

Guia Rápido Web Semântica

Breno Mazieiro

Francisco Barbosa Junior

LIS - Laboratório de Inteligência em Saúde

<i>Introdução</i>	2
<i>Conceitos necessários para entender Web Semântica</i>	2
<i>Redes de computadores</i>	2
<i>Quero saber mais sobre redes de computadores e internet</i>	4
<i>HTML</i>	4
<i>Quero saber mais sobre HTML</i>	4
<i>CSS</i>	6
<i>Quero saber mais sobre CSS</i>	6
<i>XML</i>	7
<i>Quero saber mais sobre XML</i>	8
<i>URI xURL</i>	9
<i>Quero saber mais sobre URI e URL</i>	9
<i>Grafos</i>	10
<i>Quero saber mais sobre grafos</i>	10
<i>RDF</i>	11
<i>Quero saber mais sobre RDF</i>	11
<i>Sparql</i>	12
<i>Quero saber mais sobre RDF</i>	12
<i>Quero testar Query Sparql</i>	12
<i>Web Semântica</i>	13
<i>O que é Web Semântica?</i>	13
<i>Porque é que a Web semântica é importante?</i>	14
<i>O que podemos fazer com ela?</i>	14
<i>O que ela pode fazer para nós?</i>	14
<i>Como é que a Web semântica é importante na saúde pública?</i>	15
<i>Bancos de dados na Saúde</i>	15

Introdução

Esta é uma apostila-guia, feita exclusivamente para o minicurso que será dado sobre Web Semântica para a Secretaria da Saúde, ela tem como finalidade ser um guia rápido de conhecimentos básicos necessários para se entender a Web Semântica. Todos os assuntos neste guia serão tratados de forma rasa, mas com links para quem quiser se aprofundar em cada assunto. Todos os assuntos serão aprofundados e exemplificados durante o curso.

Conceitos necessários para entender Web Semântica.

A Web semântica se olhada sem cuidado, vai parecer um dragão de muitas cabeças, mas se tomarmos o cuidado para olhar cada uma dessas cabeças, perceberemos que ela não é tão difícil assim, que podemos com um conjunto de informações desvendar esse mistério, portanto este documento pretende, além de tocar nos pontos já tocados por outros documentos disponíveis sobre Web Semântica, ser mais leve, dar a possibilidade para quem o lê de chegar rapidamente ao “Por que tanto falam nisso?”.

Mas feliz, ou infelizmente, a Web Semântica é o produto final de uma série de coisas que foram desenvolvendo-se no tempo e, portanto, sem ter um raso conhecimento sobre elas não se pode entender, nem que seja de forma superficial, a razão pela qual as pessoas responsáveis por ela pretendem fazê-la existir.

Portanto nos subcapítulos a seguir pretende-se dar ao leitor a capacidade de conhecer essas várias coisas que dão suporte à existência da Web Semântica.

Redes de computadores

Pode parecer óbvio dizer que a Web Semântica precisa de uma rede para existir, uma vez que é necessário acesso a diferentes servidores e bancos de dados, mas sempre que pensamos em rede de computadores voltamos àquela ideia de algo gigantesco e mundial, muitas vezes nos esquecemos de que ela pode ser algo pequeno, entre os objetos que estão próximos a mim, como sua tv, notebook, geladeira, smartwatch, e gadgets afins. Você já ouviu falar sobre wifi offline, por exemplo? Pois é, mas pra gente entender vamos precisar de um contexto rápido, vamos a isso.

O medo que os governos tinham na Guerra-Fria era deixar de existir ou ter suas informações comprometidas, pois com a invenção da bomba atômica era possível que rapidamente uma grande extensão de terra fosse simplesmente dizimada do planeta, com isso, mesmo que se não soubesse onde ficava o banco de informações de guerra dos EUA, mas simplesmente sabendo que ele estava em algum lugar próximo a Washington, por exemplo, eu poderia enviar uma bomba que eliminasse tudo, inclusive aquele banco de informações.

Para que isso pudesse ser evitado, as universidades estadunidenses juntamente com o governo e o exército acabaram por criar a arpanet, ou seja, uma rede que interligava inicialmente o pentágono com universidades espalhadas pelos EUA, como mostra a figura 1, assim sendo se um canto do país fosse atacado automaticamente teriam cópias de tudo no outro lado do continente, tanto norte sul quanto leste oeste. Posteriormente quando essa rede cresceu as empresas começaram a ver utilidade na rede, então o governo separou a rede militar da comercial, dando inicio, então ao que conhecemos de rede mundial de computadores.

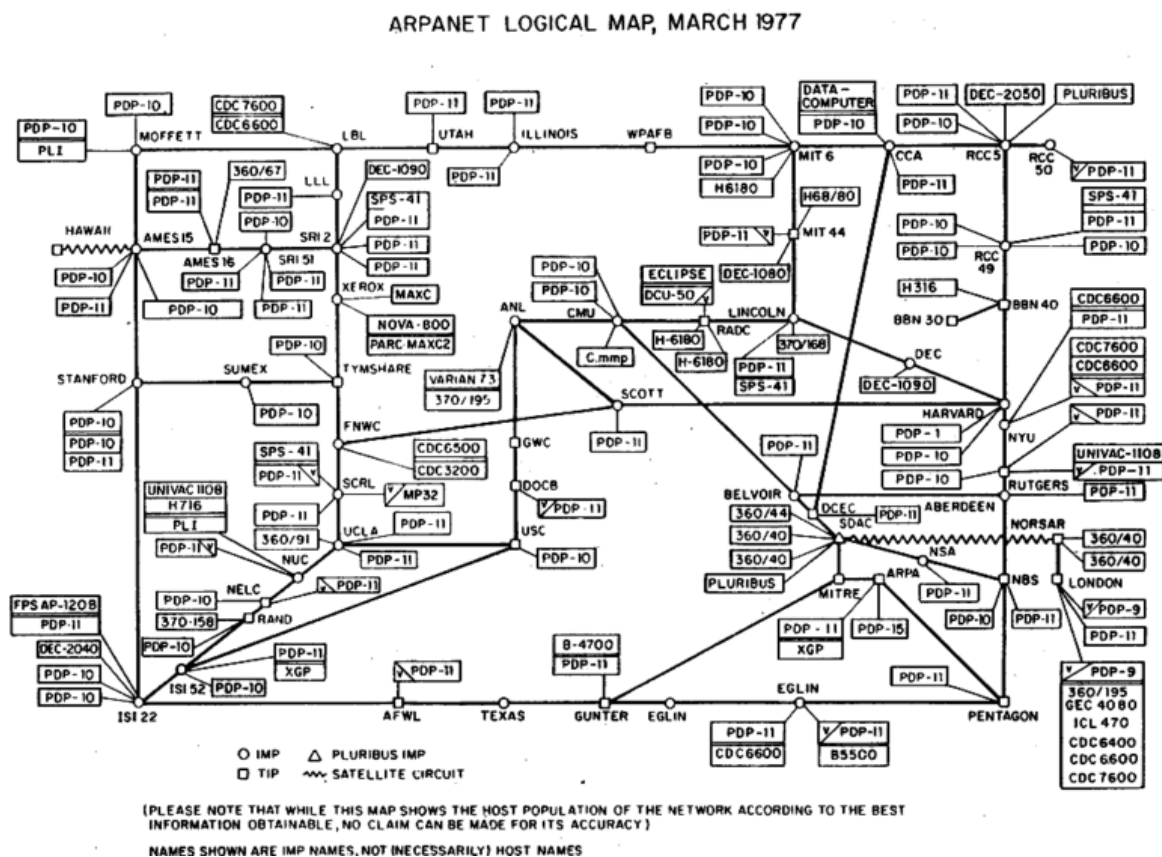


Figura 1: Mapa da Arpanet

Mas até meados de 1990, a única coisa que se podia fazer com a rede, em um contexto bem generalizado, era ter acesso a arquivos, quase como listar as pastas no seu computador atualmente, mas com o tempo as pessoas começaram a querer fazer mais com a rede, não apenas acessar coisas, mas elas queriam poder interagir com aquilo que estava lá e foi quando Tim Berners-Lee, e um jovem talento chamado Robert Cailliau do CERN, deram vida à HTML, que vamos discutir no próximo tópico.

Enfim, qual a importância da rede para a Web Semântica? Bem, a rede é o ponto de partida para Web Semântica, pois deixamos de ter sistemas que dependem de um banco ou informações locais e agora os sistemas começam a criar validações de outros sistemas que estão em outro lugar do mundo, agora existe uma ontologia para dizer para uma máquina o que é uma pessoa em toda a sua amplitude e mais ainda, podemos inclusive adicionar mais informação se pensarmos que algo falta, enfim, sem rede, sem Web Semântica.

Quero saber mais sobre redes de computadores e internet:

CASTELLS, Manuel. **The Internet galaxy: Reflections on the Internet, business, and society**. Oxford University Press, 2002.

W3C, History. Disponível em: <http://www.w3.org/History.html>

HTML

```
<a href='http://lis.fmrp.usp.br'> LIS </a>
```

Cria um link na palavra LIS, que quando clicado vai para o endereço <http://lis.fmrp.usp.br/>

```
<H1> Meu titulo </H1>
```

Faz com que a frase “Meu titulo” seja definida como um título.

Estes são apenas alguns exemplos baseados na marcação HTML, ou seja, que tem um início com uma palavra (tag) entre <> e termina com a mesma palavra (tag) adicionando-se um / no fim dela como no exemplo de <frase> </frase>.

Observe que isto é exatamente o que o criador da web fez no momento que criou a HTML, mas ele apenas criou tags específicas para apresentação de documentos que possam ser linkados, ou as famosas páginas web, para isso ele definiu uma linguagem de marcação de textos (ou hipertextos), a vontade inicial era poder fazer com que os documentos que existiam pudessem “apontar” para documentos em outros lugares. Imagine que você tenha escrito um texto sobre tuberculose e todas as vezes que alguém escrevesse a palavra tuberculose no texto dela poderia apontar para o meu texto, criando assim uma rede de arquivos linkados, ou seja, internet!

Essa primeira ideia de criar marcadores padronizados para a troca de informações foi se expandindo e criando novas formas, assim é o XML, por exemplo, que falaremos mais a frente.

Quero saber mais sobre HTML:

<http://www.w3schools.com/html/default.asp>

CSS

Para entender rapidamente o CSS imagine ele como sendo a roupa do HTML, ou seja, o HTML permanece sempre o mesmo, mas as cores e a cara que você dá pra ele podem ser trocadas automaticamente, ao trocar o CSS. Muito utilizado para que se possa apresentar o mesmo conteúdo para diferentes dispositivos.

Um exemplo é definir sua página HTML tendo um link assim:

```
<a href='http://lis.fmrp.usp.br/'> LIS </a>
```

A tag acima vai criar um link para o site do LIS na palavra LIS que está entre as tags.

O CSS vai ser algo assim:

```
a:hover {  
    color: #FF00FF;  
    text-decoration: underline;  
    background-color: #FF704D;  
}
```

O código acima vai fazer com que quando alguém passar o mouse sobre um link na sua página ele vai mudar e passar a ter a cor #FF00FF, com sublinhado e uma cor de fundo de #FF704D;

Assim sendo estamos melhorando a semântica do documento e assim começou a surgir as primeiras formas para se criar mais semânticas a documentos, um desses primeiros foi o microformats.

Quero saber mais sobre CSS:

<http://www.w3schools.com/css/default.asp>

<http://getbootstrap.com/>

XML

Metadados são Informações sobre as informações ou dados que descrevem outros dados e **se** esses metadados seguirem algum tipo de esquema eles podem tornar a buscar por informações na web mais efetivas.

Um desses esquemas que padronizam os metadados é o XML (*eXtensible Markup Language*), ele surgiu com a proposta de ser uma metalinguagem, de forma que cada pessoa define os tags que ela quer usar para definir seu próprio contexto.

“A aranha azul subiu pela parede.”

Imagine a frase acima e agora pense em como você faria para representar a frase acima de uma forma que uma máquina pudesse interpretar o que está escrito ali, bem, na verdade vamos ter milhões de saídas para isso, uma delas, por exemplo, é fazendo marcações e seria algo deste tipo:

```
<frase>
  <sujeito color='azul'> A aranha </sujeito>
  <verbo> subiu</verbo>
  <adverbio> pela </adverbio>
  <substantivo> parede </substantivo>
</frase>
<ação>
  <quem color='azul'>A aranha </quem>
  <fezOque> subiu </fezOque>
  <onde> pela parede </onde>
</ação>
```

Assim sendo, cada sujeito pode definir como representar os dados que ele tem de forma que passando a estrutura básica outras pessoas possam utilizar-se daquele conteúdo.

Vamos fazer uma definição nossa sobre pacientes da seguinte forma:

```
<xs:element name="paciente">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="name" type="xs:string"/>
      <xs:element name="birthDate" type="xs:date"/>
      <xs:element name="cpf" type="xs:string"/>
```



```
<xs:element name="size" type="xs:decimal"/>
<xs:element name="city" type="xs:string"/>
</xs:sequence>
</xs:complexType>
</xs:element>
</xs:schema>
```

Acima eu estou dizendo que o paciente terá (name, birthDate, cpf, size, city), com esse esquema eu posso validar o seguinte XML:

```
<paciente>
  <name> Breno Mazieiro </name>
  <bithDate> 01/01/2001 </birthDate>
  <cpf> 223.332.326.96 </cpf>
  <size> 1.90 </size>
  <city> Ribeirão Preto </city>
</paciente>
```

E quantos outros XMLs eu quiser definindo todos os pacientes que eu tiver. Isso faz com que apenas passando meu schema qualquer pessoa possa me passar dados, bem como eu posso receber dados padronizados dessa forma.

Mas ainda tem um problema, o XML não consegue me responder se Breno está perto de outro paciente qualquer apenas olhando a cidade e também não consigo “ligar” dois esquemas diferentes e é aí que a Web Semântica surgiu.

Quero saber mais sobre XML:

Schema: <http://www.w3schools.com/schema/default.asp>

XML: <http://www.w3schools.com/xml/default.asp>

URI xURL

O URI (*Uniform Resource Identifier*) pode ser encarado como um identificador único de algo no mundo, pode ainda ser classificado como um localizador, um nome, ou ambos. Já o termo "Uniform Resource Locator" (URL) refere-se ao subconjunto de URIs que, além de identificar uma fonte, proporcionam um meio de localizar o recurso através da descrição seu mecanismo de acesso primário (por exemplo, a sua rede de "local").

<http://www.facebook.com> é uma URL, pois está apontando para uma fonte na internet, nesta URL eu tenho acesso a uma serie de contas de rede social.

<https://www.facebook.com/bvmania> é um URI, pois identifica algo único, no caso, o usuário bvmania dentro da fonte de dados de informações sociais facebook.

Quero saber mais sobre URI e URL:

<http://www.ietf.org/rfc/rfc3986>

Grafos

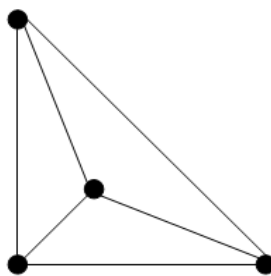


Figura 2: Um grafo com 4 vértices e 6 arestas

Um grafo é um conjunto de vértices (vamos chamar de nó) ligados por arestas, um grafo de apenas um vértice existe, mas não nos interessa para aprender Web Semântica.

Quando ligamos a teoria dos grafos ao conceito de URI, criamos a primeira noção de Web Semântica, ou seja, o vértice (nó) é único, assim sendo, se eu tenho a URI (<https://en.wikipedia.org/wiki/Tuberculosis>) e imaginar que todas as vezes que alguém quiser citar Tuberculose use essa URI como o ponto principal, por fim, ligando todos os bancos de dados com informações semânticas teremos algo parecido com isso:

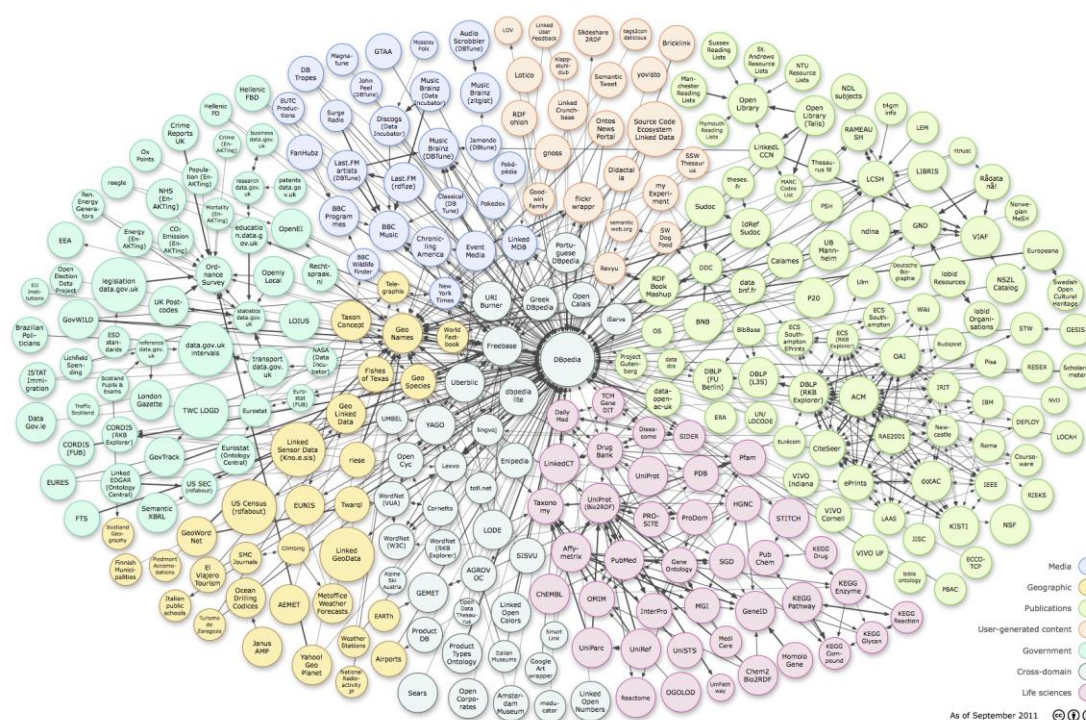


Figura 3: Bases de dados Semânticas Interligadas

Quero saber mais sobre grafos:

<http://www.ime.usp.br/~pf/teoriadosgrafos/texto/TeoriaDosGrafos.pdf>

RDF

O RDF é um **modelo de grafos que segue a regra das triplas com semântica explícitas** (cada nó é uma URI), ou seja, todo o conhecimento está ligado por:

Recurso: Aquilo que vai ser descrito, é um nó no meu grafo.

Propriedade: Propriedade daquilo que vai ser descrito, é uma aresta no meu grafo.

Valor: Valor daquilo que vai ser descrito, outro nó no meu grafo, que também pode vir a ser um recurso.

Tomamos como exemplo a seguinte droga:

<http://wifo5-04.informatik.uni-mannheim.de/sider/page/drugs/14888>

Vou chama-la apenas de 14888 e representar o nome dela Trisenox em um grafo:

14888 -> rdfs:label -> TRISENOX

Observe que neste grafo eu explicito que a droga que tem a URI (<http://wifo5-04.informatik.uni-mannheim.de/sider/page/drugs/14888>) tem um label, definido pelo próprio modelo RDF, o nome TRISENOX.

Agora quero dizer que é a mesma droga descrita na Dbpedia
(dbpedia.org/resource/Methimazole)

14888 -> owl:sameAs -> dbpedia.org/resource/Methimazole

Observe que aqui eu disse que a droga com a URI (<http://wifo5-04.informatik.uni-mannheim.de/sider/page/drugs/14888>) é a mesma (owl:sameAs, usei o modelo OWL de coisas iguais) à URI (dbpedia.org/resource/Methimazole)

Quero saber mais sobre RDF:

<http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>

http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_003-05.pdf

Sparql

É a linguagem de consulta de dados em Bancos de dados Semânticos, é a forma como podemos fazer uma consulta para retornarmos uma pesquisa que estamos fazendo em todos os bancos que estão conectados pelos grafos, como vemos na figura 3.

Em exemplo disto é a seguinte consulta:

```
SELECT DISTINCT ?presidente WHERE {  
    ?presidente a dbpedia-owl:President .  
    ?presidente dbpedia-owl:religion  
<http://pt.dbpedia.org/resource/Catolicismo> .  
}
```

Ela vai retornar todos os presidentes que são católicos e você pode observar que a consulta segue o modelo de grafos de RDF.

Quero saber mais sobre RDF:

<http://www.w3.org/TR/rdf-sparql-query/>

Quero testar query sparql:

<http://dbpedia.org/sparql>

Web Semântica

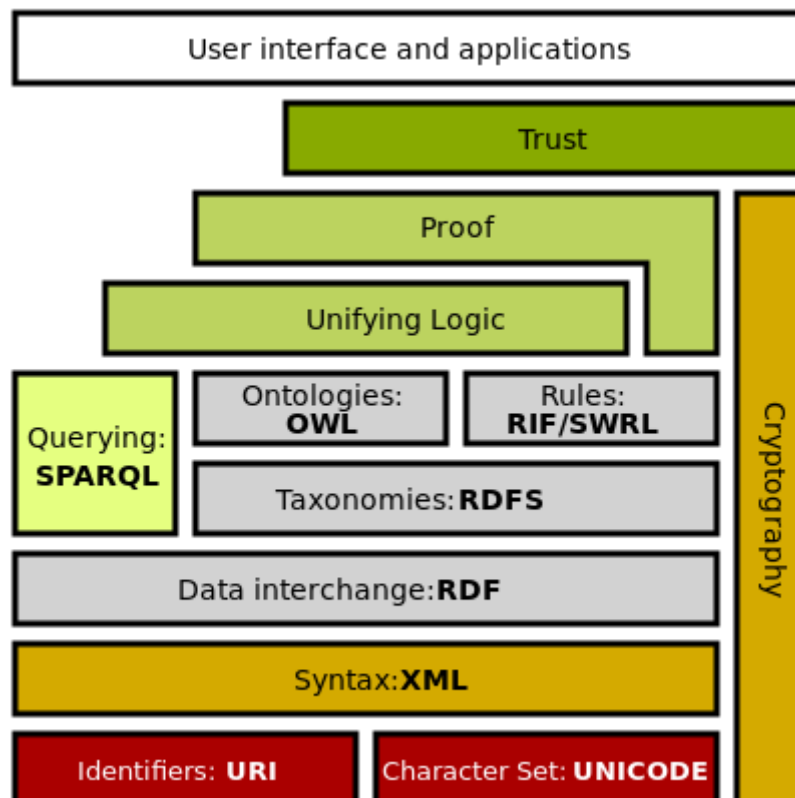


Figura 5: Estrutura da Rede semântica.

A figura acima mostra a estrutura da rede semântica, observe que ela é constituída de tudo aquilo que falamos até este momento.

A sua base está em definir que as coisas são únicas, depois estabelecer os padrões, então a forma como vamos explicar para as máquinas o que são as coisas, fazer com que essas máquinas possam fazer coisas para a gente e por fim confiar nisso.

O que é Web Semântica?

Imagine a Web Semântica como um esforço mundial para criar uma gramática que possa ser lida por todas as máquinas (ou dispositivos) do mundo. Estamos em Web Semântica ensinando as máquinas e estabelecendo padrões mundiais sobre tudo o que existe, porque assim poderemos fazer com que as máquinas façam coisas automaticamente para nós.

De acordo com a W3C: “A Web Semântica dá às pessoas a capacidade de criarem repositórios de dados na Web, construírem vocabulários e escreverem regras para interoperarem com esses dados. A linkagem de dados é possível com tecnologias como RDF, SPARQL, OWL, SKOS”.

Porque é que a Web semântica é importante?

Existem ainda motivos impensados sobre a importância dela, mas alguns descrevemos abaixo:

Padronização: Ela poderá criar um padrão mundial de identificação de tudo, conhecido como vocabulário controlado.

Acessibilidade: Com ela será possível que as máquinas possam dar suporte a pessoas com necessidades especiais.

Reuso: Ela já utiliza a HTML e com a HTML5 e CSS3 poderemos ter códigos prontos de identificação.

Personalização: O sistema consegue descobrir os fatos mais importantes sobre uma pessoa ou lugar, dando destaque para esses dados.

Relacionar coisas: Ela permitirá que as “coisas” que existem possam ser relacionadas.

O que podemos fazer com ela?

Ela nos dá a possibilidade de ser algo digital, e assim sendo eu posso ser relacionado com outras coisas que existem digitalmente e elas podem fazer relações sobre mim ou sobre qualquer coisa que existe.

Podemos querer saber sobre tuberculose, por exemplo, e ela nos vai retornar as informações de acordo com aquilo que ela souber de você, assim sendo você pode receber informações acadêmicas sobre tuberculose caso seja um professor ou informações como o guia-rápido do Ministério da Saúde caso seja um paciente.

O que ela pode fazer para nós?

Imagine que você tem tuberculose, a Web Semântica pode automaticamente perceber quais foram as pessoas que você teve contato e avisá-las pra fazer profilaxia.

Ela pode criar regras para você e quando quiser saber ela te retorna “o caminho no grafo” que ela percorreu para te dar a resposta.

Alguns exemplos sobre essas “perguntas semânticas” feitas na base de dados do Facebook podem ser encontradas em: <https://www.facebook.com/data>.

Como é que a Web semântica é importante na saúde pública?

Ligar os dados é fundamental na saúde, poderemos fazendo uma consulta no nó do paciente retornar todas as informações relacionadas a ele. Poderemos inclusive fazer um acompanhamento deste paciente.

A Web Semântica poderá deixar cada ambiente livre pra criar sua estrutura e assim mesmo poder unir as informações quando necessário.

Bancos de dados na Saúde

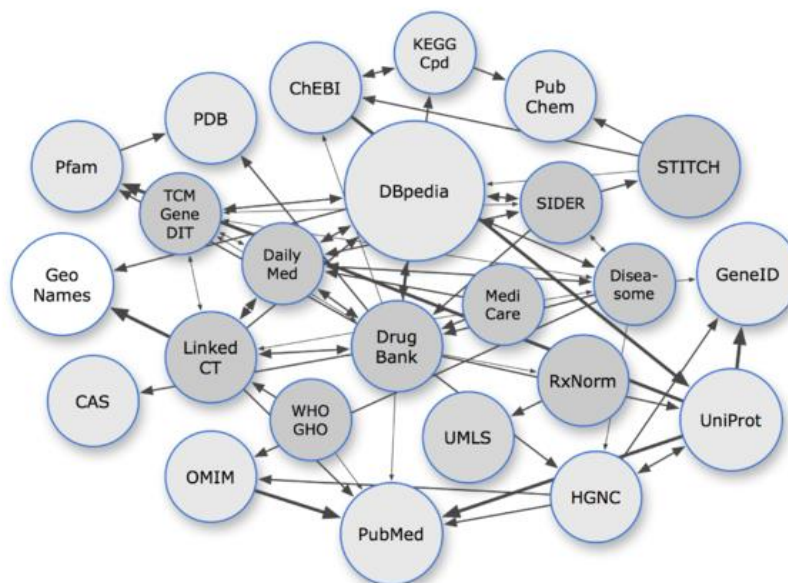


Figura 4: Principais bases sobre saúde interligadas

Acima estão listados alguns bancos de dados com informações de saúde, podemos, por exemplo, consultando o Banco da Dbpedia, retornar informações de CID de doenças específica:

```
SELECT * WHERE {  
  ?disease dbpedia-owl:icd10 ?CID10 .  
  FILTER (?disease like "%uberc%")  
}
```

<http://dbpedia.org/snorql/>

A consulta acima busca o termo “uberc” e retorna todas as doenças com esse termo e o CID10 dela.