

由机器网络到信息(文档)网络

战德臣

哈尔滨工业大学 教授·博士生导师
教育部大学计算机课程教学指导委员会委员



Research Center on **I**ntelligent
Computing for **E**nterprises & **S**ervices,
Harbin **I**nstitute of **T**echnology

由机器网络到信息(文档)网络

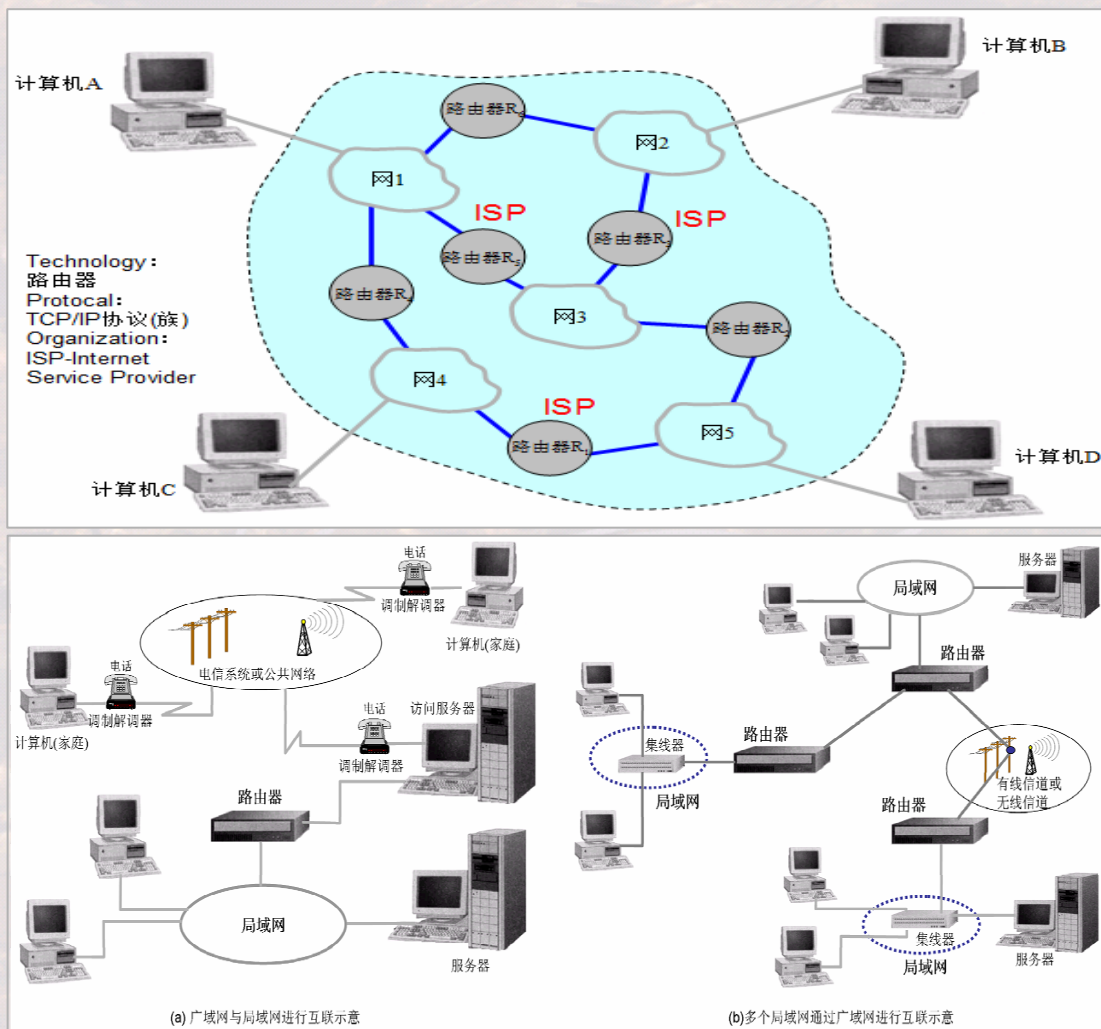
(1)回顾机器网络?

未来互联网与网络化社会

群体互动网络与社会网络

信息网络

机器网络



国际互联网

互联网

广域网

局域网

通讯与联网

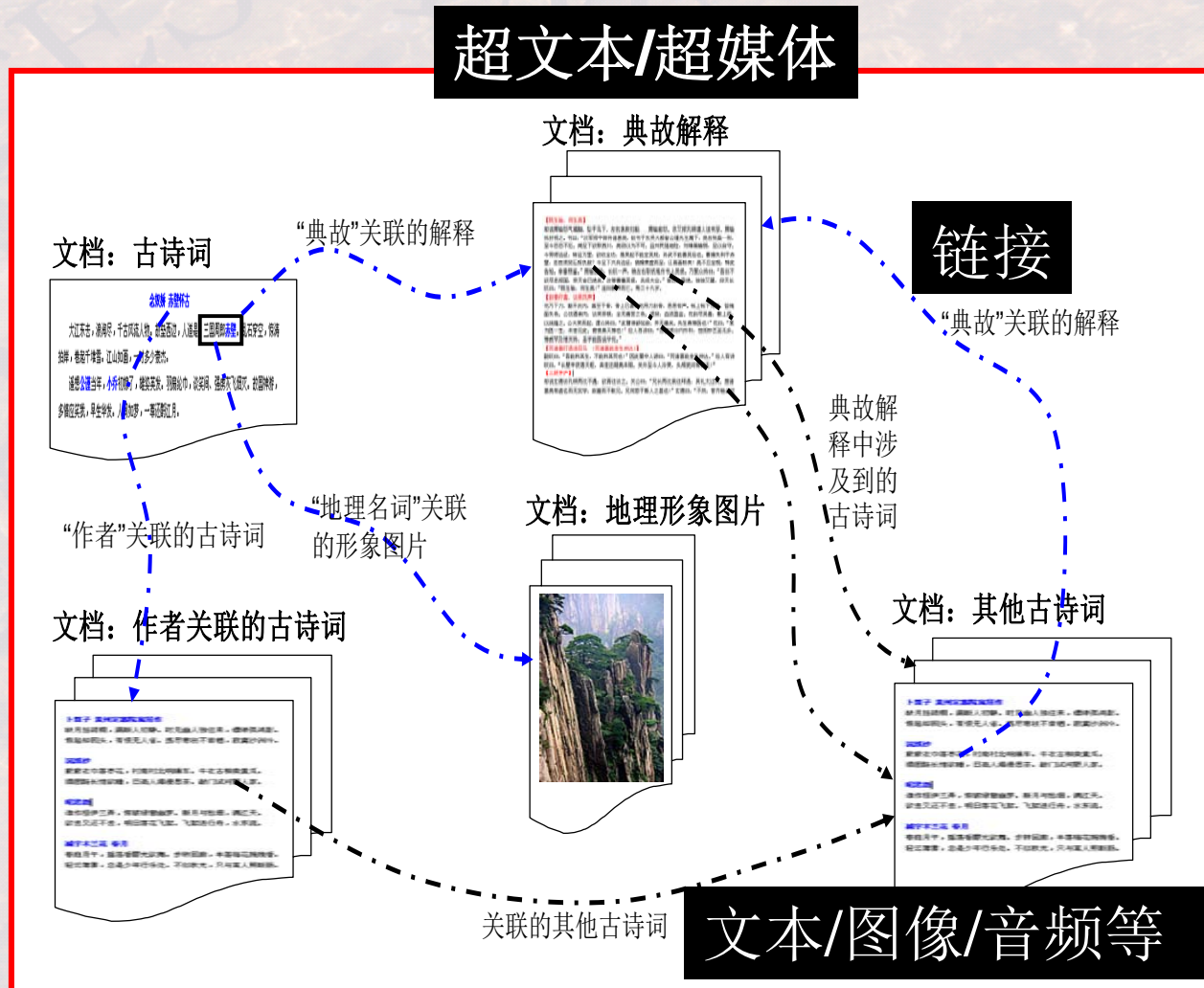
ICES 战德臣 教授



超文本/超媒体 = 普通文本/媒体 + 链接

超文本/超媒体

- 技术方面怎样实现文档之间的关联?
- 内容方面怎样建立文档之间的关联?



由机器网络到信息(文档)网络

(4)网页与HTML?

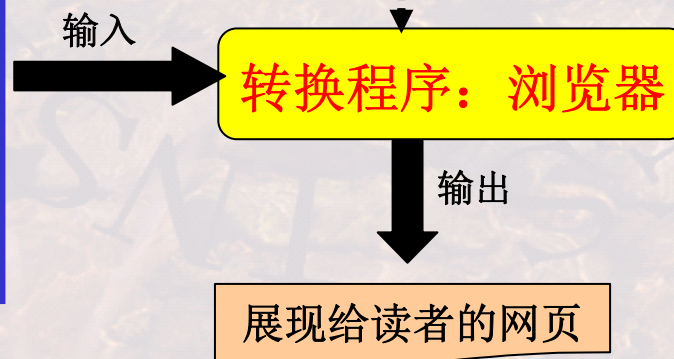
HTML与网页

- ◆ **HTML**: 超文本标注语言--一系列预定义的标记的集合
- ◆ HTML文档由HTML标记和相应的文本组成。
- ◆ 一个HTML文档又称为一个Web Page(网页)

Hello.html

```
<HTML>
  <HEAD>
    <TITLE>
      Hello World!
    </TITLE>
  </HEAD>
  <BODY>
    世界你好!
  </BODY>
</HTML>
```

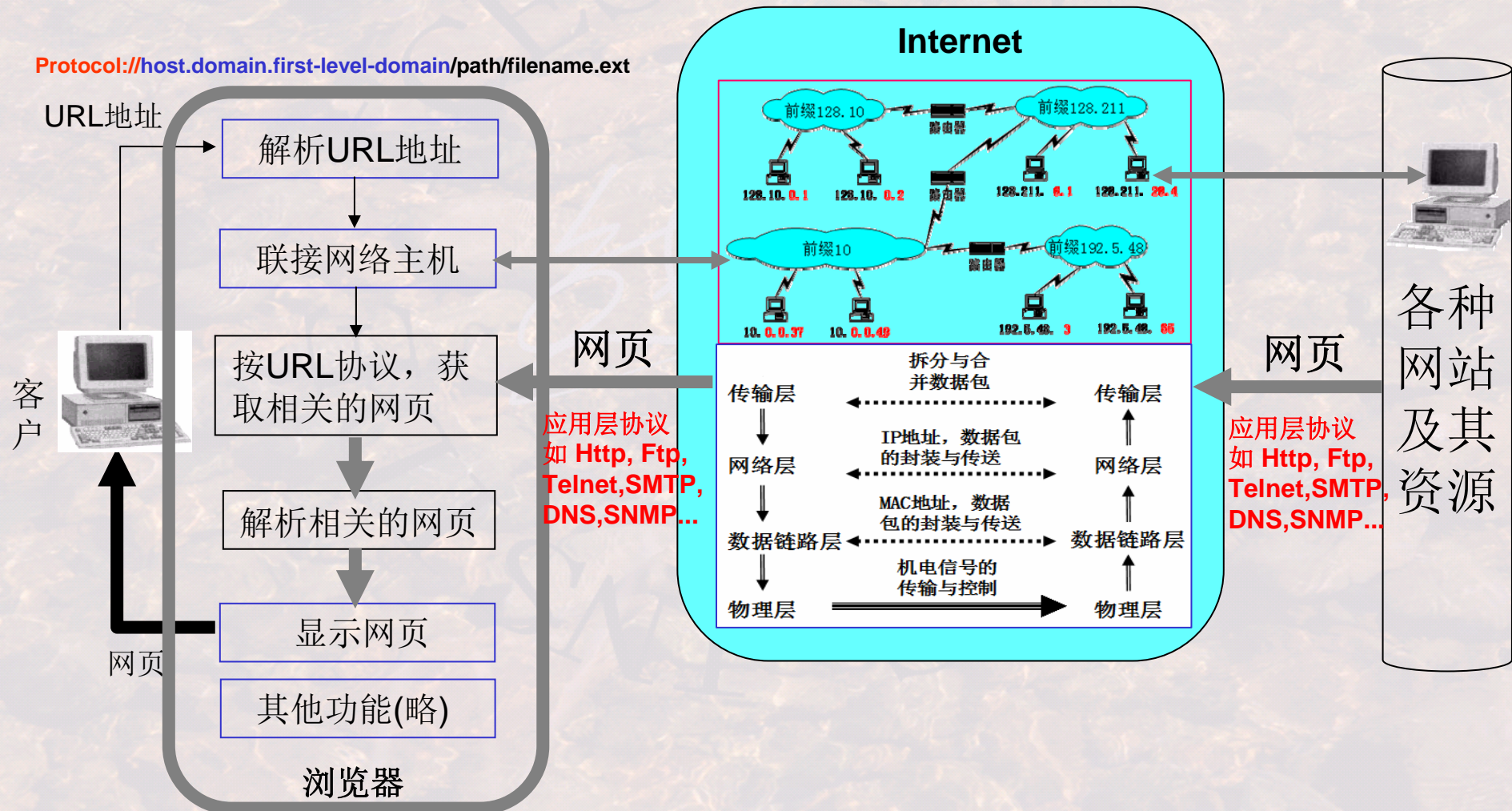
HTML语言：一套预定义的标记的集合



由机器网络到信息(文档)网络

(5)浏览器?

浏览器的核心功能解析

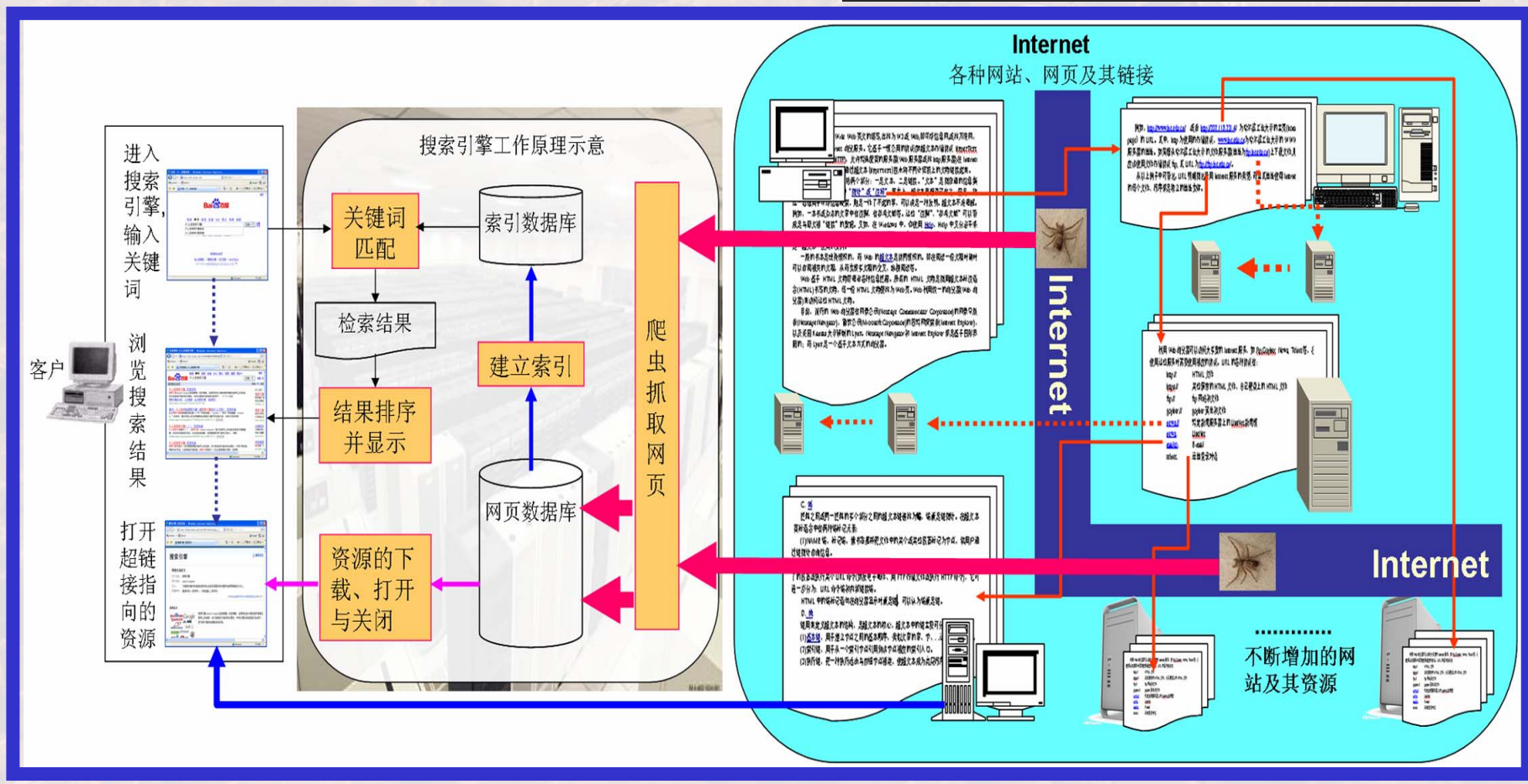


由机器网络到信息(文档)网络

(7)搜索引擎?

通用搜索引擎的功能解析

当大规模网络上的网页被链接起来后，用户怎样找到最符合期望的文档？



标记语言: HTML

战德臣

哈尔滨工业大学 教授·博士生导师
教育部大学计算机课程教学指导委员会委员



Research Center on **I**ntelligent
Computing for **E**nterprises & **S**ervices,
Harbin **I**nstitute of **T**echnology

标记语言

标记的起始

带“/” --标记的结束

<标记>文本</标记>

待处理的对象，纯文本或多媒体

待处理对象的某一方面的性质，由相应处理程序解析或利用

标记的嵌套

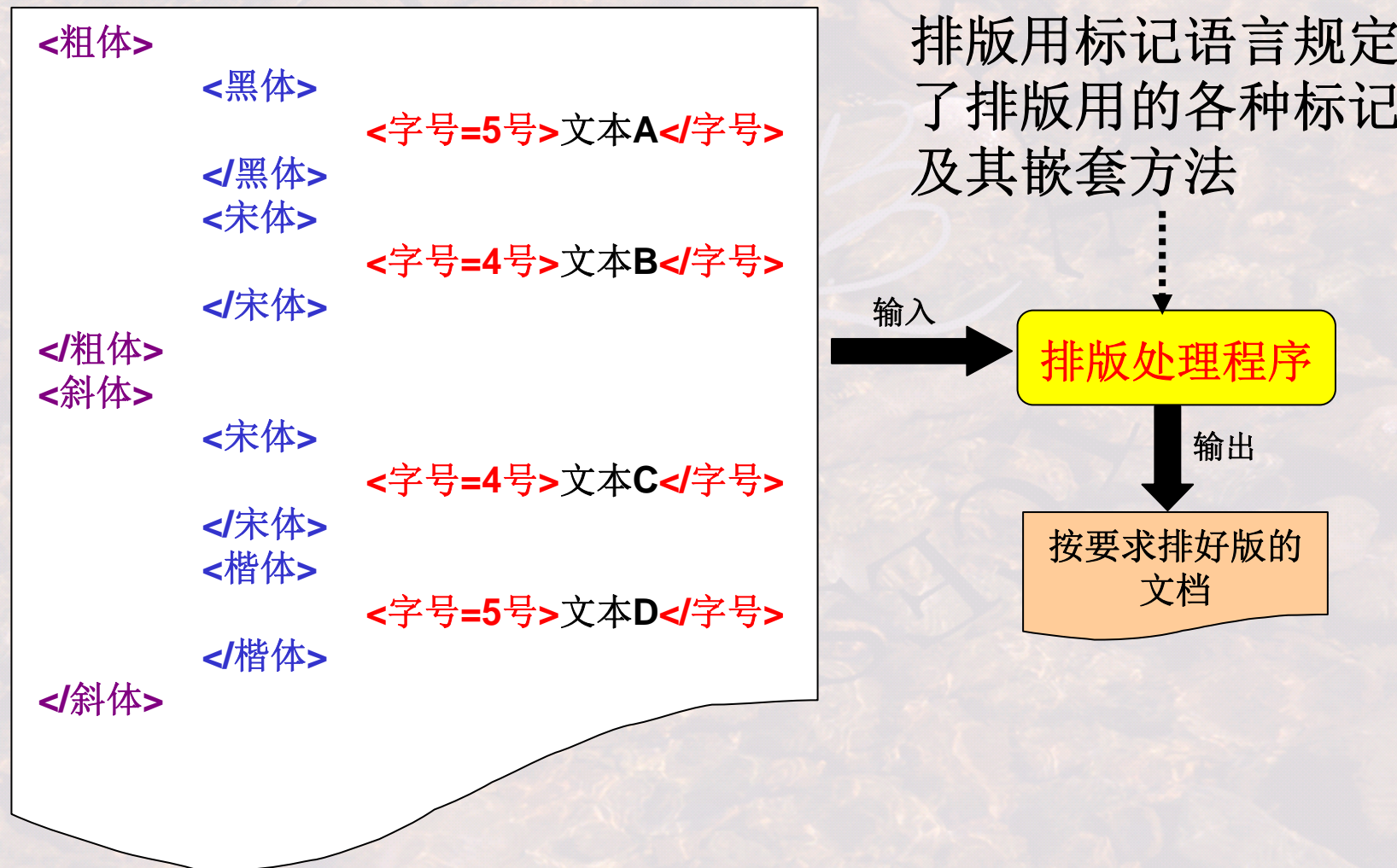
<标记F> <标记K> <标记M> 文本 </标记M> </标记K> </标记F>

<标记F>
 <标记K>
 <标记M> 文本 </标记M>
 </标记K>
</标记F>



~~<标记F>
 <标记K>
 <标记M> 文本 </标记K>
 </标记M>
</标记F>~~

排版用标记语言示意



信息交换用标记语言示意

<订单>

<订单号> 010202-10 </订单号>

<产品名> IPAD </产品名>

<数量> 20 </数量>

<价格> 4000 </价格>

<总额> 80000 </总额>

</订单>

<订单>

<订单号> 010202-20 </订单号>

<产品名> ThinkPAD </产品名>

<数量> 10 </数量>

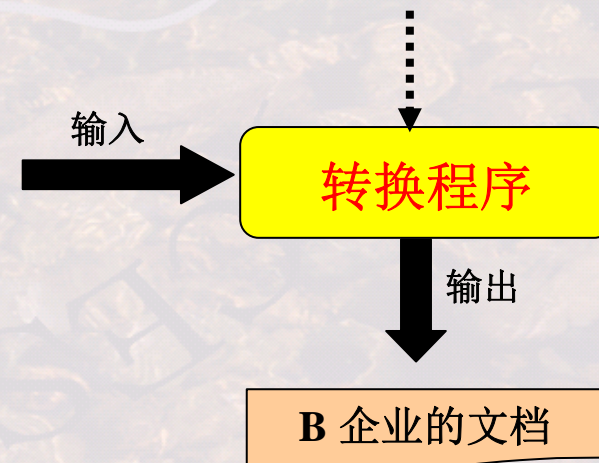
<价格> 4000 </价格>

<总额> 40000 </总额>

</订单>

A企业的文档

信息交换用标记语言
规定了大家共用的各种
标记及其嵌套方法



标记语言: HTML

(4)HTML是什么?

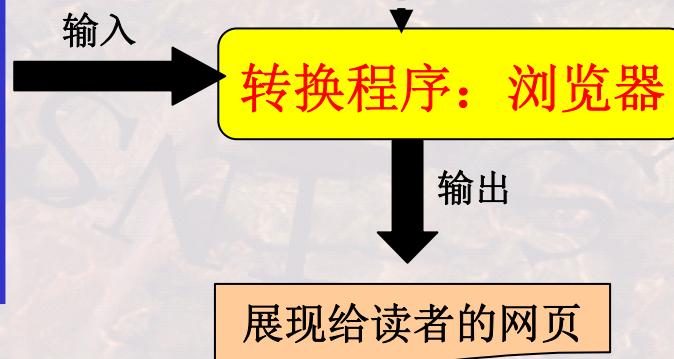
HTML与网页

- ◆**HTML**: 超文本标注语言--一系列预定义的标记的集合
- ◆HTML文档由HTML标记和相应的文本组成。
- ◆一个HTML文档又称为一个Web Page(网页)

Hello.html

```
<HTML>
  <HEAD>
    <TITLE>
      Hello World!
    </TITLE>
  </HEAD>
  <BODY>
    世界你好!
  </BODY>
</HTML>
```

HTML语言：一套预定义的标记的集合



标记语言: HTML

(4)HTML是什么?

HTML文档的基本结构

```
<HTML>
  <HEAD>
    <TITLE>
      Hello World!
    </TITLE>
  </HEAD>
  <BODY>
    世界您好!
  </BODY>
</HTML>
```

HTML 文件的起始标记, 表示下面的是 HTML 文件

文件头开始标记

标题开始标记

文件实际的标题

标题结束标记

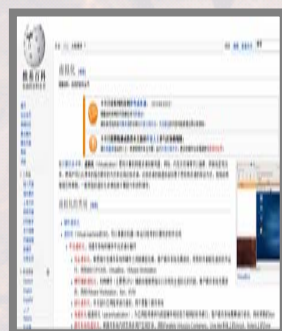
文件头结束标记

主体开始标记

文件主体内容

主体结束标记

HTML 文件结束标记



HTML文档的链接标记

文本

■绝对地址链接

Our Product

■本机内相对地址链接

Our Product

■文档内部段落链接

链源锚(链接者)

** 屏幕显示内容**

Link to another section of this document

链宿锚(被链接段落需要定义)

屏幕显示内容

Beginning of new section

HTML文档的完整处理过程示意

《黄鹤楼》

作者: 崔颢

昔人已乘黄鹤去,此地空余黄鹤楼。
黄鹤一去不复返,白云千载空悠悠。
晴川历历汉阳树,芳草萋萋鹦鹉洲。
日暮乡关何处是,烟波江上使人愁。

- 注解:
- 1、黄鹤楼: 故址在湖北武昌县, 民国初年被火焚毁, 传说古代有一位名叫费文的仙人, 在此乘鹤登仙。也有人作昔人已乘白云去。
 - 2、悠悠: 久远的样子。
 - 3、历历: 清晰、分明的样子。
 - 4、鹦鹉洲: 在湖北省武昌县西南, 根据后汉书记载, 汉黄祖担任江夏太守时, 在此大宴宾客, 有人献上鹦鹉, 故称鹦鹉洲。

(a) 纯粹的文本



(b) 与(c)中文件存储在同一个机器同一目录下的图像文件HuangHeLou.JPG

```
<HTML>
<HEAD> <TITLE>
《<A HREF = "HuangHeLou 介绍.html">黄鹤楼</A>》
</TITLE> </HEAD>
<BODY>
<P>作者: 崔颢<BR></P>
<P>昔人已乘黄鹤去,此地空余<A HREF="#注解 1">黄鹤楼。
</A><BR>黄鹤一去不复返,白云千载空<A HREF="#注解 2">悠悠</A>。<BR>晴川<A HREF="#注解 3">历历</A>汉
阳树,芳草萋萋<A HREF="#注解 4">鹦鹉洲</A>。<BR>
日暮乡关何处是,烟波江上使人愁。<BR></P>
<P>注解: <BR> <A NAME="#注解 1">1、黄鹤楼: 故址在
湖北武昌县, 民国初年被火焚毁, 传说古代有一位名叫费文的
仙人, 在此乘鹤登仙。也有人作昔人已乘白云去。<BR>
<A NAME="#注解 2">2、悠悠: 久远的样子。<A><BR> <A
NAME="#注解 3">3、历历: 清晰、分明的样子。<A><BR> <A
NAME="#注解 4">4、鹦鹉洲: 在湖北省武昌县西南, 根据后
汉书记载, 汉黄祖担任江夏太守时, 在此大宴宾客, 有人献上
鹦鹉, 故称鹦鹉洲。<A><BR></P>
<IMG SRC = "imgDir/HuangHeLou.jpg"/>
</BODY>
</HTML>
```

(c) 加入HTML标记的文档

《黄鹤楼》

作者: 崔颢

昔人已乘黄鹤去,此地空余黄鹤楼。
黄鹤一去不复返,白云千载空悠悠。
晴川历历汉阳树,芳草萋萋鹦鹉洲。
日暮乡关何处是,烟波江上使人愁。

注解:

- 1、黄鹤楼: 故址在湖北武昌县, 民国初年被火焚毁, 传说古代有一位名叫费文的仙人, 在此乘鹤登仙。也有人作昔人已乘白云去。
- 2、悠悠: 久远的样子。
- 3、历历: 清晰、分明的样子。
- 4、鹦鹉洲: 在湖北省武昌县西南, 根据后汉书记载, 汉黄祖担任江夏太守时, 在此大宴宾客, 有人献上鹦鹉, 故称鹦鹉洲。



(d) 按照HTML文档展现的超文本, 其中示意的链接是可以被执行的

黄鹤楼原址在湖北省武昌蛇山黄鹤矶头, 始建于三国时代东吴黄武二年(223年)。《元和郡县图志》记载: 孙权始筑夏口故城, “城西临大江, 江南角因矶为楼, 名黄鹤楼。”是为了军事目地而建。而据《极恩录》记载说原为辛氏开设的酒店。

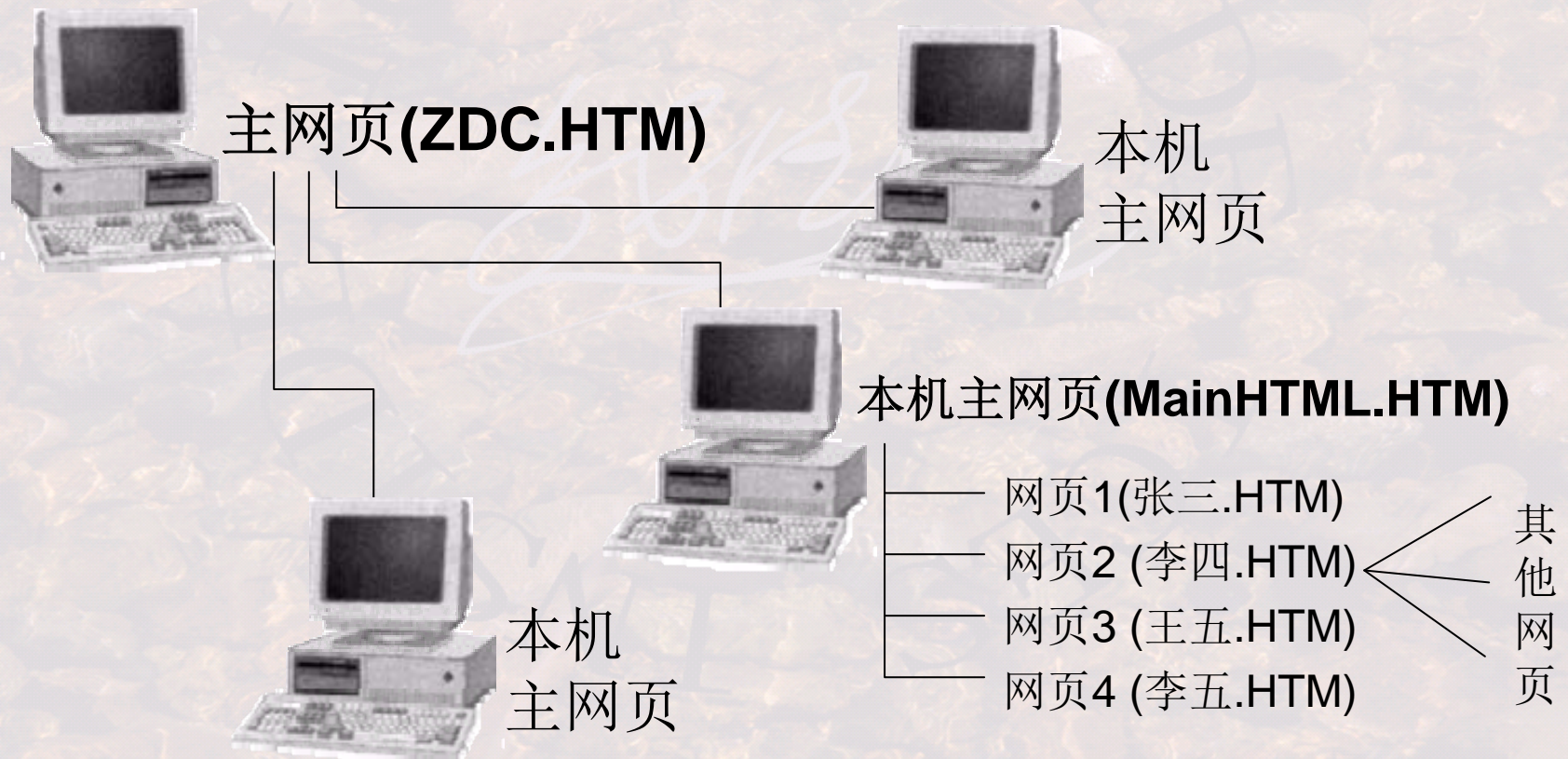
唐永泰元年(765年)黄鹤楼已具规模, 然而战火频繁, 黄鹤楼屡建屡废, 仅在明清两代, 就被毁7次, 重建和维修了10次。有“国运昌则楼运盛”之说。最后一座建于同治七年(1868年), 毁于光绪十年(1884年)。遗址上只剩下清代黄鹤楼毁灭后唯一遗留下来的一个黄鹤楼铜铸楼顶。

黄鹤楼从北宋至20世纪50年代, 还曾作为道教的名山圣地, 是吕洞宾传道、修行、教化的道场。《道藏 历世真仙体道通鉴》言: “吕祖以五月二十日登黄鹤楼, 午刻升天而去。故留成仙圣迹。”《全真晚坛功课经》中称其黄鹤楼头留圣迹。

古黄鹤楼“凡三层, 计高九丈二尺, 加铜顶七尺, 共成九九之数”。

(e) 存储在/Doc下面的HTML文档HuangHeLou介绍.html(此处隐含了其中的标记)

网站 vs. 门户网站



标记语言：由HTML到XML

战德臣

哈尔滨工业大学 教授·博士生导师
教育部大学计算机课程教学指导委员会委员



Research Center on **I**ntelligent
Computing for **E**nterprises & **S**ervices,
Harbin **I**nstitute of **T**echnology

标记语言: 由HTML到XML

(1)标记语言的本质是什么?



标记语言

标记的起始

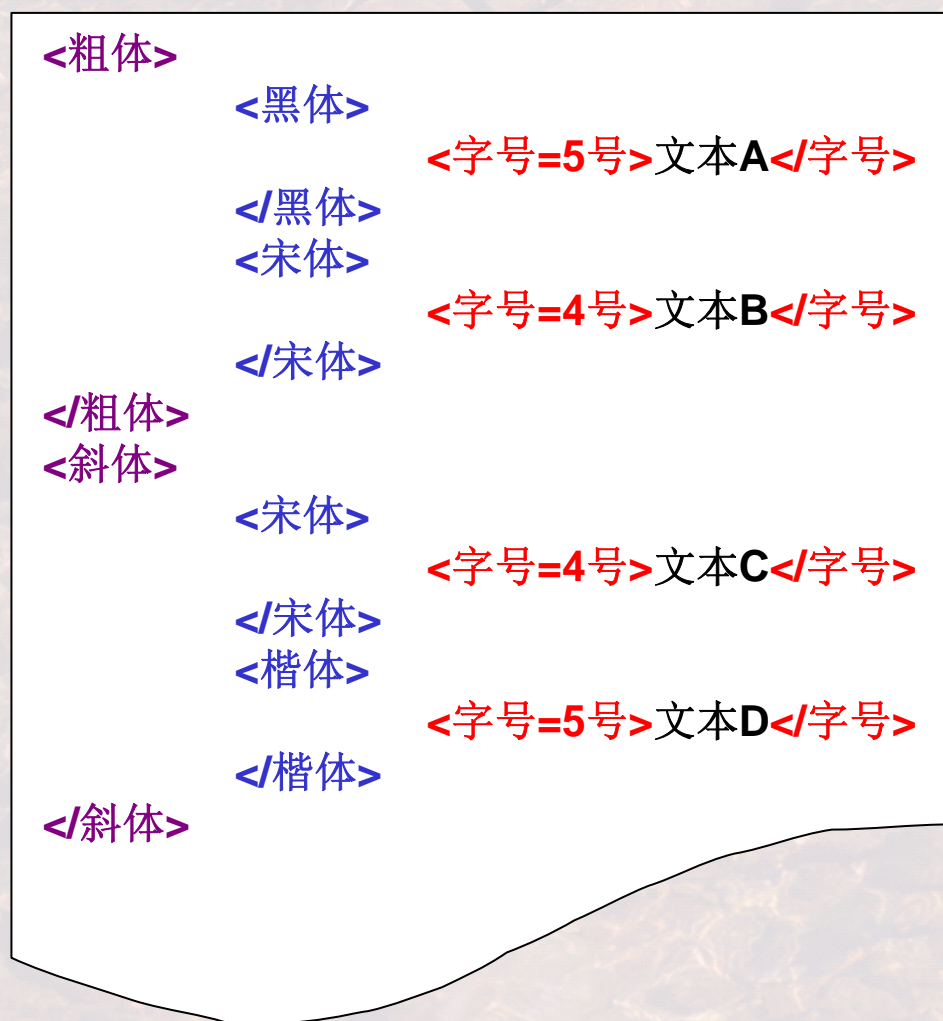
带“/” --标记的结束

<标记>文本</标记>

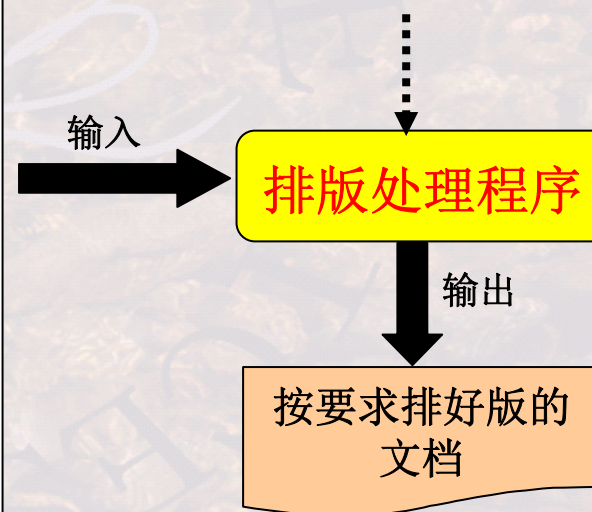
待处理的对象，纯文本或多媒体

待处理对象的某一方面的性质，由相应处理程序解析或利用

排版用标记语言示意



排版用标记语言规定了排版用的各种标记及其嵌套方法



标记语言:由HTML到XML

(7)XML是什么?



XML: eXtensible Markup Language, 扩展标记语言

HTML vs. XML

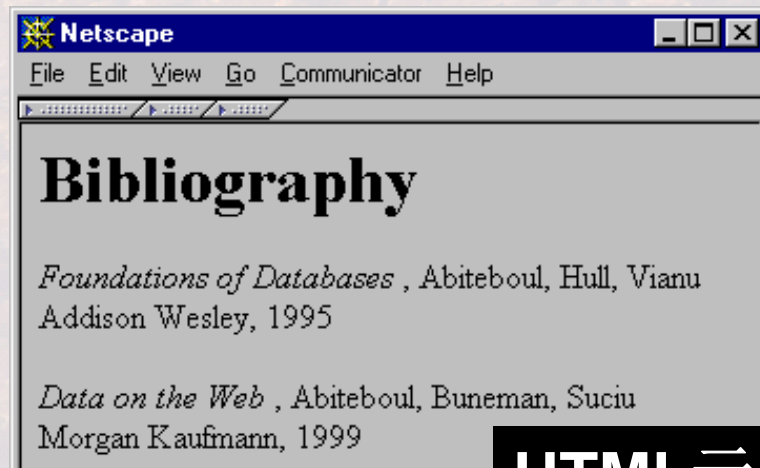
- 都是一种标记语言
- HTML语言用于格式化并显示数据；**XML**语言用于描述数据和数据结构，专为存储、传输、数据交换和数据处理而设计。
- HTML有预定义的格式化标记；**XML**允许用户定义自己的标记和自己的文档结构。
- XML不是替代HTML，而是对 HTML 的补充。
- XML是W3C(万维网联盟)的推荐标准。

标记语言:由HTML到XML

(7)XML是什么?

XML: eXtensible Markup Language, 扩展标记语言

HTML vs. XML



HTML示例

```
<h1> Bibliography </h1>
<p> <i> Foundations of Databases </i>
    Abiteboul, Hull, Vianu
    <br> Addison Wesley, 1995
<p> <i> Data on the Web </i>
    Abiteboul, Buneman, Suciu
    <br> Morgan Kaufmann, 1999
```

```
<bank>
  <account>
    <account_number> A-101 </account_number>
    <branch_name> Downtown </branch_name>
    <balance> 500 </balance>
  </account>
  <depositor>
    <account_number> A-101 </account_number>
    <customer_name> Johnson </customer_name>
  </depositor>
</bank>
```

XML示例

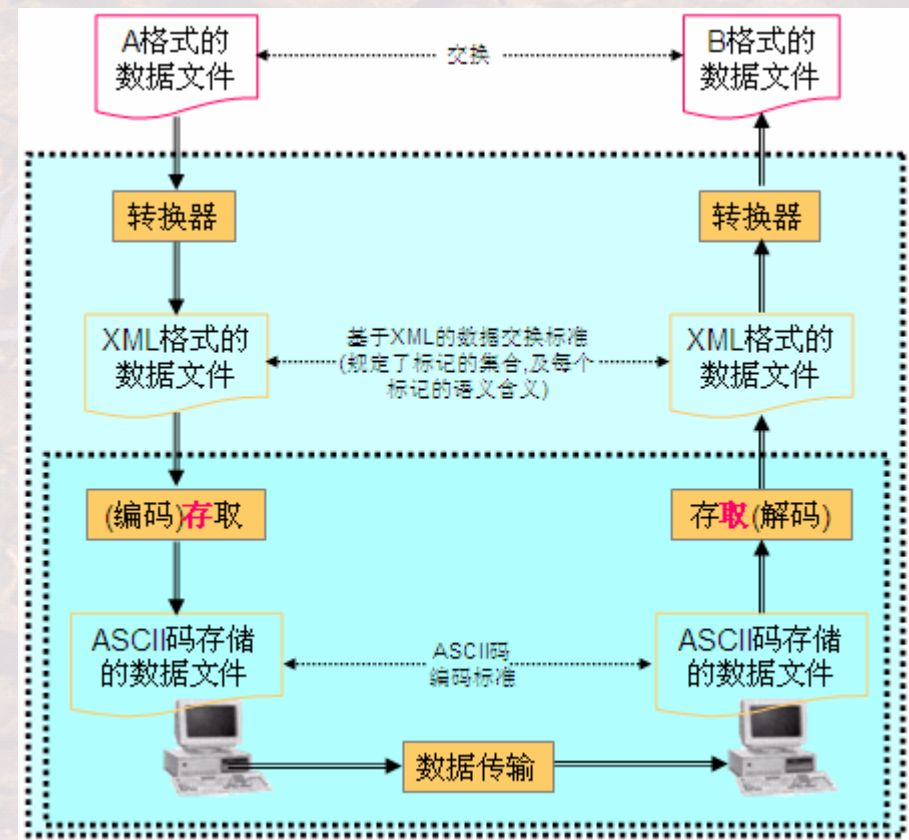
```
<book>
  <title> Foundations of DBs </title>
  <author> Abiteboul </author>
  ...
  <year> 1995 </year>
  <price> 55 </price>
  <currency> USD </currency>
</book>
```


标记语言:由HTML到XML

(8)XML的用途举例?

基于XML的信息交换

- ◆XML文件通常采用ASCII码标准进行存储，因此其易于解读。
- ◆很多数据交换标准都是基于XML制定，即确定各行各业中的XML文档的各种标记及其含义。
- ◆如此可设计基于XML的转换器，实现任意数据结构的数据交换。



基于XML的新语言及其处理

一套标记的集合
及其处理规则

新语言: 书写规范及其语法结构

<标记> 文本 **</标记>**

待处理的对象

待处理对象的某一方面的性质，由相应处理程序解析或利用

编辑器

新语言源程序

编译器

(可执行)计算机
语言源程序