

A CONECTIVIDADE E A ORGANIZAÇÃO DA INFORMAÇÃO: Uma Abordagem entre a Internet das Coisas e a Web Semântica

Débora Cristina Bonfim Aquarone

Especialista em Gestão da Informação Digital
Fundação Escola de Sociologia e Política de São Paulo
debora-bonfim@uol.com.br

Charley dos Santos Luz

Mestre em Ciência da Informação
Fundação Escola de Sociologia e Política de São Paulo
charley@feedconsultoria.com.br

Artigo de Revisão

Resumo

Aborda como se dá o armazenamento e a organização dos dados impulsionados pela conectividade proporcionada pela Internet das Coisas (IdC). Nesta análise da situação destaca-se o desenvolvimento de protocolos eficientes que possibilitem a conectividade de todos os objetos presentes na IdC. Sugere a Organização da Informação (OI) como base para a estrutura dessas informações na rede da web semântica e as ontologias como camadas de sentido para as informações nas redes neurais.

Palavras-chave

Internet das Coisas. Inteligência Artificial. Web Semântica. Ontologia. Interoperabilidade.

1 INTRODUÇÃO

Considerada a terceira onda no que se refere ao desenvolvimento de produtos e soluções na chamada Era da Internet, a Internet das Coisas (IdC) ou *Internet of Things* (IoT), em inglês, supera a primeira onda registrada nos anos 1990 em que foi possível conectar 1 bilhão de pessoas por meio de linhas e cabos telefônicos e a segunda onda, nos anos 2000 cujas tecnologias móveis conectaram outros 2 bilhões de pessoas. A IdC tem potencial para conectar bilhões de “coisas” à Internet - graças a endereços únicos de IP para cada “coisa” - que se comunicam por meio de protocolos padronizados.

Como fenômeno proveniente dos avanços tecnológicos, a IdC tem um caráter disruptivo, modificando - em certas áreas, com grande impacto - não só os aspectos

sociais nas relações humanas, mas também econômicos.

A IdC já é uma realidade e cada vez mais estaremos conectados uns aos outros e aos objetos sem que nos sintamos obrigados a aderir às facilidades que nos são ofertadas. Não nos sentimos invadidos ou incomodados com a presença - muitas vezes, quase invisível - das máquinas em nosso dia a dia.

O matemático Alan Turing previu que

A inteligência das máquinas se tornaria tão penetrante, tão confortável e tão bem integrada à nossa economia baseada em informação que as pessoas sequer conseguiriam se dar conta disso. (KURZWEIL, 2007, p. 108).

Talvez o desconforto mais sentido por alguns de nós, nos últimos anos, tenha

sido em relação à invasão da privacidade digital. Embora a IdC suscite várias questões como a segurança digital ou características relevantes como a ubiquidade, temas que emergem com frequência na literatura, podemos destacar a conexão entre objetos, e objetos e pessoas como a característica que a define.

Essas redes pervasivas têm a característica de conectar não apenas humanos a humanos, mas também humanos a objetos e objetos a objetos. A IdC corresponde à fase atual da Internet em que os objetos se relacionam com objetos humanos e animais os quais passam a ser objetos portadores de dispositivos computacionais capazes de conexão e comunicação. Nesse sentido, os objetos tendem a assumir o controle de uma série de ações do dia a dia, sem necessidade de que as pessoas estejam atentas e no comando. (SANTAELLA et al, 2013).

A análise da situação, calcada majoritariamente na revisão da literatura para a elaboração deste artigo, se justifica para ilustrar que os desafios que a IdC impõe são aqueles decorrentes da heterogeneidade e de complexas relações, portanto, dizem respeito à interoperabilidade e à geração exponencial de dados que demandam soluções para o processamento e o armazenamento e que devem possibilitar conexões eficientes, para gerar respostas eficientes, dado que a IdC, entendida como a conexão de objetos e pessoas, tem por finalidade satisfazer necessidades e oferecer soluções com rapidez.

Assim conceituamos IdC, destacamos o desafio da organização de informação nesta realidade, verificamos sua relação com a área de organização da informação da Ciência da Informação, e revisamos o conceito de dados, informação e conhecimento. A partir disso, abordamos os dados estruturados e semiestruturados, a web semântica e as redes semânticas e redes neurais e, finalmente, destacamos as ontologias como forma de organizar os dados na IdC. O objetivo desta análise é evidenciar os mecanismos que vêm sendo adotados para o processamento e a estruturação das informações geradas pela IdC e apresentar os diversos campos do

conhecimento humano que contribuem para esse arcabouço.

2 METODOLOGIA APLICADA

A metodologia utilizada para a elaboração desse artigo foi a revisão da literatura a fim de sustentar as definições dos diversos conceitos que envolvem a organização das informações no ambiente virtual, tendo a IdC como foco, bem como uma análise empírica sobre a evolução tecnológica.

Constituíram fontes da pesquisa artigos e livros em português e inglês, listados nas referências. As pesquisas foram realizadas com textos utilizados em sala de aula, na disciplina de *Arquitetura de Informação* do curso de pós-graduação em Gestão da Informação Digital da Fundação Escola de Sociologia e Política de São Paulo (FESPSP) e leitura selecionada na base de dados Pergamum da mesma faculdade e também da Escola Superior de Propaganda e Marketing (ESPM), por meio de pesquisa das seguintes palavras: internet das coisas, web semântica e ontologia.

A partir dessa revisão, criamos condições para a análise da situação sobre os desafios de processamento e da organização inteligente de dados diante do crescente número de objetos conectados fato propiciado pela IdC. Talvez uma nova arquitetura da rede se imponha como necessária e urgente, alicerçada na importante contribuição que as ontologias têm a oferecer, ou, ainda, que isso propicie o desenvolvimento de ontologias que reduzam cada vez mais as ambiguidades semânticas. Os desafios não se limitam ao processamento e ao armazenamento, mas também na análise em tempo real para que esses dados possam, de fato, serem úteis.

3 A INTERNET DAS COISAS

Com o advento da Internet, a vida das pessoas foi modificada em muitos aspectos, seja na maneira de fazer negócios, seja na maneira como os relacionamentos se estabelecem ou ainda como as pessoas se divertem. (BEHROUZ, 2010). A Internet é um sistema estruturado e organizado e, para

isso, utiliza uma linguagem comum de comunicação para que possa interligar uma rede de computadores. Castells (2003, p. 7) define rede como sendo um

[...] conjunto de nós interconectados. A formação de redes é uma prática humana muito antiga, mas as redes ganharam vida nova em nosso tempo transformando-se em redes de informações energizadas pela Internet. As redes têm vantagens extraordinárias como ferramentas de organização em virtude de sua flexibilidade e adaptabilidade inerentes, características essenciais para se sobreviver e prosperar num ambiente em rápida mutação.

O conceito da IdC começou a desenhar-se na década de 90, quando o *Transmission Control Protocol - Internet Protocol* (TCP/IP) e as interfaces gráficas baseadas na *Wide World Web* (WWW) proporcionaram a popularização da Internet. Durante a pesquisa para a elaboração desse artigo, foram encontradas diversas fontes acadêmicas e não acadêmicas que atribuem ao pesquisador britânico Kevin Ashton, do *Massachusetts Institute of Technology* (MIT), a criação do termo “Internet das Coisas”, quando em 1999, a utilizou durante a condução de uma palestra para executivos da *Procter & Gamble*. Na ocasião, o especialista falou sobre a ideia de etiquetar eletronicamente os produtos da empresa para facilitar a logística da cadeia de produção por meio de identificadores de rádio frequência (em inglês, *Radio Frequency Identification*, termo mais conhecido pela sigla RFID), sensores e atuadores. Segundo o próprio Ashton (2014, site), “a expressão Internet das Coisas pode nem ser tão brilhante, mas deu um bom título à apresentação”, e logo se popularizou.

Porém, em 1997, o termo foi utilizado na primeira edição do relatório “*Challenges to the network 1997: Telecoms and the Internet*”, da *International Telecommunication Union* (ITU), agência especializada das Nações Unidas responsável, entre outras atribuições, por criar as normas técnicas e regulamentações para as áreas de tecnologias de informação e

comunicação garantindo que as redes e tecnologias possam se interligar.

A IdC não diz respeito somente à conexão entre o ser humano e suas necessidades com os objetos tecnológicos capazes de satisfazê-las, mas também à análise de informações que essas máquinas podem e poderão realizar. Para Atzori, Iera e Morabito (2010, p. 1):

The Internet of Things (IoT) is a novel paradigm that is rapidly gaining ground in the scenario of modern wireless telecommunications. The basic idea of this concept is the pervasive presence around us of a variety of things or objects – such as Radio Frequency Identification (RFID) tags, sensors, actuators, mobile phones, etc. – which, through unique addressing schemes, are able to interact with each other and cooperate with their neighbors to reach common goals.

É certo que o volume de dispositivos, gerados pela tecnologia, aumenta exponencialmente e as previsões feitas por alguns institutos de pesquisas divergem, porém, todas elas apontam crescimentos elevados entre bilhões e trilhões de “coisas” ou objetos conectados nos próximos anos. A IdC é, assim, uma evolução tecnológica que objetiva conectar dispositivos eletrônicos utilizados cotidianamente como aparelhos eletrodomésticos, eletroportáteis, máquinas industriais, meios de transporte, entre outros, à Internet.

Segundo o *Grupo Gartner* – empresa especializada em tecnologia e que presta consultoria, realiza pesquisas, eventos e executa programas –, a estimativa é a de que 2,9 bilhões de “coisas” estejam conectadas no setor de consumo em 2015 e pode chegar a mais de 13 bilhões em 2020 (Tabela 1).

A Tabela 1 apresenta um aumento exponencial já verificado. E o volume de objetos conectados contribuirá, sem dúvida, para o aumento de dados e informações. Um estudo da Cisco (2015) aponta que haverá um crescimento no tráfego de dados cujo valor estimado para 2019 é de 292 Exabytes,

quase 10 vezes do volume atual, que é de 30 Exabytes.

Tabela 1 - Internet of Things Units Installed Base by Category

Category	2013	2014	2015	2020
Automotive	96.0	189.6	372.3	3,511.1
Consumer	1,842.1	2,244.5	2,874.9	13,172.5
Generic Business	395.2	479.4	623.9	5,158.6
Vertical Business	698.7	836.5	1,009.4	3,164.4
Grand Total	3,032.0	3,750.0	4,880.6	25,006.6

Fonte: Gartner (2014).

A forma de conectividade ressalta uma das características da IdC chamada por Atzori, Iera e Morabito (2010) de “presença pervasiva”. Vale explicar que a palavra “pervasiva” não faz parte do léxico da Língua Portuguesa; ela advém da tentativa em se traduzir o termo em inglês “*pervasive*”. Comumente, utilizamos a palavra onipresença quando nos referimos a algo cuja presença é possível em vários lugares ao mesmo tempo e muito desse uso deve-se à atribuição que a religião dá a um deus ou a uma divindade que teria essa qualidade.

Ubiquidade cujo significado é o mesmo que onipresença migrou para área da computação a partir da publicação do artigo “*The computer for the Twenty-First Century*”, publicado na *Scientific American*, em 1991, de autoria do cientista de informática norte-americano Mark Weiser (1991) a quem é atribuída a criação do conceito de computação ubíqua.

A ubiquidade está relacionada à IdC de tal forma que não seria possível dissociá-las, afinal, a existência da IdC pressupõe conectividade de pessoas e os mais diferentes objetos em vários lugares ao mesmo tempo. Etimologicamente a palavra ubiquidade vem do latim “*ubiqui*”, que é a característica daquele ou daquela que está em toda a parte ao mesmo tempo.

4 O DESAFIO DA ORGANIZAÇÃO

O volume crescente de informações tem sido o foco de estudos acadêmicos desde os anos 1940 cujos “diferentes problemas recaíram no processo de recuperação de informações.” (MACULAN, 2011, p. 62).

Porém, a IdC amplia de forma acentuada e veloz esse volume e traz à tona a preocupação com a representação semântica e a organização dos dados gerados pelas conexões dos dispositivos, além da segurança dos dados.

Volumes of diverse data are flooding in at an unimaginable rate, especially no structured contents with the development of various IoT technologies. The primary challenge for this is to find reasonable organizing models for these big data. (SUN et al., 2014, p.2).

Ou seja, o principal desafio de fato é encontrar modelos de organização razoáveis para esses grandes dados.

A preocupação em relação à organização de um volume de informação sempre em movimento ascendente não é recente, embora Logan (2012, p. 26) afirme que a

Noção de informação como algo que pode ser armazenado em, transferido ou comunicado a um objeto inanimado e a noção de informação como quantidade definida matematicamente não surgem antes do século XX.

No artigo “Organização dos documentos ou organização da informação: uma questão de escolha”, Robredo (2004) afirma que o ciclo documentário realizado, segundo ele, antigamente, por especialistas, tende cada vez mais a ser automatizado. Ele ainda afirma que a interface permite à pessoa formular perguntas a mecanismos de inferência

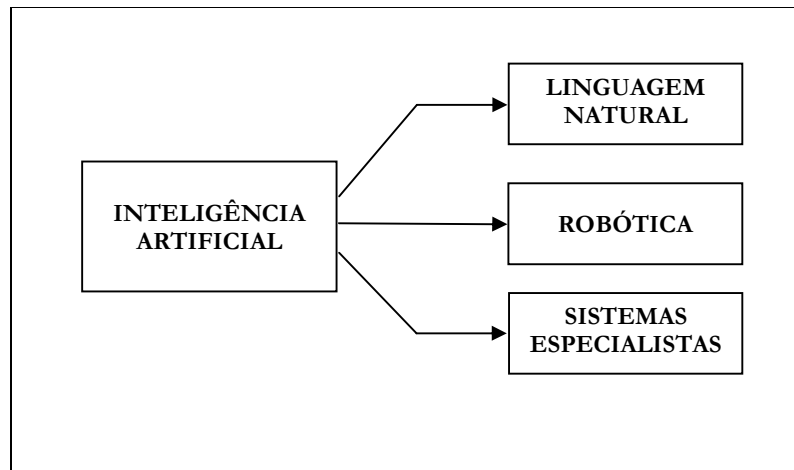
(motores de busca) que fornecem a resposta solicitada.

Em seu artigo, demonstra graficamente, como vemos na figura 1, a

[...] relação conceitual intrínseca de nossas preocupações de conserva-

ção, organização e gestão de documentos com os sistemas especialistas da inteligência artificial, e não é por acaso que já se começa a falar de ‘arquivamento inteligente da informação’.

Figura 1 – Arquivamento inteligente da informação



Fonte: Robredo (2004).

Robredo (2004) aponta que as três áreas estão implicadas na solução do problema da organização da informação e os destaca da seguinte maneira:

- o foco linguagem natural aponta para o desenvolvimento de aplicações capazes de ler, interpretar e codificar automaticamente os textos dos documentos;
- a robótica, presente numa infinidade de sistemas e aplicações sensíveis a estímulos visuais e eletromagnéticos (leitores óticos e magnéticos) está cada vez mais presente nos grandes sistemas de armazenamento/arquivamento de importantes volumes de documentos;
- [...] desenvolvimento de aplicações informáticas suscetíveis de utilizar o conhecimento simbólico da forma o mais próxima possível do comportamento humano. (ROBREDO, 2004).

A conexão entre os milhares de aplicativos existentes se dá por meio de algoritmos, que permitem que se estabeleça a co-

municação entre máquinas (*machine-to-machine*) e entre humanos. Um algoritmo pode ser definido como “uma sequência finita de instruções ou operações cuja execução, em tempo finito, resolve um problema computacional, qualquer que seja sua instância.” (SALVETTI, 1999). O algoritmo computacional é compreendido como um programa que realiza procedimentos para solucionar um problema. Algoritmos computacionais usam estruturas que ajudam o processador a chegar a um determinado resultado.

Para que os dispositivos e os objetos que fazem parte do nosso dia a dia possam ter acesso a bases de dados e estar conectados em rede e também à Internet, algumas questões precisavam ser resolvidas: uma forma de identificação eficiente em custos. Apenas assim informações sobre objetos com identidade poderiam ser coletadas e processadas automaticamente. Estamos nos referindo às tags.

As tags, enquanto elementos que inseridos em objetos cotidianos corporificam uma comunicação u-

bíqua e estabelecem um senso de ordem, atribuem informações, sendo facilmente reconhecíveis e reprodutíveis, e dentro do caráter temporário que apresentam, podem ser reprogramadas a qualquer momento. A solução RFID foi uma das que emergiram e possibilitou identificar os objetos por rádio frequência. Outra foi o QRCode (*Quick Response Code*), que permite que aparelhos celulares, através de seus algoritmos visuais, escaneiem informação digital impressa em mídias analógicas. (SANTAELLA et al, 2013).

Santaella et al (2013) ressaltam ainda que

A computação, cada vez mais invisível, salta para objetos do cotidiano, é neles implantada, dando-lhes identidade através, por exemplo, de etiquetas RFID (*Radio Frequency Identification Tags*), mas mantendo a capacidade de comunicação entre si.

Ademais, dessa importante etapa do que se constitui a IdC, temos outra, foco deste artigo: a organização da informação. Lídia Alvarenga (2006) aponta alguns desafios em relação ao tema:

A linguagem natural tem sido o meio de organização do universo da Web, pois considera-se impossível, em um âmbito mais amplo, um processamento dos objetos digitais com linguagens documentárias estruturadas, usadas para um tratamento da informação a posteriori, nas quais os termos que simbolizam conceitos (e não as palavras) seriam usados como chaves de recuperação. (ALVARENGA, 2006 p.7).

Nesse ponto, um resgate histórico se mostra adequado para contextualizarmos a questão. A organização de um conteúdo deve, obrigatoriamente, ser analisada em um primeiro estágio para que ele possa ser representado conceitualmente, ou seja, para

que possa ser classificado. Moreiro González (2011, p.13) explica que

Desde a Antiguidade clássica, têm sido procurados os conceitos mais gerais dos discursos aludidos na expressão do pensamento, conhecidos então como predicados e agora como macroestruturas. Neste ponto, é preciso voltar a Aristóteles. Dentro do estudo dos conceitos, nos “analíticos posteriores”, ele considerou os predicamentos como categoriais ou tópicos principais da expressão, de onde partem os distintos modos retóricos: Substância, Quantidade, Qualidade, Relação, Lugar, Tempo, Situação, Posse, Ação e Paixão. [...] Na medida em que as categorias remetem às formas extramentais de ser, adquirem um forte conteúdo ontológico, supondo que as coisas são captadas pela mente tal como o são na verdade.

Há diferentes maneiras de representação do conhecimento a partir da compreensão da realidade. O conhecimento tem sido discutido e definido por muitos filósofos desde Platão até os filósofos da atualidade sem que haja de fato uma conclusão. Imbricados estão os termos dado, informação e conhecimento. A distinção feita por Davenport (1998), demonstrada no quadro 1, ajuda a elucidá-los mesmo que o autor resista “em fazer essa distinção, porque ela é nitidamente imprecisa.”

O autor afirma que o conhecimento se dá quando alguém deu a ele um contexto, um significado, uma interpretação; “alguém refletiu sobre o conhecimento, acrescentou a ele sua própria sabedoria, considerou suas implicações mais amplas.”. (DAVENPORT, 1998 p. 18). Ainda sobre conhecimento, Takeuchi e Nonaka (2008) afirmam que este é formado por dois componentes dicotômicos: o conhecimento explícito e o tácito.

O conhecimento explícito pode ser expresso em palavras, números ou sons, e compartilhado na forma de dados, fórmulas científicas, recursos visuais, fitas de áudio, especificações de produtos ou manuais. O

conhecimento explícito pode ser rapidamente transmitido aos indivíduos, formal e sis-

tematicamente. (TAKEUCHI; NONAKA, 2008, p. 19).

Quadro 1 – Dados, informação e conhecimento

Dados	Informação	Conhecimento
Simple observações sobre o estado do Mundo	Dados citados de relevância e propósito	Informação valiosa da mente humana
Facilmente estruturado	Requer unidade de análise	Inclui reflexão, síntese, contexto
Facilmente obtido por máquinas	Exige consenso em relação ao significado	De difícil estruturação
Frequentemente quantificado	Exige necessariamente a mediação humana	De difícil captura em máquinas
Facilmente transferível		Frequentemente tácito
		De difícil transferência

Fonte: Adaptado de Davenport (1998).

Porém, o nosso objetivo aqui é observar a organização da informação e novamente é preciso trazer à tona também o conceito de informação e como ele se relacionava historicamente com o conhecimento embora, na prática, alguns autores não façam distinção entre os termos (BARCLAY; MURRAY, 1997).

Claude Shannon (apud KURZWEIL, 2007, p. 1948), considerado o pai da Teoria da Informação,

[...] definiu a informação como uma mensagem enviada por um emissor para um receptor. Shannon trabalhou na Bells Labs e queria resolver o problema de como melhor codificar a informação que um emissor quis transmitir para um receptor.

Para Shannon, a informação era definida matematicamente e podia ser armazenada, transferida ou comunicada a um objeto inanimado, mas “frequentemente advertia de que a [sua] teoria só deveria se aplicar a certas situações técnicas e não à comunicação em geral” (LOGAN, 2012).

Três anos após Shannon formular essa definição, Donald Mackay (apud KURZWEIL, 2007, p. 1951) defendeu “uma outra abordagem para a compreensão da natureza da informação” e foi o primeiro a usar o termo “teoria da informação” e também o primeiro a salientar que a importância da informação é o seu significado. Para ele, a informação deve ser definida como “a mudança mental em um receptor, portanto, com significado. (LOGAN, 2012).

James Gleick, em sua obra “A Informação: uma história, uma teoria, uma enxurrada” nos lembra que

A teoria da informação começou como uma ponte da matemática para a engenharia elétrica e daí para a computação. Não à toa, a ciência da computação também é conhecida pelo nome de informática. Hoje até a biologia se tornou uma ciência da informação, sujeita a mensagens, instruções e códigos. Os genes encapsulam informações e permitem procedimentos para que estas sejam lidas a partir deles e inscritas neles. A vida se expande por meio

do estabelecimento de redes. (GLEICK, 2013, p. 16).

O fato de a vida se expandir por meio de redes nos sugere que a multiplicidade de aplicativos que nos conecta nada mais é do que uma consequência natural do que somos, dado que o próprio corpo humano é um processador de informações. O nosso DNA é uma molécula de informação, a sua “função é armazenar a informação genética.” (GLEICK, 2013).

5 DADOS ESTRUTURADOS E SEMI-ESTRUTURADOS

O volume de dados gerado pela IdC é formado, em sua maioria, por dados sem estrutura ou semiestruturados, os quais Abiteboul (2000) define como “representações ‘sem esquema’ ou ‘autodescritivas’”. Ao usar dados semiestruturados, os dados são descritos diretamente, usando-se uma sintaxe simples.”.

Uma preocupação crescente entre pesquisadores na área de web semântica está relacionada à forma com que os dados da Internet estão sendo disponibilizados. A busca por mecanismos que auxiliem o entendimento tanto por aplicações quanto para o homem está se tornando frequente na área científica. A linguagem XML vem se destacando como um mecanismo de fundamental importância e tornando-se a cada dia um padrão para manipulação de dados semiestruturados (DAUM; MERTEN, 2002).

É importante uma breve reflexão sobre dados estruturados, semiestruturados e não estruturados, uma vez que a clareza desses conceitos é fundamental para o contexto analisado em que observamos o crescimento constante da informação e, conseqüentemente, impondo desafios à gestão da informação digital.

Os dados armazenados em Sistemas Gerenciadores de Banco de Dados Relacionais (SGBDRs) são considerados dados estruturados por possuírem estrutura fixa bási-

ca, ou seja, seguem um mesmo padrão, são legíveis por computadores permitindo sua recuperação. Já os dados semiestruturados não seguem padrões definidos e possuem como característica principal uma estrutura irregular, dinâmica e bastante heterogênea. Para a organização desse conteúdo composto de dados semiestruturados, a linguagem XML vem sendo adotada, pois, ela:

[...] foi projetada para ser distribuída pela Web. Os dados são representados com o uso da estrutura de um documento XML. Os elementos e atributos fornecem metadados úteis - informações sobre dados - e estruturam os dados de uma maneira na qual podem ser transferidos pela Web, manipulados por aplicativos diferentes e alterados atender a necessidades diversas. (GRAVES, 2003, p. 5).

A EdC, entendida como uma rede que interliga objetos heterogêneos e endereçáveis, torna-se possível por meio de protocolos de comunicação padronizados. Os objetos (totalmente diferentes em termos de funcionalidades, tecnologia e campo de aplicação) conectados, a qualquer tempo e em qualquer lugar, permitem a troca de informações (e também o gerenciamento dessas), por meio de sensores e de protocolos padronizados, apontados por especialistas como um dos grandes desafios para que a IdC efetivamente se estabeleça de forma abrangente ao maior número de seres humanos.

Porém, avanços têm sido conquistados permitindo que os nós da rede de sensores sem fio se conectem à Internet. Chong e Kumar (2003, p. 1253) apontam alguns de desafios e exemplificam:

A sensor field is like a database with many unique features. Data is dynamically acquired from the environment, as opposed to being entered by an operator. The data is distributed across nodes, and geographically dispersed nodes are connected by unreliable links. These features render the database view more challenging, particularly for military applications given the

low-latency, real-time, and high-reliability requirements of the battlefield.

It is important that users have a simple interface to interactively task and query the sensor network. An example of a human-network interface is a handheld unit that accepts speech input. The users should be able to command access to information, e.g., operational priority and type of target, while hiding details about individual sensors. One challenge is to develop a language for querying and tasking, as well as a database that can be readily queried. Other challenges include finding efficient distributed mechanisms for query and task compilation and placement, data organization, and caching.

Ao observarmos os aspectos que envolvem a IdC, podemos destacar como características intrínsecas: a ubiquidade, pelo fato de possibilitar a interligação entre objetos e pessoas sem restrição de lugar e tempo, os protocolos e, conseqüentemente, a semântica para fins de estruturação das informações.

6 A WEB SEMÂNTICA, REDES SEMÂNTICAS E REDES NEURAIAS

A palavra “*semantiká*”, transliterada do grego, é derivada de “*sema*”, que significa sinal. A semântica é um ramo da Linguística e refere-se ao estudo do significado atribuído por seres humanos às coisas que o cercam e aos eventos que lhe acometem, no mundo real ou abstrato, para que possam comunicar-se por meio da linguagem. Além disso, a semântica se ocupa com a relação entre significantes. A transposição da semântica para a área da computação ocorreu:

[...] nos anos 50 do século XX, com o grande investimento em tradução automática; e, uns 15 anos depois, com o processamento de língua natural. E talvez o primeiro grande livro no qual uma representação se-

mântica foi construída computacionalmente tenha sido o *Prolog and Natural-Language Analysis*, de Pereira e Shieber. Já mais modernamente, a publicação do livro *Representation and Inference for Natural Language*, de Blackburn e Bos, demonstra uma revigoração deste paradigma (FERRAREZI JR.; BASSO, 2013, p. 58).

É possível “[...] identificar o interesse pela Semântica Computacional já nas primeiras tentativas de simular o comportamento humano (como nas falsas máquinas de jogar xadrez, do final do século XVIII) [...]” (FERRAREZI JR.; BASSO, 2013, p. 58).

No entanto, foi com o artigo “*The Semantic Web: a new form of Web content that is meaningful to computers Will unleash a revolution of new possibilities*”, publicado na revista *Scientific American*, em 2001, de Tim Berners Lee, James Hendler e Ora Lassila, que o conceito se firmou. No artigo, os autores previam que a importância da Web Semântica traria estrutura para o conteúdo significativo às páginas da Web e ressaltavam que a Web Semântica não era algo separado da Web.

The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. The first steps in weaving the Semantic Web into the structure of the existing Web are already under way. In the near future, these developments will usher in significant new functionality as machines become much better able to process and “understand” the data that they merely display at present (BERNERS-LEE; HENDLER; LASSILA, 2001).

Se tivéssemos que estabelecer uma linha evolutiva da representação do conhecimento na Web, poderíamos apontar que a Web Semântica está prestes a ceder lugar à Web Semântica Cognitiva. Estudos recentes, nos campos da neurolinguística, têm contribuído para transpor desafios na relação entre

cérebro humano e linguagem e isso tem se tornado importante insumo para a Web Semântica Cognitiva.

Ao falarmos em cognição é importante trazer à reflexão dois conceitos antagônicos presentes nos estudos da Inteligência Artificial (IA) que são IA simbólica e IA conexionista. Os modelos do conexionismo

[...] dão ênfase aos processos cerebrais e procuram, através de simulações computacionais, modelar os procedimentos psicológicos e cognitivos conforme padrões de atividades distribuídos por uma rede de unidades simples - as tão afamadas redes neurais. (SOUZA, 2008)

As redes neurais constituem-se no ponto nevrálgico do modelo conexionista. K. Laudon e J. Laudon (2010, p. 339) as definem, do ponto de vista funcional, da seguinte maneira:

As redes neurais são usadas para desenvolver problemas complexos e não totalmente compreendidos, para os quais grandes quantidades de dados já foram coletadas. Elas encontram padrões e relações em gigantescas quantidades de informações que um ser humano acharia muito difícil e complicado analisar. As redes neurais alcançaram tal entendimento usando *hardware* e *software* que imitam os padrões de processamento do cérebro biológico, ou humano. Elas “aprendem” padrões a partir de grandes quantidades de dados; para tanto, “peneiram” os dados, procuram relações, constroem modelos e os revisam várias vezes, corrigindo seus próprios.

A linha conexionista foi formalizada pelo neurofisiologista Warren McCulloch e pelo matemático Walter Pitts, em 1943, em que propunham, no artigo intitulado “*A logical calculus of the idea simmanent in nervous activity*”, *Bulletin of Mathematical Biophysics*”, a modelagem da inteligência humana por meio da simulação dos componentes cerebrais, dos neurônios e suas interligações.

Porém, podemos citar que essa analogia funcional entre computadores e a mente humana também encontra a sua base de sustentação no teste de Turing “projetado para fornecer uma definição operacional satisfatória de inteligência.” (ROSA, 2011).

O teste de Turing, também conhecido como “jogo da imitação”, consiste na participação de duas pessoas e uma máquina. Uma das pessoas, em uma sala fechada, iniciava uma conversa, utilizando linguagem natural, por meio de um teclado, simultaneamente com um ser humano (que ficava em outra sala) e com uma máquina. A pessoa que iniciava a conversa assumia o papel de “juiz” ao julgar quais respostas eram dadas pela máquina e pelo ser humano. Se o “juiz” não fosse capaz de fazer essa distinção, a máquina teria vencido o jogo, “passando-se” por humano.

Assim sendo, a abordagem simbólica da cognição envolve a manipulação de representações. Porém, essa abordagem traz em seu bojo, uma questão importante que o professor Angelo Conrado Loula (2004, p.1) nos apresenta dessa maneira:

Quando falamos em representação de conhecimento, a questão central é como se representa esta informação mentalmente, e como esta informação é usada na interação com o mundo.

Para isso, devemos entender o papel das ontologias na organização do conhecimento.

7 ONTOLOGIAS

Quando analisamos o ambiente virtual, em que existe a possibilidade de estarmos lidando com inteligências artificiais, que, portanto, possuem o entendimento semântico dos termos e conseguem criar outros tipos de relações, incluindo outras formas de representações que podem se apoiar em redes semânticas, frames e redes neurais, é possível identificar a contribuição da Ontologia.

Marques (2006, p. 17) define Ontologia, a partir de sua origem etimológica,

como a “ciência ou estudo do ‘ente’ ou ‘ser’”. Ontologia é o estudo de existência de todos os tipos de entidades, abstratas ou concretas, que constituem o mundo.”. A Ontologia nasce no seio da Filosofia para migrar para outras searas à medida em que se desenvolve o saber humano. Marques (2006, p. 49) explica que:

O foco dos estudos filosóficos deslocou-se para uma nova questão, a de desvendar a capacidade do espírito humano para captar a realidade, conduzindo aos estudos que constituíram a denominada Teoria do Conhecimento, com a qual a Ontologia foi correlacionada. Não se trata de explicar o real, mas de investigar o processo cognitivo pelo qual o homem o capta, estuda e compreende, e de que modo armazena e classifica e representa o conhecimento [...].

Barry Smith atribui à Ontologia a tarefa de proporcionar uma classificação definitiva e exaustiva das entidades, em todas as esferas do ser, e assim estabelecer e possibilitar também os tipos de relações pelas quais as entidades são conectadas.

Ontology seeks to provide a definitive and exhaustive classification of entities in all spheres of being. The classification should be definitive in the sense that it can serve as an answer to such questions as: What classes of entities are needed for a complete description and explanation of all the goings-on in the universe? Or: What classes of entities are needed to give an account of what makes true all truths? It should be exhaustive in the sense that all types of entities should be included in the classification, including also the types of relations by which entities are tied together to form larger wholes (SMITH, 2003, p. 2).

Marques (2005) explica que o uso da palavra em sua forma plural e com “o” minúsculo reporta-se a formas de representa-

ções do conhecimento (MARQUES, 2005, p. 17) termo que adotaremos de agora em diante. Para Luz (2016, p. 27),

A informação digital, como forma estruturante do conhecimento contemporâneo e para ser tratada por agentes computacionais, precisa ser organizada em suas principais características, tais como os temas (taxonomias) ou a semântica (ontologias).

As ontologias desempenham um papel fundamental para a construção de vocabulários, pois é a partir delas que se definem as relações entre os diferentes conceitos, gerando linguagens que podem ser lidas por computadores, além de facilitarem o compartilhamento e reutilização de informações. As ontologias são componentes básicos para a implantação da web semântica e fazem parte do projeto gerenciado pelo *World Wide Web Consortium* (W3C), um consórcio internacional que tem por objetivo a organização de padronização da Web criando mecanismos que possam melhorar a relação homem-máquina. Conforme Luz (2016, p. 74),

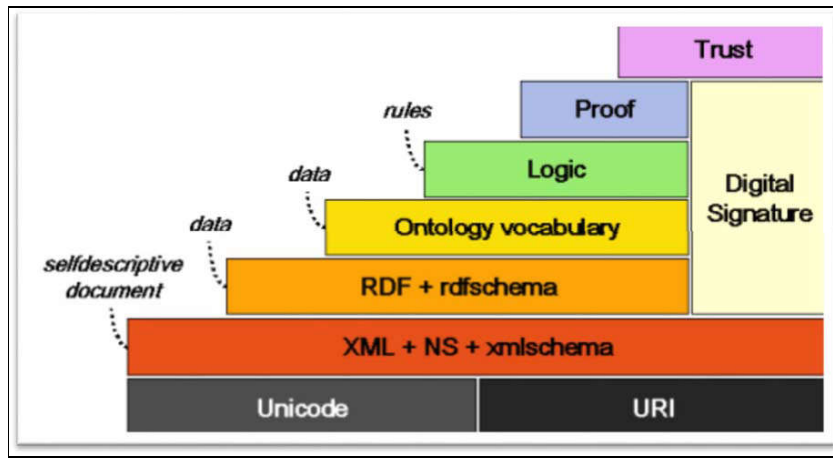
As linguagens documentárias (como as taxonomias) são elementos importantes que operam combinados, sinergicamente, com outros aspectos da Web 3.0, e fazendo parte de uma de suas camadas. Isso já estava previsto na formulação do conceito da Web Semântica que foi concebido como um conjunto de tecnologias.

É nesse ponto que se faz necessário apresentarmos as camadas semânticas (figura 2).

A camada “*Ontology Vocabulary*” dá suporte para a evolução de vocabulários permitindo a correlação de conceitos e tenta dirimir os conflitos de terminologia. Já a linguagem *RDF Schema* fornece uma semântica simplificada e poucos desses elementos podem ser utilizados para inferência. A *Web Ontology Language* (OWL) estende o vocabulário da *RDF Schema* para a inclusão de ele-

mentos com maior poder em relação à expressividade e à inferência.

Figura 2 – The Semantic Web layers



Fonte: Koivunen e Miller [201-].

A OWL é a linguagem padrão para representação de ontologias utilizada pela Web Semântica e projetada para representar o conhecimento que se baseia em lógica computacional. A OWL é recomendada pelo grupo W3C cujos recentes estudos apontam estes desafios:

The primary challenge of machine perception is to define efficient computational methods to derive high-level knowledge from low-level sensor observation data. Emerging solutions are using ontologies, such as the W3C SSN ontology, to provide expressive representation of concepts in the domain of sensing and perception, which enable advanced integration and interpretation of heterogeneous sensor data. (KOIVUNEN; MILLER, [201-]).

A OWL se presta a processar o conteúdo da informação ao invés de apenas representá-la para humanos. A OWL acrescenta mais vocabulário para descrever propriedades e classes: entre outros, relações entre classes, igualdade, tipificação mais rica de propriedades, características de propriedades e classes enumeradas.

Defining an ontology and using semantic descriptions for data will make it interoperable for users and

stakeholders that share and use the same ontology. In the IoT domain different stakeholders need to have a common agreement on ontological definitions. Most of the current ontologies and semantic description frameworks in the IoT domain are defined in the context of different projects and applications or they are currently at an early stage. To achieve global scale semantic interoperability, common semantic annotation frameworks, ontology definitions, and adaptation are key issues. Recent efforts, such as the W3C SSN ontology, are effective steps towards achieving this goal. (BARNAGHI, 2012).

A Web Semântica, em particular as ontologias, podem contribuir com as redes neurais para que a localização e a seleção das informações relevantes possam atender às necessidades dos usuários de forma personalizada, proposta da IdC, ao conectar pessoas e objetos. Para Luz (2016, p. 114) a

Ontologia atende aos objetivos das duas dimensões da essência informacional, condizentes com a dimensão descritiva, voltada aos elementos relativos à forma dos documentos e a dimensão temática, voltada aos conteúdos informacionais. Uma ontologia, portanto, é

uma estrutura que permite organizar a informação e auxiliar na sua recuperação, pois define as regras que regulam a combinação entre os termos e suas relações.

Assim, vislumbra-se um papel essencial da organização da informação e de suas ferramentas como as ontologias no universo de dados e informações usados e gerados pela IdC.

8 CONSIDERAÇÕES FINAIS

É possível que a IdC continue a contribuir para acelerar ainda mais o volume de dados no ambiente virtual permitindo às pessoas realizar atividades “por intermédio da onipresente Web de comunicações e não requerem qualquer equipamento, dispositivos ou objetos que não sejam vestíveis ou implantados.” (KURZWEIL, 2007, p. 108).

A organização da informação, em redes semânticas representadas por ontologias podem evoluir a patamares nos quais seja possível a recuperação de uma resposta analítica sintática e semanticamente, além de útil à utilização para que foi solicitada, que

ofereça dados satisfatórios. É nesse sentido que as ontologias podem representar uma importante contribuição, podem representar um importante papel.

Em realidade, a contribuição é mútua já que esse desafio também pode ser benéfico para as ontologias no sentido de que haja maior interesse na busca por soluções de modelagens que permitam um tratamento adequado para os dados gerados pela constante conexão de objetos, dispositivos e pessoas.

As ontologias podem também ser a base para a construção da estrutura da IdC a fim de que a interface possibilite cada vez ao usuário o uso de linguagem natural, que a conversa com máquina seja mais amigável, ou seja, que se torne possível uma comunicação direta, flexível entre o homem e a máquina.

A urgência de respostas a que a IdC se propõe a oferecer impõe novos modelos de armazenamento, de organização e de análise dos dados para que sejam úteis quando solicitados ou quando enviados de forma autônoma pelas máquinas.

CONNECTIVITY AND INFORMATION ORGANIZATION: An Approach Between the Internet of Things and the Semantic Web

Abstract

It addresses how the storage and data organization driven by the Internet of Things (IoT) connectivity takes place. In this analysis situation, the development of efficient protocols that allow the connectivity of all the objects present in the IoT stands out. It proposes the Information Organization (IO) as the basis for the structure of this information in the semantic web network and the ontologies as layers of meaning for information in neural networks.

Keywords

Internet of Things. Artificial Intelligence. Semantic Web. Ontology. Interoperability.

Artigo recebido em 02/04/2016 e aceito para publicação em 14/05/2017

REFERÊNCIAS

ABITEBOUL, S.; BUNEMAN, P.; SUCIU, D. **Data on the web**. São Francisco, EUA: M. Kaufmann, 2000.

ALVARENGA, L. A teoria do conceito revisitada em conexão com ontologias e metadados no contexto das bibliotecas tradicionais e digitais. **Data Grama Zero**, Rio

de Janeiro, v. 2, n. 6, dez. 2015. Disponível em: <<http://basessibi.c3sl.ufpr.br/brapci/index.php/article/download/7457>>. Acesso em: 5 maio 2015.

ASHTON, K. Internet das coisas: nova revolução da conectividade. **Em Pauta**, São Paulo, n. 18, p. 6-8, dez. 2014.

- ATZORI, L.; IERA, A.; MORABITO, G. The internet of things: a survey. **Computer Networks**, Georgia, USA, v. 54, n. 15, p. 2787-2805, oct. 2010. Disponível em: <https://www.elsevier.com/__data/assets/pdf_file/0006/97026/The-Internet-of-Things.pdf>. Acesso em: 22 jan. 2015.
- BARCLAY, R. O.; MURRAY, P. C. What is knowledge management? **knowledge praxis**. EUA, 1997. Disponível em: <http://www.providersedge.com/docs/km_articles/what_is_knowledge_management.pdf>. Acesso em: 13 fev. 2015.
- BARNAGHI, P. et al. Semantics for the internet of things: early progress and back to the future. **International Journal on Semantic Web and Information Systems**, EUA, v. 8, n. 1, p. 1-21, Jan./Mar. 2012. Disponível em: <<http://dl.acm.org/citation.cfm?id=2441528>>. Acesso em: 13 fev. 2015.
- BEHROUZ, A. F. **Cryptography and network security**. Tata McGraw Hill Education Private Limited, 2010.
- BERNERS-LEE, T. HENDLER, J.; LASSILA, O. The semantic web. **Scientific American**, EUA, p. 29-37, May 2001. Disponível em: <http://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf>. Acesso em: 20 jan. 2015.
- CASTELLS, M. **A galaxia da internet**. Rio de Janeiro: Zahar, 2003.
- CHONG, C. Y.; KUMAR, S. P. "Sensor networks: evolution, opportunities, and challenges". **Proceedings of the IEEE**, Novo México, EUA, v. 91, p. 1247-1256, 2003.
- CISCO. **Cisco visual networking index: global mobile data traffic forecast update 2014–2019 white paper**. Califórnia, EUA, 2015. Disponível em: <http://www.cisco.com/c/en/us/solution/s/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html>. Acesso em: 12 jan. 2015.
- DAUM, B.; MERTEN, U. **Arquitetura de Sistemas com XML**. Rio de Janeiro: Campus, 2002.
- DAVENPORT, T. **Ecologia da informação: porque só a tecnologia não basta para o sucesso na era da informação**. São Paulo: Futura, 1998.
- FERRAREZI JUNIOR, C.; BASSO, R. **Semântica, semânticas: uma introdução**. São Paulo: Contexto, 2013.
- GLEICK, J. **A informação: uma teoria, uma enxurrada**. São Paulo: Companhia das Letras, 2013.
- GRAVES, M. **Projeto de Banco de Dados com XML**. São Paulo: Pearson Education do Brasil, 2003.
- KIDD, S. M. **A vida secreta das abelhas**. São Paulo: Paralela, 2014.
- KOIVUNEN, M.; MILLER, E. W3C Semantic Web Activity. [201-]. Disponível em: <<http://www.w3.org/2001/12/semweb-fin/w3csw>>. Acesso em: 25 abr. 2015.
- KURZWEIL, R. **A era das máquinas espirituais**. São Paulo: Aleph, 2007.
- LAUDON, K.; LAUDON, J. **Sistemas de informações gerenciais**. São Paulo: Pearson Education do Brasil, 2010.
- LOGAN, R. K. **O que é informação?: a propagação da informação na biosfera, simbiosfera, na tecnosfera e na econosfera**. Rio de Janeiro: Contraponto, 2012.
- LOULA, A. C. **Comunicação simbólica entre criaturas artificiais: um experimento em vida artificial**. 2004. 155 f. Dissertação (Mestrado em Engenharia Elétrica) – Faculdade de Engenharia Elétrica e de Com-

putação, Campinas, SP, 2004. Disponível em:
<<http://www.dca.fee.unicamp.br/~angeloc1/msc-angelo.pdf>>. Acesso em: 15 abr. 2015.

LUZ, C. S. **Ontologia digital arquivística: interoperabilidade e preservação da informação arquivística em sistemas informatizados de arquivos e na web.** 2016. Dissertação (Mestrado em Cultura e Informação) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2016. Disponível em:
<<http://www.teses.usp.br/teses/disponiveis/27/27151/tde-03022017-154503/>>. Acesso em: 29 ago. 2017.

MACULAN, B. C. M. S. **Taxonomia facetada navegacional: construção a partir de uma matriz categorial para trabalhos acadêmicos.** 2011. Tese de Doutorado. Dissertação Mestrado – UFMG - Belo Horizonte.

MARQUES, M. L. O[o]ntologia[s]? In: SIMÃO, E.; MIRANDA, A. **O texto virtual e os sistemas de informação: nova leitura das propostas de Ítalo Calvino.** Thesaurus, 2005.

MOREIRO GONZÁLEZ, J. A. **Linguagens documentárias e vocabulários semânticos para a web: elementos conceituais.** Salvador: EDUFBA, 2011.

ROBREDO, J. Organização dos documentos ou organização da informação: uma questão de escolha. **Data Grama Zero**, v. 5, n. 1, 2004. Disponível em:
<http://basessibi.c3sl.ufpr.br/brapci/_repositorio/2010/01/pdf_6e6fdeca3e_0007605.pdf>. Acesso em: 29 ago. 2017.

ROSA, J. L. G. **Fundamentos da inteligência artificial.** Rio de Janeiro: LTC, 2011.

SALVETTI, D. D.; BARBOSA, L. M. **Algoritmos.** São Paulo: Pearson Education/Markron Books, 1999.

SANTAELLA, L. et al. Desvelando a internet das coisas. **Revista Geminis**, São Paulo, v. 1, n. 2, ano 4, p. 19-32, 2013. Disponível em:
<<http://www.revistageminis.ufscar.br/index.php/geminis/article/viewFile/141/pdf>>. Acesso em: 15 jan. 2015.

SOUZA JR., P. S. Entre Fódor e Chalmers: simbolistas e conexionistas por uma inteligência artificial. **Língua, Literatura e Ensino**, Campinas, SP, v. 3, maio 2008.

SMITH, B. **Ontology: blackwell guide to the philosophy of computing and information.** Oxford: Blackwell, 2003. p. 155-166.

SUN, Y. et al. Organizing and querying the big sensing data with event-linked network in the internet of things. **International Journal of Distributed Sensor Networks**, Nova York, EUA, v. 2014, 2014. Disponível em:
<<http://www.hindawi.com/journals/ijdsn/2014/218521/>>. Acesso em: 2 fev. 2015.

TAKEUCHI, H.; NONAKA, I. **Gestão do conhecimento.** São Paulo: Bookman, 2008.

WEISER, M. The computer for the twenty-first century. **Scientific American**, Canadá, p. 94-10, Sept. 1991. Disponível em:
<<http://www.ubiq.com/hypertext/weiser/SciAmDraft3.html>>. Acesso em: 12 abr. 2015.