

ANALYSIS OF THE FEASIBILITY OF USING DATA OBTAINED FROM SOCIAL MEDIA TO BUILD ONTOLOGIES: A LITERATURE REVIEW

Luiz Eduardo Magalhães Ferreira - UNIVERSIDADE DA REGIÃO DE JOINVILLE (UNIVILLE) JOINVILLE – SC – BRASIL - Orcid: <https://orcid.org/0000-0002-4065-8398>

Vanessa De Oliveira Collere - UNIVERSIDADE DA REGIÃO DE JOINVILLE (UNIVILLE) JOINVILLE – SC – BRASIL - Orcid: <https://orcid.org/0000-0002-7929-7753>

Social media has evolved from simple platforms for social interaction to a phenomena of dynamic global influence, which generates an immense quantity of data everyday. This event attracts the attention of institutions and entities from all fields, because this bulky and heterogeneous cluster of information can, if extracted and analysed correctly, provide singular perspectives. This article presents a Systematic Review, using the PRISMA protocol, with the intent on exploring the use of data from social media, together with Linked Open Data (LOD), for the development of ontologies. At the end of this research, it is noted that the mass of data from social media, if semantically enriched with LOD, can be used as a basis for ontology construction and that this is a recent application whose potential is still being expanded.

Keywords: Social media, PRISMA protocol, social interaction platforms, Linked Open Data, Systematic Review

ANÁLISE DA VIABILIDADE DO USO DE DADOS OBTIDOS DE MÍDIAS SOCIAIS PARA CONSTRUÇÃO DE ONTOLOGIAS: UMA REVISÃO DA LITERATURA

As mídias sociais evoluíram de simples plataformas de interação social para um fenômeno de influência global dinâmico, que gera uma quantidade imensa de dados todos os dias. Esse evento atrai a atenção de instituições e entidades de todas as categorias, pois essa volumosa e heterogênea aglomeração de informações pode, se extraída e analisada corretamente, oferecer perspectivas singulares. Este artigo apresenta uma Revisão Sistemática Literária, com o uso do protocolo PRISMA, que tem como intuito explorar o uso dos dados de redes sociais digitais, em conjunto com o uso de Dados Abertos Conectados (LOD), para o desenvolvimento de ontologias. Ao fim desta pesquisa, nota-se que a massa de dados proveniente de mídias sociais, se enriquecidos semanticamente com LOD, podem ser usados como base para a construção de ontologias e que se trata de uma aplicação recente cujo potencial ainda está sendo ampliado

Palavras-chave: mídias sociais, protocolo PRISMA, Dados Abertos Conectados, Revisão Sistemática Literária, plataformas de interação social

ANÁLISE DA VIABILIDADE DO USO DE DADOS OBTIDOS DE MÍDIAS SOCIAIS PARA CONSTRUÇÃO DE ONTOLOGIAS: UMA REVISÃO DA LITERATURA

Luiz Eduardo Magalhães Ferreira e Vanessa de Oliveira Collere

Bacharelado em Engenharia de Software
Universidade da Região de Joinville (Univille) Joinville – SC – Brasil

{luizeduardo.m.ferreira@gmail.com, vanessa.collere@univille.br}

Abstract. *Social media has evolved from simple platforms for social interaction to a phenomena of dynamic global influence, which generates an immense quantity of data everyday. This event attracts the attention of institutions and entities from all fields, because this bulky and heterogeneous cluster of information can, if extracted and analysed correctly, provide singular perspectives. This article presents a Systematic Review, using the PRISMA protocol, with the intent on exploring the use of data from social media, together with Linked Open Data (LOD), for the development of ontologies. At the end of this research, it is noted that the mass of data from social media, if semantically enriched with LOD, can be used as a basis for ontology construction and that this is a recent application whose potential is still being expanded.*

Resumo. As mídias sociais evoluíram de simples plataformas de interação social para um fenômeno de influência global dinâmico, que gera uma quantidade imensa de dados todos os dias. Esse evento atrai a atenção de instituições e entidades de todas as categorias, pois essa volumosa e heterogênea aglomeração de informações pode, se extraída e analisada corretamente, oferecer perspectivas singulares. Este artigo apresenta uma Revisão Sistemática Literária, com o uso do protocolo PRISMA, que tem como intuito explorar o uso dos dados de redes sociais digitais, em conjunto com o uso de Dados Abertos Conectados (LOD), para o desenvolvimento de ontologias. Ao fim desta pesquisa, nota-se que a massa de dados proveniente de mídias sociais, se enriquecidos semanticamente com LOD, podem ser usados como base para a construção de ontologias e que se trata de uma aplicação recente cujo potencial ainda está sendo ampliado.

1. Introdução

A interação em redes sociais tornou-se parte do cotidiano de milhares de pessoas ao redor do globo. Com o surgimento de plataformas distintas, diversas comunidades e nichos de interação são satisfeitos. O Twitter dá aos usuários a capacidade de interagir com eventos e pessoas em tempo real com o uso de mensagens curtas, denominados Tweets, esses são capazes de expressar sentimentos e demonstrar opiniões de forma dinâmica e sucinta. O Instagram permite que seus membros compartilhem memórias em formato de foto e vídeo. Postagens na linha do tempo são denominadas “stories” e

compõem a “história” de vida do usuário. Existem diversas plataformas distintas, cada uma oferece recursos de interação únicos, mas independente disso, esses geram dados, podendo ser textos, imagens, vídeos ou até mesmo relacionamentos, que podem ser extraídos e usados para diversas aplicações. Uma premissa recente explora o uso desses dados para a construção de ontologias, que são conjuntos de conceitualizações explícitas compartilhadas cuja estrutura fornece uma base de conhecimento organizada, que pode ser usada para a classificação de informações. No entanto, por se tratar de um campo de estudo recente, a comunidade científica ainda está descobrindo as possibilidades dessas aplicações e, com o intuito de colaborar com esse esforço, este artigo apresentará uma revisão sistemática literária com o objetivo de investigar as contribuições dos dados obtidos a partir das mídias sociais, em conjunto com Dados Abertos Conectados, do inglês Linked Open Data (LOD), para o auxílio na construção de ontologias e analisar a viabilidade do uso dessas plataformas como recursos de base para o desenvolvimento de Ontologias.

2. Revisão da literatura

2.1. Definições conceituais de mídia social

A concepção de rede social se estende através das épocas, Tönnies (1887, Apud Loomis, 2002) previu a existência de uma “rede” de vínculos sociais em seus estudos de grupos sociais em meados de 1890. A primeira plataforma digital a ser reconhecida “oficialmente” como uma “mídia social”, conforme as definições contemporâneas amplamente divulgadas e socialmente consolidadas, foi o SixDegrees, que consistia de um serviço de “rede social” hospedado em um website, este oferecia a capacidade do usuário de criar e personalizar um “perfil”, gerenciar uma lista de amigos e organizar afiliações “escolares”. Estes autores definem um arquétipo em formato de “colméia”, segmentado em sete partes distintas, cada uma dessas representa um aspecto específico da estrutura de uma rede social digital. Cada um dos “favos” possui particularidades das quais podem ser explicitamente exploradas de forma detalhada, este modelo foi criado com o intuito de proporcionar uma visão concreta do que Kietzmann et al. (2011) define como os “Blocos funcionais de uma mídia social”, com isso tem-se as definições:

- **Bloco “Identidade”:** O bloco localizado no interior do diagrama, denominado “Identidade” é o “coração” do conjunto, que define a extensão que os usuários revelam particularidades de sua personalidade o que pode ou não incluir informações como nome, idade, profissão, localização e quaisquer outros dados que retratem um indivíduo de uma maneira específica.
- **Bloco “Conversas”:** O bloco de “Conversas” representa a extensão a qual os usuários se comunicam uns com os outros dentro do ambiente de mídia social.
- **Bloco “Compartilhamento”:** O bloco de “Compartilhamento” representa a extensão que os usuários trocam, distribuem e recebem conteúdo. Embora o termo “social” implique troca de informações entre pessoas, Kietzmann et al. (2011) ratifica essa premissa ao explicar que embora crucial, a socialidade é construída a partir dos objetos que mediam os vínculos entre os indivíduos.

- **Bloco “Presença”:** O bloco de “Presença” representa a extensão a qual os usuários têm conhecimento em relação a acessibilidade de cada um, cada plataforma possui uma forma de representar a presença de seus usuários, algumas usam de sistemas de “status” como “disponível” ou “indisponível”, já outras oferecem a oportunidade de cada um indicar a quais lugares estão frequentando ou estão no momento.
- **Bloco “Relações”:** O contexto que representa a extensão em que os usuários se relacionam está associado ao bloco de “Relações”, este conceitualiza esta premissa através de objetos de socialização, sejam estes uma forma de conexão direta informal como “amizade” ou “seguidor” ou formal, como exemplo tem-se o LinkedIn, que constrói os vínculos através de ligações profissionais, como colegas de trabalho ou até mesmo indicações de trabalho.
- **Bloco “Reputação”:** A “Reputação” indica a extensão a qual os usuários são capazes de identificar o posicionamento de outros, inclusive a si mesmos, dentro do contexto da mídia social. O LinkedIn encarna esta abstração de forma plena, que constrói a reputação de seus usuários com base em suas ligações e seus endossos.
- **Bloco “Grupos”:** Por fim, o bloco de “Grupos” representa a extensão a qual os usuários são capazes de formar comunidades e subcomunidades.

2.2. Definições Conceituais de Ontologia

O conceito de Ontologia transcende diversas áreas de aplicações científicas. O termo original provém da filosofia, Guarino et al. (2009) explicam sua origem, que se refere a um ramo filosófico denominado metafísica, este que lida com a estrutura e organização da realidade. A Ontologia se difere das ciências experimentais pois se foca na natureza e estrutura das coisas independente de quaisquer outras possíveis considerações, inclusive a respeito da própria existência daquelas. No caso específico da Ciência da Computação, conforme Guarino et al. (2009), a Ontologia se refere a um tipo especial de objeto de informação ou artefato computacional. Ontologias Computacionais são métodos capazes de viabilizar a estruturação formal de um sistema, como por exemplo o mapeamento de entidades e relações e todos os outros conceitos que sejam úteis para o propósito desejado. Studer et al. (1998) definem uma Ontologia como uma especificação explícita formal de uma conceitualização compartilhada e a partir dessa definição, Guarino et al. (2009) extraem três abstrações:

Conceitualização: conforme Guarino et al. (2009), uma conceitualização, também denominada uma estrutura relacional intencional, pode ser definida por $C = (D, W, \mathfrak{R})$, onde D é o universo de discurso, W é um conjunto possível de Mundos e \mathfrak{R} é um conjunto de relações conceituais em um domínio de espaço $\langle D, W \rangle$

Comprometimento Ontológico: *Mundos* Pode ser definido como um conjunto totalmente ordenado de estados de *Mundo*, este definido como uma designação única de valores para todas as variáveis que caracterizam um dado sistema S . Sendo S um sistema arbitrário, D um conjunto arbitrário de elementos de S e W um conjunto de estados de *Mundo* para S , uma relação intencional pode ser definida como p^n com uma aridade n

no domínio $\langle D, W \rangle$, esta é uma função total $p^n: W \rightarrow 2^{D^n}$ do conjunto W no conjunto de n relações em D .

Modelos Pretendidos: Conforme Guarino et al. (2009), um modelo intencionado ou pretendido é definido por: “Sendo $C = (D, W, \mathfrak{R})$ uma conceitualização, L uma linguagem lógica de primeira ordem com vocabulário V e comprometimento ontológico $K = (C, \mathfrak{S})$. Um modelo $M = (S, I)$, com $S = (D, R)$ é chamado de modelo intencional de L de acordo com K se e somente se: Para todos os símbolos constantes $c \in V$ tem-se um $I(c) = \mathfrak{S}(c)$ e existe um mundo $w \in W$ de forma que, para cada símbolo predicado $v \in V$ existe uma relação intencional $p \in \mathfrak{R}$ tal que $\mathfrak{S}(v) = p$ e $I(v) = p(w)$.

O conjunto $I_K(L)$ de todos os modelos de L que são compatíveis com K são denominados modelos intencionados de L em relação à K ”.

Portanto, uma Ontologia é definida formalmente, conforme Guarino et al. (2009), como: “Seja C uma conceitualização, L uma linguagem lógica com vocabulário V e comprometimento ontológico K . Uma Ontologia O_K para C com vocabulário V e comprometimento ontológico K é uma teoria lógica que consiste de uma série de fórmulas de L , desenvolvidas de forma que o conjunto de modelos aproximam-se o máximo possível dos modelos intencionados de L de acordo com K ”.

2.3. Definições Conceituais de Dados Abertos Conectados

De acordo o W3c (2020), *World Wide Web Consortium*, Dados Abertos Conectados, do inglês *Linked Open Data* (LOD) tratam-se de um conjunto de dados de acesso livre interligados de forma nítida e com formatação estruturada. É importante que efetuadas padronizações rígidas para permitir o rápido e fácil acesso aos dados, dentre esses destacam-se os modelos RDF (*Resource Description Framework*), OWL (*Ontology Web Language*) e SKOS (*Simple Knowledge Organization System*). Dentre as principais bases de LOD tem-se o DBPedia, que indexa conteúdo da Wikipedia, e o YAGO3, que é fundamentado pela Wikipedia e a Geonet. Todas essas bases são mantidas como parte do esforço do Linked Open Data Cloud, uma gigantesca base que agrega diversos conjuntos de dados estruturados, ou datasets, pesquisadores e cientistas são encorajados a submeter suas coleções de dados à plataforma. LOD's oferecem um complemento semântico aos dados desestruturados obtidos das redes sociais digitais e, devido a sua sistematização, disponibilidade e larga escala, podem ser usados para enriquecer os domínios básicos extraídos da distinção dos conceitos que foram definidos a partir das informações obtidas de mídias sociais (Suliman et al. , 2016).

3. Procedimentos metodológicos

Neste artigo será aplicada a metodologia de Revisão Sistemática Literária (RSL), uma pesquisa qualitativa de caráter exploratório, que faz uso da literatura disponível relativa ao tema., As RSL Seguem um conjunto de métodos científicos que têm o objetivo explícito de limitar tendenciosidades, principalmente por procurar identificar, analisar e sintetizar todos os estudos relevantes (de qualquer delineamento), para

responder a uma questão específica. (PETTICREW, M.; ROBERTS, H., 2006). A busca desses dados é feita em bases acadêmicas indexadas e representam a perspectiva da comunidade científica. Uma RSL pode ser aplicada a quaisquer cenários de estudo acadêmico, mas é pertinente em casos específicos: Para revisar todos os dados acerca de uma questão específica, principalmente se existe uma incerteza em relação a resposta; Para que decisões possam ser efetuadas de forma imparcial e transparente, com base em conceitos defensáveis

É crucial para a Revisão Sistemática que sejam estabelecidos métodos delineados e concretos e que esses sejam seguidos de forma rígida para que essa cumpra seu papel na eliminação da subjetividade. Neste estudo foi usado o protocolo PRISMA (*Preferred Reporting Items for Systematic review and Meta-Analysis*), que é composto de um diagrama de fluxo de pesquisa e um conjunto de diretrizes, separadas em três seções: Informações administrativas, Introdução e Métodos. Na seção “Informações administrativas” são definidos critérios como título, contribuições e contatos, em “Introdução” é apresentado o contexto da revisão e definido a pergunta científica. Os “Métodos” inclui os critérios de elegibilidade, a definição das fontes de dados, a estratégia de busca, o gerenciamento dos dados e a síntese. O contexto a ser analisado, nesta situação, trata-se da construção de Ontologias, com isso, tem-se a seguinte questão:

Quais as contribuições das mídias sociais no desenvolvimento de Ontologias?

Esta questão, em conjunto com os critérios de elegibilidade e exclusão, guiaram os esforços deste estudo

Foram definidos construtos que constituem o arquétipo de palavras-chave a serem utilizadas na pesquisa, essas foram definidas inicialmente com base no objeto de estudo proposto.

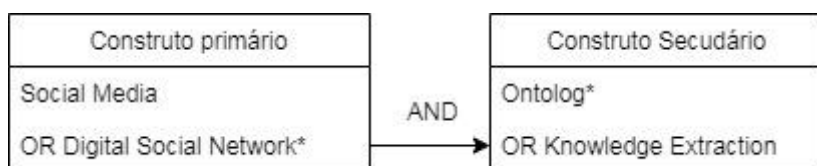


Figura 1: Construtos de busca
Fonte: Prória

São considerados elegíveis apenas os artigos que se enquadrem nos critérios: Escritos em inglês ou português; Por tipo de publicação, deve ser artigo ou revisão sistemática; Se refiram às áreas relevantes para o objeto de estudo {*Computer Science; Software Engineering; Information Technology*}; Sejam estudos que demonstrem a aplicabilidade dos dados obtidos a partir das redes sociais para construção, extração ou enriquecimento de ontologias.

Estão excluídos os artigos que se enquadram nos critérios: Refiram-se a áreas de estudo não relevantes as ciências da computação; Tenham sido publicados há mais de 10 (dez) anos; Impossibilidade de acesso ao texto completo do artigo; Não utiliza Dados Abertos Conectados

A base de dados selecionada como fonte é o Web of Science, abreviado como WoS. Essa foi selecionada por se tratar de uma base multidisciplinar internacional que tem como intuito indexar os períodos mais citados em suas respectivas áreas.

4. Análise dos dados e discussão dos resultados

Com o auxílio dos construtos foi elaborada a seguinte string de busca:

“TS = (("Social Media" OR "Digital Social Network*") AND ("Ontolog*" OR "Knowledge Extraction"))”

A imagem abaixo representa os resultados obtidos em cada etapa do fluxo de pesquisa do protocolo PRISMA

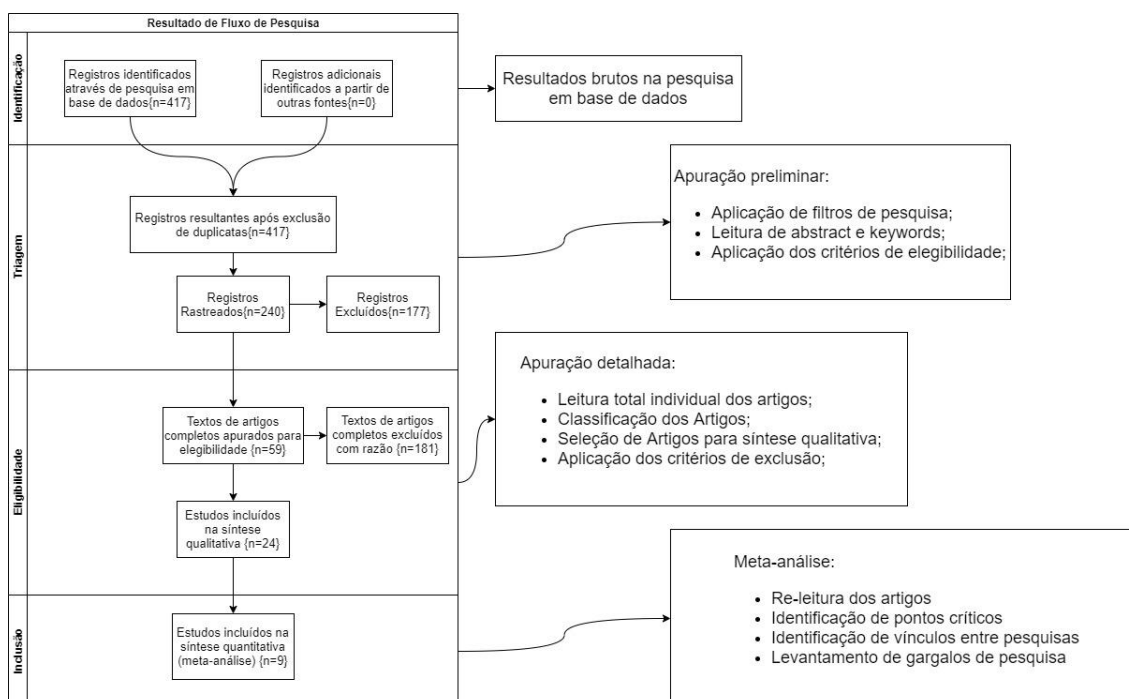


Figura 2: Fluxo de pesquisa PRISMA aplicado
Fonte: Adaptado de PRISMA(2019).

Conforme a figura 2, no processo de identificação, foi realizada a busca com o uso da string, efetuada no dia 18 de Maio de 2020. No processo de triagem foram aplicados filtros com base nos critérios de elegibilidade, primeiramente foram desconsiderados resultados de mais de 10 (Dez) anos atrás, reduzindo o número de registros do resultado à 415. Com este resultado, foi então aplicado um refinamento de áreas de estudo, as seguintes categorias foram selecionadas: *Computer Science Information Systems*; *Computer Science Theory Methods*; *Computer Science Artificial*

Intelligence; Computer Science Interdisciplinary Applications; Information Science Library Science; Computer Science Software Engineering; Computer Science Cybernetics;

Com esse refinamento, reduziu-se a quantidade de artigos a serem analisados para 240, nos quais foi feita a leitura preliminar dos resultados, a partir das informações contidas no *Abstract* e nas *Keywords* (Palavras-chave). Deste procedimento, foram extraídos 71 artigos relevantes, que foram incluídos na etapa de elegibilidade. Esses artigos foram lidos individualmente e classificados conforme a sua área específica de aplicação. Antes de efetuar a análise, foi efetuada uma busca para verificar o acesso aos artigos, com isso, notou-se que não estavam disponíveis os textos completos de 12 artigos e, portanto, não foram incluídos nesta análise. Após a leitura do texto completo de cada um dos artigos, foi elaborada uma classificação, construída com base na aplicação da pesquisa, com isso, tem-se as seguintes categorias: Análise Forense; Ciência da Computação; Análise de Sentimentos/Mineração de Opiniões; Empresarial; Rastreamento de Desastres; Educacional; Saúde; Internet das Coisas; Rastreamento de localidades; Militar.

Com esta classificação é possível separar os artigos em áreas de aplicação, considerando um total de 59 artigos, obteve-se o seguinte resultado: Ciência da Computação - 24 artigos; Empresarial - 11; Análise de Sentimentos/Mineração de Opiniões - 10; Educacional - 4; Análise Forense - 3; Rastreamento de Desastres - 2; Saúde - 2; Internet das Coisas - 1; Rastreamento de localidades - 1; Militar - 1. Dos 24 artigos da classe “Ciência da Computação”, 15 não fazem uso de LOD, o que limitou o escopo da etapa de inclusão à 9 artigos, que podem ser observados no quadro 1.

Título/Ano	Autores	Contexto
A Semantic BI Process for Detecting and Analyzing Mentions of Interest for a Domain in Tweets(2018)	Pereira, Vilmar César & Fileto, Renato & Souza, Willian Santos De & Wittwer, Matthias & Reinhold, Olaf & Alt, Rainer	O artigo propõe uma aproximação para analisar as menções de interesse de um dado conceito , combinando Web Semântica com Inteligência de Negócio. O processo tem como fundamentação um método automático de ETL (Extract, Transform e Load) que faz ligações entre tweets com LOD's, filtra as fontes LOD relevantes nas anotações utilizando pontes entre a base e uma ontologia de alto nível e adapta as LOD's hierarquicamente para servir de dimensões de análise.
Type Prediction Combining Linked Open Data and Social Media(2018)	Nechaev, Yaroslav & Corcoglioniti, Francesco & Giuliano, Claudio	O artigo apresenta uma metodologia , com o uso de LOD, em conjunto com os dados de mídias sociais para a predição de atributos de entidades , em específico, os tipos de uma determinada entidade. Os autores demonstram que os dados de uma mídia social. Neste caso o Twitter, é capaz de complementar as representações RDF (Resource Description Framework) das entidades de um grafo de conhecimento, neste caso o DBpedia, quando utilizado como entrada para um sistema supervisionado de predição de tipos.

Ontology-based approach for identifying the credibility domain in social Big Data(2018)	Wongthontham, Pornpit & Abu-Salih, Bilal.	O artigo propõe um método de extração de Big data desestruturada a partir de postagens do Twitter, esses dados passam por um processo de refinamento e então são interligados com notações de entidade do DBPedia e do Yago, para serem então armazenados em formato RDF.
Twitter mining for ontology-based domain discovery incorporating machine learning(2018)	Abu-Salih, Bilal & Wongthontham, Pornpit & Chan, Kit.	Os autores construíram um framework capaz de explorar domínios ontológicos , aplicado ao Twitter, é capaz de efetuar classificações a nível de usuário e a nível de tweets. Os autores dedicam seus esforços em definir se o dado pertence ou não ao domínio político. Para a inferência do conteúdo extraído dos dados textuais dos usuários, os autores fazem uso da integração de LOD's, entre elas, o DBPedia, O Freebase e o Yago.
Extracting Emerging Knowledge from Social Media(2017)	Brambilla, Marco & Della Valle, Emanuele & Volonterio, Riccardo & Salazar, Felix.	O principal foco da pesquisa é identificar entidades "emergentes" dado um tipo central de entidade , uma seção do grafo de conhecimento, denominado grafo de domínio, é definido, neste caso, foram escolhidos Moda(Fashion), Escrita e Exibições. Os autores fazem uso do DBPedia como fundamentação ontológica e, para extrair conteúdo de cunho social, foi selecionado o Twitter.
Ontology for cultural variations in interpersonal communication: Building on theoretical models and crowdsourced knowledge(2016)	Thakker, Dhavalkumar & Karanasios, Stan & Blanchard, Emmanuel & Lau, Lydia & Dimitrova, Vania.	O artigo apresenta um modelo ontológico , denominado AMOn, que define variações culturais e permite que esse seja usado para a inferência de menções culturais em conteúdo textual. A ontologia é definida pelos autores como híbrida, pois faz uso de operações de base empírica e teórica, esta é definida pois a fundamentação é feita com teorias de atividade e cultural, aquela é caracterizada pelo uso da base DBPedia.
Event Identification and Assertion from Social Media Using Auto-Extendable Knowledge Base(2016)	Suliman, Ahmed & Kaabi, Khaled & Wang, Di & Al-Rubaie, Ahmad & Al-Dhanhani, Ahmed & Ruta, Dymitr & Davies, John & Stincic Clarke, Sandra.	Os autores constroem uma base de conhecimento estruturado com um método que faz uso de uma ontologia compacta denominada "Ontologia Semente" esta é então alimentada com dados do Twitter e estendida com o uso de um grafo de conhecimento da ConceptNet5, uma base LOD de tamanho comparável ao DBPedia.
Learning Structured Knowledge from Social Tagging Data A critical review of methods and techniques(2015)	Dong, Hang & Wang, Wei & Liang, Hai-Ning.	O artigo apresenta uma revisão crítica , primeiramente definindo o escopo do processo de categorização de Sistemas de Organização do conhecimento, do inglês Knowledge Organization Systems (KOS), de bases desestruturadas e de semântica precária até bases altamente organizadas e de semântica rica.

Using Social Media for Ontology Enrichment(2010)	Monachesi, Paola & Markus, Thomas.	O artigo propõe um método para o enriquecimento do domínio ontológico a partir do uso de dados extraídos de uma plataforma de mídia social, heurísticas, a base DBPedia e um algoritmo responsável pela desambiguação de termos, com o intuito de providenciar ontologias dinâmicas e personalizadas que incluem o conhecimento de uma comunidade de usuários.
--	------------------------------------	---

Quadro 1: Artigos incluídos na meta-análise
Fonte: Próprio.

Nota-se que dentre os artigos analisados, às pesquisas de Monachesi e Markus (2010), Suliman et al. (2016) e Wongthontham et al. (2018) se delimitam à construção ou elaboração de métodos para extração de conhecimento estruturado das mídias sociais. As pesquisas de Brambilla et al. (2017), Nechaev et al. (2018) e Pereira et al. (2018) desenvolvem métodos para uma aplicação específica. Brambilla et al. (2017) se dedica a extrair entidades emergentes dado um domínio específico. Nechaev et al. (2018) usam os dados das mídias sociais em conjunto com LOD para prever atributos de entidades, neste caso, o tipo de uma entidade. Pereira et al. (2018) propõem analisar as menções de interesse de um dado conceito. Abu-Salih et al. (2018) propõem a construção de um framework capaz de explorar domínios ontológicos, com o intuito de afirmar se um dado pertence ou não a um domínio. Thakker et al.(2017) apresentam um modelo ontológico que é capaz de definir variações culturais, com os dados das mídias sociais como base teórica e a LOD como base empírica. Dong et al. (2015) apresentam uma revisão crítica que define o escopo do processo de categorização de Sistemas de Organização do conhecimento e analisa o processo de folksonomia, definido por Dong et al. (2015) como sistemas de marcação" social, do inglês "social tagging systems", para a extração de lista de termos e o aprendizado de relacionamentos.

Pode-se notar que o Twitter é utilizado constantemente como plataforma alvo, devido a sua natureza de microblog, que gera uma grande quantidade de informações de uma pletera de assuntos distintos expressada por indivíduos diferentes. Esse fato é explorado por Suliman et al. (2016), Wongthontham et al. (2018), Pereira et al. (2018), Abu-Salih et al. (2018), Brambilla et al. (2017) e Nechaev et al. (2018). A definição de Kietzmann et al. (2011) é capaz de explicar este fenômeno, dado os blocos fundamentais de uma rede social, o Twitter é construído com o foco em ‘Compartilhamento’ e ‘Conversas’, o que faz com que o nível de interação seja muito mais ativo do que em outras mídias sociais, como por exemplo, o Facebook, cujo foco é em ‘Relacionamentos’.

O enriquecimento semântico com o uso de LOD's é um processo fundamental para extração de conhecimento de mídias sociais, pois esses dados em sua forma bruta são desestruturados e poluídos. Conforme Guarino et al. (2009) uma ontologia é definida por uma teoria lógica formada por uma série de Linguagens Lógicas que devem se aproximar ao máximo dos modelos pretendidos conforme o comprometimento ontológico, que é um conjunto ordenado de estados de *Mundos*. Isso infere em uma estrutura sistematizada, do contrário, é impossível identificar o sentido de um dado

conceito extraído, por isso esse deve ser enriquecido com o auxílio de uma base organizada, neste caso, LOD's. Os trabalhos de Pereira et al. (2018), Nechaev et al. (2018), Wongthontham et al. (2018), Suliman et al. (2016) e Monachesi e Markus (2010) demonstram que os dados das mídias sociais, quando aplicados com anotação e enriquecimento semântico com auxílio de LOD's, podem fornecer bases de conhecimento altamente estruturadas.

A pesquisa de Monachesi e Markus (2010), demonstram o quão útil os dados de mídias sociais podem ser no processo de construção de ontologias, pois podem viabilizar o desenvolvimento de ontologias de domínio, que fornecem dados estruturados referente a um tópico que possui pouca informação relevante em bases LOD. Outro aspecto benéfico do uso de dados de mídias sociais pode ser observado na pesquisa de Brambilla et al. (2017), que afirmam que “ As mídias sociais incluem conhecimento emergente de baixa frequência que pode ser extraído utilizando métodos de semântica independentes de domínio”.

A pergunta de pesquisa proposta neste trabalho foi construída no intuito de explorar a extensão de utilidade dos dados das mídias sociais para a construção de ontologias. Diversas aplicações foram propostas, no entanto, pode-se observar que gargalos ainda existem, como exemplo pode-se citar o ramo da Inteligência Artificial, que não fora explorado de forma plena nos artigos analisados. O método de extração proposto por Monachesi e Markus (2010), se aplicado a uma mídia social de conhecimento específico, como por exemplo o Stack Overflow que contém conhecimento explícito da área da ciência da computação, poderia fornecer uma ontologia que poderia ser usada de base para uma IA responsável por empregar o uso de conhecimento técnico para a otimização de processos. A pesquisa de Suliman et al. (2016) demonstra que é possível utilizar os dados das mídias sociais para a identificação de eventos, em conjunto com o método proposto por Brambilla et al. (2017), podem ser usados para a medição e predição de conteúdos emergentes, que podem ser usados para o direcionamento de publicidade. Além disso, em nenhuma das pesquisas analisadas foi possível identificar esforço para a identificação do padrão de comportamento humano, que pode ser útil não só para o desenvolvimento de IA's mas também para a construção de sistemas mais adequados a percepção humana.

Uma outra aplicação, viabilizada com as ferramentas desenvolvidas por Monachesi e Markus (2010), é a construção de ontologias únicas cujo conteúdo não é vasto ou esteja descentralizado, como por exemplo, pesquisas médicas referentes a uma patologia específica. O modelo ontológico de Thakker et al(2017) pode ser utilizado para a identificação de ofensas de caráter cultural, além disso, o mesmo método de construção pode ser utilizado para desenvolver ontologias de assuntos específicos, principalmente temas como política e economia. O método de Pereira et al. (2018) é capaz de fazer ligações entre clips textuais, textual clips (TC) com Dados Abertos Conectados. Pereira et al. (2018) denominam Textual Clip (TC) qualquer texto livre acompanhado de uma estampa de tempo e, em alguns casos, coordenadas geográficas ou indicadores de localidade, além de outros metadados. Entre os exemplos citados por Pereira et al. (2018) tem-se postagens em redes sociais por usuários e registros médicos

de pacientes, portanto é possível utilizar este método para facilitar a análise de dados e o rastreamento de padrões, o que pode ser útil na predição de pandemias ou na identificação de padrões comportamentais.

As pesquisas analisadas demonstram que os dados das mídias sociais são capazes de contribuir de diversas formas para a construção de ontologias, as pesquisas de Monachesi e Markus (2010), Suliman et al. (2016) e Wongthontham et al. (2018) demonstram que é possível extrair conhecimento das mídias sociais e fazer uso de LOD para enriquecer este conteúdo para a construção de ontologias. Brambilla et al. (2017) mostram como é possível identificar entidades emergentes a partir de dados de mídias sociais. A metodologia proposta por Nechaev et al. (2018) mostra que a injeção de dados de mídias sociais podem complementar as representações de um grafo de conhecimento, quando utilizados como entrada para um sistema supervisionado de predição de tipos. Thakker et al. (2017) constrói uma ontologia capaz de inferir menções culturais utilizando como base dados de mídias sociais e LOD's. Abu-Salih et al. (2018) foram capazes de construir um framework que identifica se um dado conceito pertence ou não a um domínio específico.

5. Conclusão

A partir da análise feita, pode-se concluir que as mídias sociais são úteis para extrair massa de dados que podem ser refinados e tratados para oferecer conhecimento estruturado de forma dinâmica, que pode ser utilizado para o enriquecimento, construção e alimentação de Ontologias. Além disso, devido a natureza do contexto das redes sociais digitais, é possível extrair compreensões inéditas, como demonstrado nas pesquisas de Brambilla et al. (2017) e Monachesi e Markus (2010). O processo de Folksonomia, conforme Dong et al. (2015), é capaz de fornecer uma base semi-estruturada que pode ser estendida para a construção de um grafo de conhecimento. Devido à natureza imprevisível dos usuários, pode-se afirmar que os dados gerados de mídias sociais, sejam esses vídeos, imagens ou textos, são capazes de servir como base para a construção de ontologias se tratados corretamente, conforme visto nas pesquisas de Pereira et al. (2018), Nechaev et al. (2018), Wongthontham et al. (2018), Suliman et al. (2016) e Monachesi e Markus (2010). Ademais, pode-se notar que o aproveitamento dessa massa de dados é um campo de estudo recente e que pesquisas ainda estão ampliando o seu limite de aplicação. Dada essa premissa, este artigo apresenta uma estrutura sistematizada que revela o estado da arte contemporâneo do uso de dados de mídias sociais para o desenvolvimento de ontologias, o que oferece à comunidade científica uma orientação para futuras pesquisas.

Referências

ABU-SALIH, Bilal; WONGTHONGTHAM, Pornpit; CHAN, Kit Yan. Twitter mining for ontology-based domain discovery incorporating machine learning. *Journal Of Knowledge Management*, [s.l.], v. 22, n. 5, p. 949-981, 11 jun. 2018. Emerald. <http://dx.doi.org/10.1108/jkm-11-2016-0489>.

BRAMBILLA, Marco; CERI, Stefano; DELLA VALLE, Emanuele; VOLONTERIO, Riccardo; SALAZAR, Felix Xavier Acero. Extracting Emerging Knowledge from Social Media. Proceedings Of The 26th International Conference On World Wide Web, [s.l.], p. 795-804, 3 abr. 2017. International World Wide Web Conferences Steering Committee. <http://dx.doi.org/10.1145/3038912.3052697>.

DONG, Hang; WANG, Wei; LIANG, Hai-ning. Learning Structured Knowledge from Social Tagging Data: a critical review of methods and techniques. 2015 Ieee International Conference On Smart City/socialcom/sustaincom (smartcity), [s.l.], p. 307-314, dez. 2015. IEEE. <http://dx.doi.org/10.1109/smartcity.2015.89>.

GUARINO, Nicola; OBERLE, Daniel; STAAB, Steffen. What Is an Ontology? Handbook On Ontologies, [s.l.], p. 1-17, 2009. Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-540-92673-3_0.

KIETZMANN, Jan H.; HERMKENS, Kristopher; MCCARTHY, Ian P.; SILVESTRE, Bruno S.. Social media? Get serious! Understanding the functional building blocks of social media. Business Horizons, [s.l.], v. 54, n. 3, p. 241-251, maio 2011. Elsevier BV. <http://dx.doi.org/10.1016/j.bushor.2011.01.005>.

MONACHESI, Paola; MARKUS, Thomas. Using Social Media for Ontology Enrichment. Lecture Notes In Computer Science, [s.l.], p. 166-180, 2010. Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-642-13489-0_12.

NECHAEV, Yaroslav; CORCOGLIONITI, Francesco; GIULIANO, Claudio. Type Prediction Combining Linked Open Data and Social Media. Proceedings Of The 27th Acm International Conference On Information And Knowledge Management, [s.l.], p. 1033-1042, 17 out. 2018. ACM. <http://dx.doi.org/10.1145/3269206.3271781>.

PEREIRA, Vilmar César; FILETO, Renato; SOUZA, Willian Santos de; WITTEWER, Matthias; REINHOLD, Olaf; ALT, Rainer. A Semantic BI Process for Detecting and Analyzing Mentions of Interest for a Domain in Tweets. Proceedings Of The 24th Brazilian Symposium On Multimedia And The Web - Webmedia '18, [s.l.], p. 197-204, 2018. ACM Press. <http://dx.doi.org/10.1145/3243082.3243100>.

PRISMA. Disponível em: <http://www.prisma-statement.org/PRISMAStatement/FlowDiagram>. Acesso em: 19 nov. 2019.

Psychology Today: Six Degrees: Urban Myth? Disponível em: <https://www.psychologytoday.com/intl/articles/200203/six-degrees-urban-myth>. Acesso em: 19 nov. 2019.

MOHER, David; LIBERATI, Alessandro; TETZLAFF, Jennifer; ALTMAN, Douglas G.. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the prisma

statement. Plos Medicine, [s.l.], v. 6, n. 7, e1000097, 21 jul. 2009. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pmed.1000097>.

STUDER, Rudi; BENJAMINS, V.richard; FENSEL, Dieter. Knowledge engineering: principles and methods. Data & Knowledge Engineering, [s.l.], v. 25, n. 1-2, p. 161-197, mar. 1998. Elsevier BV. [http://dx.doi.org/10.1016/s0169-023x\(97\)00056-6](http://dx.doi.org/10.1016/s0169-023x(97)00056-6).

SULIMAN, Ahmed Talal; KAABI, Khaled Al; WANG, di; AL-RUBAIE, Ahmad; DHANHANI, Ahmed Al; RUTA, Dymitr; DAVIES, John; CLARKE, Sandra Stincic. Event identification and assertion from social media using auto-extendable knowledge base. 2016 International Joint Conference On Neural Networks (ijcnn), [s.l.], p. 4443-4450, jul. 2016. IEEE. <http://dx.doi.org/10.1109/ijcnn.2016.7727781>.

THAKKER, Dhavalkumar; KARANASIOS, Stan; BLANCHARD, Emmanuel; LAU, Lydia; DIMITROVA, Vania. Ontology for cultural variations in interpersonal communication: building on theoretical models and crowdsourced knowledge. Journal Of The Association For Information Science And Technology, [s.l.], v. 68, n. 6, p. 1411-1428, 5 maio 2017. Wiley. <http://dx.doi.org/10.1002/asi.23824>.

The Linked Open Data Cloud. Disponível em: <https://lod-cloud.net/>. Acesso em: 15 jun. 2020.

TÖNNIES, Ferdinand; LOOMIS, Charles Price. Community and Society. Nova Iorque: Dover Publications, Inc, 2002. 294 p.

TÖNNIES, Ferdinand et al. Gemeinschaft und Gesellschaft: abhandlung des communismus und des socialismus als empirischer culturformen. Leipzig: Fues, 1887. 294 p.

W3C. Disponível em: <https://www.w3.org/standards/semanticweb/data>. Acesso em: 15 jun. 2020.

WONGTHONGTHAM, Pornpit; SALIH, Bilal Abu. Ontology-based approach for identifying the credibility domain in social Big Data. Journal Of Organizational Computing And Electronic Commerce, [s.l.], v. 28, n. 4, p. 354-377, 2 out. 2018. Informa UK Limited. <http://dx.doi.org/10.1080/10919392.2018.1517481>.