

基于知识库及专家系统的智能信息交换系统

高毅如

(中国纺织信息中心, 北京 100742)

摘要: 研究了网站服务器之间智能信息的交换, 探讨了使用人工智能技术及专家系统构建智能信息交换模型, 并使用扩展符号语言描述非结构化数据结构模型, 构造智能交换体系, 实现信息资源和知识资源的智能共享, 通过实例说明了该方法具有可行性。
关键词: 信息智能交换; 产业联盟; 信息孤岛; 资源共享; 扩展符号语言

Intelligent Information Exchange System Based on Knowledge Base and Expert System

GAO Yi-ru

(China Textile Information Center, Beijing 100742)

【Abstract】 This paper researches on the intelligent information exchanges among Web servers. It discusses how to compose net-union intelligent information exchange model with artificial intelligence and expert system. It also uses extensible markup language(XML) to describe the non-structure data model, which composes the intelligent exchange system. The technology described is feasible and has been proved by some applications.

【Key words】 intelligent information exchange; industry ally; information island; resource sharing; extensible markup language(XML)

因特网兴起于上个世纪 90 年代, 目前已经非常普遍和发达, 但是相应形成了互联网上诸多的信息孤岛和知识孤岛, 尤其突出的是行业上下游产业链的信息和知识不能智能共享。所以, 在互联网技术方面产生了一场新的革命, 国外常用“下一代 Internet”、“Internet2”、“下一代 Web”等词语来称呼相关技术, 在国内称为网格、语义网(semantic Web)、知识管理(knowledge management)、知识本体(ontology)、智能主体(agents)、信息网格、知识网格、一体化智能信息平台等, 它们共同的研究目的是消除信息孤岛和知识孤岛。

1 智能交换系统的拓扑结构

智能交换系统的拓扑结构如图 1 所示。系统平台主要由联盟网站、智能服务器和浏览器组成。其中, 联盟网站是需要进行信息交换的、目前独立运行的服务网站, 它们都有各自的用户群体; 联盟用户(图 1 所示的浏览器)是各联盟网站的不同用户; 智能服务器是实现图 1 中联盟网站信息智能交换的服务器, 它具有信息自动接收、智能分配、信息发送以及记录系统运行状态、保存工作过程、纠正传输错误、维护系统运行等功能。

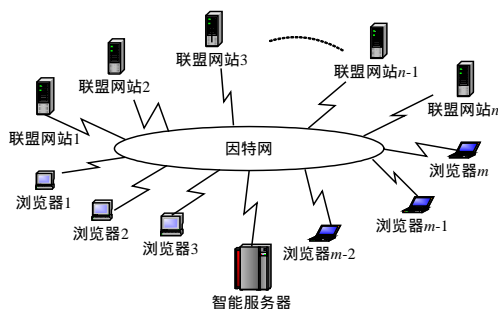


图 1 智能交换系统网络结构

从图 1 可以看出, 联盟网站及用户接入因特网原方式并没有改变, 只是增加了智能服务器。智能服务器可以放在外网, 但是为了安全, 不被黑客攻击, 一般放置在内部网段。

2 智能交换系统的实现

2.1 方案的确定

研究发现, 人工智能技术及专家系统正向着广义知识表达式、综合知识库系统(知识库、方法库、模型库)、搜索推理技术以及面向应用的开发工具等方向发展。由于实现信息智能交换的信息知识库所包含的内容广泛、数据类型多样, 因此建造信息知识库必须采用非结构化的数据结构模型。在选择非结构化的数据结构的过程中, 自然把目标锁定在扩展符号语言(extensible markup language, XML)上。它由 W3C 国际组织制定, 并于 1999 年确定为国际推荐标准, 此后很快得到广泛的承认。

XML 是典型的非结构化数据结构描述语言。它采用树型结构, 可实现多维数据描述, 树型结构中的节点内容可用多种数据类型表示, 同时具有良好的可读性。W3C 组织还制定了功能十分强大的 XPath 查询语言标准, 并明确要求, XML 语言可以采用递归的方法实现搜索功能。这极大地简化了程序的设计工作。

目前, XML 语言已经获得广泛的应用, 而且主流浏览器都支持这一标准。经过反复论证, 本文认为 XML 符号语言完全能够满足构建信息知识库的基本要求, 采用 XML 可方便地构建知识表达式、描述广义知识、构建知识模型, 是建造知识库较为理想的方案。

作者简介: 高毅如(1965 -), 女, 高级工程师, 主研方向: 网络技术, XML 技术

收稿日期: 2006-12-20 **E-mail:** gyr@ml.ctei.gov.cn

2.2 运行模型的建立

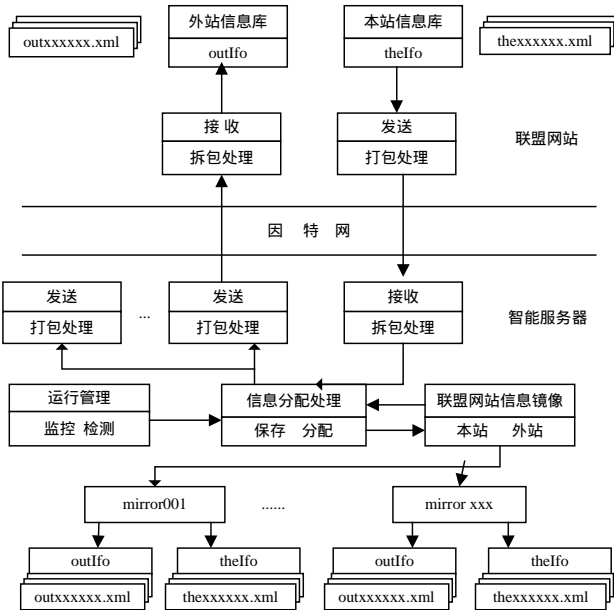


图2 智能信息交换及存储运行

图2清楚地表示出了联盟网站与智能服务器之间信息交换的过程，从中可以看出，运行的关键部件是智能服务器，它所包含的功能模块主要有访问协议、发送接收、信息分析、数据访问、数据仓库等。其中，运行管理模块将依次按时地检测各网站的运行状态，控制信息的发送和接收。另外，还为智能服务器设计了状态检测、网站工作显示以及各项操作命令等辅助模块，以方便操作者及时观察或干预系统的运行。

智能服务器的设计充分考虑了它的可靠性、安全性、保密性以及今后的可扩充性。所以在总体设计上，数据的读、写操作均采用本站程序完成，绝对禁止人员的远程操作^[1]。

2.3 建立信息分类文件及信息知识表达式

信息的统一分类是实现信息智能交换的基础。根据不同的联盟范畴可以形成不同的信息分类文件(classIfo.xml)，信息分类文件用XML描述，制定编码规则，为每个信息分类进行编码。分类可以单层，也可以多层，根据需要而定。

信息分类的每个末级类都可以定义不同的信息知识表达式，如新闻类信息的信息知识表达式为标题、内容、发布人、发布网站、发布日期、流媒体信息、信息性质等内容，知识表达式构成的信息知识节点不仅包括说明型的文字信息，还可以以附件的形式包括图片、影片、动画等多媒体信息，并可以集成信息处理及表现方法^[2]。

2.4 数据目录设计

图3清楚地显示了信息的分类存储。联盟网站一侧有本站信息库(theIfo)和外站信息库(outIfo)，本站信息库存储的信息为本站发布生成的信息知识资源，外站信息库存储的是接收联盟网其他网站发布的共享信息知识资源。在智能服务器一端分别对联盟网站的本站信息库和外站信息库进行镜像存储。镜像存储就是为每个联盟网站设置镜像文件夹mirrorXXX，内含一个本站文件夹和一个外站文件夹，这2个文件夹中保存信息文件与其对应网站的信息文件完全相同。它主要用于数据统计、全文检索以及信息恢复和信息维护^[3]。

图4给出联盟网站和智能服务器信息存储的目录设计。

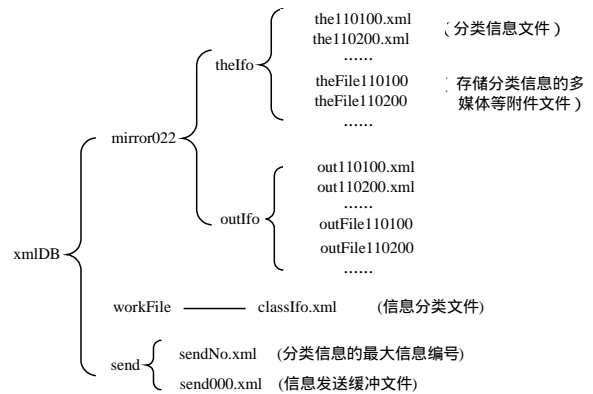


图3 联盟网站数据目录

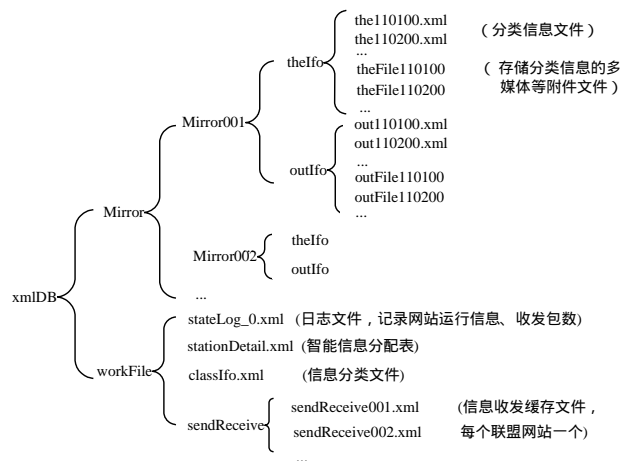


图4 智能服务器数据目录设计

2.5 信息包格式定义

信息交换是以信息包为基本单位的，信息包中除了需要交换的信息内容外，还集成了信息交换协议。信息包的通用格式如下^[3]（XML文件形式）：

```
<?xml version="1.0" encoding="gb2312"?>
<root xmlns:dt="urn:schemas-microsoft-com:datatypes">
  <commStatus> true | false </commStatus>
  <requestType> 0 | 1 | 2 | 8 | 9 </requestType>
  <sendCode> 0000000020000 </sendCode>
  <sendTime> 2005/03/15 </sendTime>
  <oprType> 0 | 1 | 2 | 8 | 9 </oprType>
  <publishCode> 00000000 </publishCode>
  <dataSort> 000000 </dataSort>
  <dataXML> ...
  <!-- xml 数据体 -->
  </dataXML>
  <fileName> binName_01.exp </fileName>
  <fileName> binName_02.exp </fileName>
  ...
  <binContent dt:dt="bin.base64">
    ( 二进制流信息_01 )
  </binContent>
  <binContent dt:dt="bin.base64">
    ( 二进制流信息_02 )
  </binContent>
  ...
</root>
```

其中，

(1)节点<commStatus>说明返回包的状态,成功为 true ,失败为 false 。

(2)节点<requestType>说明由智能服务器发来信息包的工作性质,其值可以是

0|1|2|8|9

其中,0 为运行检测;1 为提交信息;2 为请求信息;8 为维护申请;9 为维护提取。

(3)节点<sendCode>为发送代码,用于确定信息包的属性,编码并按以下方式分组:000 000000 2 00000,其中,000 表示网站编码;000000 表示信息分类码;2 表示信息外发指示码;00000 表示该类信息发送序列码。

(4)节点<sendTime>表示发送的时间。

(5)节点<oprType>说明信息包的操作方法,其中,0 表示直接写入;1 表示覆盖写入;2 表示指定作废,即设置作废标志;9 表示返回诊断信息。

(6)节点<publishCode>说明信息的来源,编码格式为 000 0000 00

其中,000 表示网站码;0000 为企业码;00 为发布人员代码。

(7)节点<dataSort>由发布人员指定的信息分类码。

(8)节点<dataXML>存放 XML 数据,信息包中只有一个节点。

(9)节点<fileName>存放二进制文件的文件名 xxx.exp。

(10)节点<binContent>存放二进制文件体,它与二进制文件名<fileName>顺序对应,可以有多个。

2.6 智能信息交换系统主要功能的实现

智能服务器掌握着通信的主动权。它定时循环检测各个联盟网站数据目录中的发送缓存文件(xmlDB/send/sendxxx.Xml)是否有信息包,如果有,它将激活相应网站的发送知识节点,经发送处理发送给智能服务器。智能服务器随后也激活自己接收知识节点,作接收及相应的拆包处理。智能服务器根据智能分配表(xmlDB/workFile/stationDetail.xml)将接收到的信息包分别写入智能服务器数据目录中的每个联盟网站收发缓存文件(xmlDB/sendReceive/sendReceive xxx.xml),然后激活智能服务器的发送知识节点。发送知识节点则激活相应联盟网站的接收知识节点,网站一端的接收模块将该信息包接收并作拆包,根据信息包中 dataSort 节点内容找到数据存储文件,并根据信息包中的 oprType 内容进行增加、覆盖、删除操作。智能服务器在分发的同时进行镜像存储操作,将信息存入相应的网站 mirror,信息来源网站写入 theIfo 目录的类文件,待分发的网站写入 outIfo 目录的类文件。

以上是信息智能交换的基本过程。在信息传输的过程中由于网络故障等一些不确定因素,免不了发生信息丢包和重包,作为完整的信息交换系统,智能服务器提供了处理丢包和重包的功能,它根据镜像文件中信息包的编码节点 sendCode 的连续性判断是否有丢包,然后激活相应联盟网站,重新发送丢包的内容,执行信息交换过程。如果镜像文件中信息包的编码节点 sendCode 重复,删除多余信息包,并生成 oprType=2 的信息包,进行分发过程。

3 应用实例

根据上述智能交换原理所建立的纺织产业网联盟系统已于 2006 年 7 月通过专家测试验收,现简述此实例。

3.1 纺织产业网联盟的需求

纺织产业网联盟由一系列的地区性纺织网站组成,它们在供应链的关系上能自动提供并获取所需要的信息。具体需

求如下:

(1)各纺织专业网站信息互通共享。

(2)信息发布功能。在某一网站发布信息,可同时在相关网站指定栏目发布。

(3)广告发布管理功能。多个网站同时发布,对点击率、客户分布及各网站贡献率等作出统计分析。

(4)全文检索功能。可在登录网站检索,也可自动到其他网站检索。

(5)登录管理。在一家网站注册,能自动获得联盟网站的注册资格。

(6)系统管理。对联盟网站整体运行状况进行监测,对系统功能实施维护。

3.2 拓扑结构的确定

在图 1 智能交换系统的网络结构中,根据实际需要,指定联盟网站中国纺织经济信息网(www.ctei.gov.cn)为中心服务器,负责为联盟中的其他网站提供权威的统计信息、协会动态、会展资讯、以及专业的市场分析等多种信息知识。智能服务器放置在纺织工业协会局域网内。

3.3 纺织产业网联盟体系的建立

(1)在认真分析纺织产业网联盟信息特点的基础上,制定了分类规则,对信息进行了分类,形成如下信息分类文件 classIfo.xml(片段):

```
<?xml version="1.0" encoding="GB2312" ?>
```

```
- <信息分类>
```

```
  <名称>纺织信息分类代码表</名称>
```

```
  <文件>classIfo.xml</文件>
```

```
- <分类 01>
```

```
  <名称>新闻报道</名称>
```

```
  <编码>100000</编码>
```

```
- <分类 02>
```

```
  <名称>行业报道</名称>
```

```
  <编码>110000</编码>
```

```
- <分类 03>
```

```
  <名称>纺织工业协会</名称>
```

```
  <编码>110100</编码>
```

```
</分类 03>
```

```
...
```

```
</分类 02>
```

```
...
```

```
</分类 01>
```

```
...
```

```
</信息分类>
```

分类编码规则:信息编码由 6 位数字组成,格式为

0 0 00 00

其中,第 1 位表示大类;第 2 位表示中类;第 3 位、第 4 位表示小类;第 5 位、第 6 位表示分组。第 1 位码值为 0 时,属于维护信息,第 1 位~第 9 位表示应用信息。

(2)建立信息表现标准,即 sx_500000.xml 和 pubLocate.xml。sx_500000.xml 对网站的各个网页及栏目进行编码,通过建立信息与栏目的关联关系实现本站信息在不同栏目上的智能显示。pubLocate.xml 根据统一的信息分类编码设置从不同网站、不同发布人员与表现编码的对应关系,实现外站信息的智能显示。

(3)建立信息的属性标准,一条信息除了它的分类归属外,本身具有多个属性,如地域属性、等级属性、行业属性
(下转第 171 页)