

硕士研究生课程 《智能信息处理》

# 语义网基础理论

研究生教材

宋炜, 张铭. 语义网简明教程

高等教育出版社, 2004

大连海事大学信息科学技术学院

## 第1章

# 语义网导论

万维网

语义网

相关的技术

Internet

# 网络服务

WWW

E-mail

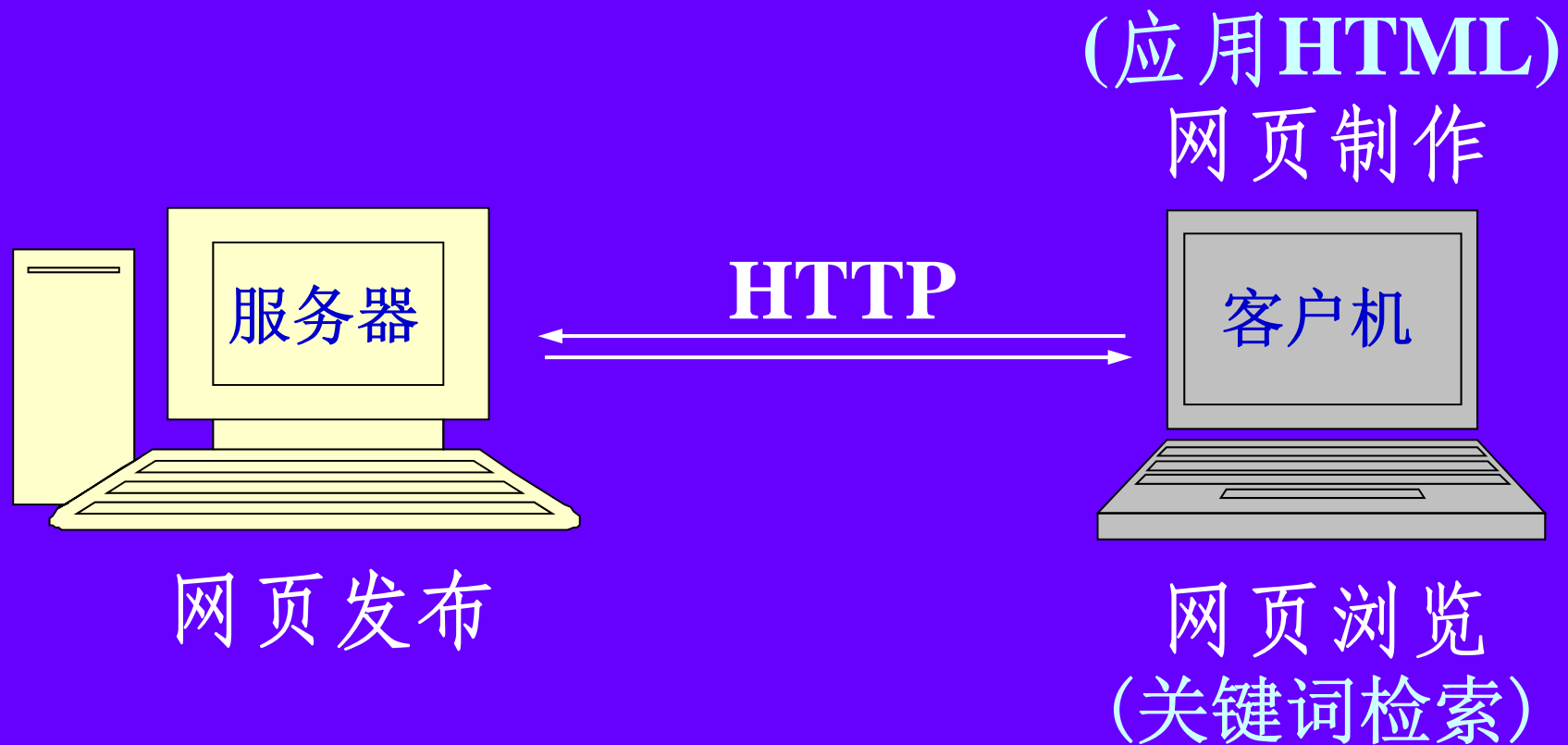
FTP

BBS

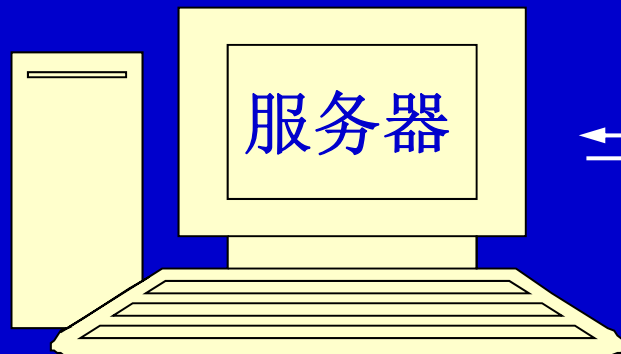
网络通信



# WWW(World Wide Web, 万维网)



## 万维网的局限(P2)

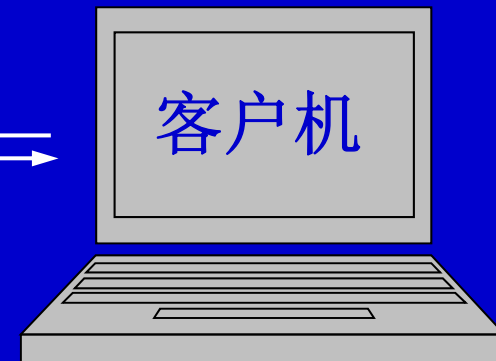


网页发布  
海量信息  
缺乏元数据

HTTP



缺乏语义  
(应用HTML)  
网页制作



网页浏览  
(关键词检索)  
检索效果差

万维网



HTML网页



信息有格式描述  
信息无语义描述



不能语义匹配

只能语法匹配



关键字检索

效果差

## 万维网根本缺陷



# HTML: 信息的显示格式描述(P5)

```
<html>
```

```
  <head>
```

```
    <title> 红楼梦 </title>
```

```
  </head>
```

```
  <body>
```

```
    <h1> 书名:《红楼梦》 </h1>
```

```
    <h2> 作者: 曹雪芹 </h2>
```

```
    <h3> 描述: 中国四大古典名著之一 </h3>
```

```
  </body>
```

```
</html>
```

标记 元数据

## XML: 信息的语义描述 (P5)

```
< ? XML VERSION="1.0"  
      ENCODING="GB2312"  
      standalone="no" ? >
```

```
< ! DOCTYPE Book SYSTEM  
      http://db.pku.edu.cn/Book.dtd >
```

```
<书>
```

```
<标题> 《红楼梦》 </标题>
```

```
<作者> 曹雪芹 </作者>
```

```
<描述> 中国四大古典名著之一 </描述>
```

```
</书>
```

元数据  
标记  
/



(一篇文献)  
信息资源

描述

(作者/时间/主题)  
关键信息点

元数据

(metadata)

内容元数据  
管理元数据  
参考元数据  
载体元数据



XML

信息的内容  
信息的历史  
信息的链接  
信息的外观

HTML

## 元数据的重要作用 (P4)

- 描述信息资源
- 增强各种资源之间的可交换性
- 提供资源的可访问性
- 沟通不同的数据格式

HTML网页



格式显式  
语义隐式



人可理解格式  
人可理解语义

机器可理解格式  
机器不理解语义

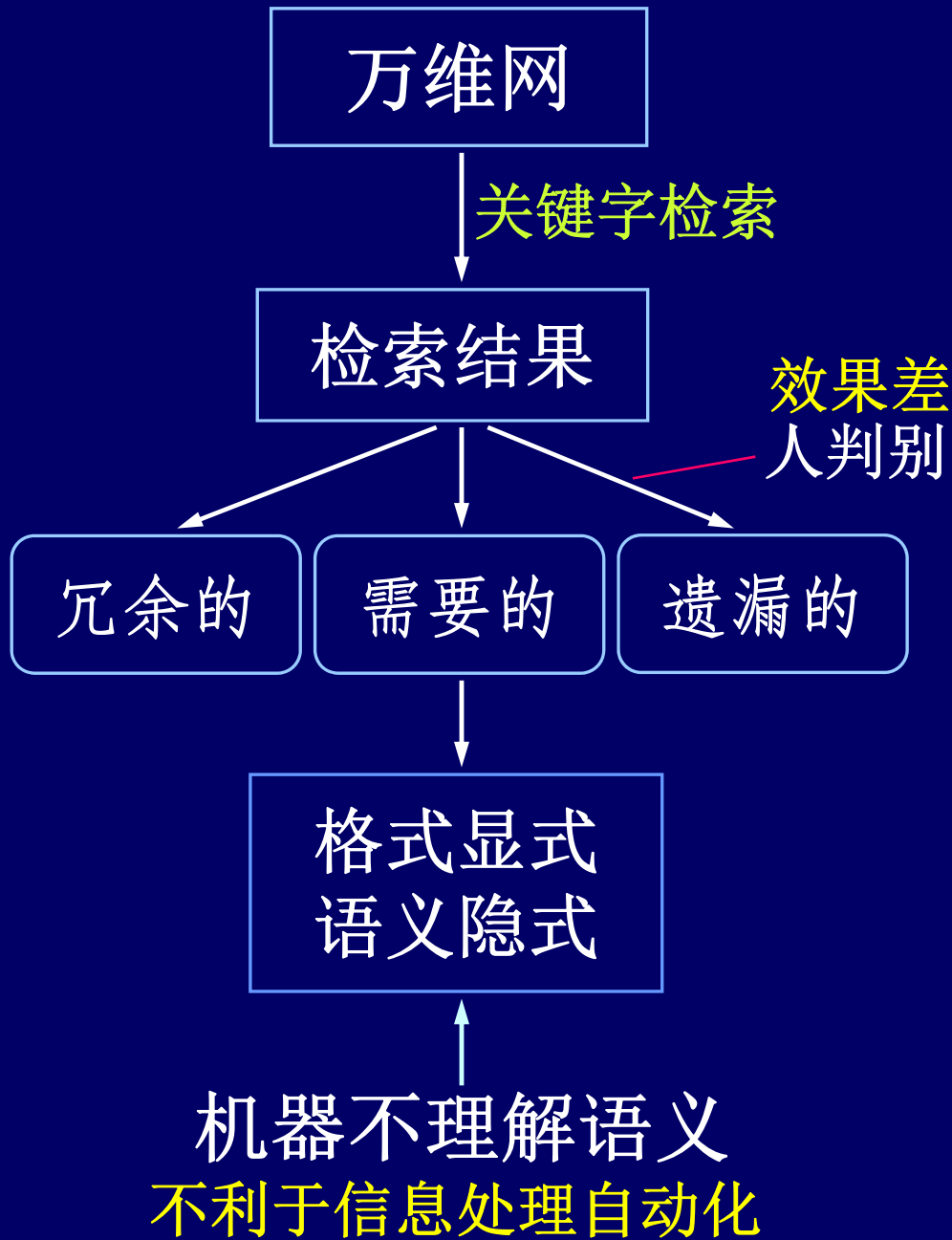
信息处理自动化



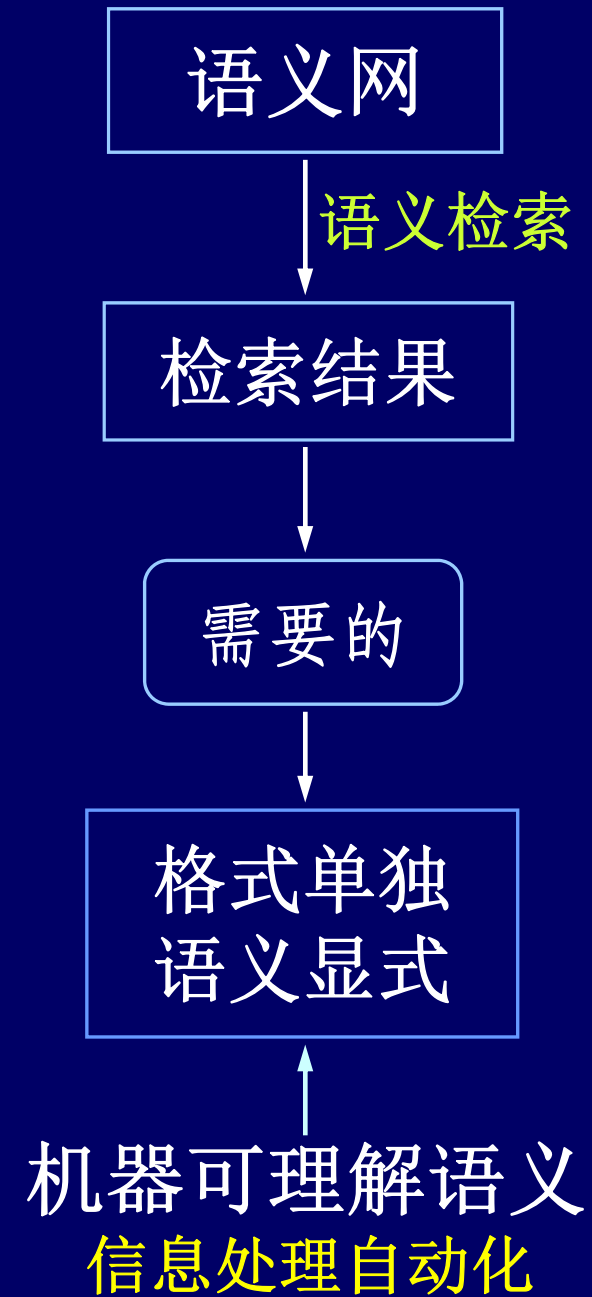
更严重缺陷

```
<html>
  <head>
    <title> 红楼梦 </title>
  </head>
  <body>
    <h1> 书名:《红楼梦》 </h1>
    <h2> 作者:曹雪芹 </h2>
    <h3> 描述:中国四大古典名著之一 </h3>
  </body>
</html>
```

## Tim Berners-Lee, 1990



## Tim Berners-Lee, 2000



## 语义网的目标 (P2)

为网上信息提供具有计算机可理解的语义，  
满足Agent对异构、分布信息的有效检索和访问，  
实现网上信息资源在语义层上的全方位互联，  
实现网上信息的更高层、基于知识的智能应用。



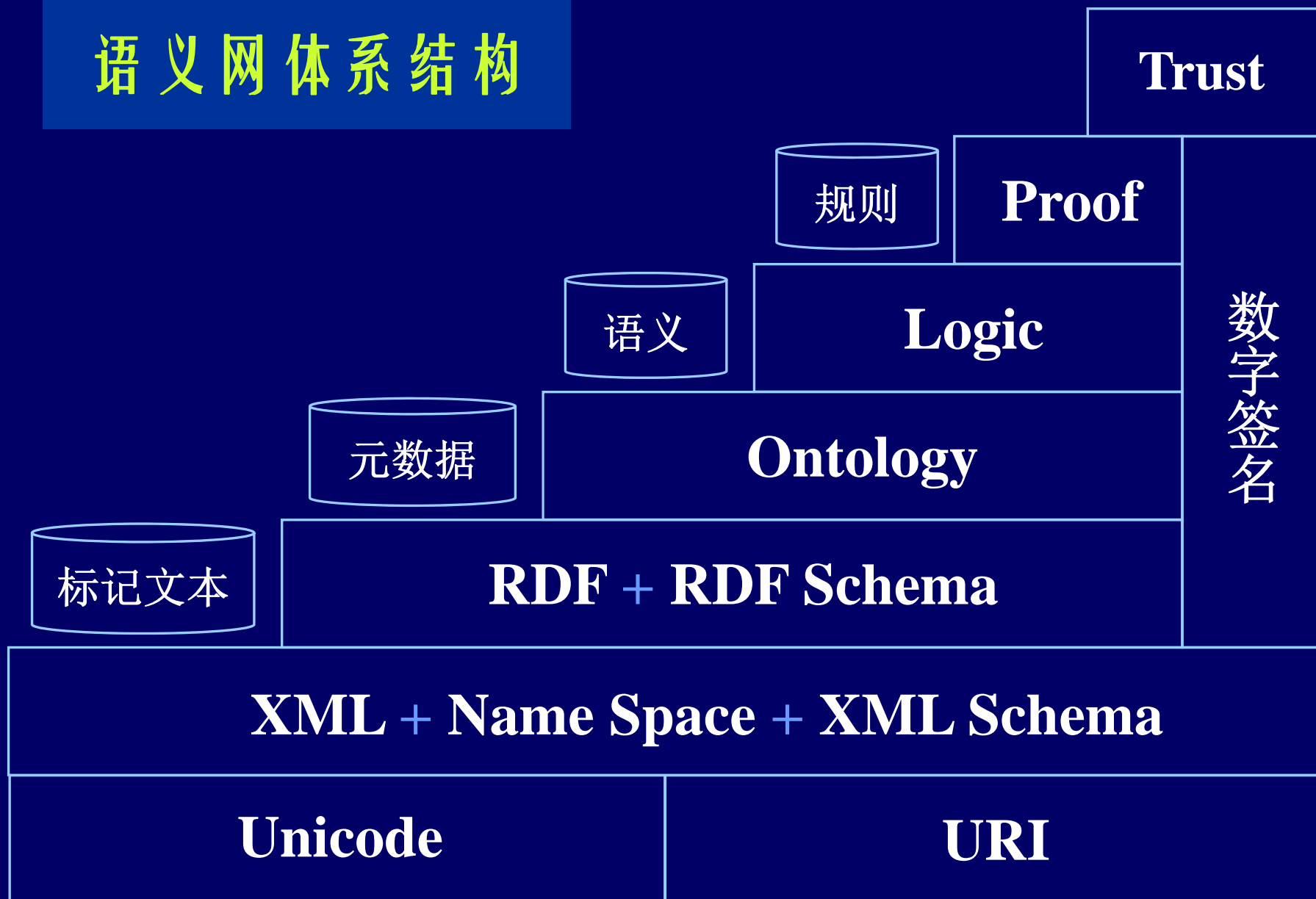
网上信息具有语义  
有效检索/访问/应用

## 语义网设计原则(P10)

---

- 所有资源都能用**URI**来标识
- 资源与链接可以有类型
- 部分/片断/不完整信息是容许的
- 信息不必是绝对真的
- 能支持、反映信息的变化与演化
- 最小设计原则

# 语义网体系结构





第七层 信任

第六层 证明

第五层 逻辑

第四层 本体

第三层 **RDF**

第二层 **XML**

第一层 **Unicode/URI**

信息(可信的)

信息(基于本体的推理结果)

信息(被本体中的关系所联系)

信息(被本体中的概念所标记)

信息(作为属性用于描述资源)

信息(被元数据所标记)

信息(统一字符/统一定位)

信息

# 超文本标记语言

(HyperText Markup Language, HTML)

- 当前主导的Web标记语言
- 用户不能定义自己的标记
- 缺乏对内容的描述
- 结构不规范



# 可扩展的标记语言

(Extensible Markup Language, XML)

- 信息内容与信息显示分离
- 信息显示使用CSS或XSL ■
- 可扩展的, 即用户可自定义标记
- 文档结构: 文档头+文档体+文档尾
- 支持以统一标准定义自描述数据
- 支持异构数据的交换
- 支持多源数据的集成
- 是简单的、开放的



## 层叠样式单 CSS

---

(cascading stylesheet)

定义浏览器显示文本的格式

## 可扩展样式单语言 XSL

---

(extensible stylesheet language)

实现创建目录、编辑索引等

# 资源描述框架

(resource description framework, RDF)

原理



方式

三元组

{ 属性, 资源, 属性值 }

图形化

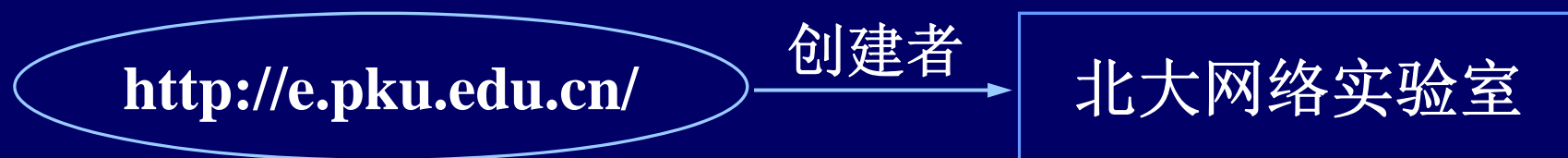


XML

```
< rdf: Description about = “资源” >  
  <属性> 属性值 </属性>  
< /rdf: Description >
```

## 举例：RDF描述资源 <http://e.pku.edu.cn/>

{ 创建者, [<http://e.pku.edu.cn/>], “北大网络实验室” }



```
< ? xml version = "1.0"? >
```

```
< rdf: RDF xmlns: rdf = "http://www.w3.org/
                                1999/02/22-rdf-syntax-ns#"
        xmlns: s = "http://description.org/schema/" >
  < rdf: Description about = "http://e.pku.edu.cn" >
    <s: Creator> 北大网络实验室 </s: Creator>
  < /rdf: Description >
```

```
< /rdf: RDF >
```



# 本体(Ontology)

本体定义

本体构造准则

本体语言OWL

## 本体的定义 (Gruber, 1993)

**An ontology is an explicit specification of a conceptualization.**

本体是概念模型的一种显式规范说明。

本体是共享的概念化的、明确的、形式化的规范说明。



## 本体的作用(P17)

- 领域知识显式化、明确化
- 领域知识与操作性知识分离
- 领域知识的一致理解

## 构造本体的准则 (Gruber, 1995)

1. 清晰性 (Clarity)
2. 完全性 (Completeness)
3. 一致性 (Coherence)
4. 可扩展性 (Extendibility)
5. 最小承诺 (Minimal ontological commitment)  
最小编码偏好 (Minimal encoding bias)

## 本体语言应具有的特点(P18)

- 语法和语义丰富
- 可描述半结构化的知识
- 使用共有的、一致的字典
- 提供一种领域知识

Web Ontology Language

# W3C语义网行动计划

```
graph TD; A[W3C语义网行动计划] --> B(RDF核心工作组); A --> C(万维网本体工作组); B --> D[RDF标准]; C --> E[OWL标准];
```

**RDF核心工作组**

**RDF标准**

**万维网本体工作组**

**OWL标准**

# 课间休息



再会

