumentos terminológicos mais adaptáveis e centrados nos seus utilizadores.

REFERÊNCIAS

- AL-KHALIFA, H. S.; DAVIS, H. C. FolksAnnotation: a semantic metadata tool for annotating learning resources using folksonomies and domain ontologies. **Innovations in Information Technology**. 2006.
- ALLAM, H. et al. Collaborative ontologies in social tagging tools: A literature review of natural folksonomy. **Proceedings of the Seventh International Conference on Information Technology Trends**, p. 126-130, 2020.
- ALMEIDA, M. B. Uma abordagem integrada sobre ontologias: Ciência da Informação, Ciência da Computação e Filosofia. **Perspectivas em Ciência da Informação**, v. 19, n. 3, p. 242-258, 2014.
- ALRUQIMI, M.; AKNIN, N. Bridging the Gap between the Social and Semantic Web: Extracting domain-specific ontology from folksonomy. **Journal of King Saud University - Computer and Information Sciences**, v. 31, 2017.
- ALVES, H; SANTANCHÈ, A. Folksonomized Ontologies – from social to formal. In: **SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB (WEBMEDIA)**, 17., 2011, Florianópolis. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2011 . p. 58-65.
- ANGELETOU, S. et al. Bridging the Gap Between Folksonomies and the Semantic Web: An Experience Report. In: **Proceedings of the ESWC workshop: Bridging the Gap between Semantic Web and Web 2.0, European Semantic Web Conference**. 2007.
- ARAÚJO, C. A. Á. Fundamentos teóricos da classificação. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 11, n. 22, p. 117-140, 2006.
- ASSIS, J. **A interoperabilidade a partir de consensos e semânticas emergentes: um modelo para elaboração e atualização de sistemas de organização do conhecimento**. 2015. 183f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.
- ASSIS, J. Folksonomias e pós-verdade: desafios para a organização do conhecimento. **Liinc em Revista**, v. 17, n. 1, p. 1-15, 2021.
- BEVILACQUA, C. R.; FINATTO, M. J. Bocorny. Lexicografia e Terminografia: alguns contrapontos fundamentais. **ALFA: Revista de Linguística**, v. 50, n. 2, p.43-54, 2006.
- BISCALCHIN, R. Os Sistemas de Organização do Conhecimento e os desafios frente a geração Google. **Páginas A&B: arquivos e bibliotecas**, p. 3-9, 2019.
- BREITMAN, K. **Web semântica: a internet do futuro**. Rio de Janeiro: LTC, 2005. 190 p.
- CAMPOS, M. L. de A. Elaboração de modelos de domínio em ontologias: a abordagem onomasiológica e a função da definição. **Ciência da Informação**, v. 46, n. 1, p. 23-33, 2017.
- CLEVERDON, C. The cranfield tests on index language devices, **Aslib Proceedings**, v.19, n.6, p. 173-194, 1967.
- DEL CASTANHEL, L.; TACLA, C. A.; FREDDO, A. R. Integração da web social e semântica: das folksonomias às ontologias. In: **SEMINÁRIO DE INICIAÇÃO CIENTÍFICA E TECNOLÓGICA DA UTFPR**, 2011, Ponta Grossa. Anais [...]. 2011.
- DJUANA, E. Gold-standard evaluation of a folksonomy-based ontology learning model, **Journal of Physics**, v. 971, 2018.
- DOTSIKA, F. Uniting formal and informal descriptive power: Reconciling ontologies with folksonomies. **International Journal of Information Management**, v. 29, n. 5, p. 407-415, 2009.
- FRANCO, M.R. O efeito Long Tail e Fat Head nos motores de busca. 2018. Dissertação (Mestrado em Marketing) - Faculdade de Economia, Universidade do Porto, Porto, 2018.
- FREDDO, A. R. **Folkconcept: método de suporte à modelagem conceitual de ontologias a partir da aquisição de conhecimentos de folksonomias**. 2010. 225 f. Tese (Doutorado em Engenharia Elétrica e Informática Industrial) – Programa de Pós-Graduação em Engenharia Elétrica

e Informática Industrial, Universidade Tecnológica Federal do Paraná, Curitiba, 2010.

GUERGOUR, H-E; BOUFAIDA, Z. SocioOntoProcess: um processo centrado no usuário para engenharia participativa de ontologies. **International Journal of Intelligent Information and Database Systems**, v. 9, n. 2, p. 155-196, 2016.

GRUBER, T. **TagOntology**: a way to agree on the semantics of tagging data. Presentation to Tag Camp, 2005.

GRUBER, T. Ontology of folksonomy: A mash-up of apples and oranges. **International Journal on Semantic Web and Information Systems**, v. 3, n. 2, p. 1-13, 2007.

HERNÁNDEZ, D. de la C. R.; LABRADA, J. V.; HERNÁNDEZ, Yadira de las Mercedes Rivero. Ontologías: Integración de esquemas. **Tlatemoani**: revista académica de investigación, n. 17, p. 16-32, 2014.

KAPETANIOS, E. SCHAAL, M. An algebra and conceptual model for semantic tagging of collaborative digital libraries. In: **The second workshop on foundations of digital libraries in conjunction with 11th European conference on research and advanced technologies on digital libraries**, 2007, Budapeste, Hungria. Anais [...]. Budapeste: ECDL, 2007.

MOUSSELLY-SERGIEH, H. et al. Tag relatedness in image folksonomies. **Document numérique**, v. 17, n. 2, p. 33-54, 2014.

OLIVEIRA, J. G.; VIEIRA, T. T.; TÔRRES, V. A.; SARMENTO, D.; MAIA, L. C. G.; PARREIRAS, F. S. Criação e ampliação de ontologias por folksonomias: uma revisão sistemática da literatura. In: **ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO**, 2017, Marília. Anais [...]. Marília: UNESP, 2017.

OHMUKAI, I., HAMASAKI, M.; TAKEDA, H. A proposal of community-based folksonomy with RDF metadata. **Proceedings of the International Conference on Web Intelligence**, p.700-703. 2005.

O'REILLY, T. **What is Web 2.0**: design patterns and business models for the next generation

of software. O'Reilly Media: Sebastopol, 2005.

PICKLER, M. E. V. Web semântica: ontologias como ferramentas de representação do conhecimento. **Perspectivas em Ciência da Informação**, v. 12, n. 1, p. 65-83, 2007.

SALGADO, I. da M. R.. Um modelo de ontologias folksonomizadas para a auditoria em instituições financeiras no Brasil. In: **SEMINAR ON ONTOLOGY RESEARCH IN BRAZIL**, 7., 2019, Porto Alegre. Proceedings [...]. Porto Alegre: UFRGS, 2019.

SANTOS, R. F. dos.; ALBUQUERQUE, M. E. B. C. de. Aplicabilidade da folksonomia na construção e atualização de sistemas híbridos de organização do conhecimento. **Ponto de Acesso**, v. 15, n. 3, p. 381-397 2021.

SHARIF, A. Combining ontology and folksonomy: an integrated approach to knowledge representation. In: **World Library and Information Congress**. 2009.

SOUZA, R.R.; ALVARENGA, L. A Web Semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, v. 33, n. 1, 2004.

SPECIA, L.; MOTTA, E. Integrating Folksonomies with the Semantic Web. In: Franconi, E., Kifer, M., May, W. (ed.). **The Semantic Web: Research and Applications**. Anais [...]. Berlin, Heidelberg: Springer, 2007.

TURCATEL, I. O. Aquisição social do conhecimento na construção colaborativa de ontologias. In: **SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS**, 2012, Porto Alegre. Anais [...]. Porto Alegre: UFRGS, 2012.

VAN DAMME, C.; HEPP, M.; SIORPAES, K. Folksonomy: an integrated approach for turning folksonomies into ontologies. In: **ESWC 2007 Workshop on bridging the gap between Semantic Web and Web 2.0**, 2007.

VANDER WAL, T. Folksonomy coinage and definition. 2007.

VALENTE, P. G. **Aplicações híbridas para a criação de conteúdo jornalístico na internet**. 2007. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) - Universidade Federal de Santa Catarina.

WASSERMAN, S.; FAUST, K. **Social network analysis**: Methods and applications. Cambridge: Cambridge University Press, 1994. 857 p.

YADAV, U.; KAUR, J.; DUHAN, N. Semantically related tag recommendation using folksonomized ontology. **Proceedings of the International Conference on Computing for Sustainable Global Development**, 2016.