

## **Uma Visão da Web Semântica**

*Júnio César de Lima      Cedric Luiz de Carvalho*

Technical Report - RT-INF\_001-04 - Relatório Técnico  
March - 2004 - Março

The contents of this document are the sole responsibility of the authors.  
O conteúdo do presente documento é de única responsabilidade dos autores.

**Instituto de Informática**  
**Universidade Federal de Goiás**  
*www.inf.ufg.br*

# Uma Visão da Web Semântica

Júnio César de Lima \*  
junio@inf.ufg.br

Cedric L. de Carvalho †  
cedric@inf.ufg.br

***Abstract.** The World Wide Web, or just Web has many deficiencies. To eliminate these deficiencies, new models have been proposed, among them the Semantic Web. According to Tim Berners-Lee, the Semantic Web is the evolution of the current Web. It aims to provide structures and give semantic meanings to Web pages, building an environment where software and users can work in a cooperative way. There is a long way to make this evolution a reality. Many efforts have been applied in pattern and technologies development, in order to support new tasks necessary to make the Semantic Web work properly. The basis of the Semantic Web, the proposed technologies, the projects that have been developed and how the scientific community has reacted to the birth of the Semantic Web are some of the topics considered in this text.*

**Keywords:** Internet, Semantic Web.

***Resumo.** A World Wide Web, ou simplesmente Web, tem uma série de deficiências. No intuito de tentar eliminar tais deficiências, novos modelos têm sido propostos. Neste contexto surge a “Web Semântica”. Segundo Tim Berners-Lee, um dos seus idealizadores, a Web Semântica é uma evolução da Web atual. Ela visa fornecer estruturas e dar significados semânticos ao conteúdo das páginas Web, criando um ambiente onde softwares e usuários possam trabalhar de forma cooperativa. Para que realmente ocorra esta evolução, muito trabalho ainda deve ser feito. Com este objetivo, muitos esforços vêm sendo empregados no desenvolvimento de padrões e tecnologias para suportar as novas tarefas necessárias ao bom funcionamento da Web Semântica. Os fundamentos da Web Semântica, as tecnologias que estão sendo propostas, os projetos que estão sendo desenvolvidos atualmente e como a comunidade científica está reagindo ao nascer da Web Semântica são alguns tópicos considerados neste texto..*

**Palavras-Chave:** Internet, Web Semântica.

## 1 Introdução

A WWW (*World Wide Web*), frequentemente chamada de *Web*, guarda uma enorme quantidade de informações criadas por uma grande variedade de diferentes organizações, comunidades e indivíduos. Ela também dispõe de uma enorme variedade de serviços que facilitam a vida

---

\*Mestrando em Ciência da Computação - GEApIS/INF/UFG

†Orientador - GEApIS/INF/UFG

das pessoas, tais como: comércio, turismo, notícias, ensino, movimentações bancárias, etc. A simplicidade de uso é um aspecto chave que fez a WWW tão popular. Todavia, o estrondoso crescimento do número de páginas disponíveis tem dificultado, cada vez mais, a localização, o acesso e a manutenção de informações nesta rede. Estes problemas vêm do fato que usuários, com diferentes níveis de conhecimento, disponibilizam uma vasta quantidade de recursos, sem nenhuma padronização. Surge daí a insatisfação dos próprios usuários, pois quando estes realizam uma pesquisa na *Web*, geralmente demoram muito para obter um resultado. O mais grave, entretanto, é que os resultados das pesquisas trazem informações, que em grande parte, não satisfazem às necessidades do usuário. Fica a cargo deste a filtragem destas informações para obter a resposta às suas perguntas.

A principal propriedade da WWW é a sua universalidade, onde “qualquer coisa pode ser ligada a qualquer coisa”. As tecnologias *Web*, portanto, não discriminam entre um texto bem escrito e um rascunho, entre informações comerciais e acadêmicas ou entre culturas e línguas diferentes. Além disso, estas informações podem estar expressas em diversos formatos. Alguns destes formatos podem ser mais adequados ao entendimento humano e outros ao processamento por máquinas.

Hoje, o conteúdo das informações é disponibilizado principalmente para o entendimento humano, havendo portanto, uma grande dificuldade na execução dos serviços de recuperação processados por máquina.

Embora a *Web* apresente muitas deficiências, ela tem um papel muito importante para a sociedade como um todo. Portanto, seu desenvolvimento constitui um desafio para a comunidade científica que se esforça na tentativa de que a integração, o intercâmbio e o entendimento semântico das informações ocorram com sucesso o mais breve possível. Várias iniciativas, como as desenvolvidas pelo Consórcio W3C (*World Wide Web Consortium*), buscam, por intermédio da criação de padrões, arquitetura de metadados, serviços de inferências e ontologias, uma melhor forma de tornar as informações processáveis pelas máquinas [5].

O uso de metadados proporciona a descrição abrangente de recursos na *Web*, sendo uma solução adequada para promover a recuperação de serviços com mais eficiência.

Há vários projetos direcionados para produção de arquiteturas de metadados com o objetivo de prover suporte à codificação, ao transporte e ao intercâmbio de uma variedade de metadados desenvolvidos de forma independente [5]. Porém, o intercâmbio destes metadados deve ser efetuado levando-se em consideração três aspectos de interoperabilidade: semântico, sintático e estrutural. A interoperabilidade semântica possibilita compreender o significado de cada elemento do recurso descrito, juntamente com as associações nele encontradas. Para se assegurar este tipo de interoperabilidade, o uso de vocabulários específicos, ontologias e padrões de metadados são essenciais.

A interoperabilidade sintática determina como os metadados devem ser codificados para a transferência de informações. A linguagem XML (*eXtensible Markup Language*) é a tecnologia que tem sido empregada com o objetivo de fornecer este tipo de interoperabilidade aos recursos na *Web*. A interoperabilidade estrutural especifica como os recursos estão organizados, juntamente com os tipos envolvidos e os possíveis valores para cada tipo. Estas características podem beneficiar-se dos recursos providos pela arquitetura RDF (*Resource Description Framework*) e do Esquema RDF (*RDF Schema*), que são tecnologias que compõem um poderoso mecanismo para prover facilidades de descrição de forma geral.

Recentemente Tim Berners-Lee (cientista britânico, atualmente diretor do W3C) vislumbrou o desenvolvimento da Web Semântica que, a partir do uso intensivo de metadados, visa prover acesso automatizado às informações através de máquinas, com base no processamento semântico dos dados e de heurísticas [5]. Segundo Berners-Lee explica, a Web Semântica

visa construir uma rede que, além de ligar documentos entre si, seja capaz de reconhecer o significado desses documentos e, através de um processamento via máquina, inferir novos conhecimentos.

A seguir, na Seção 2, é apresentada uma introdução à Web Semântica, mostrando os problemas encontrados na *Web* hoje, as soluções propostas para estes problemas, usando-se a infra-estrutura da Web Semântica e a sua arquitetura em camadas. Na Seção 3, é descrita a sintaxe usada na Web Semântica, apresentando as tecnologias XML e RDF. Na Seção 4, é apresentada uma das camadas mais importantes da Web Semântica, a camada da Ontologia. Na Seção 5, são apresentadas as camadas da Prova e da Validação, duas camadas que ainda estão em fase de estudos. Nesta seção, também se discute, a importância do uso dos agentes de software na Web Semântica. Na Seção 6, é discutido o poder das linguagens da Web Semântica, enfocando que estas linguagens devem ser poderosas, mas simples. Na Seção 7, é discutido como a Web Semântica está evoluindo, ou seja, em quais aspectos ela vai bem e em quais ela precisa melhorar. Na Seção 8, são apresentados alguns trabalhos que estão sendo feitos na tentativa da difusão da Web Semântica. Finalmente, na Seção 9, são apresentadas algumas considerações finais.

## 2 A Web Semântica

Com o aperfeiçoamento da Internet, procura-se fazer com que os computadores ligados à rede cumpram tarefas mais complexas na busca de informações. Um dos desafios é que eles compreendam e vinculem dados colocados na rede com semântica, isto é, os computadores, através de ontologias e regras de inferências, encontrem respostas mais precisas para as consultas, descartando o que é irrelevante para o usuário.

O objetivo da introdução de semântica na *Web* é tornar a informação “compreensível” para o computador. Ela representa uma revolução no processamento da informação e, por consequência, uma revolução na maneira de se obter e organizar os conhecimentos. Pretende-se que os recursos disponibilizados sejam expressivos o bastante para que máquinas e/ou agentes de software sejam capazes de processar e “entender” o real significado dos dados [5].

“A Web Semântica não é uma *Web* separada, mas uma extensão da atual. Nela, a informação é dada com um significado bem definido, permitindo melhor interação entre os computadores e as pessoas” [3], diz Berners-Lee, idealizador da WWW, URIs, HTTP e HTML. Ele ainda diz que, “Adequadamente projetada, a Web Semântica pode auxiliar a evolução do conhecimento humano como um todo”. Já Mark Frauenfelder, colaborador de diversas publicações sobre Internet, explica que “acrescentar semântica modificará radicalmente a natureza da rede, de um lugar onde as informações são meramente oferecidas, para outro onde elas são interpretadas, intercambiadas e processadas” [3].

Existe hoje um dedicado grupo de pessoas do W3C trabalhando para melhorar, estender e padronizar sistemas e, além disso, muitas linguagens, publicações e ferramentas já foram desenvolvidas. Todavia, as tecnologias da Web Semântica estão ainda na infância e, embora o futuro deste projeto pareça ser brilhante, ainda há pouco consenso sobre as prováveis direções e características neste começo da Web Semântica [6].

Nas Subseções 2.1 e 2.2, a seguir, são mostrados os problemas encontrados na *Web* hoje, as soluções propostas para estes problemas usando-se a infra-estrutura da Web Semântica e a sua arquitetura em camadas.

## 2.1 Problemas e Soluções

Para a Web Semântica funcionar, computadores deverão ter acesso à coleções de informações estruturadas e conjuntos de regras de inferências que poderão ser usados para conduzir a um raciocínio automático. Recursos de Inteligência Artificial têm sido utilizados na Representação de Conhecimento mesmo antes da Web Semântica. Esta contém a semente de importantes aplicações a serem amplamente utilizadas.

Os sistemas tradicionais de Representação de Conhecimento costumam ser centralizados, ou seja, requerem que todos os sistemas, que estão interligados, possuam as mesmas definições sobre os mesmos conceitos. Entretanto, este controle central é uma asfixia, pois o incremento do tamanho e do escopo destes sistemas, rapidamente os tornam impossíveis de serem gerenciados. Além disso, estes sistemas limitam as questões que podem ser consultadas e isso não é bom.

Pesquisas na Web Semântica, em contraste, admitem que contradições e questões sem respostas são o preço que deve ser pago para se alcançar a versatilidade [2]. As linguagens para as regras deverão ser tão expressivas quanto possíveis para permitir que se possa raciocinar sobre a *Web* tão amplamente quanto necessário.

A Web Semântica pode ser vista então como sendo uma solução prodigiosa para resolver os problemas da *Web* atual, mas ela é mais do que isto. Imagina-se que ela facilitará a publicação de dados em vários formatos. Então, mais pessoas irão querer publicar seus dados com semântica gerando-se um efeito dominó [6]. Com isto, imagina-se que um grande número de aplicações da Web Semântica possam ser desenvolvidas.

## 2.2 As Camadas da Web Semântica

Os princípios da Web Semântica são implementados em camadas de tecnologias e padrões *Web*. A Figura 1 apresenta estas camadas. As camadas Unicode e URI estabelecem um conjunto de caracteres internacionais e fornece meios para a identificação de objetos na Web Semântica. A camada XML, com definição de *namespace* e *schema*, estabelece que se pode integrar definições da Web Semântica com outros padrões baseados em XML. A camada RDF e RDF Schema possibilita estabelecer declarações sobre objetos com URIs e definir vocabulários que podem ser atribuídos por URIs. Esta é a camada onde se tem a capacidade de oferecer tipos para recursos e *links*.

A camada da Ontologia suporta a evolução de vocabulários assim como pode definir relações entre conceitos diferentes. Já a camada da Assinatura Digital detecta as alterações em documentos. Esta é a camada que está sendo progressivamente padronizada por grupos de trabalho do W3C.

As camadas da Lógica, da Prova e da Validação, estão sendo progressivamente pesquisadas e aplicações demonstrativas simples já estão sendo construídas. A camada da Lógica possibilita a escrita de regras. A camada da Prova executa estas regras e avalia, juntamente com a camada da Validação, os mecanismos que permitem às aplicações confiar ou não nas provas realizadas.

## 3 A Sintaxe

Duas importantes tecnologias de desenvolvimento da Web Semântica são XML e RDF, já citadas na Seção 1. A linguagem XML permite que qualquer um possa criar suas próprias *tags*. Os *scripts* ou programas podem fazer uso destas *tags* segundo as suas necessidades. Em outras

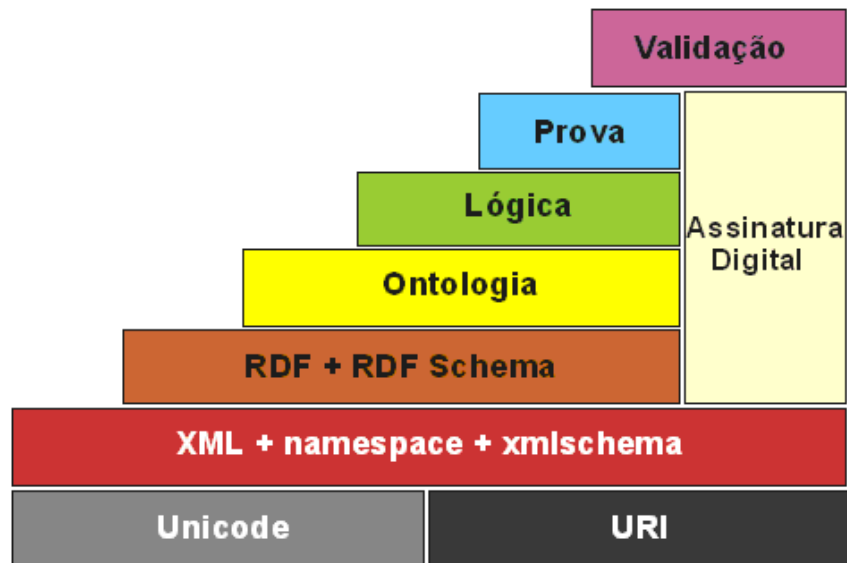


Figura 1: A visão da Web Semântica pelo W3C. A arquitetura da Web Semântica é dividida em várias camadas [8]

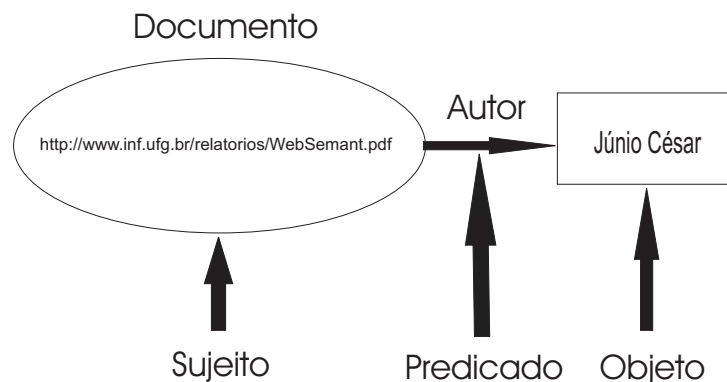


Figura 2: Exemplo de uma representação de um documento RDF

palavras, XML permite que usuários adicionem uma estrutura arbitrária em seus documentos mas sem nada a dizer sobre o que a estrutura significa.

A sintaxe da Web Semântica é feita geralmente com o uso de URIs (*Universal Resource Identifier*) para representar os dados. Os URIs permitem de identificação única de recursos, bem como os relacionamentos entre eles. Usualmente, os dados em URIs são representados por estruturas baseadas em triplas. Muitas triplas de dados URIs podem ser mantidas em bancos de dados ou trocadas na WWW usando a arquitetura RDF. Portanto, o significado é expresso em RDF, que é codificado em um conjunto de triplas. Cada tripla é formada por sujeito, predicado e objeto de um elemento de uma sentença [2]. Estas triplas podem ser escritas usando-se *tags* XML.

A Figura 2 possui a representação de um documento em RDF da declaração “Júnio César é o autor do recurso <http://www.inf.ufg.br/relatorios/WebSemant.pdf>”.

O diagrama da Figura 2, utilizando-se *tags* XML, pode ser representado da forma mostrada na Figura 3.

```
</xml version= 1.0"?>
<rdf:RDF
  xmlns:rdf= http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc= http://purl.org/dc/elements/1.1">
  <rdf:Description about = http://www.inf.ufg.br/relatorios/WebSemant.pdf"
    <dc:Creator>Júnio César</dc: Creator>
  </rdf:Description>
</rdf:RDF>
```

Figura 3: Exemplo de uma representação de um documento RDF em XML.

Em RDF, um documento faz afirmações sobre coisas particulares (pessoas, *Web sites*) tendo propriedades (tal como: “é mãe de” ou “é autor de”) com certos valores (outra pessoa, outro *Web site*). Esta estrutura é uma forma natural de descrever dados a serem processados por máquinas.

Um URI permite a identificação única de recursos, bem como os relacionamentos entre eles. Qualquer um pode criar um URI e tornar-se seu proprietário. Logo se forma uma base tecnológica ideal para construção de uma *Web* global. Os URIs são componentes fundamentais na *Web* atual e são a base da Web Semântica.

Sean B. Palmer, em seu artigo [6], explica dois motivos pelos quais se usa RDF. Primeiramente, o benefício de se conseguir projetar uma linguagem em RDF é que as informações são mapeadas diretamente e sem ambigüidade para um modelo, sendo que este modelo é descentralizado e para o qual há muitos analisadores disponíveis. O segundo motivo, é que se imagina que os dados RDF se tornarão parte da Web Semântica.

### 3.1 Esquema RDF (*RDF Schema*)

Um esquema (*schema*) é simplesmente um documento ou pedaço de código que controla um conjunto de termos em outros documentos ou em pedaços de códigos.

O esquema RDF foi projetado para ser um modelo de tipo de dados simples para RDF. Por exemplo, usando-se o esquema RDF, pode-se dizer que “Tupi” é um tipo de “Cachorro” e que “Cachorro” é uma subclasse de “Animal”.

A seguir, na Seção 4, é apresentada uma das camadas mais importantes da Web Semântica, a camada da Ontologia.

## 4 As Ontologias

Quando se imagina que existam dois bancos de dados que possuem diferentes identidades para um mesmo conceito e que se tem um programa que queira comparar ou combinar informações através destes dois bancos de dados, este programa deverá saber que existem dois termos sendo utilizados com o mesmo significado. Idealmente, o programa deve ter um caminho para descobrir os significados destes termos quando for manipulá-los. Uma solução para este problema é fornecida por um dos componentes básicos da Web Semântica, uma coleção de informações chamada de ontologia.

Na filosofia, uma ontologia é uma teoria sobre a existência da natureza, sobre que tipos de coisas existem. Os pesquisadores da inteligência artificial e da *Web* converteram este termo

para o seu próprio jargão, sendo que, para eles, ontologia é um documento ou arquivo que formalmente define as relações entre termos.

Com páginas na *Web* usando ontologias, soluções para problemas de terminologia começam a emergir. O significado de termos ou códigos XML usados nas páginas *Web*, podem ser identificados através do uso de uma ontologia. Naturalmente, alguns problemas agora surgem. Por exemplo, se uma aplicação X aponta para uma ontologia que define uma loja que vende “carros” e uma aplicação Y aponta para uma loja que vende “veículos”, logo se percebe o problema. Este problema é bastante natural para o ser humano e difícil para uma máquina. Este tipo de confusão pode ser resolvido se as ontologias proverem relações de equivalência, ou seja, se uma ou as duas ontologias, contiverem informações dizendo que “carro” da aplicação X é equivalente ao “veículo” da aplicação Y.

O computador não entende verdadeiramente essas informações, mas ele pode agora manipular os termos mais eficientemente em caminhos que são úteis e significativos para os usuários humanos.

As ontologias podem melhorar e elevar as funcionalidades da *Web*. Elas podem ser usadas em uma aplicação na tentativa de melhorar a exatidão de buscas na *Web* - o programa de busca somente retornará as páginas relevantes para o usuário, ao invés de retornar também as páginas com conteúdo irrelevante. Algumas das aplicações mais avançadas no futuro usarão ontologias para inferir novos conhecimentos, elevando as funcionalidades da *Web*.

Nas Subseções 4.1, 4.2 e 4.3, a seguir, são discutidas a lógica empregada na camada da Ontologia e as linguagens DAML e OWL, linguagens usada para representar ontologias.

## 4.1 Lógica

Por meio de um processo de inferência, é possível derivar novos dados a partir de dados que já são conhecidos. A inferência é um dos princípios que move a Web Semântica.

A adição de lógica na *Web* (isto é, o uso de regras para inferências, escolha de métodos de ações e respostas a perguntas) é um dos principais desafios da comunidade da Web Semântica no momento e é essencial para a sua consolidação. A mistura de matemática e máquinas de decisão complicam esta tarefa. A lógica deve ser poderosa o bastante para descrever as propriedades complexas de objetos mas não tão poderosa que os agentes de software não consigam buscar as informações requeridas ou mesmo serem enganados quando analisarem uma contradição.

Para a Web Semântica tornar-se expressiva o bastante para poder fazer inferências em uma ampla série de situações, será necessário uma poderosa linguagem lógica para fazer inferências.

## 4.2 DAML

A DAML (*DARPA Agent Markup Language*) é uma iniciativa da DARPA (*Defense Advanced Research Projects Agency*) que está sendo desenvolvida como uma linguagem de ontologia e inferência baseada em RDF.

A mais recente iniciativa da DARPA é oriunda da combinação de DAML e OIL, denominada DAML+OIL, uma linguagem proposta como padrão para representação de ontologias e metadados pelo W3C.

Têm sido feitos grandes investimentos em pesquisas nesta área. Somente as pesquisas sobre a DAML receberam, ultimamente, investimentos na ordem de 70 milhões de dólares [9].



### 4.3 OWL

A OWL (*Web Ontology Language*) foi projetada para ser usada por aplicações que necessitem processar o conteúdo de informações, ao invés de somente apresentar a visualização destas informações. Ela pode ser usada para explicitamente representar o significado de termos em vocabulários e os relacionamentos entre estes termos. A OWL tem mais facilidades para expressar significados e semânticas do que XML, RDF e RDF Schema. A OWL é uma revisão da linguagem DAML+OIL, incorporando as lições aprendidas de projetos e aplicações da DAML+OIL [7]. A OWL, recentemente, passou a fazer parte das recomendações do W3C para a Web Semântica. A OWL possui três sub-linguagens incrementais projetadas para serem usadas por diferentes comunidades de implementadores e usuários[7]:

- OWL Lite: usa somente algumas das características da linguagem OWL e possui mais limitações do que OWL DL ou OWL Full.
- OWL DL: é usada por usuários que queiram o máximo de expressividade, com completude (todas as conclusões são garantidas para serem computáveis) e decidibilidade (todas as computações terminarão em um tempo finito) computacional. Ela inclui todas as construções da linguagem OWL, mas estas construções somente podem ser usadas sob certas restrições.
- OWL Full: é usada por usuários que queiram o máximo de expressividade e independência sintática de RDF sem nenhuma garantia computacional.

A seguir, na Seção 5, são apresentadas as camadas da Prova e da Validação, duas camadas que ainda estão em fase de estudos. Nesta seção, também se discute, a importância do uso dos agentes de software na Web Semântica.

## 5 A Camada da Prova e da Validação

Sabe-se que nenhuma informação encontrada na *Web* hoje pode ser dita absolutamente verdadeira sem primeiro se fazer uma análise destas informações. A Web Semântica não vai mudar este fato. A análise da veracidade das informações vai evoluir para que as aplicações as processem e decidam se confiam nestas informações ou não. As camadas da arquitetura da Web Semântica que tratam disto são camada da prova e a camada da validação.

Há muito pouco escrito sobre estas camadas, apesar de que elas são muito importantes para a Web Semântica. Um exemplo da importância destas camadas: se uma aplicação diz que X é vermelho e outra aplicação diz que X não é vermelho, logo toda a Web Semântica poderia cair em pedaços. Mas, por dois motivos isto não ocorre: a) as aplicações na Web Semântica, no momento, dependem do contexto; b) e as aplicações, no futuro, contarão com mecanismos de conferências de provas e assinaturas digitais.

### 5.1 Contexto e Assinatura Digitais

As aplicações na Web Semântica dependem do contexto, ou seja, para as aplicações confiarem nos dados que elas recebem, o contexto é importante. Por exemplo, quando uma pessoa recebe dados em RDF informando que um amigo assistiu um determinado filme e o avaliou bem, esta pessoa poderá confiar que aquele filme realmente deve ser bom. Um outro exemplo

recai para grupos de pessoas que operam em contextos compartilhados. Se um grupo está desenvolvendo um portal usando a infra-estrutura da Web Semântica para serviços de artes, então a confiança no grupo dependerá de como se confia nas pessoas do grupo.

Portanto, o contexto é importante uma vez que ele deixa as pessoas operarem intuitivamente em escalas locais, sem ter que confiar nas validações de autenticações complexas e de sistemas de conferência. Todavia, acontecem os casos em que se têm dados RDF de uma pessoa que se conhece e confia, mas que não se tem certeza que estes dados venham realmente desta pessoa. Aqui se aplica a assinatura digital.

A assinatura digital é um código que é incluído em uma mensagem ou texto, para identificar o remetente da mensagem. A assinatura digital também é aplicada para dados RDF. As pessoas vão criar seus documentos, como sendo dados RDF, e enviá-los a um destinatário usando a assinatura digital.

## 5.2 Agentes de Software

O real poder da Web Semântica só poderá ser constatado quando forem criadas muitas aplicações que sejam capazes de obter informações, oriundas de diversas fontes na *Web*, de processar estas informações e de trocar os resultados deste processamento com outras aplicações. A Web Semântica permitirá que até os agentes, os quais não foram propriamente projetados para trabalharem juntos, possam transferir dados entre si, quando a esses dados, estiver associada alguma semântica.

Uma importante face das funcionalidades dos agentes de software será a troca de prova escrita na unificação das linguagens da Web Semântica. Quando agentes de software encontrarem uma nova regra, eles trocarão esta regra entre si na tentativa de provar que esta regra é válida. Um outro aspecto vital será a assinatura digital (ver a Seção 5.1). Computadores e agentes de software poderão usar a assinatura digital para comprovar que as informações anexadas são de uma fonte específica e de confiança.

Os agentes deverão ser cuidadosos com as afirmações que eles conduzirão na *Web* até que eles chequem a fonte das informações.

Muitos serviços automáticos baseados na *Web* já existem sem semântica. Alguns destes serviços automáticos, chamados de descoberta de serviço (*trading*), somente são utilizados quando existe uma linguagem comum, compartilhada e entendida por todos que a utilizam. Esta linguagem é usada para descrever um serviço, em um caminho que permita que os agentes de software descubram quais as funções oferecidas e como utilizá-las. Mas, estes agentes de software não conseguem “raciocinar” e descobrir novos serviços, eles somente trabalham com funções pré-definidas. Os serviços e os agentes de software podem informar suas funções, por exemplo, depositando-as em um diretório de descrição, analogamente como ocorre nas “Páginas Amarelas” das listas telefônicas.

A Web Semântica, em contraste, é mais flexível. Os agentes consumidores e produtores podem obter um entendimento comum pelo compartilhamento de ontologias, que fornecem o vocabulário necessário para a troca de informações. Quando os agentes descobrirem novas ontologias, eles poderão desenvolver novas capacidades de raciocínio. Será um processo típico que envolverá a criação de uma corrente de valores em que pedaços de informações são passadas de um agente para outro, sendo que cada um adiciona um novo valor para construção do produto final requisitado pelo usuário.

### 5.3 Linguagens de Prova

Uma linguagem de prova é simplesmente uma linguagem que permite provar que uma declaração é ou não verdadeira.

Uma instância de uma linguagem de prova geralmente consiste de uma lista de itens de inferências que são usados para derivar as informações em questão e a validação das informações para cada um dos itens que podem ser conferidos.

Na Seção 6, a seguir, é discutido o poder das linguagens da Web Semântica.

## 6 O Poder das Linguagens da Web Semântica

O principal poder das linguagens da Web Semântica é que qualquer pessoa pode criar uma, simplesmente por uma publicação RDF que descreve um conjunto de URIs, o que elas fazem e como devem ser usadas.

### 6.1 O Princípio do Menor Poder

A Web Semântica trabalha com o princípio do menor poder: menos regras, melhor. Isto tendo em vista que ela permite que se diga qualquer coisa sobre qualquer coisa.

Este princípio torna-se óbvio e necessário. Se a Web Semântica possuir muitas regras, será mais difícil a construção um amplo leque de aplicações. Desta forma, ela se tornará pouco útil.

As aplicações que serão construídas com a infra-estrutura da Web Semântica vão ser projetadas para executar tarefas específicas e como tal, muito bem definidas.

Uma das vantagens que se consegue publicando informações em formato RDF é que, uma vez publicadas em um domínio público, estas informações poderão ser usadas com outros objetivos, com muita facilidade [6].

A seguir, na Seção 7, é discutido como a Web Semântica está evoluindo, ou seja, em quais aspectos ela vai bem e em quais ela precisa melhorar.

## 7 A Evolução

Embora a Web Semântica completa esteja ainda distante, algumas aplicações já foram desenvolvidas dentro deste novo contexto. Estas aplicações estão começando a publicar informações usando o formato RDF, e através disto, fazendo com que os próprios desenvolvedores das aplicações se adaptem ao modelo da Web Semântica.

Segundo Palmer [6], não tem sido feito o bastante para ligar informações. Em outras palavras, a parte “Semântica” da Web Semântica está indo bem, mas a parte “Web” não. As aplicações não estão usando as ontologias de outras aplicações efetivamente. Ainda segundo Palmer, quando as aplicações usam termos de determinadas ontologias, como por exemplo um termo que se refere ao código postal *zip*, elas freqüentemente o fazem porque estão tentando ajudar no desenvolvimento da Web Semântica. Mas, estas aplicações estão justamente gerando um erro neste processo. Se um desenvolvedor de uma aplicação quer usar dados de outras aplicações, deve tentar primeiro encontrar quais as vantagens disto.

Uma outra parte do problema pode ser devido a um problema similar ocorrido no começo da WWW: porque se preocupar em publicar um sítio *Web* quando não há outros sítios para ligar? Porque se preocupar em publicar sítios *Web* quando poucas pessoas possuem *browsers*? Porque

se preocupar em escrever um *browser* quando há poucos sítios *Web*? Algumas pessoas têm que dar saltos para que as coisas ocorram, e isso é certamente um processo lento, mas necessário.

Um princípio que se emprega muito bem para as aplicações da Web Semântica é o uso de esquemas já definidos. Se uma determinada pessoa tem construído um esquema que é compreensivo e usa um conjunto de termos que outra pessoa também necessita para usar na sua aplicação, então não há porque refazer o trabalho que já foi feito. Em alguns pontos, isso pode levar a uma forma de “guerra de esquemas”, mas a sobrevivência dos melhores deve ser levada em consideração, isto é, os melhores esquemas serão mais utilizados. Isto é o que Berners-Lee [6] teve em mente quando disse que termos bons emergiriam, ou seja, as aplicações usariam o mesmo termo “zip” (por exemplo), ao invés de uma anotação que o seu termo “zip” é equivalente a outro termo “zip”, que também é equivalente a outro termo “zip”. Todos juntos usariam o mesmo URI do termo “zip”, portanto, a interoperabilidade seria melhorada.

A seguir, na Seção 8, são apresentados alguns trabalhos que estão sendo desenvolvidos na tentativa da difusão da Web Semântica.

## 8 Aplicações para Web Semântica

As aplicações para a Web Semântica estão sendo construídas, mas ainda em pequena escala. Palmer, em seu artigo [6], diz que há muitas maneiras das pessoas contribuírem para a criação da Web Semântica. Algumas são citadas a seguir:

- Publicando alguns dados globalmente úteis em RDF;
- Escrever uma máquina de inferência em uma linguagem qualquer;
- Difundir as palavras empregadas na Web Semântica;
- Ajudar no desenvolvimento de RDF Schema e/ou DAML;
- Contribuir na representação de estados em RDF, ao contrário de omitir campos de pesquisas;
- Aplicar os próprios conhecimentos de desenvolvimento para a Web Semântica, dando a todos novos ângulos para serem considerados;
- Em vez de usar algum sistema proprietário para alguma aplicação, fazer um projeto para a Web Semântica em seu lugar.

A seguir são mostrados alguns trabalhos que estão sendo desenvolvidos usando a Web Semântica ou mesmo para a disseminação da mesma.

### 8.1 *Pedantic Web*

Atualmente está sendo difundido pela comunidade da Web Semântica o termo *Pedantic Web* [6]. Este termo se refere a um documento que foi idealizado para ajudar a esclarecer algumas idéias erradas sobre a Web Semântica que estão sendo disseminadas [6].

Todos os tipos de compreensões erradas estão sendo colecionados em um guia. Este guia possui uma coleção de conselhos e sugestões sobre a Web Semântica para as pessoas poderem consultá-lo quando tiverem dúvidas.

Por exemplo, uma interpretação errônea da proposta da Web Semântica que poderia ser disponibilizada neste documento seria que a Web Semântica capacitará as máquinas a compreenderem documentos e dados com semântica e não compreender a fala e a escrita humana.

A proposta ao longo do documento é parte de um movimento para trazer, para as pessoas, o conhecimento sobre a Web Semântica.

### **8.1.1 A *Semantic Web Activity***

A *Semantic Web Activity* é parte do *W3C Technology and Society Domain*. O objetivo das suas atividades é de projetar tecnologias que ajudem as máquinas na troca de conhecimentos globais.

A *Semantic Web Activity* oferece um ambiente para colaborações e cooperações. Alguns grupos de trabalhos e alguns projetos são discutidos sucintamente a seguir.

### **8.1.2 *RDF Core Working Group (RDFCore)***

O padrão corrente do RDF provê um *framework* para representação de metadados através de muitas aplicações. O objetivo do RDFCore é elucidar e melhorar o modelo abstrato do RDF e a sintaxe XML, de acordo com as implementações originais [8].

Este grupo está patenteando a descrição completa do vocabulário RDF. Em suas publicações, esse grupo de trabalho também explica os relacionamentos entre os componentes básicos do RDF e das recomendações da grande família XML. Em exposições adicionais, incluem uma teoria semântica precisa destes padrões para suportarem trabalhos futuros, bem como uma cartilha fornecendo um entendimento básico de RDF e de suas aplicações.

### **8.1.3 *Web Ontology Working Group (WebOnt)***

O grupo de trabalho WebOnt busca padronizar os conhecimentos que podem ser usados para definição de ontologias na *Web* para descreverem os conceitos das estruturas [8]. A DAML+OIL é desenvolvida com a ajuda da *DARPA DAML Initiative* que está entrando neste grupo de trabalho. O WebOnt está construindo um esquema RDF (classes e subclasses, propriedades e subpropriedades) enquanto estende estas construções para permitir relacionamentos mais complexos entre as entidades.

### **8.1.4 *Semantic Web Advanced Development (SWAD)***

O SWAD é uma iniciativa do W3C que conta com a colaboração de um grande número de pesquisadores e empresas [1]. Ele estimula o desenvolvimento de infra-estruturas de componentes da Web Semântica. A seguir são discutidos alguns destes trabalhos.

- SWAD DAML - A proposta do projeto SWAD DAML é para contribuir com o desenvolvimento da Web Semântica, pela construção de uma boa infra-estrutura para a mesma e pela demonstração de como esta infra-estrutura pode ser usada para o trabalho e aplicações orientadas a usuários [1].

A SWAD DAML é projetada para construir uma infra-estrutura usando a linguagem DAML, provendo um intercâmbio entre duas ou mais aplicações diferentes. O seu principal objetivo é o compartilhamento de conteúdos semânticos.

- SWAD-Europe - O SWAD-Europe destacará exemplos práticos de onde valores reais podem ser adicionados para a *Web* através de tecnologias na Web Semântica [1].

- SWAD Símile - O W3C está trabalhando juntamente com a HP, a Biblioteca do MIT e o Laboratório de Ciência da Computação do MIT em Símile, buscando elevar a interoperabilidade entre propriedades, esquemas, metadados e serviços, através das distribuições individuais, para as comunidades e instituições e através de correntes de valores que provêm serviços
- SWAD Oxygen - O projeto MIT/LCS Oxigen está modelado para possibilitar a penetração da computação centrada no ser humano através da combinação de usuários específicos e sistemas tecnológicos [1]. O Oxigen usará tecnologias diretamente endereçadas às necessidades humanas.

## 8.2 O Caminhar da Web Semântica

As ciências modernas continuam em seu crescimento exponencial, em termos de complexidade e escopo. As necessidades para mais colaborações entre os cientistas de diferentes instituições, comunidades e países está se tornando cada vez mais importante. Os trabalhos em algumas pesquisas com certos níveis de análise podem necessitar de resultados de outros níveis, de outros campos, mesmo de campos científicos totalmente diferentes.

Na *Web*, todavia, isso não ocorre com muita frequência. Para a *Web* atingir seu potencial, os cientistas da informação devem moldar novos modelos de cooperações para a disseminação desta próxima geração de ferramentas científicas na *Web*. Esta nova geração de tecnologias para a *Web* é a Web Semântica.

Recentes *workshops* têm focado o uso da Web Semântica com suporte para a biociência e para a ciência ambiental [4]. Outros exemplos incluem o *National Virtual Observatory*, que está explorando o uso das tecnologias da Web Semântica para ligar numerosos recursos astronômicos, e também há o projeto da British MONET, que usa a Web Semântica para a criação de algoritmos matemáticos acessíveis na *Web* para uma variedade de pacotes de software.

Infelizmente, muitos cientistas estão desatentos ao esforço da Web Semântica. As principais correntes de desenvolvimento estão caminhando separadamente desta empreitada científica. Esta é uma situação paralela do desenvolvimento da *Web* original, onde os cientistas serviam como clientes e usuários das tecnologias *Web*, ao contrário de tentar expandir as tecnologias somente em direção às necessidades de seus campos de trabalho [4]. De fato, muitos dos recursos financeiros das aplicações em tecnologia da informação para a ciência têm ido para tecnologias que não podem competir com a *Web*.

A iniciativa da *e-Science* na Representação de Conhecimento é um bom exemplo de como os recursos de cientistas podem trabalhar juntos para o melhoramento da ciência. Hendler em seu artigo [4], diz que os cientistas ao redor do mundo deveriam se unir com seus colegas nos campos da Ciência da Computação e da Tecnologia da Informação intensificando as similaridades de seus programas de pesquisa. Além disso, estes cientistas deveriam unir seus trabalhos na tentativa de fazer com que as tecnologias da Web Semântica sejam incluídas em programas como o *U.S. National Science Foundation's* ou o *National Center for Research Resources's BioInformatics Research Network* nos Estados Unidos ou ainda outros programas internacionais com infra-estrutura científica similar [4].

O sucesso da Web Semântica não será completo se o conteúdo e as ferramentas não forem amplamente compartilhados, pelo menos no começo da exploração da Web Semântica. A revolução da WWW original foi propiciada pelo código aberto, software livre, e a ampla disseminação de tecnologias de computação de baixo custo. A Web Semântica requer essa mesma atitude de códigos abertos e disseminados, para que ela realmente venha revolucionar o conhecimento humano.

## 9 Considerações Finais

A Web Semântica é hoje mais do que um avanço da *Web* atual, ela é uma necessidade para que a *Web* possa continuar exercendo seu papel fundamental na vida de todos.

Quando a Web Semântica estiver em seu estado completo, ela melhorará a interoperabilidade de bancos de dados, provendo ferramentas para a interação com coleções de multimídias e fornecendo novos mecanismos para suporte aos agentes de software, onde pessoas e máquinas trabalharão com mais interatividade. Os agentes de software poderão usar os seus conhecimentos para buscar as informações requisitadas, filtrar informações irrelevantes e estudar estas informações, na tentativa de recuperar os dados exatos requisitados pelos usuários, e ainda tentar inferir novos conhecimentos.

De acordo com especialistas, o funcionamento da Web Semântica terá um impacto extraordinário na compilação, estruturação e transmissão do conhecimento. Com ela estará disponível uma nova classe de enciclopédia inteligente que, graças aos processadores avançados, a soma total do conhecimento humano estará ao alcance de qualquer computador conectado à rede.

A Web Semântica está começando a se disseminar. Grupos de trabalhos estão sendo criados sob a anuência do W3C, como o *Semantic Web Activity* que oferece um ambiente para colaboração e cooperação sobre as tecnologias e padronizações da Web Semântica. Já se percebe que outras áreas científicas como a biociência, dentre outras, já começam a perceber o poder que a Web Semântica pode trazer para as mais variadas áreas do conhecimento.

Nas três camadas inferiores (veja a Figura 1), as tecnologias e padronizações já estão bem avançadas, existindo diversas aplicações já desenvolvidas para estas camadas. Já nas camadas superiores pouco foi desenvolvido até agora. É para elas que os cientistas estão começando voltar suas preocupações, uma vez que elas são fundamentais para o desenvolvimento por completo da Web Semântica. A ontologia é atualmente o grande enfoque das pesquisas e dos projetos desenvolvidos nesta área e é, com certeza, o grande passo a ser dado para a consolidação da Web Semântica.

Há ainda grandes desafios a serem ultrapassados até que a Web Semântica venha realmente a revolucionar o conhecimento humano. Um destes desafios é fazer com que os cientistas e pesquisadores de todo o mundo compartilhem recursos e informações em favor da Web Semântica. Observa-se que os pesquisadores estão preocupados somente com a expansão das tecnologias em seus campos de pesquisa específicos e não com uma expansão por todos os campos e áreas. Somente com o trabalho em conjunto entre os cientistas, compartilhando conteúdos e ferramentas, a Web Semântica começará a deixar de ser uma ficção para se tornar uma realidade que revolucionará a *Web*.

## 10 Agradecimento

A Profa. Dra. Ana Paula Laboissière Ambrósio, pela avaliação do presente texto e pelas sugestões feitas, as quais muito contribuíram para a melhoria do texto original.

## Referências

- [1] BERNERS-LEE, T. MILLER, E. **The Semantic Web Lifts Off.** [http://www.ercim.org/publication/Ercim\\\_News/enw51/berners-lee.html](http://www.ercim.org/publication/Ercim\_News/enw51/berners-lee.html), acessado em Março de 2004, 2002.
- [2] BERNERS-LEE, T. H; J. LASSILA, O. **The Semantic Web.** . Scientific American , May 2001.
- [3] DE OLIVEIRA, J. J. **Web Semântica.** <http://limbo.ime.usp.br/mac339/index.php/IcicWebSemantica>, acessado em Março 2004, 2003.
- [4] HENDLER, J. **Enhanced: Science and the Semantic Web.** <http://www.sciencemag.org/cgi/content/full/299/5606/520> , acessado em Março de 2004, 2003.
- [5] MOURA, A. **A Web Semântica: Fundamentos e Tecnologias.** <http://www.udabol.edu.bo/biblioteca/congresos/cicc/cicc2001/datos/Tutoriales/Tutorial4/T4.pdf>, acessado em Março 2004, 2003.
- [6] PALMER, S. **The Semantic Web: An Introduction.** <http://infomesh.net/2001/swintro/#whatIsSw>, acessado em Março 2004, 2001.
- [7] PEREIRA, G. **OWL Web Ontology Language Guide.** <http://www.w3.org/TR/2004/REC-owl-features-20040210/>, acessado em Março 2004, 2004.
- [8] W3C. <http://www.w3.org/2001/12/semweb-fin/w3csw>, acessado em Março de 2004, 2003.
- [9] WEB., W. **Sistema de Gerenciamento de Documentos e Coleções na Web (SGDC-W).** <http://wonderweb.semanticweb.org/index.shtml>, acessado em Março 2004, 2003.