



Web Semântica

Uma nova forma de lidar com a Web

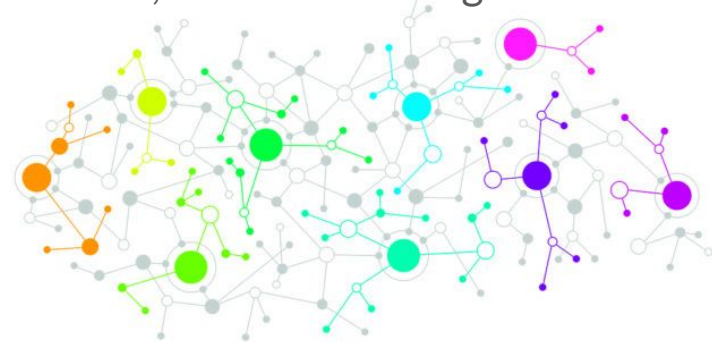


Web Semântica

- A Web Semântica é uma visão, proposta por Tim-Berners-Lee, James Hendler e Ora Lassila no artigo The Semantic Web, de como a Web seria no futuro.
- Os objetivos principais dessa proposta é que a informação seja facilmente encontrada e compreendida por mecanismos de busca, ser acessível para pessoas com deficiências visuais utilizando leitores de voz, e de forma mais utópica, ser compreendida por computadores.
- A Web Semântica (ou Web 3.0) não é uma Web separada, na verdade, é uma extensão da atual.

Dados “Linkados”

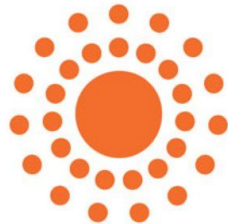
- O grande diferencial da web semântica é a forma pela qual é buscada a integração em larga escala dos dados na web, logo, as relações estabelecidas entre estas informações são o que permitem a categorização e desambiguação de termos através dos conteúdos.
- Aqui entra em cena um dos principais aliados desta visão, os Dados Interligados (Linked Data).





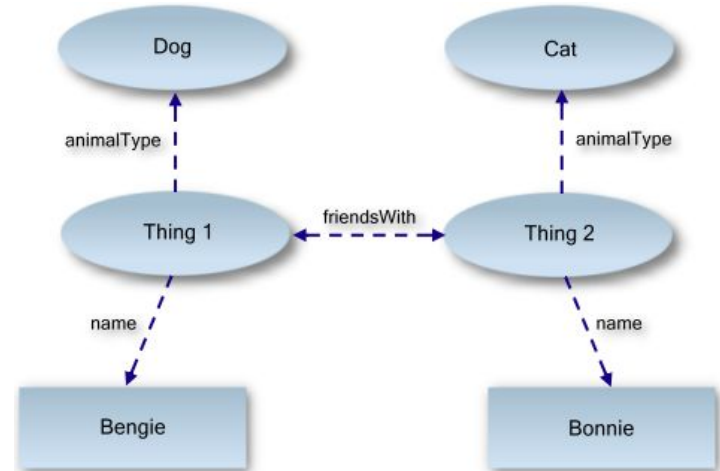
Metadados

- Para que seja possível implementar essa extensão da web é necessário conectar páginas da web(html) com recursos para identificar os dados(metadados). A ideia básica é organizar os dados através de outros dados.
- Linguagens para descrever metadados: RDF, Dublin Core, WarWick, etc



RDF(Resource Description Framework)

Uma das tecnologias usadas na modelagem de informação, definindo um tipo de banco de dados em formato de grafo(graph database), representando informações como grafos. Utiliza XML para representar os metadados.





RDF - Continuação

O RDF em si não descreve o significado(semântica dos dados). Para isso existem outras tecnologias: RDFS(RDF Schema) e OWL(Web Ontology Language).

RDF Uso:

Para a linguagem python encontramos a biblioteca RDFLib:

<https://github.com/RDFLib/rdfliib>



Vocabulário x Ontologias

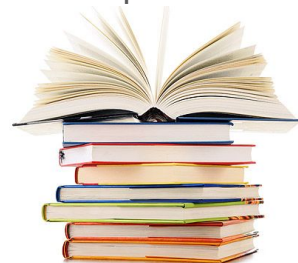
- Na Web Semântica, vocabulários definem os termos usados para descrever e representar áreas de interesse. Vocabulários são usados para classificar os termos que podem ser utilizados em determinada aplicação, caracterizar possíveis relacionamentos e definir possíveis restrições no uso destes termos.
- Não há divisão clara entre o que se refere a “vocabulários” e “ontologias”. A tendência é o uso da palavra “ontologia” para coleções de termos mais complexas e possivelmente mais formais.
- Seja vocabulário ou ontologia, é fato que a ferramenta é uma base para técnicas de inferência na Web Semântica.



Uso prático de ontologias

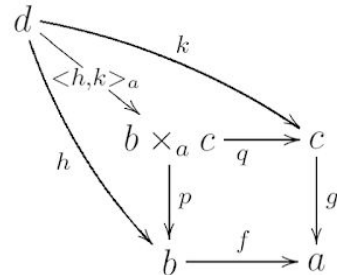
Exemplo

Suponhamos que um livreiro queira integrar dados vindos de diferentes editoras. Os dados podem ser importados para um modelo RDF comum através do uso de conversores. Entretanto, um banco de dados pode utilizar o termo “autor”, enquanto outros podem fazer uso do termo “criador”. Para tornar a integração efetiva, uma definição extra deveria ser adicionada aos dados RDF, descrevendo o fato que a relação descrita como “autor” é a mesma que “criador”



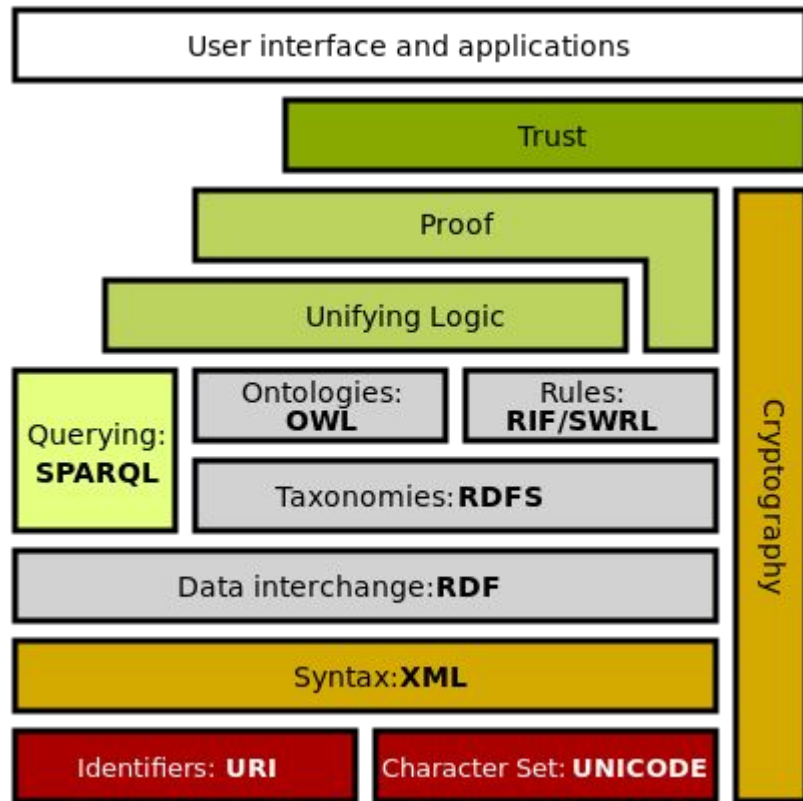
Teoria das Categorizações

- A Teoria das Categorizações é uma teoria matemática geral de estruturas e sistemas de estruturas, nascida de uma generalização da Teoria dos Conjuntos.
- Tem sido um objeto de estudo para as mais diversas áreas, pois está em constante evolução.
- É uma ferramenta poderosa pois permite a visualização de componentes de uma família de estruturas e como estruturas de diferentes tipos podem estar relacionadas. Logo, estuda “objetos” e “morfismos” entre eles.



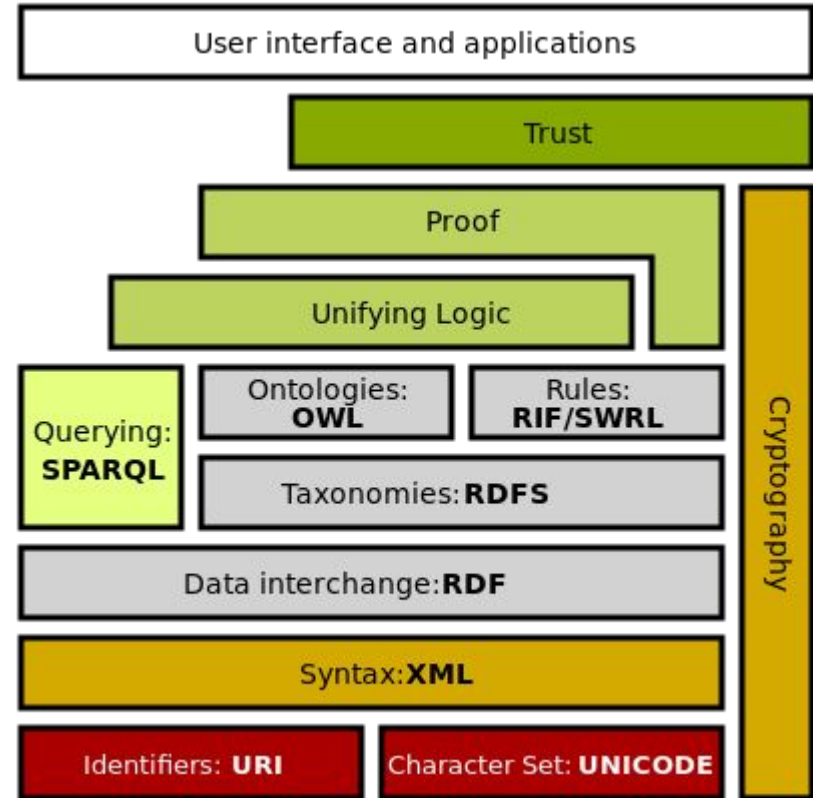
A Pilha da Web Semântica

Como mostrado na Pilha da Web Semântica, as seguintes tecnologias são usadas para o desenvolvimento da web 3.0. As tecnologias na parte inferior da pilha até OWL têm sido padronizadas e aceitas no desenvolvimento de aplicações com web semântica. Ainda não é claro como o topo da pilha será implementado no futuro, mas todas as camadas devem ser implementadas visando respeitar as visões da web semântica.



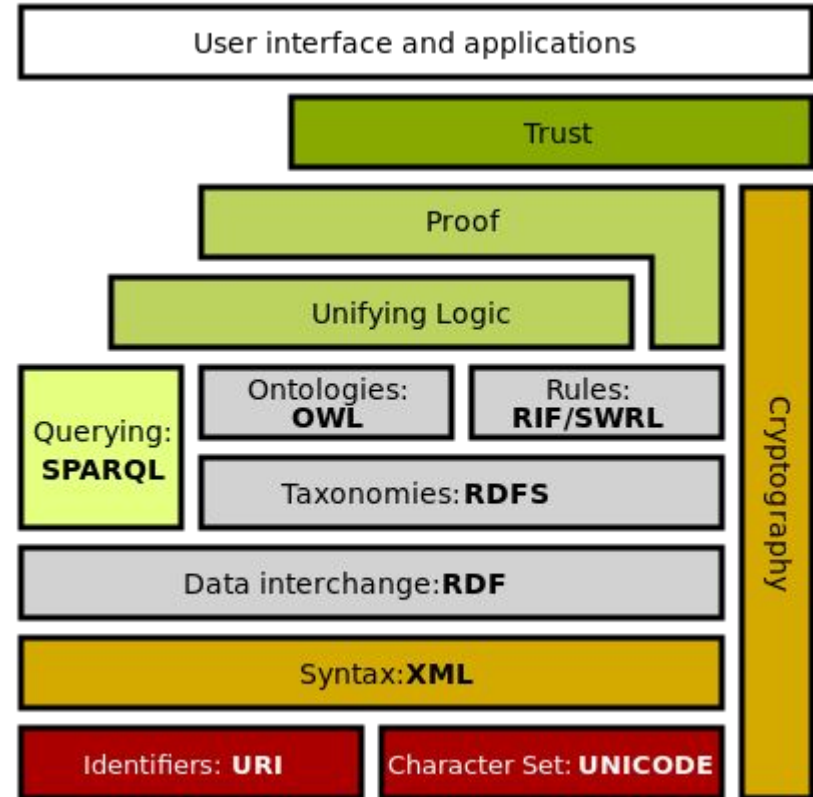
Tecnologias de Hipertexto

- **URI** - É um identificador uniforme de recursos na web.
- **UNICODE** - Representa e manipula texto em diversas linguagens.
- **XML** - Linguagem de marcação que permite a criação de documentos compostos de dados estruturados.



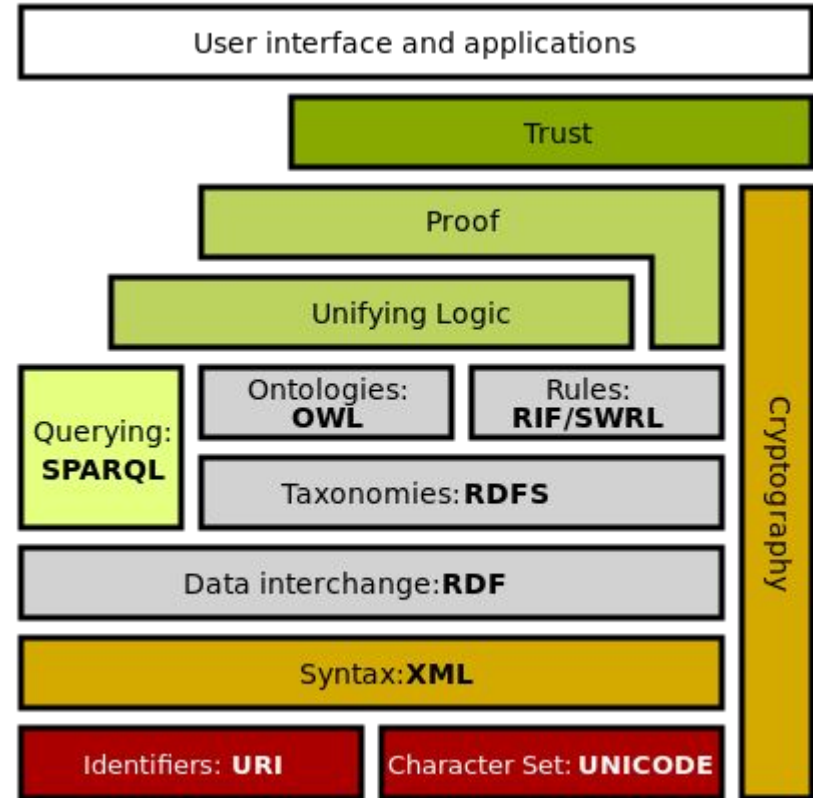
Tecnologias Padronizadas

- **RDF** - é um framework que permite representar informações sobre recursos em grafos.
- **RDF Schema (RDFS)** - Oferece vocabulário básico para RDF, permitindo a criação de hierarquia de classes e propriedades.
- **OWL** - Estende RDFS ao adicionar construções mais avançadas para descrever as semânticas de declarações RDF.



Tecnologias Padronizadas

- **SPARQL** - É uma linguagem de Query em RDF.
- **RIF** - Permite descrever relações que não podem ser diretamente descritas através da lógica usada em OWL.



A cobertura do bolo...

- **Criptografia** - É importante verificar que declarações semânticas estão vindo de fontes confiáveis, através da assinatura digital do RDF.
- **Confiança** - Acreditar que as premissas são confiáveis e confiar na lógica formal ao adquirir novas informações.
- **Interface de Usuário** - A cereja do bolo! É a camada final que permite aos usuários o uso das aplicações.

