

A evolução da Web rumo à Web Semântica

The Web evolution towards the Semantic Web

Marcello Peixoto Bax

Escola de Ciência da Informação da Universidade Federal de Minas Gerais, Brasil
bax.ufmg@gmail.com

Resumo

Definir o que é uma URI (*Uniform Resource Identifier*) é um dos problemas fundamentais por trás da interoperabilidade semântica na Web. Trata-se de uma questão crucial para a sua próxima geração: o que significa uma URI? A Web é uma rede universal que nomeia recursos informacionais usando URIs. O problema surge quando na Web Semântica, URIs são usadas também para referir a recursos no mundo - coisas no universo do discurso e não apenas arquivos eletrônicos. Como referir-se a coisas fora da web? Que natureza de recurso uma URI denotaria nesse caso? Seria a URI a identidade da coisa no universo do discurso, ou um arquivo descritivo dessa coisa? Como determinar a identidade daquilo que é designado por uma URI? Esse problema de identidade de um recurso tem sido debatido na comunidade da Web Semântica por mais de uma década. Pensar o problema como relativo ao "significado" do recurso referido nos remete a questões filosóficas clássicas (de significado e referência), longamente debatidas em filosofia da linguagem. Essas questões teóricas devem ser revisitadas se considerarmos a Web como base da representação do conhecimento, como é o caso na Web Semântica. Os debates sobre o significado e status referencial de uma URI que ocorrem na Web Semântica são análogos aqueles sobre o significado e referência de nomes que tem lugar na filosofia da linguagem. Este artigo discute o problema da identidade na Web Semântica sob o olhar da filosofia da

Abstract

Defining what is a URI (Uniform Resource Identifier) is one of the fundamental problems behind the semantic interoperability on the Web. This is a crucial issue for their next generation, which means a URI? The Web is a universal network naming information resources using URIs. The problem arises when, in the Semantic Web, URIs are also used to refer to the resources in the world - things in the universe of discourse and not just electronic files. How to refer to things outside the web? What nature resource URI denote in this case? URI would be the identity of the thing in the universe of discourse, or a file description of this thing? How to determine the identity of what is called a URI? This problem of identity of a resource has been debated in the Semantic Web community for over a decade. Thinking the problem as concerning the "meaning" of the appeal referred us back to classical philosophical questions (of meaning and reference), discussed at length in the philosophy of language. These theoretical issues must be revisited when considering the Web as the basic knowledge representation, such as the Semantic Web. The debates about the meaning and referential status of a URI that occur in the Semantic Web are analogous ones on the meaning and reference of names takes place in philosophy of language. This article discusses the problem of identity in the Semantic Web from the perspective of philosophy of language.

linguagem¹.

Palavras-chave: URI, Identidade, Web semântica.

Keywords: URI, Identity, Semantic Web.

“Ontologies get around the difficulty of categorizing the Web by building rigid logical relationships between opaque chains of symbols (URIs) and strings of characters borrowed from natural languages. But there are several disconnected ontologies.” (Pierre Lévy, 2009)

1. Introdução

A comunicação é um processo complexo. Um elemento óbvio e básico desse processo, que possibilita a comunicação entre pessoas, é a capacidade de dar nome às coisas. Só falamos e nos entendemos porque somos capazes de “identificar” as coisas sobre as quais falamos usando nomes (substantivos ou nomes próprios). Normalmente chamamos de sentido, significado ou semântica, o mapeamento ou a relação existente entre as ideias (que temos na mente) e as coisas tal como se encontram no mundo. Apenas quando ideias consolidam-se por seu uso prático e convencional, que se dá ordinária e cotidianamente num dado contexto linguístico social, é que então aparece a noção de “conceito”. Um conceito decorre da um consenso que surge a partir de ideias.

Desde a virada do século novos entes passam, cada vez mais, a participar de nossas conversas, de nossos contextos linguísticos. São entes artificiais, artefatos computacionais, agentes (programas), que criamos para nos ajudar a realizar tarefas cotidianas, tais como: pesquisar um tema, organizar uma viagem, trabalhar em conjunto etc. Como se sabe, a Web Semântica vai constituindo-se aos poucos nesse ambiente, onde pessoas e agentes de software interagem de forma colaborativa.

O uso de URI's (Uniform Resource Identifier) para identificar ou nomear recursos informacionais na Web é notório. Porém na Web Semântica as URI's são usadas também

¹ Alguns dos elementos dessa reflexão se inspiraram e se fundam principalmente no trabalho Halpin (2011).

para identificar coisas no mundo e não apenas recursos de informação. Surge assim a questão: como saber se uma dada URI refere-se a um arquivo que contém uma informação sobre uma coisa no mundo ou à própria coisa no mundo, ou ambos? No primeiro caso a URI é um endereço localizador de informações sobre um recurso, no segundo caso é simplesmente o nome de algo no mundo. Ao longo do artigo ilustraremos com mais detalhes as razões desse problema crucial para a evolução da Web rumo à Web Semântica. O foco é colocado sobre o chamado “problema da identificação”, detalhado mais a frente no texto. O problema ocorre nos processos de comunicação entre programas (aqui considerados agentes de software) que estão por trás do funcionamento dos serviços de informação de sítios da Web Semântica.

Procuramos mostrar neste artigo que existe uma forma fundamental, mais filosófica de enquadrar o problema de identificar recursos por meio de URI's. Acredita-se que essa maneira de enxergar a questão pode iluminar possíveis soluções de engenharia, que encontram-se ainda no estágio de propostas técnicas no seio do W3C. Dito de outra forma: examinaremos o problema da identidade de recursos (URIs) a partir da perspectiva filosófica na busca de compreendê-lo em toda a sua amplitude. Procuramos também mostrar que essa questão é fundamental para a evolução da Web atual em direção à uma Web cada vez mais semântica.

Sob a perspectiva filosófica três posições, explicadas mais adiante no texto, se apresentam como base de reflexão sobre o problema da identidade na Web: (1) primeiramente a posição lógica, exemplificada pela Teoria das Descrições originária de Bertrand Russell (1905); (2) em segundo lugar a posição inspirada pela Teoria Causal da Referência de Kripke (1972), reforçada pela Teoria da Referência Direta de Putnam (1975); e finalmente (3) em terceiro lugar uma posição que poderia denominar-se Linguagem Pública de inspiração wittgensteiniana, que vê as URIs como termos constitutivos de uma linguagem pública. No caso da Web, trata-se de uma linguagem compartilhada que vai se constituindo aos poucos graças ao papel central de uso dos grandes motores de busca hoje.

Por outro lado, interessa à filosofia notar que essas diferentes posições mostram que os debates dentro do campo mais específico da filosofia da linguagem estão bem vivos e

podem ser relevantes para orientar soluções de engenharia; assim como pode a filosofia beneficiar-se de um vasto campo empírico para suas experimentações teóricas que é a Web.

Veremos mais adiante que a terceira posição indicada acima é uma novidade que decorre do advento dos grandes mecanismos de busca (agentes) que vem possibilitando pensar o problema da identidade de coisas no contexto de nossas conversas com os mais diversos agentes de software. Procurando clarear especificamente esta posição, atualmente, considerações de filósofos que estudam a Web Semântica alinham-se às já clássicas e influentes reflexões de Ludwig Wittgenstein e concluem que tais agentes estão aos poucos entrando em nosso “jogo de linguagem”, para usar esta expressão do reconhecido filósofo da linguagem.

Deve-se também notar, em contraponto, que existe uma outra forma de pensar o problema de identidade no contexto da Web Semântica. Alguns autores defendem que o debate sobre o problema da identidade (usando URIs) não avançará se derivar para discussões filosóficas “subjetivas”. Se isso ocorre, dizem eles, é meramente porque o problema não está ainda suficientemente claro. Ou seja, não está ainda fundamentado em “casos de uso” adequadamente especificados, cf. p. ex. Booth (2012).

Independente da visão que se tem sobre o problema, de inspiração filosófica ou de engenharia, sua resolução abrirá fronteiras até então inimaginadas para a inteligência artificial e poderá melhor legitimar, filosófica e cientificamente, a nosso ver, o uso do termo “Web Semântica”, criticado por diversos autores. Com efeito, se por um lado as máquinas continuam exclusivamente manipuladoras de símbolos, por outro, como elas estão “escutando” (inputs de dados) milhões de pessoas perguntar em suas buscas e pesquisas na Web, nada as impede de usar esses dados para entrar cada vez mais profundamente em nosso “jogo de linguagem” (ou melhor, nada impede seus programadores e analistas de fazê-lo por meio de teorias inovadoras). Expressando de forma bastante simplificada nosso ponto de vista: assim como uma criança humana aprende a falar imitando seus pais, uma máquina o faria, pelo menos rudimentarmente, observando milhões de pessoas se referirem a coisas no mundo e usando ferramentas linguísticas e estatísticas, implementadas nos algoritmos de busca. Obviamente que uma tal afirmação deve ser relativizada sob a luz de inúmeros trabalhos também em filosofia da linguagem, como p. ex. os trabalhos em

linguística cartesiana de N. Chomsky, além de muitos outros em fenomenologia hermenêutica.

De toda forma, na perspectiva filosófica a possibilidade de existir algum grau de semântica, mesmo que rudimentar, no jogo da linguagem que temos com máquinas na Web é um ponto crucial, pois é exatamente nele que os filósofos que criticam a possibilidade de uma inteligência artificial se apegam (John Searle, Luciano Floridi, Dreyfus e muitos outros), refutando cabalmente tal possibilidade. Com efeito há um desafio óbvio na comunicação com tais agentes visto que esses não estão “plugados”, “acoplados” ao mundo físico real, espaço-temporal, como nós estamos (cf. p. ex. “Fenomenologia da Percepção” de Merleau-Ponty). O insight central da filosofia de Merleau-Ponty sobre o modo como se dá a percepção de significados é que ele se dá de maneira essencialmente corporal; e não apenas contingencialmente corporal. Ou seja, não há como falar de significado (semântica) sem que haja corpo orgânico. Enquanto seres artificiais, os agentes não possuem corpo orgânico natural e não compartilham conosco os cinco sentidos humanos. Como não possuem um corpo natural orgânico e sim artefactual, para esses filósofos (e para todos nós), máquinas são seres exclusivamente sintáticos. Não podem ser semânticos (vide John Searle e seu exemplo do “quarto chinês”, e Dreyfus que também se baseia na fenomenologia da percepção de Merleau-Ponty para criticar as iniciativas de pesquisadores da inteligência artificial).

Somos "máquinas naturais" processadoras de símbolo infinitamente mais poderosas em criatividade do que as máquinas artificiais que construímos (La Mettrie apud Chomsky (1966)). Mas isso não significa, entretanto, que estas últimas não possam participar, em graus mesmo que extremamente simplificados (mas ainda úteis) de nossa linguagem. Isso não significa que não possam existir canais de comunicação entre homens e máquinas, via a constituição de linguagens artificiais especificamente voltadas para essa comunicação. Isso claro partindo já de uma certa aceção (bastante genérica) do termo “comunicação”. Para ilustrar esta aceção genérica de “comunicação”, vale citar Simondon² que nos oferece uma definição de informação em oposição à quantificação do sinal e ruído dada por Claude

² Gilbert Simondon foi um filósofo francês, mais conhecido por sua teoria da individuação, fonte de inspiração para Gilles Deleuze e Bernard Stiegler.

Shannon e a teoria da informação. Para Simondon, a informação é "a tensão entre duas realidades diferentes". Desta forma, a informação precede a individuação e nesse contexto o conceito "informação" se aproxima do conceito "dado" como ocorre na teoria da informação. Entretanto a diferença é que aqui a informação gera resoluções, porém parciais, de tal tensão. A informação precede a individuação³, e é através de processos de transdução que a individuação ocorre.

Não é objetivo aqui entretanto aprofundar a natureza da comunicação homem/homem ou homem/máquina para além de simplesmente ilustrar a relevância do problema da identidade nesse contexto. Assim, de volta ao tema fulcro do texto, sua exposição será estruturada apresentando na seção que segue uma ilustração do grau de evolução atual do diálogo usuário/mecanismos de busca cada vez mais "semânticos"; a Seção 3 discute sobre sentido e referência na Web Semântica; na Seção 4 o problema da identidade é apresentado na perspectiva filosófica; Finalmente na Seção 5 tecemos considerações finais.

2. A busca semântica: um diálogo com a máquina?

Iniciemos revendo mesmo que muito brevemente um pouco da história da web... Trinta anos se passaram desde as primeiras ideias de objetificação da informação com o padrão SGML e a engenharia de documentos até a linguagem de ontologias OWL e os "Grafos de Conhecimento" do Google utilizados amplamente hoje (Figura 1) (Bax, 2001). Essa tecnologia Web mais recente representa um grande avanço no tratamento e na busca por informações sobre entes do mundo real, pessoas, eventos, coisas etc. e organização de suas propriedades. Pesquise por um personagem qualquer na história e obtenha imediatamente seu nascimento, morte, marido, filhos, outras pessoas relacionadas... etc. Por trás da busca, estruturando a informação, há um grafo ligando nomes, classes, propriedades etc. (Figura 1).

³ Correspondente às relações das realidades díspares, a transdução pode ser rastreada através das estruturas e padrões que emergem da resolução de um determinado conjunto de relações, especificamente padrões e estruturas que não estavam presentes antes de transdução.

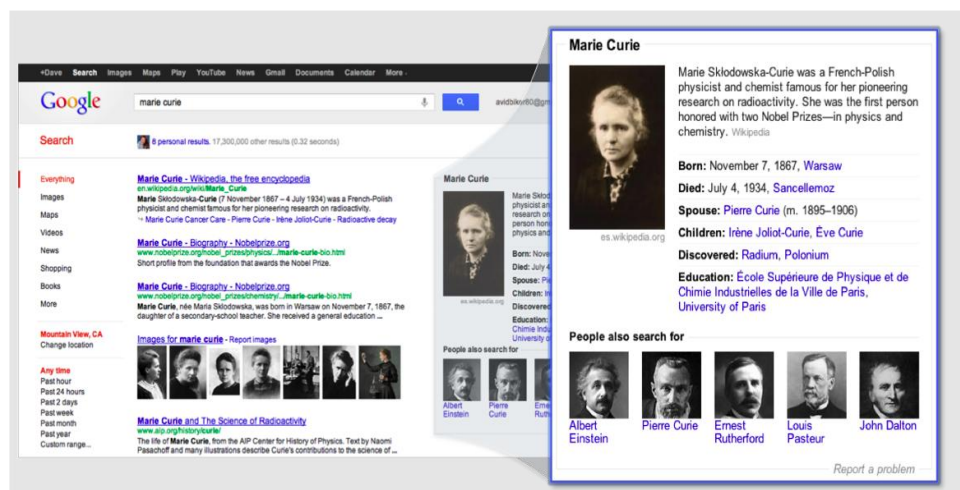


Fig. 1 - Grafos de Conhecimento
Fonte: o autor.

Outros buscadores como o Wolfram Alpha estabelecem sofisticados diálogos com o usuário. Por exemplo, à pergunta: "máquina, diga-me qual é o lago mais profundo dos Estados Unidos?" a resposta retorna precisamente: Crater Lake. Nota-se um sofisticado tratamento semântico por trás da interface, que é realizado pelo algoritmo que é capaz de codificar a pergunta feita originalmente em linguagem natural e interpreta-la (Figura 2).



Fig. 2 - Tratamento semântico de uma pergunta feita em linguagem natural.
Fonte: o autor.

Pode-se perguntar pelo que está acontecendo de fundamental nessas interações de buscas. Estamos compartilhando uma linguagem comum com essas máquinas de busca? O que mais precisamente chamamos “semântica” no termo Web Semântica?

Sabe-se que a Web é um espaço de informações “identificadas” por nomes, URIs ou *Uniform Resource Identifiers*. Sintaxe inventada por T. Berners Lee. Um recurso é "qualquer coisa que pode ser identificada por um URI" (Berners-Lee et al. 2005). As URIs são usadas para acessar recursos de informação na web. A Web é um espaço virtual construído sobre a infraestrutura física da Internet. São os protocolos que realmente enviam os bits que compõem os hipertextos que fluem da aplicação servidora para a aplicação cliente, o browser ou navegador. Imagine-se vários dispositivos conectados pela Internet (Cf. Figura 3). Em cada um deles pode-se configurar um servidor web (o popular Apache p.ex.) que permite o acesso por outras pessoas usando uma URI.



Fig. 3 - Dispositivos conectam informações usando URIs
Fonte: o autor.

Coloca-se então páginas em cada um desses servidores móveis ou fixos e elas podem ser interligadas constituindo e integrando-se num gigantesco grafo. Cada recurso do grafo tem um nome único global, uma URI. A comunicação entre cliente e servidor na web se dá pela utilização do protocolo HTTP. Note na Fig. 4 a mensagem enviada pelo cliente ao servidor, do tipo GET / HTTP 1.1 e mensagem em resposta do servidor contendo a página HTML que o servidor entende como padrão para a URI requisitada.

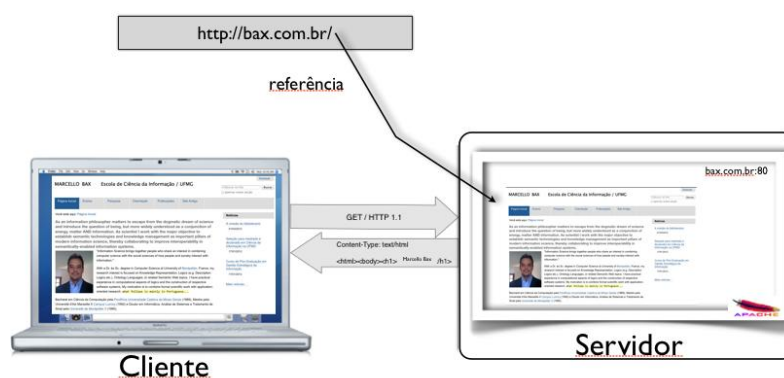


Fig. 4 - Mensagem HTTP enviada pelo cliente ao servidor e resposta do servidor.

Fonte: o autor.

O tempo passou e mais de 1 bilhão de usuários fizeram da Web a maior revolução da virada do século XXI (dizem alguns: “Tim Berners-Lee tornou-se ‘Sir’ Berners-Lee”).

A Web como a descrevemos até agora consistiu de um conjunto de URIs referindo-se exclusivamente a recursos de informação, documentos. No caso de documentos, o sentido da informação não estruturada neles contida só pode ser dado pela interpretação humana. Assim, como ilustra a Fig. 5, não se faz distinção entre sentido e referência na Web original. Ou seja, os sentidos só encontram ou casam com suas referências gerando uma interpretação na mente do ser humano que interpreta o documento, mas obviamente não é assim para as máquinas.



Fig. 5 - Sentido e referência são uma coisa só na Web original.

Fonte: o autor.

Entretanto, vale notar a distinção entre sentido e referência, um dos legados de G. Frege ao final do século XIX. Para Frege, sentido é diferente de referência. Essa questão baseou todo o trabalho, no século XX, da tradição em filosofia analítica e filosofia da linguagem (Bertrand Russell, 1o. Wittgenstein, Carnap, o círculo de Viena, Quine, Tarski, Putnam etc.).

A seguir encontra-se uma rápida explicação sobre sentido e referência na Web Semântica que é básica para que se entre no tema principal do artigo

3. Sentido e Referência na Web Semântica

Parece consenso que o caso de uso paradigmático da tecnologia de Web Semântica é a integração de dados em larga escala. Comparado com abordagens baseadas em sistemas de banco de dados relacional tradicionais, três aspectos da Web Semântica facilitam a integração de dados distribuídos globalmente a partir de fontes heterogêneas e autônomas: (1) um modelo de dados simples, mas expressivo e extensível que é o *Resource Description Framework* (RDF); (2) a identificação por URIs usadas para nomear recursos globalmente, e (3) a possibilidade de raciocínio baseado em Lógica Descritiva.

Com efeito, a proposta da Web Semântica é permitir que aplicações combinem e processem dados e informações disponíveis na rede (Berners-Lee, 2001). A Web Semântica usa RDF para representar a informação processável por máquina e os conjuntos de dados vão sendo criados de forma independente um do outro. O padrão RDF usa URIs para nomear recursos, assim o caso de uso fundamental da Web Semântica que ilustra o problema da identidade pode ser resumido assim:

Um agente semi-atômico deve ser capaz de fundir de maneira "sensata" dois conjuntos de dados, criados independentemente um do outro, usando URIs comuns para relacionar os dados.

Na Web Semântica uma URI não serve apenas para identificar um recurso de informação, ela pode também identificar qualquer coisa fora do mundo virtual. A Web Semântica representa um avanço por permitir explicitar a distinção existente entre sentido e referência. A

referência passa a ser o objeto em si, presente no mundo, e o sentido é a descrição RDF desse objeto. Como exemplo note-se que a URI na Fig. 6 que se refere ao autor desse artigo, mapeia também para um documento que o descreve formalmente por um grafo (aí representado no padrão FOAF -- *Friend-Of-A-Friend*).

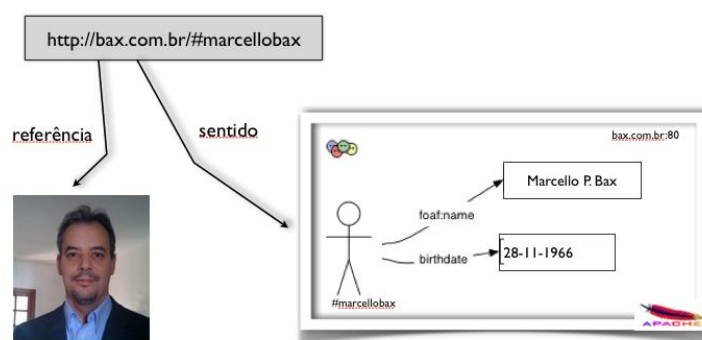


Fig. 6 - Sentido e referência na Web Semântica.
Fonte: o autor.

Não é mais novidade que a Web Semântica, ou web das coisas, difere da web de documentos no sentido em que nela à informação pode-se atribuir significado bem definido. É isso que permite que computadores e pessoas trabalhem em cooperação cada vez mais avançada. As informações, antes voltadas exclusivamente para a interpretação humana, viram objetos que agentes descentralizados se acordam em referenciar. Inclusive referenciando coisas que não são acessíveis na Web, como pessoas, lugares e conceitos abstratos etc.

Com base nesses avanços tecnológicos, de repente a questão de definição do que é um "recurso", e como diferentes agentes (humanos e máquinas) podem identificar e tratar tais recursos, pode ser vista como uma questão "puramente teórica" tal como é tratada em filosofia da linguagem. *Insights* teórico-filosóficos iluminam soluções práticas importantes para o desenvolvimento da web.

3.1 RDF

Existem vários padrões definidos pelo W3C que compõem o conjunto de tecnologias denominadas “tecnologias semânticas”. O RDF é um desses padrões, talvez o mais importante hoje para a Web Semântica. É a primeira linguagem de representação do conhecimento na Web Semântica. RDF foi influenciado diretamente pelos desenvolvimentos em representação de conhecimento originados das pesquisas em “redes semânticas”, na área de IA. A diferença com relação às “redes semânticas” (Fig. 7) é que RDF recebeu uma semântica formal. Isso faz com que grafos RDF possam ser processados de diversas maneiras por máquinas. Nesse sentido são bem mais do que meros desenhos que organizam ideias. A formalização torna-os bases de conhecimento tais como entendidas em Inteligência Artificial.

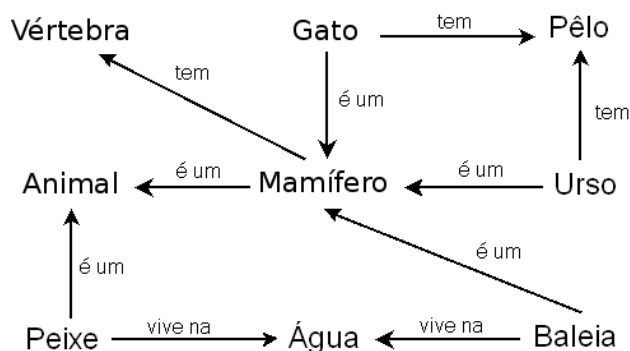


Fig. 7 - Exemplo de uma rede semântica.

Fonte: Wikimedia Commons (http://commons.wikimedia.org/wiki/File:Rede_Sem%C3%A2ntica.png).



Fig. 8 - O modelo RDF estrutura-se por triplas: *objeto*, *propriedade*, *valor*.

Fonte: Santos Neto (2013).

Declarações RDF são feitas na sintaxe ou formato de triplas (*objeto*, *propriedade*, *valor*) (Fig. 8). Nas triplas, *objeto* e *valor* são entes, *propriedade* é relação. Simplificando um pouco, cada elemento da linguagem, ente ou relação, é identificado por uma URI. Pode-se perceber então como as URIs substituem os termos da linguagem natural usados nas redes

semânticas. Durante a fusão de conjuntos de dados, ao compartilharem URIs as triplas formam grafos. Grafos podem ser mesclados para formar grafos maiores (Fig. 9).

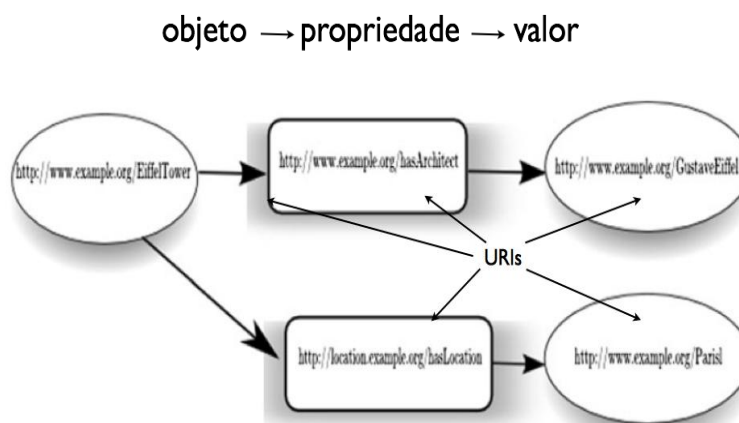


Fig. 9 - Grafos RDF são formados pela fusão de triplas.

Fonte: adaptado de Halpin (2011).

A partir da relação de referência, pode-se explicar o que seria "semântica" na Web Semântica. Trata-se da relação de mapeamento entre uma URI e seu referente, a relação que une identificadores -- não nomes, como no caso da língua natural, mas URIs -- a coisas. A relação semântica na Web Semântica une adequadamente URIs ao mundo real. 'Sintaxe', por outro lado, é a maneira como os nomes ou termos podem ser combinados para produzir frases bem formadas.

Devido à natureza propositadamente primitiva do modelo de dados do RDF, na Web Semântica imagina-se o mundo a ser modelado como composto por coisas e relações binárias (entre as coisas). Isso explica o porque o modelo usa triplas, constituídas de simples relações binárias entre coisas. Dessa maneira tem-se mais facilmente o efeito lego ("*building blocks*") que permite incrementar a representação (o grafo RDF) de forma modular (e orgânica).

A parte superior da Figura 10, composta de três URIs sucessivas (tripla) forma uma frase RDF. São apenas cadeias de caracteres e mera sintaxe. A semântica, ou a função que estabelece a relação com o mundo, é o que satisfaz a sentença de acordo com uma dada interpretação. A semântica é representada pelas setas na Fig. 10.

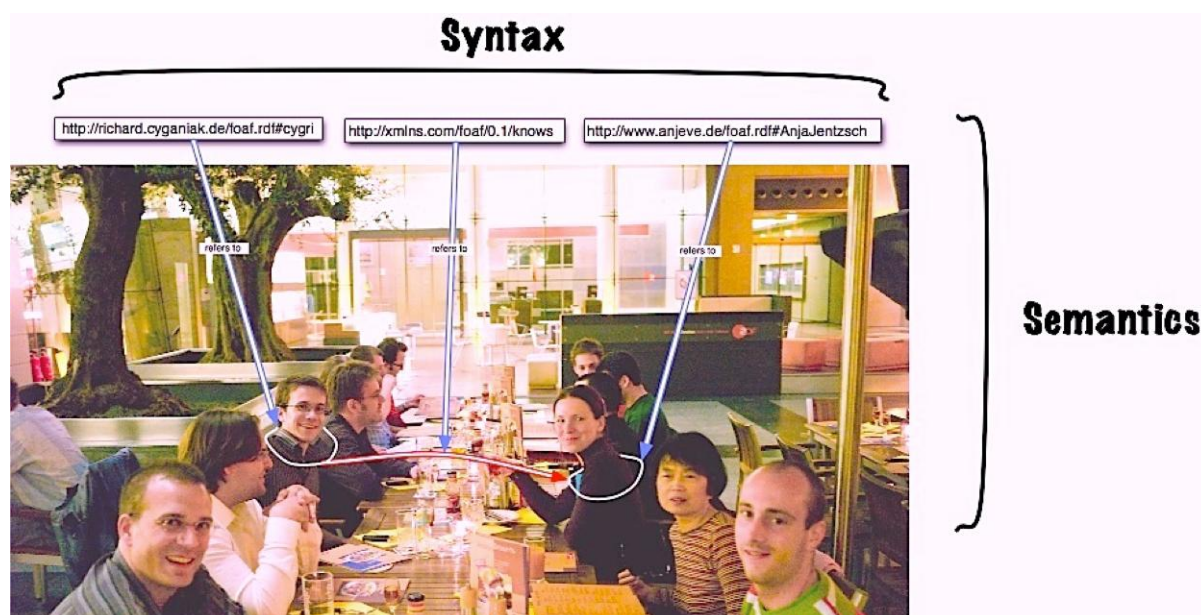


Fig. 10 - Sintaxe e Semântica.

Fonte: Henry Story (2010).

Por um lado parece estranho falar em ‘semântica’, pois sabemos que os computadores são muito bons em manipular símbolos, mas não em interpretá-los. Como eles conseguiriam fazer essa relação com o mundo real? Como eles conseguiriam alguma ‘semântica’? Como essa relação com o mundo pode se dar? Alguns filósofos dizem que ela não é possível, um conhecido filósofo da informação como L. Floridi, no artigo “Web 2.0 vs. the Semantic Web - A Philosophical Assessment”.

“Semantic Web, [...] unrealistic and inflated hype or just unreliable and shameless advertisements.” Floridi, L. “Web 2.0 vs. the Semantic Web - A Philosophical Assessment”, 2010.

A seguir examinar-se-á mais detalhadamente esta questão.

Em um *workshop* ocorrido no início do século sobre Web Semântica Tim Berners-Lee propõe o dito: “A Web Semântica é um mapeamento entre URIs e significados”. Porém, como pode se dar um tal mapeamento constitutivo do ‘sentido’ de uma URI? Na sessão seguinte o artigo analisa as propostas existentes hoje na literatura para realizar o mapeamento entre URIs e seu significado.

Para mapear URIs a significados existem atualmente duas propostas mais conhecidas em discussão. Elas dividem a comunidade que está por trás da construção da Web Semântica. Na proposta que tem como principal porta-voz um dos mais conhecidos teóricos da representação do conhecimento em lógica, Patrick Hayes, esse mapeamento deve ser estabelecido por uma ‘descrição’ lógica (expressa em RDF). Para Berners-Lee entretanto, defendendo a proposta antagônica, o mais indicado seria que o mapeamento URI/significado se desse por uma referência ou apontamento direto da URI à coisa. Eis aí a essência dos dois posicionamentos do debate referido acima como “Crise de Identidade” na Web Semântica.

Para os objetivos deste artigo, o que interessa notar é que tais propostas ou posicionamentos podem ser colocados lado a lado numa analogia com linhas de pensamento teóricas mais fundamentadas, debatidas em filosofia da linguagem.

Além disso, o artigo dá ciência de uma terceira proposta mais recente que seria inspirada não na lógica pura nem no apontamento direto, mas no uso social que se faz das URIs na Web. Essa proposta também pode ser contrastada com desenvolvimentos teóricos da filosofia da linguagem. O objetivo principal desse artigo é tecer uma reflexão sobre o estado de evolução da Web Semântica dadas estas três propostas. Faremos isso à luz do que se conhece em filosofia da linguagem.

4. O problema da identidade na perspectiva filosófica

Dois problemas principais aparecem: (1) o problema da ambiguidade que está sempre presente no sentido de um conceito: como se poderia garantir que uma URI para a Torre Eiffel, por exemplo, será sempre usada para se referir à “Torre Eiffel” em Paris? e não p.ex. à parada de metrô de mesmo nome? (2) o problema da representação: a URI deveria apenas representar o objeto no mundo físico ou apontar também para uma representação informacional sobre ele? ou ambas? Embora haja uma posição oficial do órgão padronizador da Web (o W3C) para responder a essas questões, ela ainda não é amplamente aceita pela comunidade.

Referir seria descrever logicamente o recurso ou meramente nomeá-lo diretamente (*i.e.*, batizá-lo)? Duas respostas mais conhecidas à essa pergunta aparecem na comunidade. A primeira equivale, em filosofia da linguagem, à uma posição inspirada na "teoria das descrições" atribuída a Russel, Wittgenstein (em sua primeira fase) e Carnap. Teoria que encontrou em Tarski a sua forma mais acabada. Nessa teoria, as descrições são consideradas declarações lógicas. Mostra-se que um nome é na verdade uma descrição lógica disfarçada⁴. O referente do nome é equivalente ao conjunto de coisas possíveis, dado por um modelo matemático, de tal modo que todas as declarações que contêm o nome sejam satisfeitas.

Interessante notar que tal teoria foi mais fortemente desafiada por outra mais recente: a "teoria causal de referência" de Kripke⁵ e Putnam. Esta última pode ser usada para justificar filosoficamente o posicionamento oficial do W3C. É curioso notar que as duas visões acima equivalem a dois posicionamentos antagônicos em resposta à uma pergunta clássica na filosofia da linguagem: o sentido está no objeto apontado (*i.e.*, na referência) ou na descrição lógica que se faz do objeto⁶?

Entendendo uma URI como uma descrição (lógica)

O posicionamento, defendido por P. Hayes, pesquisador que está por trás das formalizações lógicas dos padrões do W3C, equivale ao cenário apresentado pela Figura 11 abaixo. Este posicionamento é defendido sobretudo por lógicos, que enxergam a Web Semântica como uma camada lógica totalmente distinta da camada hipertextual que constitui a web original. A enxergam sobretudo como uma base de conhecimento gigantesca e representada em formalismo lógico.

⁴ Vide o exemplo clássico da análise da declaração "O atual rei da França é calvo" (Russell *apud* BAX, M. P. e COELHO, 2012).

⁵ A primeira tese fundamental de Kripke diz: nomes próprios são – diferentemente da maioria das descrições – designadores rígidos.

⁶ Os dois posicionamentos têm em comum o fato de definirem o sentido por extensão e não por intensão, mas voltaremos sobre esse ponto específico mais tarde.

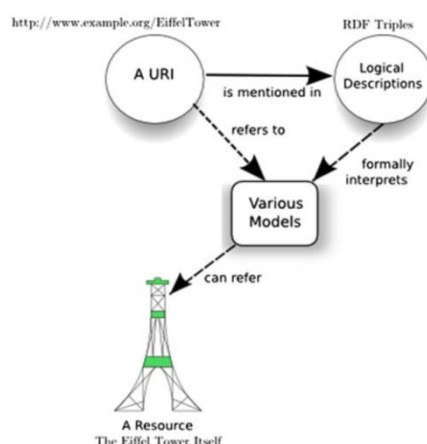


Fig. 11 - Identidade de um recurso por sua descrição lógica (em RDF).

Fonte: Halpin (2011).

Ou seja, por essa abordagem lógica, uma URI, quando mencionada em um conjunto de tripas RDF (uma descrição lógica), se refere exclusivamente à uma constante⁷ presente nos modelos possíveis que satisfazem a descrição. Essa constante por sua vez deveria referir-se a um objeto do mundo real (como a Torre Eiffel na figura), mas não há como garantir já que não há “pontos de contato” necessariamente determinados *à priori* entre a camada lógica e a camada ontológica. Entretanto isso não impede o acesso a páginas da web por meio de URIs. O acesso a representações hipertextuais (páginas) da Web é simplesmente ortogonal à questão da referência de uma URI. Isso ocorre porque é "a arquitetura da Web que determina o acesso, mas isso não tem influência direta sobre a referência" (Hayes e Halpin 2008). Ou seja, problema de conectar essas descrições lógicas com o mundo fora da Web Semântica formal é desconsiderado nessa abordagem, abrindo um abismo intransponível entre as descrições lógicas e dados sensoriais. De forma geral, o argumento de Hayes contra Berners-Lee, no tocante ao problema de identidade (de URIs) equivale à posição de manter a semântica formal separada da Web (como planos ortogonais que não se tocam).

Entendendo URI como referência direta (batismo): a teoria causal da referência

Em contraposição à abordagem lógica descrita brevemente acima, o posicionamento oficial do W3C sobre a questão da identidade é que a URI deve apontar para um único recurso. O

⁷ Na linguagem da lógica de predicados de primeira ordem, o que desempenha o papel de nomes são as constantes individuais.

significado deve ser dado exclusivamente por aquele que cunhou a URI, seu proprietário, i.e. quem registrou a URI.

Entretanto essa visão é incompatível com a formalização semântica dada por Hayes para o RDF. Como vimos, em lógica os planos dos termos lógicos e aquele das páginas, supostamente referenciadas por eles, são ortogonais. Assim, simplesmente não há como afirmar que uma URI irá designar um único recurso. Essa é a crítica que faz Hayes a Berners-Lee.

Apesar da crítica de Hayes quanto à possibilidade (em lógica) de uma URI identificar univocamente algo no mundo real, o dito de Berners-Lee de que "URIs identificam uma coisa" (Berners-Lee 2003b) não deixa de parecer intuitivo. No exemplo dado por Halpin (2011), quando se diz "eu fui visitar a Torre Eiffel", acredita-se estar-se referindo à uma coisa muito particular no mundo real chamado de 'Torre Eiffel', não a um conjunto de descrições ou modelo abstrato do mundo (como no caso da posição lógica defendida por P. Hayes).

Assim como a posição lógica da teoria das descrições (comentada acima) serve de suporte teórico à posição de P. Hayes, a posição da referência direta de Berners-Lee tem um paralelo na filosofia, "a teoria causal da referência" de Saul Kripke (Kripke 1972). Seu argumento contra a teoria descritivista da referência, i.e., contra a posição de Hayes (onde o referente do nome é determinada pela interpretação de descrições lógicas) é o de que o nome refere-se ao objeto por uma cadeia causal que leva diretamente à este (Kripke 1972).

Nessa abordagem do problema, pelo registro do domínio "example.org", mas sobretudo ao cunhar a URI "example.org/EiffelTower" o seu proprietário estabelece o nome para a coisa em si, a "Torre Eiffel". Como ilustra a figura abaixo.

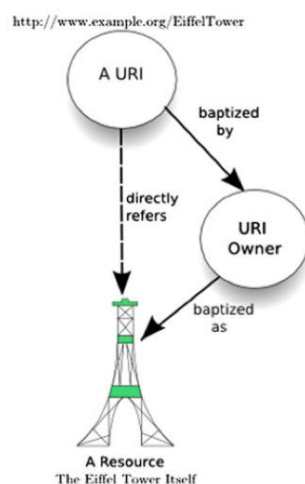


Fig. 11 - Identidade de um recurso por referência direta ou "por batismo".
Fonte: Halpin (2011).

Dessa forma, se alguém procura nomear um recurso deve meramente cunhar sua própria URI para designar o recurso que deseja nomear. Caso haja necessidade de enriquecer a descrição do recurso pela adição de outro conjunto de dados que também se refere ao mesmo recurso, poder-se-á utilizar a propriedade "sameAs" entre as duas URIs que assim passariam a apontar para o mesmo objeto no novo conjunto de dados resultante.

Um problema é que nesse cenário as URIs se multiplicariam ao infinito (motivo do dito de Pierre Levy na epígrafe do artigo), ou seja ao longo do tempo o risco é que existiriam uma infinidade de nomes para designar o mesmo objeto. O que minaria qualquer expectativa de coesão entre os dados da Web Semântica.

Uma solução possível para evitar a multiplicação de URI designando o mesmo recurso, poderia considerar a existência de um dicionário gigantesco contendo todas as coisas do mundo. O dono da URI a mapearia para tal dicionário. Contudo, seria realista um tal dicionário? questões sociais complicadíssimas se imporiam. Tal dicionário seria uma ontologia universal que conteria todas as coisas (instâncias) referenciáveis no mundo. Quando um autor se perguntar que URI deveria usar para um determinado objeto este dicionário seria o local a consultar. Dado o objeto como entrada ele retornaria a URI mais apropriada.

Além disso, o que a consulta a esse dicionário retornaria? uma página web? mas como o agente do usuário (um aplicativo qualquer) diferenciaria que se trata do nome de algo fora

da web ou informação sobre esse algo? Trata-se do segundo ponto referido no início da seção.

Na abordagem em questão, esse ponto é parcialmente equacionado pela proposta denominada "redireção 303". Uma URI é retornada para a coisa-em-si e caso o acesso seja feito por um navegador, então este é redirecionado para outra URI (para uma URL). Entretanto este mecanismo foi considerado difícil de usar e de entender.

A distinção fregeana entre "sentido" e "referência"

Segundo Halpin (2011), a raiz filosófica do problema poderia se explicar pelo fato de que tanto Russell e Kripke e assim como Berners-Lee e Hayes rejeitam a noção intensional de 'sentido'. Essa rejeição teria levado tanto Russell quanto Kripke a construir suas teorias do significado exclusivamente baseadas em referência extensiva.

Como exemplo singelo de uma tal abordagem, exclusivamente extensionalista, seja a frase: "A classe dos objetos que estão sobre a mesa". Essa classe de objetos é representada pelo conjunto de todos os objetos que estão de fato sobre a mesa (pode-se conta-los). E não pela proposição intencional "os objetos que estão sobre a mesa" (quando não se pode conta-los). Que corresponderia a uma classe intensionalmente definida e não extensionalmente definida (Bax & Coelho, 2012).

Porém, conforme nos lembra Halpin (2011), em seus trabalhos Frege considerou que o significado de qualquer termo de uma língua é "o sentido" das frases que usam o termo, ao invés da referência direta do termo" (Frege 1892). Frege define 'sentido' em termos de um misterioso "modo de apresentação": "pensá-los ligados a um signo (sign), além do que o signo refere-se [...], também o que eu gostaria de chamar o sentido do signo, onde o modo de apresentação está contido" (Frege 1892).

Halpin (2011) sugere que a razão pela qual essa noção de sentido teria sido discriminada por muitos filósofos, como Russell e Kripke, poderia a ser a de que sentido, nessa acepção, é visto como uma "noção misteriosa", uma noção singular e individual, como a noção lockeana

de idéia. Sobre o termo 'Ideia' John Locke nos diz: "... significa tudo que é objeto do entendimento quando um homem pensa." (Ensaio I, 1, 8, p. 47). No entanto, Frege diferencia a noção de sentido, daquela de ideia subjetiva na cabeça de um indivíduo quando pensa em um referente.

Sentido como noção objetiva

Sentido é o mesmo que semântica, significado ou significação. Pode-se entender o sentido como algo interno na mente de um agente? ou seja, como parte de uma linguagem privada? Para Frege não, pois segundo ele próprio:

"sentido é atributo comum a muitas pessoas [...] não é um modo da mente individual [...] a humanidade tem um estoque comum de pensamentos que são transmitidos de uma geração para outra". Frege 1892.

Para Frege 'sentido' seria algo objetivo, na medida em que é compartilhado na linguagem comum pública. Ou seja, na prática o sentido é algo objetivamente compartilhado entre os agentes no uso dos termos de uma língua.

Para Halpin (2011) essa aceção ao mesmo tempo objetiva e intensional do significado pode inspirar ainda um terceiro posicionamento possível (além dos dois comentados acima), no qual o sentido não estaria nem no objeto apontado, nem na sua descrição lógica.

Como um possível novo posicionamento, cabe perguntar como este poderia ser implementado na prática da Web Semântica. Como essa visão poderia funcionar na Web? Citando Dummett (1993): "a tese de Frege de que 'sentido' é algo objetivo é uma antecipação da doutrina de Wittgenstein de que 'o significado é o uso'". Pode-se assim pensar numa terceira posição, a posição que busca operacionalizar o uso de uma "linguagem pública" que toma a noção objetiva do "sentido" e a análise de Wittgenstein do "sentido como uso". Algo análogo ao sentido como "significação" em Quine (e não como "significado").

Segundo Halpin (2011) estas noções estariam enraizadas no uso real de URIs na Web. Ou seja o sentido poderia ser resolvido observando o comportamento dos usuários na Web.

Entendendo URI como uma construção social pública (linguagem pública)

A abordagem baseada em uma ideia de “linguagem pública” afirma que a Web Semântica pode ser vista como uma forma de linguagem compartilhada entre múltiplos agentes, homens e máquinas. O significado de uma URI equivale ao uso comunitário que os agentes fazem dessa URI. O significado de uma URI não seria então determinado por qualquer referência atribuída a ela pelo seu proprietário. Segundo essa visão, a atribuição de um significado para uma URI por seu dono não seria suficiente para designar o significado realmente compartilhado pela comunidade. Seria necessário antes, que pesquisas nos mecanismos de busca por determinado termo/nome retornasse a URI nas primeiras posições. A posição mostraria o grau de “certeza” que aquela comunidade considera uma dada URI o nome ou a identidade efetiva de um conceito (i.e., uma ideia pública). Assim a posição que pleiteia pensar o problema por analogia a uma ‘linguagem pública’ para resolver a “identidade de uma URI” não daria ao proprietário da URI qualquer privilégio especial. Por outro lado, essa posição também teria que conviver com a ambiguidade dos significados, já que estes constituem-se implicitamente pelo comportamento dos usuários na Web.

Ou seja, o jogo de linguagem em nível conceitual não se reduz a definir nomes para referências (coisas no mundo), mas ele é antes de tudo saber compartilhar uma convenção socialmente estabelecida o suficiente para realizar tarefas ou resolver problemas. A ambigüidade é geralmente resolvida pelo contexto implícito na linguagem. Essa posição corrobora o dito da segunda fase do pensamento de Wittgenstein: o significado de qualquer termo é dado pelo seu uso social. Assim o significado da URI é “seu uso” por uma comunidade de agentes (humanos e máquinas) e não apenas por uma referência direta, nem apenas por uma teoria formal de modelos (razão pura).

Para aplicar o pensamento de Wittgenstein à Web Semântica, o primeiro movimento seria considerar que a Web Semântica constitui um novo jogo de linguagem. Não há nenhuma razão para que jogos de linguagem, no sentido wittgensteiniano, se restrinjam às línguas

naturais. O próprio Wittgenstein observa o aspecto dinâmico das linguagens: "novos tipos de linguagem, novos jogos de linguagem, passam a existir e outros se tornam obsoletos e são esquecidos" (Wittgenstein 1953).

Dado que os agentes podem ser máquinas e humanos, com suas próprias normas de linguagem e protocolos, tomando o significado como o uso compartilhado dos termos parece não haver nenhuma razão para que computadores, ou combinações de computadores e seres humanos, não possam criar e usar novos jogos de linguagem. Contra a posição de Berners-Lee da 'referência direta', apenas em raros jogos de linguagem a referência deve ser especificada de forma 'inequívoca' (Berners-Lee diz que deveria existir apenas uma URI para cada coisa na web).

Resta-nos entender um pouco mais profundamente como uma teoria de significado desse tipo poderia funcionar na Web Semântica. Segundo Halpin (2011) a resposta pode advir dos motores de busca da Web, como o Google. A recuperação de informações por esses motores, e sua metodologia estatística orientada a dados, é análoga à uma filosofia da linguagem neo-wittgensteiniana à qual é dada operacionalização computacional. Ou seja, motores de busca e suas abordagens estatísticas, podem ser o substrato operacional para se enquadrar o problema da identidade na perspectiva de uma filosofia da linguagem neo-wittgensteiniana.

Em seu artigo Halpin (2011) levanta o quanto Wittgenstein influenciou a área de RI. Se o dito de Wittgenstein de que o "significado é uso" parece complicado de ser formalizado, Halpin lembra que próprio Wittgenstein já perguntava: "Como a noção de "significado como uso" poderia ser operacionalizada em um método que fundamentaria uma ciência da linguagem?"

Contemporaneamente, a resposta ao problema da identidade poderia vir do estudo da estrutura da linguagem de forma empírica, baseada em dados estatísticos implícitos nos algoritmos dos mecanismos de busca. O uso de termos na caixa de busca resulta em uma lista não ordenada de URIs relevantes, dentre milhares ou milhões. Então pode-se dizer a

princípio que o significado de uma URI, sua identidade, equivalha ao conjunto dos termos que designam aquela URI.

5. Conclusão

No ímpeto de buscar uma compreensão mais aprofundada do estado atual de desenvolvimento da Web, ou da evolução da Web rumo à Web Semântica, discutiu-se neste artigo três posições ou modelos de referência para pensar o uso de nomes próprios na web semântica. O modelo de referência direta onde uma URI é apenas um identificador, sem descrição "essencial" ligado a ela (à la Kripke). O modelo de referência lógico onde uma URI sempre implica uma descrição (à la Russell). O modelo de referência segundo práticas estabelecidas pelo uso dos termos em uma comunidade (à la Wittgenstein).

Como vimos, determinar o que é o significado de uma URI é um problema fundamental que está por trás de toda e qualquer possibilidade de interoperabilidade semântica na web. Trata-se de um problema ainda em aberto. Várias discussões e artigos científicos foram publicados sobre o problema. Para citar apenas alguns: Ferrara et. Al 2008; Halpin e Place 2010; Booth 2012. Entretanto, como Booth (2012) deixa claro, ainda não há uma solução definitiva e padronizada adotada pelo W3C. Uma vez resolvido este impasse, a expectativa é a de que a Web Semântica evolua e cresça rapidamente, já que parte dos entraves ao seu crescimento advêm desse problema fundamental ainda em aberto. Questão debatida ainda hoje em filosofia da linguagem. Talvez por "tropeçar" no difícil problema filosófico de referência e significado, o sucesso da Web Semântica, um dos projetos mais ambiciosos de representação do conhecimento até agora, tenha obtido resultados ainda tímidos.

Nesse artigo discutiu-se os elementos teóricos principais desse contexto. O objetivo foi mostrar como uma abordagem filosófica, que analisa mais a fundo na busca pelo entendimento do problema, pode iluminar uma possível solução de engenharia ainda não concretizada. Na linha semelhante ao trabalho de Halpin (2011), argumentou-se neste artigo que os debates sobre o significado e referência de URIs podem ser vistos como um renascimento do mesmo debate entre as teorias causal e descritivista da referência na

filosofia da linguagem. Deste modo percebe-se como, no decurso da prática da informática, problemas filosóficos surgem.

O impasse entre Hayes e Berners-Lee sobre o significado de URIs reflete um debate de longa data dentro da filosofia da linguagem. Mas também pode-se perguntar o que a prática da informática na Web tem a oferecer a estes debates? A máxima de Wittgenstein "significado é o uso" parece poder ser operacionalizada na Web. O que pode fornecer uma maneira de ir além dos impasses sobre as teorias de referência na Web.

Como sugere Halpin (2011), vimos que os motores de busca na Web podem ser usados para construção de uma teoria wittgensteiniana da linguagem concorrente às duas outras propostas. Uma teoria que tenta alavancar o significado de URIs não em seu apego tanto a descrições lógicas ou referências diretas, mas em seu sentido tal como dado pelos termos utilizados pelos usuários para descobrir estas URIs. Dado o enorme sucesso dos motores de busca como o Google em estruturar a Web, pode ser que seja a hora de reviver os debates filosóficos em torno de significado e referência de uma forma computacionalmente mais madura e informada.

Embora as técnicas de recuperação de informação e de aprendizagem de máquina possam não ser cognitivamente transparentes, é talvez um erro de ambas as teorias descritivistas e causal de referência considerar que o significado pode ser reduzido tão explicitamente à lógica simbólica. As complexas interações sócio-técnicas exemplificadas pelos motores de busca podem de fato ser mais propensas a fazer justiça ao problema da atribuição de significado às palavras.

Deve-se notar entretanto, que pretender enquadrar o problema como um problema filosófico, ou próprio ao tratamento da filosofia da linguagem, não é consenso na comunidade. Segundo D. Booth (2012), na mesma linha de Quine, o significado seria algo que não se pode testar facilmente. Para este autor, enquadrar o problema como uma questão de definir o "significado" de uma URI seria um engano de abordagem. Para ele o caso de uso baseado na comunicação entre agentes que tentam estabelecer o significado dos termos não especifica adequadamente o problema real. Nas palavras de Booth "...nós

não temos nenhum meio objetivo de verificar se o destinatário da mensagem, na verdade, obteve o significado pretendido do remetente".

Booth defende que o caso de uso fundamental da Web Semântica envolveria a fusão de duas bases de dados RDF que foram criadas de forma independente. Segundo o mesmo autor, o problema arquitetônico que este caso de uso representa deveria ser enquadrado em termos de como definir uma URI e não em termos de significado. Em termos de fusão de conjuntos de dados de forma independente de autoria, em vez de em termos de comunicação entre agentes. Este autor considera que para que o caso de uso fundamental da Web Semântica seja implementado adequadamente é importante a padronização de um protocolo para especificação de URIs (URI Definition Discovery Protocol (UDDP)).

Entretanto, ainda assim, um protocolo de definição de descoberta de URIs será útil apenas para aqueles que o seguem. Mas o efeito de violar o protocolo não é pior do que o efeito da publicação de dados ruins, e a web é projetada para ser resistente a dados ruins. O mercado separará aqueles que "jogarem bem o jogo" - seguindo protocolos padrão e publicando dados de qualidade - daqueles que não o fizeram.

6. Referencias Bibliográficas

BAX M.P. "Introdução às linguagens de marcas"...

BAX, M. P. e COELHO, E. M. P. Compromissos Ontológicos e Pragmáticos em Ontologias Informacionais: Convergências e Divergências. Revista Datagramazero, aceito para publicação, junho de 2012.

BERNERS-LEE, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American (2001).

<http://www.scientificamerican.com/article.cfm?id=the-semantic-web> Retrieved 2012-06-08.

BERTRAND RUSSELL. On Denoting. Mind, New Series, Vol. 14, No. 56. (Oct., 1905), pp. 479-493.

CHOMSKY, Noam. Lingüística cartesiana. 1966.

DUMMETT, M. (1993). What is a Theory of Meaning. In The Seas of Language, pages 1–33. Oxford University Press, Oxford, United Kingdom. Originally published in Truth and Meaning: Essays in Semantics in 1976.

FERRARA, A., Lorusso D., Montanelli S. Automatic Identity Recognition in the Semantic Web.

In Proc. of the 1st ESWC Workshop on Identity and Reference on the Semantic Web (IRSW2008), Tenerife, Spain. 2008.

HALPIN H., Place B. When owl:sameas isn't the same: An analysis of identity links on the semantic web. In Linked Data on the Web. LDOW2010. 2010.

HALPIN, H. Social meaning on the web: from Wittgenstein to search engines. IEEE Xplore Digital Library, v. 24, p. 27-31, 2009.

HALPIN, H. Sense and reference on the web. Minds & Machines, v. 21, p. 153-178, 2011.

HALPIN, H.; LAVERENKO, V. Relevance feedback between hypertext and semantic search. In: Proceedings of Semantic Search Workshop at the World Wide Web Conference, 2009.

HAYES, P. Message to www-rdf-comments@w3.org. 2003a. Disponível em: <<http://lists.w3.org/Archives/Public/www-tag/2003Jul/0147.html>>. Acesso em: 24 jul. 2012.

HAYES, P. Message to www-rdf-comments@w3.org. 2003b. Disponível em: <<http://lists.w3.org/Archives/Public/www-tag/2003Jul/0198.html>>. Acesso em: 24 jul. 2012.

HAYES, P. In defense of ambiguity. In: Proceedings of the identity, reference, and the web workshop at the WWW conference, 2006. Edinburgh, Scotland, 2006.

HENRY STORY. "Philosophy and the Social Web". Talk given at the 1st Web and Philosophy Conference. 2010.

KRIPKE, S. (1972). Naming and Necessity. Harvard University Press, Cambridge, Massachusetts, USA.

PUTNAM, H. (1975). The meaning of meaning. In Gunderson, K., editor, Language, Mind, and Knowledge. University of Minnesota Press, Minneapolis, Minnesota, USA.

SANTOS NETO, Antonio Laurindo dos et al. Tecnologias de dados abertos para interligar bibliotecas, arquivos e museus: um caso machadiano. Transinformação, Campinas, v. 25, n. 1, Apr. 2013. Available from <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862013000100008&lng=en&nrm=iso>. access on 07 May 2013. <http://dx.doi.org/10.1590/S0103-37862013000100008>.