

## Introdução a Web Semântica

**Veja neste artigo uma introdução a Web Semântica, indo desde seu histórico até sua implementação, procurando fazer analogias com os princípios de inteligência Artificial, mostrando os benefícios e as possibilidades do uso.**

---

### 1. Introdução

A Web como conhecemos está mudando. Atualmente, os principais meios de localização de informação são os sistemas de busca. Mas, da forma como os conhecemos, os resultados de busca passam a não ter certa confiabilidade, ou seja, nem sempre encontramos aquilo que realmente desejamos.



**Figura 1:** Buscadores web – principais meios de localização de informação

Partindo deste princípio que surge a [Web Semântica](#), tendo como base que a Web pode nos trazer diversas informações, onde tanto pessoas como computadores sejam capazes de entender esses dados. Neste paradigma, a Web passa a ser vista como um grande banco de dados.

Exemplificando: imagine-se em busca de um determinado produto. Com nossas atuais tecnologias, deveríamos acessar um sistema de e-commerce ou ainda um site de busca, localizar o produto e, através do sistema disponibilizado, realizar o cálculo do frete e verificar os prazos de entrega, pois bem, esta não é a forma mais prática de se realizar este procedimento. A Web Semântica trata as informações disponíveis na Web como informações mesmo, ou seja, com estes recursos se você procurasse o produto, teria em suas mãos não somente os locais disponíveis, mas também o valor do frete e prazos de entrega de forma rápida e ágil.

Desta forma a Web passa a ser encarada como uma grande base de dados. Com os atuais paradigmas, temos informações das mais variadas, mas que estão simplesmente inacessíveis e este é um dos fundamentos sugeridos por Tim Berners-Lee, um dos responsáveis pelo surgimento desta metodologia e/ou forma de a Web ser vista, onde a informação deve ser livre e fora de controle, onde, por exemplo, um servidor sequer deveria armazenar informações a cerca da navegação de seu usuário sem que o interesse expresso nesse tipo de controle.

Podemos encarar a Web Semântica, ou ainda Web 3.0, como sendo um assistente particular. A idéia é que esta grande rede, ou ainda seu browser, auxilie o usuário em seus problemas. Por exemplo: uma pessoa navega constantemente e seu navegador recolhe essas informações, ou seja, através de levantamento de dados o browser pode identificar as preferências e ainda as coisas das quais você não gosta. Em conseqüência imagine a situação onde o utilizador do sistema pode fazer perguntas ao seu próprio computador, como por exemplo, após uma coleta de dados satisfatória, o browser pode ser capaz de responder questões como “Onde vou almoçar hoje?”. Isto se torna possível a partir do momento em que o navegador identifica as preferências do usuário e ainda as suas preferências.

Outro problema relacionado é o não entendimento da pesquisa por parte dos sistemas de busca atuais. Ou seja, quando realizamos uma busca, o sistema retorna as informações não com base em um entendimento, mas sim com base em palavras chaves que podem ou não estarem relacionadas com o que desejamos. Por exemplo, ao digitarmos o termo “saturno” em um sistema de busca, ele nos retorna tanto informações a cerca do planeta Saturno ou ainda sobre a montadora de carros Saturno.

Através deste novo paradigma podemos fazer com que a rede responda as nossas dúvidas. Podemos imaginar a seguinte situação, onde você faz perguntas ao seu computador e através da Grande Rede pode-se obter a resposta. Um exemplo de pergunta que a Web Semântica visa responder é “Gostaria de ir visitar a África, com custo abaixo de R\$ 4.000” e com base nesta pergunta, a rede se encarrega de pesquisar e retornar somente as informações relevantes.

## **2. Histórico**

A idéia de [Web Semântica](#) surgiu em 2001, após a publicação de um artigo através da revista Scientific American denominado “Web Semântica: um novo formato de conteúdo para a Web que tem significado para computadores vai iniciar uma revolução de novas possibilidades.”. Este artigo foi elaborado por Tim Berners-Lee, James Hendler e Ora Lassila.

Desta lista de autores o primeiro, Tim Berners-Lee, é atual presidente da W3C (World Wide Web Consortium), professor do MIT no Laboratório de Inteligência Artificial e Ciência da Computação e é tido como um dos maiores gênios ainda vivos. James Hendler trabalha com Inteligência Artificial no Instituto Politécnico Rensselaer (Rensselaer Polytechnic Institute), nos Estados Unidos. Ora lassila é arquiteto e estrategista de tecnologia, trabalha atualmente na Nokia Services e é membro do Nokia's CEO Technology Council.

A Web Semântica tem como um de seus principais objetivos tornar a Web mais acessível em nível de informações. Esta idéia surge para facilitar a vida dos usuários, ou seja, em princípio não deve haver mudanças a nível visual nas páginas, mas sim na forma como o conteúdo está organizado e apresentado em termos de código.

Em outras palavras, a web Semântica é uma forma de os computadores entenderem o que desejamos. Com este ponto de vista, o problema passou (e passa) a ser criar uma tecnologia capaz de fazer com que, tanto homens quanto computadores, entendam o que se deseja e o principal, o significado das palavras e termos.

Segundo James Hendler, um dos criadores deste paradigma, no início de sua utilização, a Web Semântica será restrita a ilhas a “ilhas de conhecimento” onde um determinado problema em uma pequena porção de usuários irão se beneficiar de suas facilidades.

### **3. Implementações**

Mesmo em uma fase inicial, já existem diversos serviços que contemplam Web Semântica ou que ainda trazem recursos para os desenvolvedores e/ou usuários destes serviços.

Para a implantação de sistemas semânticos, da forma como atualmente está sendo feito, podemos ter dois tipos de abordagem: Bottom Up e Top Down.

Na primeira abordagem deve-se inserir Tags nos textos e dados. Isso faria com que fosse necessário um grande esforço para que as informações disponíveis fossem adaptadas a este novo conceito.

A abordagem Top Down também necessita de um grande esforço, pois, nesse conceito, faz-se necessário processar os dados já existentes e ainda criar as conexões e estabelecer o significado entre eles.

#### **3.1. Exemplos de recursos que utilizam Web Semântica**

Freebase.com – Base de dados aberta e gratuita a comunidade com a finalidade de interligar conteúdos de forma semântica.

Hakia.com – Visto como um dos principais concorrentes ao Google faz a análise do termo pesquisado antes de efetivamente efetuar a pesquisa.

Triplt.com – Sistema que permite ao usuário gerenciar e planejar suas viagens, interligando os mais diversos conteúdos como, por exemplo, criar suas viagens e interligar as informações com os serviços do Flickr, Google Calendar e Wikipédia.

Swotti.com – Sistema de busca semântico que visa facilitar ao usuário localizar seus produtos, com o diferencial de localizar avaliações do produto em sites e blogs.

Siri.com – Aplicativo disponível para o aparelho celular Iphone da Apple que tem por finalidade de interagir e se adaptar ao usuário de uma forma diferente, onde o aplicativo tenta se adaptar e aprender com as ações realizadas.

Muitos outros serviços estão em desenvolvimento ou em fase de testes, mas algo a ser salientado é que a Web Semântica não possui uma forma definitiva de utilização, mas, podemos ter em mente que a sua finalidade é tornar as redes e resultados mais inteligentes.

Alguns recursos podem facilitar esse tipo de implantação. Atualmente a internet dispõe de recursos novos e interessantes que são os widgets e mashups. Widgets são pequenos aplicativos que se utilizam de API's de serviços e podem ser incluídos em sites e facilitar a vida dos internautas. Um mashup é a união de dois ou mais aplicativos, ou seja, quando se combinam esses recursos passamos a ter um mashup. Este último é visto como sendo um forte mecanismo de apoio a web semântica

## **4. Implementando Web Semântica**

A implementação de recursos semânticos ainda não obedece a uma estrutura definida. Como não se sabe o que o futuro reserva ainda não se sabe qual será a abordagem utilizada para Web Semântica, visto que, alguns especialistas cogitam uma adaptação de recursos existentes, como por exemplo, utilização de Tags em documentos HTML, em contra partida, outros especialistas alegam a necessita de de uma nova tecnologia que começaria do básico, mas é importante saber que uma abordagem definitiva ainda não foi definida.

### **4.1. Tecnologias e Definições**

Muitas tecnologias têm tido um diferencial considerável neste início de mudança. Como já vimos, algumas já existem, no caso, mashups e códigos HTML, algumas outras definições e tecnologias são pouco conhecidas, como veremos nos próximos tópicos.

#### **4.1.2. Ontologia**

Pode ser definido, em Ciência da Computação, como sendo um modelo de dados, que visa representar um conjunto de conceitos em um determinado domínio e como estes conceitos se relacionam. [Ontologias](#) são amplamente utilizadas em tecnologias de inteligência artificial como uma das formas de representar conhecimento sobre o mundo ou alguma parte deste.

Como se pode ver, Ontologias têm por finalidade se referir a conceitos e termos que descrevem alguma área do conhecimento. Essa descrição é feita utilizando-se dos seguintes elementos:

- Indivíduos: os objetos em sua visão básica;
- Classes: conjuntos, coleções ou ainda os tipos de indivíduos;
- Atributos: podem ser entendidos como sendo os as propriedades dos indivíduos;
- Relacionamentos: é a forma como os indivíduos se relacionam.

#### **4.1.2.1 Ontologia e Inteligência Artificial**

Para que possamos recuperar de forma correta uma informação, podemos lançar mão dos conceitos de revocação e precisão. A revocação é a capacidade de um sistema de localizar o máximo possível de registros com base em uma consulta, como por exemplo, se efetuarmos uma busca através de um sistema operacional, a revocação seria o máximo de arquivos que esta busca efetuada através sistema operacional pode nos trazer. Em paralelo temos a precisão, que nada mais é que a capacidade de esta busca retornar assuntos que realmente importam e tem relevância.

Através das idéias expostas, pode-se perceber que a precisão não é tão simples. Isto se deve ao fato de ela ter a necessidade de trazer os itens de acordo com o que realmente é buscado, ou seja, é necessário que o sistema possua certo grau de inteligência, em conseqüência adentramos a área de inteligência artificial.

Em termos de Web Semântica, podemos ter a interligação das páginas utilizando ontologias. Estes elementos definem o significado exato do conteúdo da página, e ainda passa a ser possível uma conexão lógica entre um ou mais termos.

O principal problema das ontologias passa a ser a sua criação automatizada. Em uma página que possui o termo “manga rosa”, por exemplo, passa a ser complicado determinar em qual contexto isso se aplicado, ou seja, se faz referência a roupa ou a fruta. Atualmente podem-se criar as ontologias manualmente, isso garante uma precisão bastante grande, pois independe de sistemas informatizados, que, atualmente, não são muito precisos.

Para auxiliar nesta tarefa estão surgindo tecnologias, como no caso o Processamento Probabilístico da Linguagem. Esta tecnologia faz a análise de documentos e cria estatísticas sobre os assuntos contidos no mesmo e aprende através de exemplos como classificar um documento. A partir desta análise ele sugere uma área de conhecimento para o assunto em questão, por exemplo, se num mesmo documento constam as palavras manga, solo e

plantação, passa a se tornar de baixa probabilidade que o termo manda fala referência a roupas e, portanto, faz mais sentido se encaixar em um contexto de fruta.

Outra área que têm despertado interesse é a abordagem probabilística para processamento de linguagem natural. Esta área tem atingido grande percentual de confiabilidade, Koller e Sahami em suas pesquisas conseguiram atingir o grau de 98,6% de precisão em sistemas de classificação de documentos.

#### 4.1.2. RDF (Resource Description Framework)

É uma linguagem para representar informações na internet. É uma tecnologia aprovada pela W3C e se utiliza de arquivos RDF que nada mais são que arquivos de dados, ou metadata, tendo como um dos principais objetivos criarem um modelo simples para armazenamento de informações. Em outras palavras é um Padrão da W3C para representar conhecimento, onde este pode ser decomposto em partes pequenas mas com um certo nível de relacionamento entre essas partes.

Os arquivos RDF são compostos por três elementos básicos:

- Recurso: Qualquer coisa que pode conter um URI, ou seja, é o campo que especifica o endereço ou cadeia de caracteres que identifica um recurso na internet;
- Propriedade: Armazena um recurso que possua um nome e possa ser utilizado como uma propriedade;
- Indicação: Consiste na união de um recurso, uma propriedade e um valor.

Segundo Joshua Tauberer em seu artigo What is RDF, o termo semântica na Web Semântica não significa que os computadores realmente serão capazes de entender o significado das palavras e termos, mas que o conhecimento poderá ser mecanicamente manipulado por computadores, visando trazer maior facilidade aos humanos.

#### Listagem 1: Exemplo de arquivo RDF apresentado por Tauberer em What is RDF

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ex="http://www.example.org/">
  <rdf:Description
rdf:about="http://www.example.org/vincent_donofrio">
    <ex:starred_in>
      <ex:tv_show
rdf:about="http://www.example.org/law_and_order_ci"/>
    </ex:starred_in>
  </rdf:Description>
  <rdf:Description
rdf:about="http://www.example.org/the_thirteenth_floor">
    <ex:similar_plot_as
rdf:resource="http://www.example.org/the_matrix"/>
  </rdf:Description>
</rdf:RDF>
```

Esse exemplo diz que o ator Vincent Donofrio apareceu no programa de TV Law & Order: Criminal Intent. Também diz que o filme The Thirteenth Floor tem um enredo similar ao de The Matrix.

Segundo Joshua, um padrão para representação distribuída de conhecimento, como por exemplo, o RDF, deve levar em consideração alguns fatores:

- O conhecimento é flexível: Nem todo conhecimento pode ser armazenado apenas em tabelas ou em estruturas hierárquicas, e para isso existem recursos como, por exemplo, o uso de mapas conceituais;
- É necessário que possam ser criadas relações entre as informações na Web Semântica: Para que isso aconteça faz-se necessário que os nomes dos sujeitos e objetos passem a ser convencionados.
- Serão necessários vocabulários específicos para domínios específicos: Alguns termos só fazem sentido se utilizados em um determinado contexto, ou seja, de alguma forma será necessário que estes termos sejam explicados para que possam ser efetivamente utilizados.

### 4.1.3. Representação Utilizando HTML

Uma das linguagens de tags mais utilizadas atualmente é a HTML. Praticamente quase todos os sites fazem uso desta linguagem. Mas, mesmo de forma básica e, por muitos, considerada antiga, ela pode nos trazer uma série de informações úteis a cerca de uma página. Por exemplo, a tag H1, comumente é utilizada para textos ou termos que necessitam de um maior destaque, em conseqüência este é um trecho que pode ser considerado mais importante dentro de uma página. A tag P, é utilizada para a construção de parágrafos e portanto nos diz que, se tiver uma tag H1 e em seguida uma tag P, pode ser que tenhamos um título e em seqüência sua definição, claro que isso não é obrigatório, mas já é um início.

Este tipo de abordagem é bastante útil, pois trata de uma tecnologia já existente que apenas precisa passar por uma padronização, e não há a necessidade de simplesmente refazer tudo que já existe e trazer novas tecnologias. Para isso, faz-se necessário alguns ajustes, vejamos um exemplo de HTML utilizável do ponto de vista semântico e um exemplo de HTML de difícil utilização.

#### 4.1.3.1. HTML utilizável

**Listagem 2:** Exemplo de código útil do ponto de vista semântico

```
<html>
<head><title>Web Semântica</title></head>
<h1>Título da página</h1>
<p>Texto descrevendo o título acima</p>
</html>
```

O código demonstrado na figura 1 facilita a identificação e localização dos termos como título e sua definição, ou seja, através de um código limpo e de

tags simples e utilização de classes podemos ter a localização e identificação de termos e seus significados.

#### **4.1.3.1. HTML não utilizável**

**Listagem 3:** Exemplo de código não útil do ponto de vista semântico

```
<html>
<head><title>Web Semântica</title></head>
<font size="14">Título da página</font>
<br><br>Texto descrevendo o título acima<br><br>
</html>
```

Através da figura 2 podemos visualizar um código sem muita padronização e de difícil identificação de seus elementos, visto que a tag font pode ser utilizada no meio de textos e a tag br não traz grandes informações sobre o motivo desta quebra de linha.