

# Information Sharing on the Semantic Web

Heiner Stuckenschmidt

Frank van Harmelen

## Prefácio

Os problemas de produção, armazenamento e transmissão de informações em formato digital têm sido resolvidos pelas tecnologias de hardware e software disponíveis. A produção de grande parte de documentos, mesmo os destinados a impressão, ocorre na forma digital. As tecnologias existentes permitem o armazenamento de um grande volume de dados em computadores de uso pessoal com acesso a internet de alta velocidade. A WWW torna disponível um volume cada vez maior de informações.

A grande quantidade de informação disponível é a causa de dois novos problemas:

- busca de informação adequada: é cada vez mais difícil encontrar uma informação que atenda a uma necessidade específica;
- integração de informações: as informações quando encontradas podem estar em formatos inadequados para uso imediato, devendo ser previamente transformadas.

Estes dois problemas em conjunto definem o problema de compartilhamento de informação (*information sharing*).

A principal tese do livro é de que o problema de compartilhamento de informação é solucionável somente se for dado ao computador acesso a semântica da informação. Em arquivos textuais, além de informações comuns relativas ao metadado, tais como autor, título, resumo, etc. de um documento seria necessário fornecer também acesso as informações relativas aos conceitos abordados no documento, seus relacionamentos com conceitos disponíveis em outros documentos e com conceitos gerais definidos sobre um domínio de conhecimento.

Com base nesta tese, são discutidos no livro tópicos de pesquisa relativos a:

- Como representar a semântica dos itens de informação de modo que possam ser acessados por um computador.
- Como usar este “metadado semântico” na busca e integração de informações.

O compartilhamento do significado intencionado deste “metadado semântico” entre os criadores e usuários do metadado pode ser realizado através de ontologias, modelo formalizado de um domínio particular, que pode ser processada por computadores.

No livro são investigados o uso da ontologia na integração semântica de informação em ambientes “fracamente” estruturados, particularmente na WWW. Neste contexto são analisadas soluções para os seguintes problemas:

Modelos conceituais ~~incompletos~~ ausentes (*Missing conceptual models*): Não existe acesso aos modelos conceituais de uma fonte de informação da Web.

Limites de sistemas imprecisos (*Unclear system boundaries*): A dinâmica da Web, dificulta a identificação de um limite no qual uma informação possa estar contida.

Representação heterogênea (*Heterogeneous representations*): Na Web podem existir diferentes representações de uma ontologia. Uma integração no nível ontológico também será necessário.

Parte I – Introdução ao problema de compartilhamento de informação e a necessidade do uso de representação explícita da semântica da informação para possibilitar o seu compartilhamento.

Parte II – Métodos para criação de representações explícitas da semântica de informações

Parte III – Descreve o uso da infra-estrutura representacional no compartilhamento de informações

Parte IV – Revisão do uso de ontologias no compartilhamento de informações aplicado no contexto de um ambiente distribuído e ontologias distribuídas. Introdução ao conceito de ontologia modular.

## Capítulo 1

### Integração Semântica

O problema de integração de informação heterogênea pode ser dividido em 3 categorias:

- problemas de integração sintática: gerados por exemplos na integração de informação existentes em diferentes formatos.
- problemas de integração estrutural: como por exemplo nos termos sinônimos, homônimos ou com diferentes atributos em uma tabela de banco de dados.
- problemas de integração semântica: diferenças de significado intencionado de um termo de acordo com contexto ou aplicação.

A existência de um grande número de formatos padrões (tais como XML, HTML, SVG definidos pelo W3C) têm permitido a solução do problema de integração de informação no nível sintático.

#### 1.1 – Padrões sintáticos

##### 1.1.1. HTML

##### 1.1.2. XML

##### 1.1.3. RDF - Resource Description Framework

O formato XML descreve modelos de dados de uma aplicação no nível sintático. O RDF, por outro lado, foi proposto como uma solução permitindo a descrição semântica de recursos disponíveis na Web, mais particularmente a representação dos metadados destes recursos. A idéia básica da RDF está na identificação de recursos através de URI's (*Uniform Resource Identifiers*) e da descrição destes recursos em termos de propriedades e valores de propriedades. A abordagem utilizada pela RDF pode ser generalizada para descrever não somente recursos da Web, mas entidades quaisquer para as quais se possa atribuir um identificador e propriedades, mesmo que não estejam disponíveis na Web. A descrição de um recurso em RDF pode ser realizada por meio de expressões na forma:

(*sujeito, predicado, objeto*)

O *predicado* (ou *propriedade*) expressa uma relação binária entre um *sujeito* (ou *recurso*) e um *objeto* (valor da propriedade). O *objeto* pode ser um tipo de dado, um outro *sujeito* ou um valor literal. O *objeto* pode ser substituído por uma variável representando um outro *recurso*, que por sua vez é descrito por um conjunto de *propriedades*.

(recurso, propriedade, X)  
(X, propriedade1, valor1)  
...  
(X, propriedadeN, valorN)

Para identificar de maneira única cada um dos elementos desta expressão; sujeito, predicado e objeto; a cada um deles está associado uma URIref (*Uniform Resource Identifier reference*), ou seja, uma URI adicionada de um fragmento identificador em seu final. As expressões em RDF podem ser representadas ou na forma de documentos XML, permitindo o seu processamento automatizado (W3C, 2007b), ou na forma de grafos ou de triplas.

Com o objetivo de simplificar as expressões em RDF, o URIref é substituído pelo recurso de identificação de nomes qualificados, ou *QName*.

Como ilustração a sentença “A página <http://www.exemplo.com.br/index.html> foi criada por João da Silva em 29-10-2006 na” pode ser descrita usando as seguintes triplas RDF.



Figura 1 – Exemplo de triplas RDF

Usando a generalização destes termos, associada com nomes qualificados podemos reescrever as triplas acima como ilustrado na figura 2.

|               |                       |              |
|---------------|-----------------------|--------------|
| ex:index.html | dc:criador            | exfunc:4114  |
| ex:index.html | extermos:criação-data | "29-10-2006" |

Figura 2 – triplas usando recurso de nomes qualificados

Onde *ex* é o prefixo para a URI *http://www.exemplo.com.br/*, *dc* para *http://purl.org/dc/elements/1.1/*, *extermos* prefixo de *http://www.exemplo.com.br/termos*, e *exfunc* é o prefixo para *http://www.exemplo.com.br/func*. O conjunto de URIref que podem ser definidos para referenciar os termos de uma expressão RDF define um vocabulário. Este vocabulário representa um conjunto de conceitos cuja semântica está bem definida dentro de um domínio específico. Por exemplo "index.html", "criação-data" e "4114" são termos de um vocabulário cujo o significado intencionado será correto dentro da instituição que os definiu, utilizando a URI base *http://www.exemplo.com.br/*. Neste exemplo o valor literal "João da Silva" foi substituído pela URIref *exfunc:4114*, que permite identificar de maneira única a entidade que criou o documento <http://www.exemplo.com.br/index.html>.

A Figura 5 apresenta a representação em RDF da sentença. O atributo *rdf:about* da tag XML *rdf:Description* indica o sujeito da sentença. As tags XML *extermos:criação-data* e *dc:creator* descrevem os valores das propriedades (texto em azul).

```

1 <?xml version="1.0"?>
2 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3   xmlns:extermos="http://www.exemplo.com.br/termos/">
4   <rdf:Description rdf:about="http://www.example.org/index.html">
5     <extermos:criação-data>August 16, 1999</extermos:criação-data>
6     <dc:creator
7   rdf:resource="http://www.exemplo.com.br/func/8574
8   0"/>
   </rdf:Description>
 </rdf:RDF>

```

Figura 3 – Representação de expressões RDF em XML

Através do mecanismo de reificação (“reification”)

**Referências**

W3C, *RDF Primer*, Disponível em <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>

W3C, *RDF/XML Syntax Specification*, Disponível em <http://www.w3.org/TR/rdf-syntax-grammar/>