

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Infraestrutura tecnológica Internet, Web e Web Semântica

Módulo 2 – parte 4

Prof. Mauricio B. Almeida
Escola de Ciência da Informação UFMG
<http://mba.eci.ufmg.br>
mba@eci.ufmg.br

1

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Apresentação

Assunto abordado:
Internet, Web e Web Semântica

Temas abordados:
Internet e *World Wide Web*: visão geral, protocolos, linguagens e considerações;
Web Semântica: linguagens de marcação, padrões eletrônicos de metadados, Web Services, protocolos, pilha de aplicações;
Estruturas de organização da informação: taxonomias, tesauros e ontologias;

2

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

A Internet

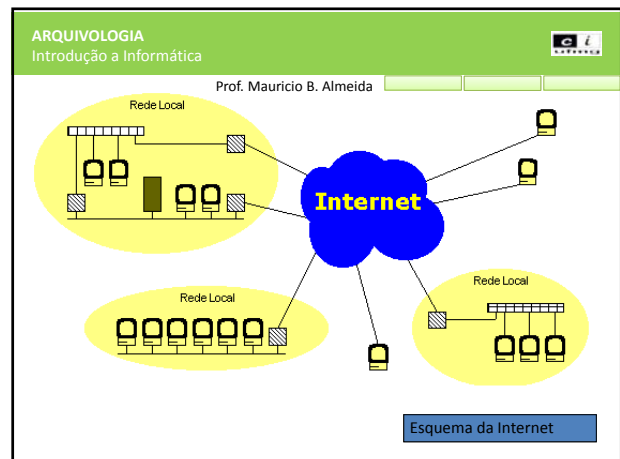
O que se designa por *Internet* é um conjunto de redes e computadores que se interligam em nível mundial.

Os meios usados para garantir esta interligação são muito diversos, recorrendo às redes públicas de comunicações, baseadas em cabos elétricos ou ópticos, terrestres ou submarinos e ligações via rádio terrestres ou via-satélite.

Com tecnologias de transmissão tão diversas é necessário um elo comum que permita a transferência de dados entre qualquer equipamento ligado à *Internet*;

Esse elo é o protocolo de rede, no caso da *Internet*, designado por *Internet Protocol* e abreviado para *IP*.

3



4

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Endereçamento na Internet

Uma necessidade em uma rede é a forma de identificação de cada uma das máquinas que estão ligadas. É um problema idêntico ao que acontece quando se pretende enviar uma carta ou telefonar a alguém;

A característica fundamental de um endereço é ser único e isso pode ser um problema numa rede global como a *Internet*.

O formato dos endereços de rede depende do protocolo de rede, atualmente o mais divulgado é o *IP* versão 4;

Este protocolo utiliza apenas 4 octetos (conjuntos de 8 bits que permitem representar números de 0 a 255) para identificar cada destino.

5

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Este número de octetos é reduzido para a expansão da *Internet*. Não existem endereços suficientes para a quantidade de máquinas que lhe estão ligadas.

Por essa razão foi desenvolvido o *IP* versão 6. Entretanto foram encontradas outras soluções e o *IPv6* não é ainda muito usado.

Seja qual for o protocolo de rede, os dados transferidos pela rede devem conter dois endereços: **origem e destino**. Isto é fundamental por duas razões:

- O destinatário pode querer responder, ou simplesmente confirmar que os dados chegaram.
- Se a transferência dos dados não é bem sucedida, a rede deve avisar o remetente de tal fato.

6

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Modelo cliente-servidor

Este é o modelo que serve de referência à maioria das comunicações em rede, baseia-se no conceito de prestação de um serviço e define um diálogo típico de pedido-resposta:

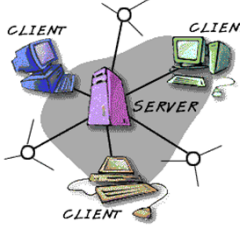
Servidor - aplicação passiva que aguarda o contato pelo cliente que, uma vez contactado, recebe o pedido de serviço, executa o serviço e devolve ao cliente uma resposta.

Cliente - aplicação que toma a iniciativa, contata o servidor, envia-lhe o pedido e aguarda pela resposta.

7

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida



Deste modelo pode-se tirar algumas conclusões imediatas:

- O cliente é em muitos casos manuseado diretamente pelo usuário.
- O cliente tem de conhecer previamente o endereço do servidor.
- O servidor não necessita de saber previamente o endereço do cliente.

Esquema do modelo cliente-servidor

8

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Nomes de máquinas

Devido à impossibilidade de os usuários lidarem com endereços de rede, foi desenvolvida uma alternativa que é o nome de máquina. Consiste em batizar as máquinas dando-lhes nomes com significado "humano" depois é necessário um mecanismo que efetue a tradução automática.

Os clientes recebem do usuário o nome da máquina onde está o servidor e utilizam o referido "mecanismo automático" para obter o respectivo endereço de rede. Esta operação de obtenção do endereço de rede a partir do nome da máquina designa-se por "resolução do nome".

9

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Se numa rede local, com algumas dezenas de máquinas a implementação de um mecanismo de resolução de nomes é simples e pode ter várias abordagens, na *Internet* torna-se muito complicada;

Se existisse uma base de dados contendo um registro para cada máquina ligada à *Internet* ela seria enorme e a pesquisa sobre ela muito demorada.

A solução adotada consiste na distribuição desta base de dados segundo nomes designados por domínios e, para facilitar ainda mais a resolução, os domínios estão organizados em forma hierárquica;

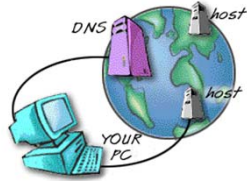
10

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Esta base de dados e respectivos serviços têm a designação **Domain Name System (DNS)**.

Por exemplo, no endereço web <http://www.ufmg.br/> o *br* é um domínio que indica que a página web encontra-se no Brasil.



Visão esquemática do serviço de DNS

11

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Serviços

Cada máquina pode ter em funcionamento simultâneo vários servidores de diferentes de tipos.

Cada máquina pode ter múltiplos clientes em funcionamento. Por exemplo, uma máquina pode disponibilizar serviços *www* para servidores *web* (páginas Web), *ftp* (transferência de arquivos), *mail* (correio eletrônico) etc.

Uma vez que a ligação à rede é geralmente única, há necessidade de definir um mecanismo lógico que permita separar e evitar interferências entre as diferentes comunicações em curso através da ligação física.

12

ARQUIVOLOGIA
Introdução a Informática

Prof. Maurício B. Almeida

Este tipo de operação é designado por **multiplexagem**.

A solução consiste em atribuir um número de identificação ao dados;

Assim os dados que chegam a uma máquina têm associados a si um identificador numérico e no interior da máquina, este identificador é usado para fazer chegar os dados à aplicação correta, por exemplo um servidor.

Este processo é usado em diversos níveis, permite por exemplo que uma máquina use vários protocolos de rede. Nos protocolos de rede estes identificadores são conhecidos por portas (*ports*).

13

ARQUIVOLOGIA
Introdução a Informática

Prof. Maurício B. Almeida

Para cada tipo de servidor/serviço está estabelecido um número de porta fixo, tanto o servidor como o cliente sabem qual é esse número.

No caso do servidor, é nessa porta em que é recebido o pedido; no caso do cliente para ele saiba para qual porta tem de enviar o pedido.

Os usuários não necessitam de conhecer estes números de porta pois eles estão implícitos pela aplicação cliente que é usada.

14

ARQUIVOLOGIA
Introdução a Informática

Prof. Maurício B. Almeida

A World Wide Web

As principais tecnologias que suportam a *Web* são

- Protocolo HTTP-*Hypertext Transfer Protocol*;
- URL-*Uniform Resource Location*;
- Protocolo TCP/IP;

15

ARQUIVOLOGIA
Introdução a Informática

Prof. Maurício B. Almeida

Esquema básico de funcionamento da Web

Fonte: Vanini, 2001

16

ARQUIVOLOGIA
Introdução a Informática

Prof. Maurício B. Almeida

Protocolo HTTP

HTTP é o protocolo usado na comunicação entre o servidor Web e o Browser; constitui a base sobre a qual o servidor Web realiza suas operações.

A ideia do HTTP é a distribuição de **conteúdo estático** (tipicamente páginas escritas em HTML).

Evolução: distribuição de **conteúdo gerado dinamicamente**, em resposta a dados fornecidos pelo usuário (por exemplo, consulta a um banco de dados).

CGI-*Common Gateway Interface* é a forma mais antiga de prover conteúdo dinâmico; atualmente tem-se outros recursos como as páginas dinâmicas (asp, jsp, etc).

17

ARQUIVOLOGIA
Introdução a Informática

Prof. Maurício B. Almeida

Esquema básico de funcionamento do HTTP

18

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Uma conexão **HTTP** é feita através de comandos **Telnet**, que são formados por mensagens texto trocadas através de uma conexão TCP.

Por exemplo:

```
%> telnet www.umsite.com 80
%> GET /index.html HTTP/1.0
```

19

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Comandos (ou métodos) do HTTP

Method	Description
GET	Request to read a Web page
HEAD	Request to read a Web page's header
PUT	Request to store a Web page
POST	Append to a named resource (e.g., a Web page)
DELETE	Remove the Web page
LINK	Connects two existing resources
UNLINK	Breaks an existing connection between two resources

Fonte: Vanini, 2001

20

```
> telnet www.w3.org 80
Trying 18.23.0.23
Connected to www.w3.org
Escape character is '^]'.
> GET /hypertext/WWW/TheProject.html
HTTP/1.0
HTTP/1.0 200 Document follows
MIME-Version: 1.0
Server: CERN/3.0
Content-Type: text/html
Content-Length: 8247
<HEAD>
<TITLE> The World Wide Web
Consortion (W3C) </TITLE>
</HEAD>
<BODY>
...
</BODY>
```

- 1-Browser interroga DNS pelo IP de www.w3.org;
- 2-DNS responde: 18.23.0.23;
- 3-Browser estabelece a conexão na porta 80 de 18.23.0.23;
- 4-browser envia comando: `GET/hypertext/WWW/TheProject.html`
- 5-O servidor www.w3.org envia o arquivo `TheProject.html`;
- 6-A conexão TCP é encerrada;
- 7-O browser apresenta o texto contido em `TheProject.html`;
- 8-Browser carrega e apresenta o conteúdo contido no arquivo `TheProject.html`;

Exemplo de mensagens HTTP entre servidor e cliente

21

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

URL

As URLs constituem a forma de endereçamento dos *recursos* na Web. É formado por quatro partes:

- o protocolo utilizado
- nome do domínio
- caminho de busca
- nome do arquivo

O protocolo define o tipo do servidor onde está o recurso. O *browser* determina a porta e o tipo de protocolo utilizado. Os principais protocolos são FTP, HTTP, MAILTO, TELNET;

22

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

O nome do domínio define a localização do servidor que hospeda o recurso desejado. O nome do domínio é "resolvido" pelo DNS-*Domain Name System*;

Pode ser fornecido como um nome de domínio ou como um endereço IP.

O par `<protocolo>:// <nome do domínio>` dá acesso à *home page* ou *diretório raiz* do servidor.

O caminho de busca e o nome do arquivo dependendo da localização e do tipo de informação podem ser omitidos; quando utilizados seguem a nomenclatura do UNIX para diretórios e arquivos;

23

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Name	Used for	Example
http	Hypertext (HTML)	http://www.cs.vu.nl/~ast/
ftp	FTP	ftp://ftp.cs.vu.nl/pub/minix/README
file	Local file	/usr/suzanne/prog.c
news	News group	news:comp.os.minix
news	News article	news:AA0134223112@cs.utah.edu
gopher	Gopher	gopher://gopher.tc.umn.edu/11/Libraries
mailto	Sending email	mailto:kim@acm.org
telnet	Remote login	telnet://www.w3.org:80

Exemplo da composição do endereço Web: protocolo, domínio e arquivo acessado

Fonte: Vanini, 2001

24

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Protocolo TCP/IP

O TCP/IP é um conjunto de protocolos que dá suporte a comunicação em rede, e, em particular, na Internet. As principais características do TCP/IP são:

1)Endereçamento lógico:
Em redes de grande porte não há possibilidade de um sinal ser enviado a cada computador componente;
As redes são então segmentadas em sub-redes, via software, através de endereçamentos lógicos, ou seja, endereços IP;

2)Roteamento:
O roteador é um dispositivo que redireciona dados na rede evitando tráfego; o TCP/IP é um protocolo roteável;

25

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

3)Tradução de nomes
O serviço DNS-*Domain Name System* faz a tradução entre os endereços IP numéricos em nome de domínio amigáveis;

4)Verificação de erro e controle de fluxo
O protocolo oferece recursos para fornecimento confiável de dados pela rede através da camada de transporte; inclui verificação da transmissão e confirmação de recebimento com sucesso;

5)Suporte a aplicações
O protocolo oferece interface para acesso a softwares do computador, através de um sistema de canais denominados portas;

26

Camada Application	Camada Application	Comparação das camadas do Modelo OSI e do TCP/IP
Camada Transport	Camada Presentation	
Camada Internet	Camada Session	
Camada Network Access	Camada Transport	
	Camada Network	
	Camada Data Link	
TCP/IP	OSI	

O protocolo TCP/IP não tem todas as camadas do modelo OSI, mas se tornou um padrão de fato para a transmissão de dados Internet em função de sua capacidade de atender a diversas plataformas.

27

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Considerações finais sobre a Web

Como era a *Web* a cerca de 10 anos atrás?

- Predominância de documentos HTML;
- Documentos voltados apenas para uso por pessoas;
- Muitos documentos gerados automaticamente por aplicativos;
- Dificuldade de intercâmbio de dados em função da heterogeneidade das fontes e dados;
- Dificuldades para recuperação automática eficiente em função da inexistência de semântica legível por máquinas;

Essas questões, dentre outros, tem levado à pesquisa sobre a *Web Semântica*;

28

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

A Web Semântica

Essa seção está dividida de forma a apresentar as principais tecnologias que suportam a Web Semântica:

- Conceitos preliminares: linguagens de marcação - SGML-*Standard Generalized Markup Language*, HTML-*Hypertext Markup Language*, XML-*Extended Markup Language* - e tecnologias correlatas;
- RDF-*Resource Description Framework* e RDFS- *Resource Description Framework Schema*
- Web Services
- Ontologias
- Usos e aplicações

29

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Linguagens de marcação – SGML, HTML e XML

Historicamente, a palavra “marcação” descreve anotações em um texto, que fornecem instruções ao datilógrafo sobre a maneira como uma parte do texto deve ser representada;

Com a evolução da TI, o termo foi estendido para os códigos de marcação em textos eletrônicos.

Codificar ou “marcar” um texto para processamento por computadores é também indicar como o conteúdo do texto deve ser interpretado.

Linguagem de marcação é um conjunto de convenções para a codificação de textos; deve especificar as marcas permitidas, as exigidas, como distingui-las do texto e qual o significado da marcação.

30

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

SGML-Standard Generalized Markup Language é um padrão internacional para definição de estrutura e conteúdo de diferentes tipos de documentos eletrônicos. A SGML pode ser chamada de “língua mãe”;

HTML-Hypertext Markup Language consiste de uma aplicação específica do SGML utilizada na Internet. O HTML define um tipo de documento simples, com marcações fixas projetadas por exemplo para cabeçalhos, parágrafos, listas e algumas possibilidades de multimídia.

XML-Extended Markup Language é versão abreviada do SGML, que possibilita especificar dados e a descrição da estrutura do documento, através do DTD-Data Type Definitions (gramáticas que conferem estrutura ao documento XML).

31

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

As linguagens de marcação são representantes dos dados semi-estruturados, aqueles com “alguma” estrutura;

Os dados semi-estruturados representam um componente importante de ambientes heterogêneos como a Internet, e, a XML, em função da marcação flexível, aceita bem variações na estrutura, características dos tipos de dados.

Classificação de dados em relação a estrutura

32

Exemplos de fragmentos HTML e XML

```
<h1>Pessoas que estudam na UFMG </h1>
<p> <b> João </b>, 30 anos, <i>joao@ufmg.br </i> </p>
<p> <b> Claudia </b>, 25 anos, <i>claudia@ufmg.br </i> </p>
<p> <b> Jose </b>, 27 anos, <i>jose@ufmg.br </i> </p>
```

```
< Pessoa >
  < nome > João </ nome >
  < idade > 30 </ idade >
  < email > joao@ufmg.br </ email >
</ Pessoa >
```

33

Um DTD-Document Type Definition é uma gramática para o documento XML e a possibilita ao usuário definir suas marcações.

Exemplos de XML e DTD correspondente

```
<bd>
  < Pessoa >
    < nome > João </ nome >
    < idade > 30 </ idade >
    < email > joao@ufmg.br </ email >
  </ Pessoa >
  ...
  < Pessoa >
  ...
</bd>
```

```
< DOCTYPE bd [
  < ELEMENT BD ( Pessoa * ) >
  < ELEMENT Pessoa ( nome, idade, email ) >
  < ELEMENT nome ( # PCDATA ) >
  < ELEMENT idade ( # PCDATA ) >
  < ELEMENT email ( # PCDATA ) >
] >
```

34

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Outras considerações sobre XML e HTML

O HTML é uma linguagem de **apresentação**; o XML é uma linguagem de **conteúdo**;

Para que os dados sejam apresentados em um documento XML é preciso agregar um arquivo CSS-Cascade Style Shets, ou construir um arquivo XSLT-XML Style Language Tranformation;

Um conceito importante na linguagem XML são os *namespaces*, que tem papel fundamental na definição dos termos por indicarem um endereço único da Web.

35

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

```
< ?xml version="1.0" encoding="ISO-8859-1" ? >
< rdf: RDF xmlns: daml="http://www.daml.org/2001/03/daml+oil#"
  xmlns: dc="http://purl.org/dc/elements/1.1/"
  xmlns: oil="http://img.cs.man.ac.uk/oil/oiled#"
  xmlns: rdfs="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns: rdfs="http://www.w3.org/2000/01/rdf-schema#" >
```

Exemplos de namespaces XML indicando endereços na Web

36

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Padrões eletrônicos de metadados – RDF/ RDFS

O RDF possibilita a codificação, o intercâmbio e a reutilização de metadados; consiste de uma aplicação do XML que impõe limitações estruturais proporcionando métodos para expressão semântica.

O RDF possibilita meios para a publicação de vocabulários legíveis por pessoas e computadores, além de proporcionar a reutilização e a distribuição dos metadados entre comunidades de informação.

O modelo de dados RDF é composto por: Recursos, Propriedades, Valores

37

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

O *recurso* é um objeto identificável de forma única por uma URI - *Uniform Resource Identifier*.

Os atributos ou características dos recursos são chamadas "*propriedades-tipo*" e expressam os relacionamentos dos "*valores*" associados aos recursos correspondentes.

Os valores podem ser "*atômicos*" (caracteres de texto, números, etc) ou outros recursos, os quais, por sua vez, podem ter suas próprias propriedades.

A coleção destas propriedades que se refere ao mesmo recurso é chamada "*descrição*".

38

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Esquema de uma descrição RDF com recursos, propriedades e valores

39

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

```
< ? xml : namespace ns = "http://www.w3c.org/rdf/rdf" prefix = "RDF" ? >
< ? xml : namespace ns = "http://www.eb.ufmg/CI" prefix = "CI" ? >
<RDF : RDF >
  <RDF: Description RDF : HREF : http://URI_documento_1>
    <CI: Creator> João Silva </CI: Creator>
  </RDF : description>
</RDF : RDF>
```

40

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

O RDFS define propriedades características ou restrições de valores das propriedades-tipo de uma descrição RDF. O *namespace* XML identifica os esquemas RDF.

O RDFS estende o RDF em situações que demandam maior número de informações (metadados) para a identificação dos dados.

O RDFS é expresso em RDF e permite verificar a integridade dos dados, estabelecendo uma hierarquia de tipos, tais como classes, subclasses, relacionamentos, restrições, dentre outras.

41

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Pode-se necessitar de mais informações sobre um recurso, "por exemplo, se uma propriedade representa um aniversário, é preciso verificar que o valor é uma data e não um tipo de animal" (DODDS et al., 2001).

O RDF/RDFS é processado através de *parsers* que fornecem como resultado uma tripla de dados; os dados da tripla podem ser utilizados por uma aplicação através de uma *API-Application Program Interface*;

Essa implementação tem grande relevância para os propósitos da Web Semântica, pois a estrutura semântica (ontologia) pode ser expressa em RDFS.

Exemplo de parser RDF *on-line* do W3C em:
<http://www.w3.org/RDF/Validator/>

42

ARQUIVOLOGIA
Introdução a Informática

Prof. Maurício B. Almeida

Web-Services

Definições para Web-services

“Web-services são interfaces que descrevem uma coleção de operações que são acessíveis via rede através de mensagens XML padronizadas, atendem a tarefas específicas e fornecem detalhes para acesso a um serviço. [...] Ocultam detalhes de implementação dos serviços de forma que podem ser utilizadas independente da plataforma de hardware, software e da linguagem de programação.” (IBM)

“Um web-service é uma aplicação programável que é acessada através de protocolos padrão da Internet e combinam as melhores características do desenvolvimento baseados em componentes e da Web.” (Microsoft)

43

ARQUIVOLOGIA
Introdução a Informática

Prof. Maurício B. Almeida

“Web-Services são aplicações de software que podem ser descobertas, descritas e acessadas a partir de XML ou protocolos padrão da Web em intranets, extranets e Internet.”(Daconta, 2003)

Onde:

“...são aplicações de software...”, são aplicações disponíveis na Internet que executam funções específicas;

“...que podem ser descobertas, descritas e acessadas a partir de XML ou protocolos padrão da Web...”, são construídos com sintaxe XML, enfatizam interoperabilidade; utilizam o http como protocolo de envio de mensagens; utilizam outras tecnologias como: o UDDI, o ebXML, o WSDL e o SOAP

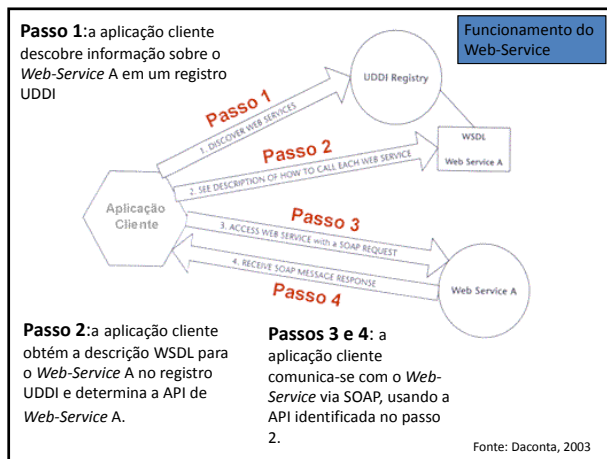
“...em intranets, extranets e Internet”, podendo ser serviços públicos disponíveis na Internet, ou em Intranets privadas

44

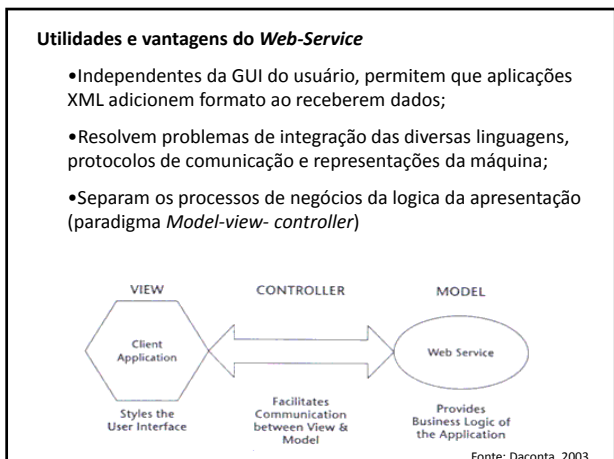
Descobrir (UDDI, ebXML, registries)	UDDI-Universal Description, Discovery and Integration e ebXML
Descrever (WSDL)	WSDL Web-Service Definition Language
Acessar (SOAP)	SOAP-Simple Object Access Protocol
XML	XML como formato de intercâmbio
Camada de Comunicação (HTTP, SMTP, outros)	Protocolos padrão Web para comunicação

Esquema de Web-Service em camadas

45



46



47

ARQUIVOLOGIA
Introdução a Informática

Prof. Maurício B. Almeida

Componentes do Web-service

1)SOAP-Simple Object Access Protocol (ACESSO)

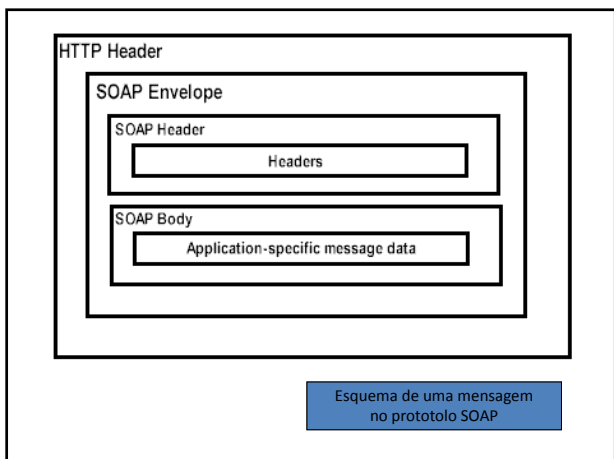
O Protocolo SOAP é um “envelope” que empacota as mensagens XML enviadas por HTTP entre diferentes Web-Services.

Uma aplicação envia um pedido SOAP-request a um Web-Services e este retorna uma resposta em forma de um SOAP-response;

Uma mensagem SOAP contém os seguintes elementos:

- Um envelope SOAP que embrulha a mensagem;
- Uma descrição de como os dados estão codificados;
- Um Body-SOAP que contém mensagens que são interpretáveis pela aplicação;

48



49

```
<SOAP-ENV:Envelope
  xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
  SOAP-ENV:encodingStyle="http://schemas.xmlsoap.org/soap/encoding/"
  <SOAP-ENV:Body>
    <m:GetLastTradePrice xmlns:m="Some-URI">
      <symbol>DIS</symbol>
    </m:GetLastTradePrice>
  </SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```

Exemplo de um SOAP-request

O objetivo desse código é obter o último preço de um "DIS".
 Na linha 3, um atributo do envelope SOAP, *encodingStyle*, mostra a codificação da mensagem, para que o *Web-Services* possa lê-la.
 Nas linhas de 4 a 7 está o *body*-SOAP da mensagem que empacota a informação da aplicação (chamada *GetLastTradePrice* no corpo).
 Um *Web-Services* recebe esta informação, processa o pedido no corpo e pode assim retornar um SOAP-response.

50

```
<SOAP-ENV:Envelope
  xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
  SOAP-ENV:encodingStyle="http://schemas.xmlsoap.org/soap/encoding/"
  <SOAP-ENV:Body>
    <m:GetLastTradePriceResponse xmlns:m="Some-URI">
      <Price>34.5</Price>
    </m:GetLastTradePriceResponse>
  </SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```

Exemplo de um SOAP-response

O código mostra a SOAP-response para o pedido de preço do estoque da mercadoria DIS;
 A mensagem é sintaticamente a mesma consistindo de um envelope que empacota a mensagem, descreve seu estilo de codificação na linha 3;
 As linhas de 4 a 9 mostram o conteúdo da mensagem empacotado no *body*-SOAP;
 A mensagem do corpo é um pouco diferente. Nas linhas 5 até 7, vê-se a mensagem é inserida na marcação *GetLastTradePriceResponse*, como o resultado do preço mostrado na linha 6.

51

ARQUIVOLOGIA
Introdução a Informática

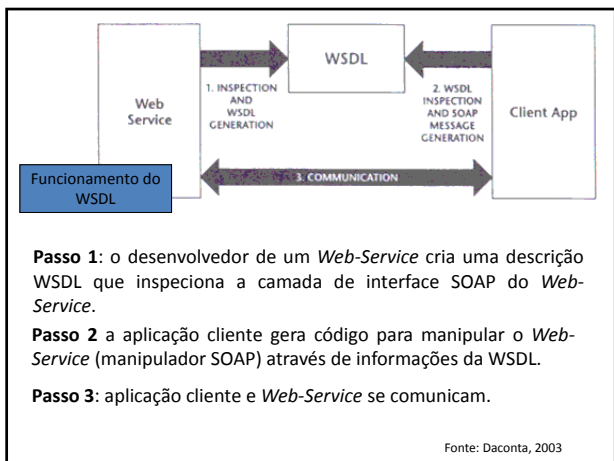
Prof. Mauricio B. Almeida

2) WSDL-Web-Service Definition Language (DESCRIÇÃO)

O WSDL descreve os detalhes da comunicação e as mensagens específicas que podem ser enviadas via SOAP;
 O WSDL também utiliza a sintaxe da XML;
 Descreve a informação operacional:

- Onde o serviço está localizado;
- O que ele faz;
- Como invocá-lo;

52



53

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

3) UDDI-Universal Description, Discovery and Integration (DESCOBERTA)

A UDDI funciona como um "catálogo telefônico" para os *Web-Services*; permite descobrir *Web-Services* da mesma forma que os serviços de busca de recursos na rede;
 Registram informações públicas sobre os *Web-Service* e tipos de serviço de forma que aplicações podem achar informações sobre o funcionamento daquele *Web-Service*.
 A UDDI é composta por três componentes: páginas amarelas, páginas brancas e páginas verdes;
 Uma opção do UDDI é o padrão ebXML, voltado para transações entre empresas (B2B).

54

UDDI REGISTRY

WHITE PAGES

- Business Name
- Contact Information
- Description

YELLOW PAGES

- Services and Products
- Industry Codes
- Geographic Index

GREEN PAGES

- ebusiness Rules
- Service Descriptions
- WSDL

Esquema do registro UDDI

Páginas brancas: informações de contato da instituição (diferentes línguas, contatos por e-mail, números de telefone, links que descrevem);

Páginas amarelas: classificam as atividades através de taxonomias padrão (tipos de informação que os serviços fornecem);

Página verdes: contem a documentação técnica sobre os serviços oferecidos (como utilizar o *Web-Service*, lista regras de negócio e especificação sobre como invocar os *Web-Service*, ou seja, a WSDL).

55

ARQUIVOLOGIA
Introdução a Informática

Prof. Maurício B. Almeida

Estruturas de organização e ontologias

Conceitos preliminares – taxonomias, tesaurus, ontologias e breve comparativo

1) Taxonomias

De forma simples, uma taxonomia é uma forma de classificar ou categorizar um conjunto de coisas em uma hierarquia;

Uma taxonomia é uma classificação de entidades informacionais em forma de hierarquia, de acordo com relações presumidas a partir dos objetos do mundo real que tais entidades representam;

O tipo de relação em uma taxonomia é *subclasse-de* ou *superclasse-de*, denotando a uma **relação semântica fraca**;

56

Exemplo de taxonomia na tela do sistema Windows

57

ARQUIVOLOGIA
Introdução a Informática

Prof. Maurício B. Almeida

2) Tesaurus

Um tesaurus é um instrumento especializado e pós-coordenado de controle de vocabulário que fornece informações para indexadores e pesquisadores sobre a linguagem adotada no sistema de informação.

Auxiliando com respeito a qual termo usar, o tesaurus proporciona melhorias de qualidade na representação e recuperação da informação.

O controle de vocabulário é conseguido pela determinação do significado de descritores, pela ligação entre termos sinônimos através de relações de equivalência (BT, NT e RT) e pela solução de polissemias.

58

Fragmento – Thesaurus for graphic materials

59

ARQUIVOLOGIA
Introdução a Informática

Prof. Maurício B. Almeida

3) Ontologias

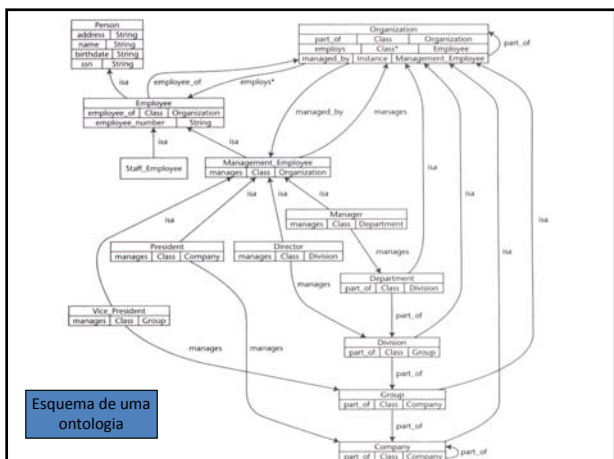
As ontologias são estruturas que possibilitam uma compreensão comum e compartilhada de um domínio do conhecimento e, assim, têm papel importante no intercâmbio de informações.

Possibilitam a comunicação entre os agentes envolvidos no processo de recuperação da informação ao reduzir as diferenças conceituais.

Podem ser consideradas extensões das taxonomias, contendo vocabulários e seus significados, com semântica expressiva e bem definida;

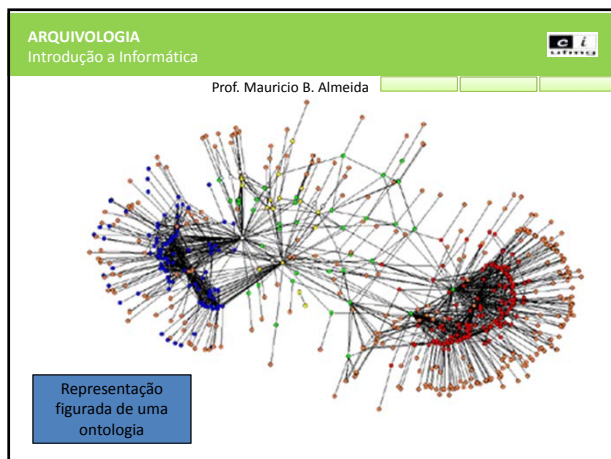
São as estruturas utilizadas para expressar semântica legível por máquinas no contexto da Web Semântica.

60



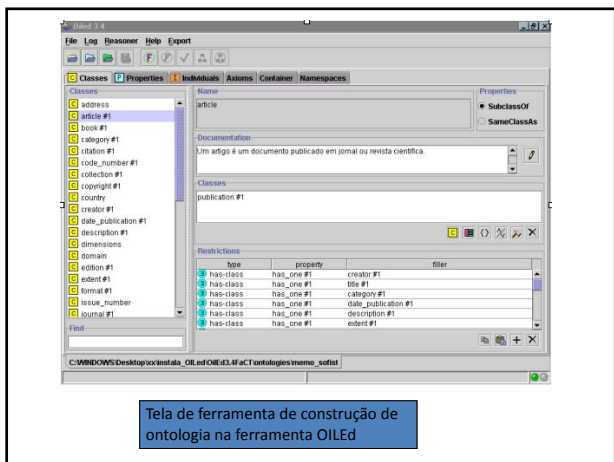
Esquema de uma ontologia

61



Representação figurada de uma ontologia

62



Tela de ferramenta de construção de ontologia na ferramenta OILED

63

ARQUIVOLOGIA
Introdução a Informática

Prof. Maurício B. Almeida

4) Tesouros e ontologias

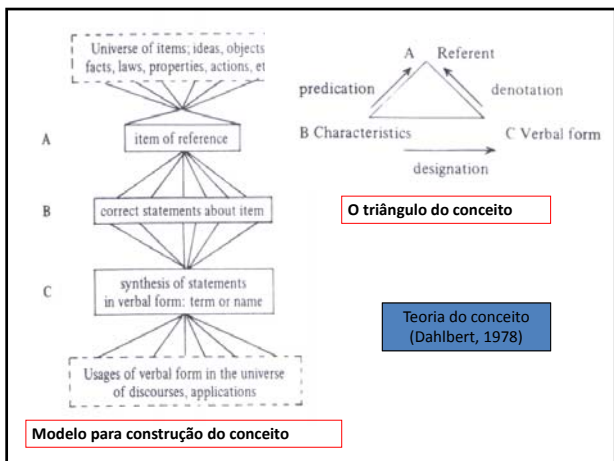
As duas estruturas de representação e organização diferem em vários aspectos, dentre eles:

- Capacidade de expressar semântica formal;
- Possibilidade e tipos de relações;
- Possibilidade de inferência automáticas;

Algumas demonstrações podem facilitar o entendimento das principais diferenças como:

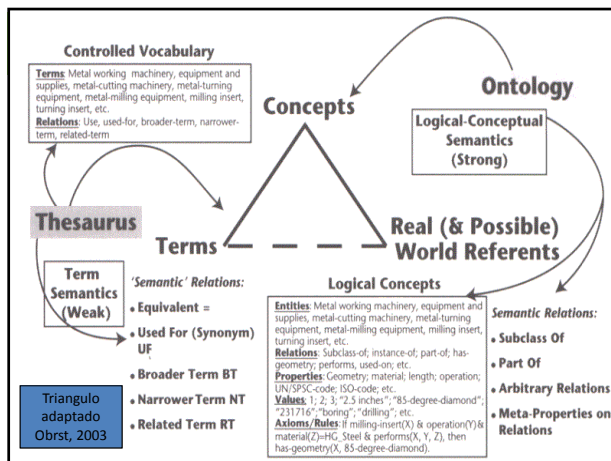
- Triângulo de significação;
- Espectro semântico de diversas estruturas;

64



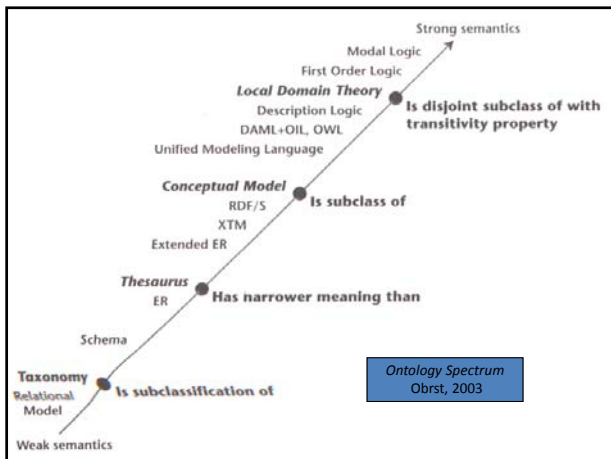
Modelo para construção do conceito

65



Triângulo adaptado Obrst, 2003

66



67

ARQUIVOLOGIA
Introdução a Informática

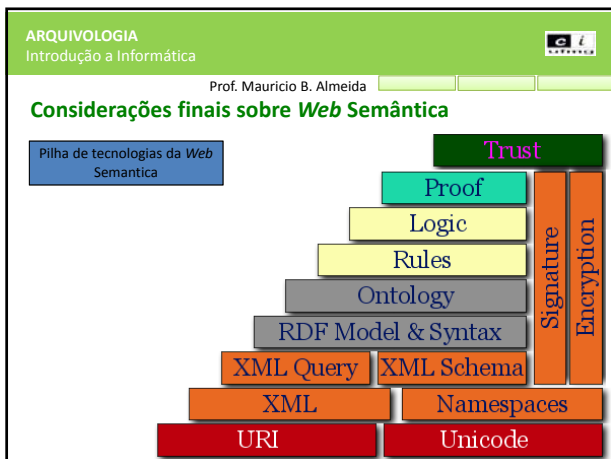
Prof. Mauricio B. Almeida

5) Considerações finais sobre ontologias

Em função de sua maior capacidade de expressão semântica e de outros fatores, a ontologia é a estrutura utilizada para representar significado na Web Semântica. Além disso:

- Ontologia é um termo controverso com diversas definições possíveis;
- Existem diversas metodologias para desenvolvimento;
- Existem diversas ferramentas para desenvolvimento;
- Ontologias são escritas em diversas linguagens;

68



69

ARQUIVOLOGIA
Introdução a Informática

Prof. Mauricio B. Almeida

Usos e aplicações

A tecnologia Web é promissora em novas aplicações a um custo baixo. Alguns exemplos são:

- Intranets e extranets;
- Portais corporativos;
- Bibliotecas digitais;
- Ferramentas para intercâmbio e integração de dados;
- Web Semântica corporativa;
- ...

70