

分布式智能答疑系统的知识库构建与维护研究

张以利, 刘亚军

(东南大学 计算机科学与工程系, 江苏 南京 210096)

摘 要: 构建知识库是搭建分布式环境下智能答疑系统平台的关键环节。文中针对如何表示各站点知识库中的知识以及如何构建和维护知识库问题, 提出用 XML 表示知识来构建知识库, 通过使用 DOM 树来维护知识库, 并给出了知识库的组织结构图及 XML 的 DTD 文件。实验表明, 用该方法构建的知识库便于在系统平台下实现资源共享和知识库维护。

关键词: XML; 移动代理; 智能答疑系统; DOM

中图分类号: TP18

文献标识码: A

文章编号: 1673-629X(2006)07-0015-02

Research of Setting up and Support of Distributed Intelligent Question Answer System

ZHANG Yi-li, LIU Ya-jun

(Department of Computer Science and Engineering, Southeast University, Nanjing 210096, China)

Abstract: A distributed intelligent question answer system is a platform sharing resource among the sites. It is presented due to meaning the knowledge and set up database, which is based on the XML. It uses the DOM tree to support the base. And the diagram of organization structure of the knowledge base and the DTD document are given. Experimental results show that the base setting up in this way is easy to share and support the knowledge.

Key words: XML; mobile agent; intelligent question answer system; DOM

0 引言

近年来, 在远程教育领域中, 智能答疑系统成为人们研究的热点之一^[1]。分布式智能答疑系统(简称 DIQA), 就是针对远程教育智能答疑系统中的知识资源不足问题, 采用 Mobile Agent 技术, 构建分布式环境下多站点之间的智能答疑平台, 通过对各站点知识库的知识资源搜索, 完成答疑任务, 并补充本地资源库, 实现多站点间的知识库资源共享。

但是, 由于这些站点知识库面对的可能是大量的形式各异的文档数据, 不能进行相互间交换数据, 也不便同网上其他的教育资源进行数据交换与共享。因此, 如何构建和维护知识库是顺利成功搭建 DIQA 原型系统平台的关键环节。

1 XML 的特点

XML 主要包含以下 3 个元素: DTD/Schema、XSL 和 XLink。DTD 规定了 XML 文件的逻辑结构, XML Schema

给 XML 加入了描述数据类型的能力; XSL 定义了 XML 的表现方式; XLink 进一步扩展了目前 WEB 上已有的简单链接。其主要优点如下^[2]:

1) 良好的可扩展性。XML 允许各个不同的行业根据自己的需要制定自己的一套标记。这就使得该领域中的人们可以自由交换信息。

2) 内容和形式相分离。XML 文档用于描述数据的内容, XSL 和 CSS 描述其显示方式, 这样, 同一内容可以根据实际要求呈现多种表现形式, 用户可以自行定制数据显示方式。

3) 便于不同系统之间传输信息。不同企业、不同部门中存在着许多不同的系统。由于 XML 是非专有的并易于阅读和编写, 使得它易于在这些异构的平台、数据库之间传输信息, 成为在不同的应用间交换数据的理想格式。

4) 具有较好的保值性。XML 的保值性源自 SGML 语言。SGML 最初设计的一大目标就是要为文档提供 50 年以上的寿命。SGML 和 XML 不但能够长期作为一种通用的标准, 而且很容易和其它格式的文档转换。

2 DIQA 知识库在系统中的角色及其功能需求

图 1 是 DIQA 原型系统功能模块中的知识库及其相关部分。在原型系统平台下, 用户通过浏览器, 用自然语言向站点服务器提问; 经分词模块预处理后, 由搜索 A-

收稿日期: 2005-11-14

基金项目: 江苏省基金资助项目(20031184)

作者简介: 张以利(1972-), 男, 安徽霍邱人, 硕士研究生, 研究方向为数据库应用、网络教育; 刘亚军, 硕士生导师, 教授, 研究方向数据库应用、网络教育等。

gent 携带问题向量在本地站点或远端站点的知识库依次查找相匹配的答案^[3];并将查找到的答案通过消息 Agent 返回给用户,同时把从远端站点搜索到的答案传给维护代理,补充到本地知识库。若搜索 Agent 在以上均未找到答案,则问题将转发给人工处理,然后补充到知识库。

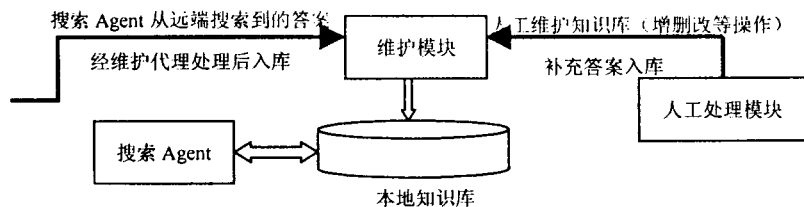


图 1 DIQA 知识库及其相关功能模块

从 DIQA 原型系统的工作机制可以知道,实际上,这些站点知识库在原型系统平台下进行数据交换,实现知识资源共享。DIQA 知识库有如下的功能需求:知识库中的知识表示应根据需要制定自己的一套标准格式,而且是与平台、应用软件都无关的标准化的文档格式,使得该领域中的人们可以自由交换信息,便于各知识库间以及与其他教育资源间实现知识共享;同一答案可以根据实际需要呈现多种表现形式,以解决答案结构不合理,呈现方式不够丰富等所带来的不便;知识库应该是开放式的,在保持知识库良好的基础,能够方便地对知识库进行操作,包括人工的和自动的。由 XML 的特点和优势可以知道,XML 能够很好地满足这些需求。因此,XML 是构建和维护 DIQA 知识库的理想工具。

3 DIQA 知识库的构建及其维护模块的设计

3.1 知识库的结构分析

DIQA 知识库由多个领域学科组成,学科下存放可用材料。对于每段材料,存放可用的中心句、关键词和材料内容。整个知识库是按学科分类的,是树状的数据结构。当用户的问题句经过分词后,由 Agent 携带问题向量到各站点知识库,与其中的中心句关键词进行匹配。其组织结构如图 2 所示。

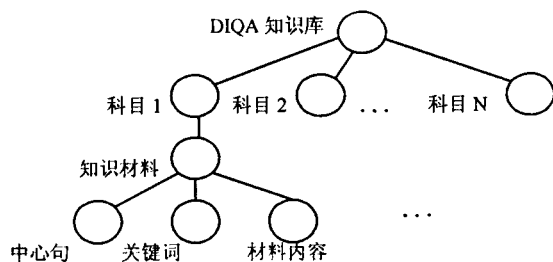


图 2 DIQA 知识库的组织结构

3.2 构建 XML 知识库

以上对 DIQA 知识库的功能需求及其数据结构进行了分析,下面定义 XML 的 DTD 文件。其中,“学科”包含子元素:ID(学科号),Title(学科名称),Knowledge(材料信息);“材料信息”包含子元素:KID(材料 ID),Headword(材料中心句),Keyword(材料关键词),Content(材料内

容)。

```
< ! ELEMENT DIQA (Subject + ) >
```

```
< ! ELEMENT Subject (ID,Title,Knowledge + ) >
```

```
< ! ELEMENT ID # PCDATA >
```

```
< ! ELEMENT Title # PCDATA >
```

```
< ! ELEMENT Knowledge (KID, Headword, Keyword + ,Content) >
```

```
< ! ELEMENT KID # PCDATA >
```

```
< ! ELEMENT Headword # PCDATA >
```

```
< ! ELEMENT Keyword # PCDATA >
```

```
< ! ELEMENT Content # PCDATA >
```

例如,下面是“数据库系统”学科中的一段材料信息:“外模式和内模式的区别是什么”。

```
<DIQA>
```

```
< Subject>
```

```
< ID>0101< /ID>
```

```
< Title>数据库系统< /Title>
```

```
< Knowledge>
```

```
< KID>070102< /KID>
```

```
< Headword>外模式和内模式的区别是什么< /Headword>
```

```
< Keyword>外模式< /Keyword>
```

```
< Keyword 内模式< /Keyword>
```

```
< Content>外模式是个别用户涉及到的数据的逻辑结构,而内模式是数据库中全部数据的物理结构。< /Content>
```

```
< /Knowledge>
```

```
.....
```

```
< /Subject>
```

```
.....
```

```
< /DIQA>
```

3.3 知识库维护模块的设计

DIQA 知识库应该是开放式的,从远端站点搜索到的答案可以按本地知识库的结构自动插入到本地库中;并且,教师或专业人员也可以在离线的情况下对知识库进行操作。这个功能由知识库维护模块来完成。如图 1 所示,知识库维护模块包括维护代理和人工处理模块两大部分。

维护代理通过 DOM,使用微软的 MSXML 解析器^[4],将知识库中的 XML 文档信息转化为一棵节点层次树,从而可以方便地实现对 XML 文档中的数据进行增加、修改和删除等操作,使知识库不断得到更新和完善。如果要向知识库中增加答案材料,则维护代理会自动启动相应程序,找到 DOM 树中的相应节点,然后,自动把材料进行编号,并按本地知识库的组织结构插入到 DOM 树相应节点中。

人工处理模块是人工维护知识库的入口模块,可以在离线的情况下进行,负责回答搜索代理在本地站点或远端站点没有搜索到的用户问题,并将该答案材料通过维护代理存入到知识库。如果要浏览、修改或删除知识库中的答案材料,管理员用户可以通过系统提供的接口,先浏览答

(下转第 170 页)

```
input[width-1:0] din;
output[width-1:0] dout;
assign dout = din ^ {1'b0, din[3:1]};
endmodule
```

程序 2 格雷码到自然码的转换:

```
module norm_to_gray(din,dout);
parameter width=8;
input[width-1:0] din;
output[width-1:0] dout;
integer i,j;
reg temp;
reg[width-1:0] buff;
always @(din)
for (i=width-1;i>=0;i=i-1)
begin
temp=0;
for(j=width-1;j>=i;j=j-1)
begin
temp=temp ^ din[j];
end
buff[i]=temp;
end
assign dout=buff;
endmodule
```

程序 3 写地址产生(只给出地址产生部分):

```
always @(posedge wr_clk or negedge rst)
//rst 是复位信号
begin
if(!rst)
wadd<=0; //初始自然码写指针
else
begin
wadd<=wadd+1;
norm_to_gray(wadd,wadd_gray);
//wadd_gray 是格雷码写指针
end
end
```

end

程序 4 满标志产生:

```
always @(radd_gray or wadd_gray)
//radd_gray 是读指针格雷码
begin
gray_to_norm(wadd_gray,wadd);
gray_to_norm(radd_gray,radd);
if(radd[width-1]! = wadd[width-1]
&& radd[width-2:0] = wadd[width-2:0])
full<=1;
else
full<=0;
end
```

类似的,可以写出读地址产生模块和空标志产生模块^[5]。

4 结 论

文中针对异步 FIFO 设计中的难点和要点,提出了具体的解决方案,并用 verilog 语言给出了电路的设计,利用 Altera 的 Cyclone II 系列 FPGA 实现,并被应用于多种电路中,在实际应用中取得了较好的效果。

参考文献:

- [1] KANOPOULOS N, HALLENBECK J J. A First - In, First - out Memory for Signal Processing Applications[J]. IEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, 1986, CAS-33(5):556-558.
- [2] 夏宇闻. Verilog 数字系统设计教程[M]. 北京:北京航空航天大学出版社,2004.
- [3] 罗 昊. 一种异步 FIFO 的设计方法[EB/OL]. <http://www.dianzi.net/show.asp?id=1733>, 2004-12-19.
- [4] 任爱锋, 初秀琴, 常 存, 等. 基于 FPGA 的嵌入式系统设计[M]. 西安:西安电子科技大学出版社,2004.
- [5] 杜建国. Verilog HDL 硬件描述语言[M]. 北京:国防工业出版社,2004.

(上接第 16 页)

案材料,然后进行删除或修改。

4 结束语

DIQA 知识库,是用 XML 表示知识,各知识库拥有统一的标准化文档格式,并与平台、应用软件无关,通过 DOM 处理文档信息,便于维护。实验表明,用该方法构建起来的知识库,极大地方便了在系统平台下实现资源共享和知识库维护。DIQA 知识库作为原型系统平台下存储知识资源的容器,在“信息高速公路”中会有广阔的应用前景^[5]。

参考文献:

- [1] 刘亚军,徐 易. 一种基于加权语义相似度模型的自动答疑系统[J]. 东南大学学报(自然科学版),2004,34(5):609-612.
- [2] 吕俊峰,潘会友. 基于 XML 的格式化网络课件[J]. 太原理工大学学报,2002,33(2):204-207.
- [3] 徐 易. 智能答疑系统的研究与实现[D]. 南京:东南大学,2002.
- [4] 侯晓强,徐春荣,勾海波. Java 服务器编程实例[M]. 北京:清华大学出版社,2003. 131-142.
- [5] 张丽晖. Mobile Agent 原型系统的研究与开发[D]. 南京:东南大学,2004.