

信息对象开放式采集系统的形式化描述

赵正旭, 张常有, 刘立嘉

(石家庄铁道大学 信息科学技术学院, 河北 石家庄 050043)

摘 要: 为获得研究对象的描述特征, 设计一种面向 Internet 的开放式采集系统. 该系统的主要功能包括信息对象的登记、登记信息查询、对象信息录入、对象修改、对象浏览等. 在此基础上, 使用 Z 规范对系统需求进行了形式化描述, 能够有效避免二义性, 并有利于系统的进一步模型检测.

关键词: 信息对象; 采集系统; Z 规范; 形式化

中图分类号: TP311

文献标识码: A

文章编号: 1000-7180(2010)08-0020-04

The Formalization of Open Collection System of Information Objects

ZHAO Zheng-xu, ZHANG Chang-you, LIU Li-jia

(School of Informatics Science & Technology, Shijiazhuang Tiedao University, Shijiazhuang 050043, China)

Abstract: In order to discover the description properties of the research objects, we design an open system for information objects collecting over Internet. The major functions of this system include object registering, registration information querying, object information inputting, object updating, objects browsing, and so on. We formalize the system in Z notations to avoid the misunderstanding in semantics in the requirement specification. This formalization will be used in further model checking.

Key words: information object; collection system; Z notation; formalization

1 引言

信息世界中的每一个实体, 称为一个信息对象 (Info). 信息对象由若干特种属性 (attribute) 集合表示. 信息对象往往与一种或多种计算机软件相关联. 研究计算机软件之间的关系, 能够观察信息对象在软件节点之间的转换规律, 得到信息存储与转换的最优化路径. 信息世界中信息对象的数量巨大, 通过互联网, 基于 Internet 的开放式信息对象采集系统, 能够汇聚全球爱好者的贡献, 数据数量和质量得到保障.

统一建模语言 UML (Unified Modeling Language) 是一种面向对象的可视化系统分析建模语言. 它在面向对象的软件工程领域的重要性正在递

增^[1]. 使用 UML, 系统分析员能够正确捕捉用户的需求, 避免自然语言描述软件需求文档中的理解歧义^[2]. 研究表明, UML 在许多领域软件开发中得到成功应用, 得到研究者的继续研究^[3-6].

2 信息对象

2.1 信息对象的定义

信息对象由若干特征来表示. 这些特征称为信息对象的属性 (attribute). 全部属性构成属性集合 attr. 信息对象定义为

$$\text{Info} ::= \text{attrs} = \{\text{attribute}_1, \text{attribute}_2, \dots, \text{attribute}_i, \dots, \text{attribute}_N\}$$

2.2 信息对象之间的关系

不同的信息对象之间存在某种客观联系. 定义

这种联系为对象之间的关系,则关系 r 表现为该属性集合的幂集($\mathbb{P} \text{ attrs}$)上的函数 $f(\mathbb{P} \text{ attrs})$. 即

$$r = f(\mathbb{P} \text{ attrs})$$

函数 f 的性质决定关系 r 的性质. 因此,通过分析函数 f 的性质,分析信息世界中信息对象之间的静态和动态规律,从而分析庞大而复杂的信息世界,以发现其符合特定约束条件的客观规律,用于探索信息的安全、快速处理方法.

由此可见,信息世界由信息对象 InfO 和其间的关系 r 构成,如图 1 所示.

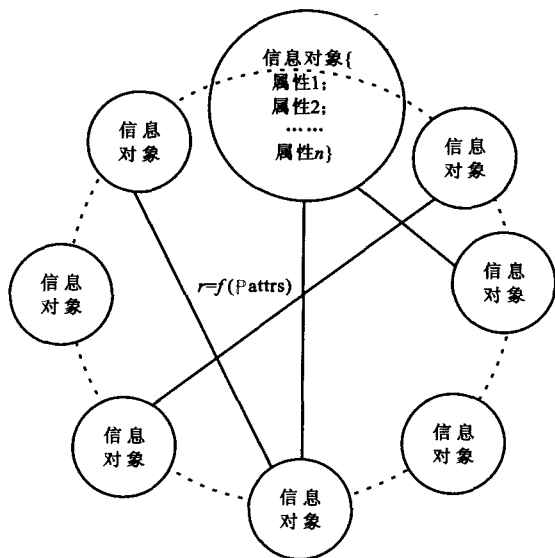


图1 信息对象的结构与关系

本文以计算机软件为研究对象,构建信息对象的描述方法. 考虑到信息对象的属性数量不尽相同,其表示方法应该支持可变的属性条目数量,且属性取值可多值并存.

2.3 对象描述的层次结构

根据信息世界中的事实,本系统描述的信息对象的数量上限很大,是个“很大”的数字. 信息对象千差万别,用于描述不同对象的属性个数也不尽相同,其属性数据也很难收齐. 所以,系统中用于描述信息对象的属性个数也是一个“较大”的数字.

对象描述结构采用树形结构,从上到下的各层依次为:信息世界→信息对象→属性→值. 从上到下为一对多的关系,如图 2 所示.

2.4 外部存储方案

为描述树形结构分支的差异性,本文采用 XML 文档格式存储属性数据. 在 XML 文档中,不仅信息对象个数自由添加、对象的属性信息条目自由添加、条目的先后顺序不影响文件的整体语义、稀疏数据文档占用少量空间,并且 XML 文档中的数据没有

数据类型,避免数据库的结构化数据类型带来的系统实现的麻烦. 另外,XML 文档具有标准性、开放性、跨平台性,容易与其他开放系统接口.



图2 信息对象描述的树形结构

文档结构对应于信息对象描述的树形结构,根节点为“InfoObjectDescription”. 根节点只有一个. 二级节点为属性名,如 title、openFlies、saveFlies、importFlies、exportFlies 等. 三级节点的元素名称统一采用“items”,其中的定语(attributed)的以不同的修饰语区分属性取值的不同含义. 因为属性的相同取值,有可能表现为不同的含义.

这种相同取值有不同含义的现象普遍存在. 以“MS Word”办公软件生成的文档扩展名为例.“MS Word”默认保存的文件格式为“*.doc”. 无论 MS Word 2003 版,还是 2007 版,甚至早期版本,默认文档扩展名均相同,但 MS Word 2003 版不能正确打开 MS Word 2007 版生成的 doc 文档.

综上,本文设计的 XML 文档格式形如下:

```
<InfoObjectDescription>
  <title>
    <item attributed="English">AutoCAD</item>
  </title>
  <saveFlies>
    <item attributed="Auto CAD 2008 图形">*.
dwg</item>
    <item attributed="Auto CAD 2004 / LT 2004 图形">
*.dwg</item>
    <item attributed="Auto CAD 2008 DXF">*.dxf</
item>
    .....
  </saveFlies>
  .....
</InfoObjectDescription>
```

3 系统模型

根据系统的功用,本系统命名为 InfoTools,用于收集各种软件的属性信息,并分析其间可能存在的关系. 所有能够登录网络,并使用浏览器浏览和输入信息的网络用户,都可能是本系统的潜在用户.

系统采用 B/S 结构. 用户可以通过 HTTP 协议

登录 Web 服务器,通过本地网页输入并提交选定软件的相关信息.系统模型如图 3 所示.

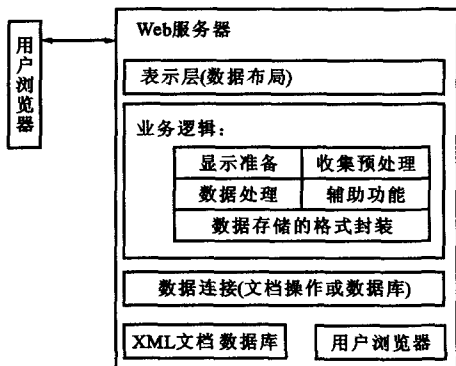


图 3 信息对象属性数据采集系统模型

根据信息对象数据采集的需要,系统将实现五个主要功能:(1)信息对象登记:录入信息对象的基本信息,为信息对象建立唯一标识;(2)检索对象注册信息:根据软件名称和版本号,查询该软件的登记信息;(3)对象信息录入:逐条属性录入;(4)对象信息预览:浏览已经输入的对像信息;(5)对象信息修改:只允许对象的创建者修改软件信息.

4 基于 UML 的系统设计

针对系统需要的功能,系统的设计从用例图开始,经过事件流、顺序图、状态图、类图、部署图等阶段,完成了系统设计.限于篇幅,本文的图 4 和图 5 分别列举了针对对象登记子系统和对象操作子系统的类图.

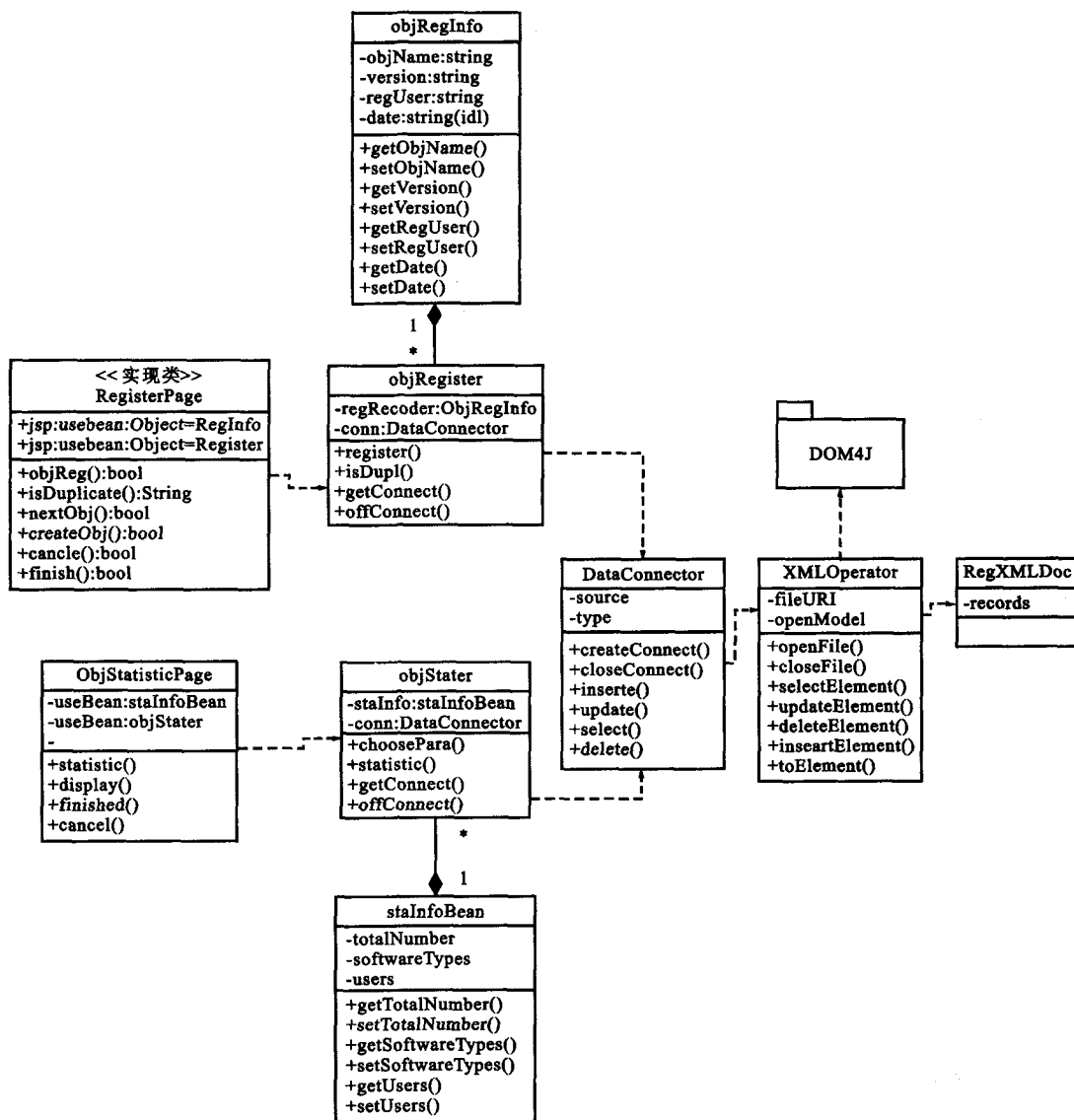


图 4 对象登记子系统类图

图4中用于信息对象登记的主要类有: Register、ObjStatistic、ObjRegInfo、ObjRegister、ObjIn-

foBean、ObjStater、DataConnector、DataConnector、XML Operator等。

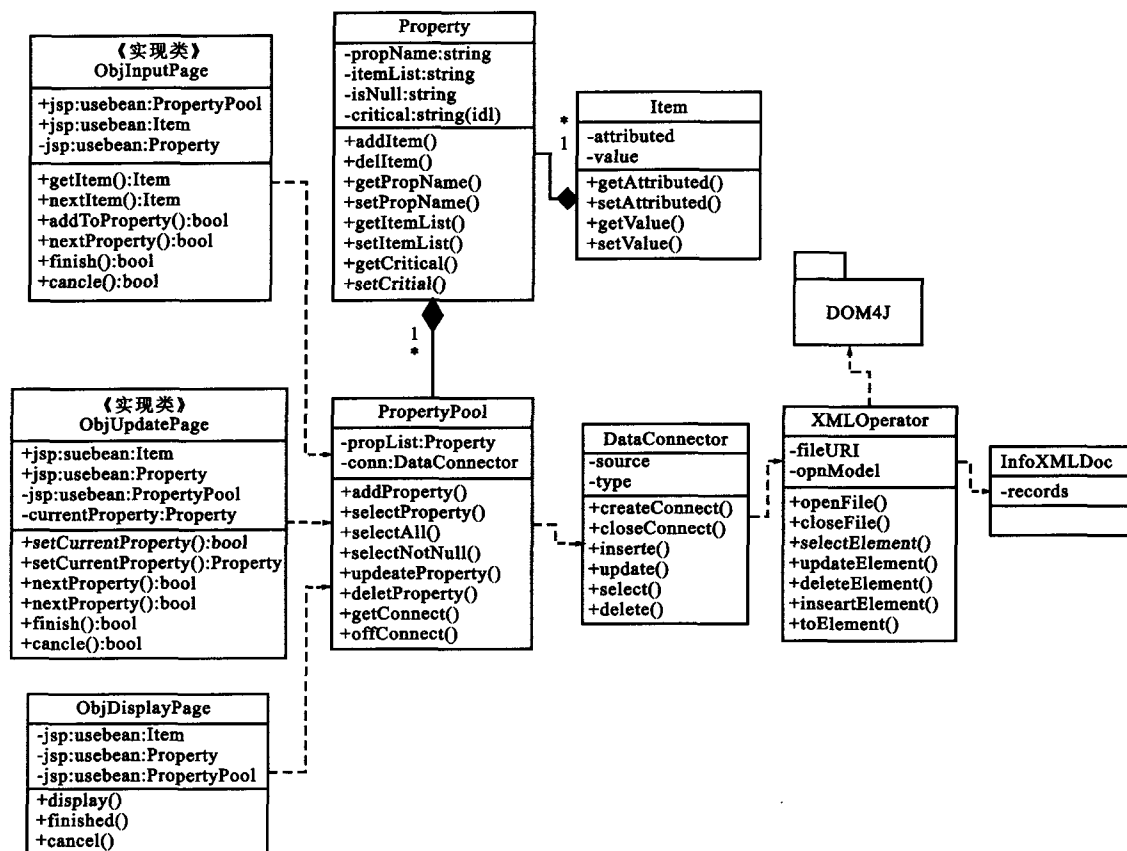


图5 对象操作子系统类图

图5中的用于信息对象操作的主要类有: ObjInput、ObjUpdate、ObjDisplay、Item、Property、PropertyPool等,还有与对象操作系统共用的 DataConnector、DataConnector、XML Operator等。

5 结束语

本文首先为信息对象的属性采集设计了一个开放式系统。系统主要实现信息对象的登记、查重、信息录入、浏览、修改等功能。针对这些功能,本文用Z规范描述了系统的状态和各种操作对系统状态的影响,从而描述了系统的运行规律。对于系统来说,还有许多细节需要完善。比如系统的统计功能、细粒度的权限管理、信息对象属性条目的融合优化等。另外,Z规范中还有表达能力更强的符号,本文中只使用了其中基本的符号,也能够达到了描述本系统的目的。

参考文献:

[1] Object Management Group. Unified model language(UML):

superstructure version 2.0[EB/OL]. [2005-07-04].

<http://www.omg.org/docs/formal/05-07-04.pdf>.

- [2] Cui Qifeng, Liu Chengliang, Zha Xuan F. Intelligent drug delivery system using UML diagrams analysis[J]. Journal of Shanghai Jiaotong University(Sci.), 2008, 13(3): 312-317.
- [3] 罗开平, 李一军, 姜雄. 基于UML的弹道导弹预警仿真系统建模[J]. 系统仿真学报, 2010, 22(4): 845-849.
- [4] 鲁鹏程, 易小琳, 毛国君, 等. 基于的服装试穿衣系统设计与实现[J]. 北京工业大学学报, 2008, 34(9): 1001-1004.
- [5] 张岩, 梅宏. UML类图中面向非功能属性的描述和检验[J]. 软件学科, 2009, 20(6): 1458-1469.
- [6] 程亮, 张阳. 基于UML和模型检测的安全模型验证方法[J]. 计算机学报, 2009, 32(40): 699-708.

作者简介:

赵正旭 男, (1960-), 教授. 研究方向为网络信息系统、虚拟现实.