

# 1

## Introdução

A internet é uma ampla fonte de disseminação de informações, abrangendo praticamente todas as áreas de conhecimento. A maioria das informações disponíveis hoje para a grande parte dos usuários não estão estruturadas e só podem ser processada por seres humanos. Hoje a principal forma de acesso a essas informações é através de navegadores web. Como a quantidade de dados existentes é grande e bastante desorganizada, o acesso a essas informações é feita com a ajuda de buscas e índices. As buscas existentes hoje indexam o conteúdo existente de uma forma sintática, desconhecendo a real semântica das informações ali contidas. Os índices são difíceis de serem feitos e mantidos, uma vez que a classificação do conhecimento precisa ser feita manualmente.

A proposta da web semântica [Berners-Lee et. al, 2001] é trazer sentido para as informações disponíveis na internet de forma que seres humanos e computadores possam trabalhar em conjunto. Para isso, grande parte das informações deve passar a ser disponibilizada de uma forma mais estruturada, seguindo ontologias [Zhang, 2007] que fazem com que os dados disponibilizados possuam semântica definida.

Ontologias são modelos de dados que descrevem um domínio. Através de uma ontologia é possível representar conhecimento sobre o domínio que a ontologia descreve. Um computador pode processar e realizar inferências sobre um conhecimento representado através de uma ontologia previamente conhecida. Ou seja, uma vez que as informações passem a seguir uma ontologia, seus dados passam a possuir a semântica que existe na definição da ontologia, de forma que possam ser processados tanto por seres humanos quanto máquinas.

A forma padrão adotada pela comunidade, que é uma especificação da W3C<sup>1</sup>, para a representação de dados que seguem uma ontologia e possuem semântica é através do modelo RDF [Manola et al, 2004]. A sigla RDF significa

---

<sup>1</sup> World Wide Web Consortium - <http://www.w3c.org>

Resource Description Framework, um modelo simples de representação de dados baseado em recursos e afirmações sobre esses recursos para descrever qualquer tipo de informação. Todos os recursos são representados através de uma URI [Berners-Lee et al, 1998]. Uma base de dados em RDF consiste em um conjunto de triplas. Cada tripla pode ser interpretada como uma afirmação sobre um recurso. A tripla possui três elementos, e a posição dos elementos define a semântica da afirmação. O formato utilizado é: sujeito, predicado e objeto. Na posição do sujeito, está o recurso sobre qual a afirmação está sendo feita. O predicado é outro recurso que denota uma propriedade do sujeito e o relaciona através dessa propriedade com o objeto. O objeto pode por sua vez ser outro recurso ou possuir um valor literal, como um número ou uma sequência de caracteres. Um exemplo de uma afirmação representada através de uma tripla RDF pode ser: “Carla é mãe de Patrícia.”, que seria representada da seguinte forma:

- Sujeito: [HTTP://example-domain.com/people/carla](http://example-domain.com/people/carla) (URI do recurso que representa Carla)
- Predicado: [HTTP://family-ontology/predicates/isMotherOf](http://family-ontology/predicates/isMotherOf) (URI do recurso que define uma relação de maternidade)
- Objeto: [HTTP://example-domain.com/people/patricia](http://example-domain.com/people/patricia) (URI do recurso que representa Patrícia)

Através de diversas triplas, como a demonstrada acima, uma base RDF pode conter qualquer tipo de informação sobre qualquer domínio.

As ontologias também podem ser descritas através de triplas RDF. Para isso existem extensões do padrão RDF para a descrição de ontologias. Uma das extensões mais utilizadas é o RDFS [Brickley et al, 2004], que significa Resource Description Framework Schema. RDFS é uma extensão de RDF que possibilita a descrição de ontologias. RDFS funciona como uma meta-ontologia, sua utilização possibilita a descrição, por exemplo, de classes, subclasses e atributos de um domínio. RDFS é utilizada para descrever ontologias sobre os diversos domínios, e essas ontologias são utilizadas para representar informações sobre esses domínios. Outra extensão de RDF bastante utilizada para a descrição de ontologias é conhecida como OWL [Dean et al, 2004], que significa Web Ontology Language. OWL possui uma expressividade maior que RDFS. Ambas as extensões são padrões recomendados pela W3C.

Para exemplificar a representação de uma ontologia simples que descreve uma árvore genealógica, utilizando a extensão RDFS, podem ser utilizadas três triplas:

```
Sujeito: <http://family-ontology/classes/Person>
Predicado: <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
Objeto: <http://www.w3.org/2000/01/rdf-schema#Class>
```

```
Sujeito: <http://family-ontology/classes/Person>
Predicado: <http://www.w3.org/2000/01/rdf-schema#label>
Objeto: Pessoa
```

```
Sujeito: <http://family-ontology/predicates/isMotherOf>
Predicado: <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
Objeto: <http://www.w3.org/1999/02/22-rdf-syntax-ns#Property>
```

```
Sujeito: <http://family-ontology/predicates/isMotherOf>
Predicado: <http://www.w3.org/2000/01/rdf-schema#label>
Objeto: Mãe de
```

```
Sujeito: <http://family-ontology/predicates/isFatherOf>
Predicado: <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
Objeto: <http://www.w3.org/1999/02/22-rdf-syntax-ns#Property>
```

```
Sujeito: <http://family-ontology/predicates/isFatherOf>
Predicado: <http://www.w3.org/2000/01/rdf-schema#label>
Objeto: é pai de
```

A primeira tripla afirma que a URI na posição do sujeito representa uma classe no domínio e a segunda tripla define o nome da classe como “Pessoa”. Diferente da primeira tripla, o objeto da segunda tripla é um literal, ou seja, não se trata de outro recurso.

As duas triplas seguintes afirmam que existe uma propriedade no domínio chamada “Mãe de”. A semântica dessa propriedade é de maternidade. De forma análoga, as duas últimas triplas afirmam a existência de outra propriedade, com semântica de paternidade, chamada de “é pai de”.

Uma base RDF que utilize a ontologia apresentada como exemplo e possua a afirmação “Carla é mãe de Patrícia” poderia ser composta das seguintes cinco triplas:

```
Sujeito: <http://example-domain.com/people/carla>
Predicado: <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
Objeto: <http://family-ontology/classes/Person>
```

```
Sujeito: <http://example-domain.com/people/carla>
Predicado: <http://www.w3.org/2000/01/rdf-schema#label>
Objeto: Carla
```

```
Sujeito: <http://example-domain.com/people/patricia>
Predicado: <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
Objeto: <http://family-ontology/classes/Person>
```

```
Sujeito: <http://example-domain.com/people/patricia>
Predicado: <http://www.w3.org/2000/01/rdf-schema#label>
Objeto: Patrícia
```

```
Sujeito: <http://example-domain.com/people/carla>
Predicado: <http://family-ontology/predicates/isMotherOf>
Objeto: <http://example-domain.com/people/patricia>
```

A primeira tripla afirma que a URI na posição do sujeito representa uma instância da classe Pessoa. A segunda tripla define como “Carla” o nome dessa pessoa. De forma análoga, a terceira e a quarta tripla definem uma pessoa chamada Patrícia. A última tripla define a relação de maternidade entre Carla e Patrícia.

A ontologia apresentada é muito simples, uma ontologia completa pode descrever inúmeras propriedades do domínio, permitindo até a validação de bases RDF que a utilizam.

O entendimento dos objetivos da web semântica, assim como os conceitos de RDF, RDFS e ontologias são importantes, pois essa é a área de estudo desta dissertação.

Através do avanço da web semântica, cada vez mais bases de dados em RDF contendo todo tipo de informações nos mais variados domínios estão disponíveis na internet para serem acessadas por diversos usuários. O conhecimento armazenado desta forma traz novas possibilidades e desafios, a começar pela forma de consumo dessas informações.

Uma das possibilidades que surgem é o cruzamento de dados do mesmo domínio de diferentes fontes que utilizem as mesmas ontologias. O uso dessas

bases de forma integrada é algo que se torna possível, mas ainda existem poucas ferramentas adequadas para que ele seja realizado.

A utilização do navegador tradicional para explorar informações na internet deixa de ser a única opção e não é mais necessariamente a opção mais interessante, uma vez que a informação agora pode ter conteúdo semântico. Novas formas mais eficientes de adquirir conhecimento através das informações disponíveis no mundo digital vêm sendo pesquisadas, e essa é a frente de pesquisa explorada nesse trabalho.

Um navegador que possua conhecimento da ontologia utilizada para descrever a informação acessada, por conhecer a semântica dos dados, pode interpretá-la e auxiliar o usuário na tarefa que ele deseja realizar, seja uma busca, um cruzamento de informações ou uma navegação livre. Com diversas bases sendo disponibilizadas com informações semânticas, criou-se a necessidade do desenvolvimento de ferramentas adequadas para auxiliarem os usuários no acesso a essas informações. Esse trabalho foca a pesquisa de soluções que facilitem esse acesso aos dados contidos em bases semânticas.

Nesse primeiro capítulo foi apresentada a área da pesquisa. No próximo capítulo é discutido o problema existente na área apresentada que motivou o trabalho realizado. No capítulo três são apresentadas algumas soluções existentes, assim como uma breve discussão do estado da arte na área de interesse. O quarto e quinto capítulo possuem respectivamente a solução proposta e detalhes da sua implementação. Finalmente, nos últimos capítulos é discutida a solução apresentada e os resultados finais do trabalho.