

文章编号: 1000-5013(2012)05-0513-05

刻面和规约描述相结合的构件检索

陈杜英, 刘韶涛

(华侨大学 计算机科学与技术学院, 福建 厦门 361021)

摘要: 通过分析基于刻面分类和构件规约语法的检索方法,发现刻面分类及其检索方法侧重于构件的静态特征描述,没有考虑构件的行为描述等,而构件规约语法描述部分偏重于构件的行为描述,构件规约对构件中每个功能的语法描述具有独立性.设计一种基于字符串匹配的软构件刻面检索算法,同时考虑到刻面检索没有支持组装这一需求,提出在检索过程中,插入基于构件规约的语法匹配过程,以提供较准确的构件组装信息和提高复用效率.

关键词: 构件检索; 刻面分类; 规约描述; 字符串匹配

中图分类号: TP 311.13 **文献标志码:** A

软件复用被认为是解决软件危机的有效途径,而基于构件的软件开发方法是软件复用的重要手段,因此需要构件库提供合理的构件描述组织方式,并提供方便的构件检索机制.复用者从构件库中检索构件,组装成新的应用系统,从而达到提高软件开发效率的目的,所以构件的描述应能够满足组装的需要,反映构件组装信息.刻面分类方法包括刻面模式、同义词辞典和概念距离图,从不同的具有代表性的刻面维度对构件进行描述,是目前应用最广泛的构件检索法方法.基于构件规约语法描述的检索方法具有较高的查准率.然而,刻面描述方法不具有对构件行为方面的描述,对于软构件复用者的组装需求无法给予很好的满足.本文采用以刻面分类为主,语法规约描述相结合的方法对构件进行分类^[1].

1 基于刻面的构件描述模型

1.1 基于 XML 技术的构件表示^[2]

XML(可扩展标记语言)是万维网联盟于 1998 年 2 月发布的一种标准^[3].XML 主要用于描述数据,在定义标记方面具有较大的灵活性和伸缩性.使用 XML 描述软构件信息体现了一些优越性,主要表现在:可以任意选择具有代表性的属性进行描述,较清晰地对软构件信息进行分层和组织,以及定义软构件实体之间的约束关系等.因此,文中采用基于 XML 的构件刻面描述方法,对软构件的刻面进行描述,把每一个软构件描述成一棵软构件描述树,将查询信息转换成一棵查询树.

根据参考树匹配模型^[4-5],采用松弛匹配模型以达到较高的查全率.松弛匹配模型,只要求软构件描述树与查询树的叶子节点之间保持阈值范围内即可,且通过节点的逻辑操作,包括插入、删除、修改计算匹配代价,要求查询树的所有叶子节点在描述树中找到对应的匹配节点.

1.2 构件的描述

为了检索和组装的方便,给出其刻面和构件的规约语法描述信息.刻面描述是对应构件的语义描述,构件的规约是其语法描述,具有唯一性.给出构件的 5 个刻面,分别是使用环境(AE),功能(FUNn),应用领域(AA),表示方法(REP)和层次(LAY).使用环境包括硬件环境、操作系统、数据库和编程环境等方面;功能是对构件所提供的功能描述;应用领域是对该构件可能应用到的领域的名称;表示方法是用来描述构件内容的表达方法;层次是表示构件描述的抽象层次.每个刻面都有相应的标准刻

面值,这些刻面值还具有一个达到一定阈值的语义同义词表.展示的一个构件的 XML 描述文档,如图 1 所示.

为了满足组装的需求,即满足第二次规约语法匹配的需要,对入库的每个构件要求提供对规约语法的描述,包括填写构件对外需求的服务,以及对外提供的服务,包括填写接口描述信息,服务个数,每个服务的服务名、参数名、类型列表信息和返回值名及返回值类型等信息.

1.3 为构件建立后缀排列构件库

用字符串代表某刻面属性路径,即将路径匹配转换为字符串的匹配,应用相应的高效字符串查询方法来实现构件查询的目的.把构件的每一条路径用字符串来表示,用字符串代表刻面路径,把“#”,“—”两个特殊符号用于构件树的路径匹配转换.“#”用来分隔不同的路径,“—”用来连接不同的字符^[6],如对应图 1 所示,软构件描述转换成字符串形式为 COM-AA-MIS#COM-AE-DB-Oracle#COM-AE-Memory-1G#COM-AE-OS-linux#COM-FUN-fileDownload#COM-LAY-design#COM-REP-IDL.

为后续构件刻面查询方便,在每次构件入库或有构件修改时,实时更新后缀排列(stuffx array)构件库,将构件库中的所有构件按照从根节点到叶节点的路径顺序组成字符串,并按字典顺序路径串组成后缀排列.

对构件库中的所有构件树产生后缀排列构件库算法^[6]有如下几个步骤:

输入为构件库 D,输出为后缀排列构件库 SD,并保存更新后的 SD;

- 1) $SD=\varnothing$;
- 2) for 软件库中的每一棵构件描述树;
- 3) {;
- 4) 按照先序遍历构件树的所有从根节点到叶节点的路径,用字符串表示,将所有路径字符串按分隔符分开;
- 5) 按照字典顺序对所有路径长度字符串进行后缀排列;
- 6) 把所有的字符串连接成一个长的字符串 S,把 S 加入到 SD 中;
- 7) };
- 8) return (SD);
- 9) update(SD).

2 基于刻面描述的构件规约语法匹配

2.1 规约语法匹配方法

构件规约是对构件复用信息的形式化描述,服务作为语法描述信息的重要部分,每个服务都是不可分割的一个独立存在,如图 2 所示.因此,将服务作为语法匹配的原子单位^[7].

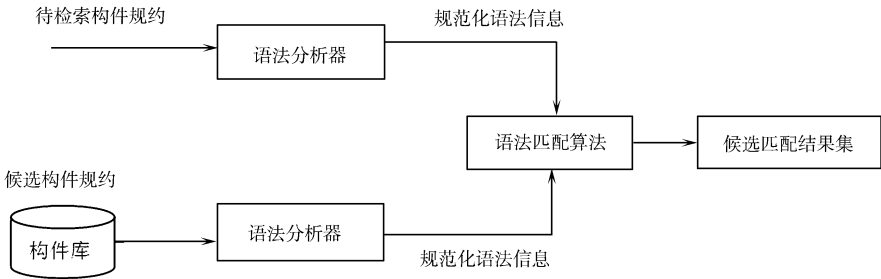


图 2 基于规约语法匹配的原理

Fig. 2 Principle based on the specification syntax matching

服务中,每个 Function 代表构件提供或请求的一种功能.由于服务参数类型和返回值类型的不可

```
<component>
<AE>
  <Database> oracle </Database>
  <Operation System> windows </Operation>
  <Ems memory> 1G </Memory>
</AE>
<FUN> fileDownload </FUN>
<AA> MIS </AA>
<REP> IDL </REP>
<LAY> design </LAY>
</Component>
```

图 1 构件的 XML 文档描述形式
Fig. 1 Document description
form of the component's XML

更替性,被作为服务匹配的基调. 基调匹配是基于基本类型匹配的,类型主要分为 4 种,即简单数据类型;结构类型(结构、联合和枚举);数组类型和自定义类型. 任意替换两个等价类型不会影响程序的运行,基调匹配正是建立在类型等价原理之上的,而规约语法匹配是基于参数类型和返回值类型的匹配.

2.2 基于刻面描述的构件匹配过程

结合刻面和规约描述的构件匹配流程图,如图 3 所示. 刻面检索过程中,首先用户输入刻面描述查询信息,系统将查询信息自动生成构件查询树,并组成按字典顺序排好序的后缀排列查询字符串. 其次,检索系统将调用刻面检索字符串匹配算法,将查询字符串与后缀排列构件库中的每一棵构件树的字符串进行匹配处理. 最后,返回满足匹配条件的字符串序列代表的构件,用户选择是否继续进行规约语法匹配查询,如选择继续匹配,用户输入规约语法查询信息,系统对检索出的构件进行按规约语法匹配结果返回构件集合.

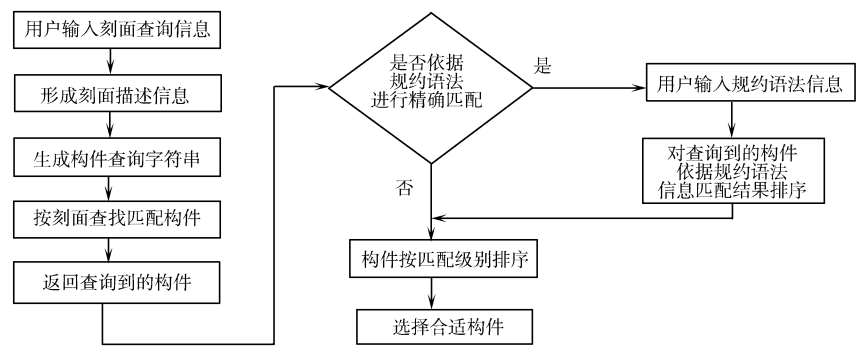


图 3 结合刻面和规约描述的构件匹配流程图

Fig. 3 Matching flow chart combined with facet and specification description

2.3 构件的检索

将用户输入的刻面查询信息生成查询树 Q , 并生成字符串 T , 与后缀排列构件库 SD 中的每一棵构件描述树的字符串 S 进行比较. 如果字符串匹配不成功, 则退出循环; 否则, 结合软构件树的插入、删除、修改节点操作来计算构件匹配代价. 按照匹配代价, 输出满足查询条件的软构件集.

构造查询树路径字符串算法^[6]主要有如下几个步骤:

输入为查询树 Q , 输出为查询字符串 T ;

- 1) T 为空;
- 2) 按先序遍历查询树 Q 的所有从根节点到叶节点的每条路径, 建立字符串, 并用分隔符分开路径字符串;
- 3) 按照字典顺序对所有路径长度字符串进行后缀排列;
- 4) 把所有的字符串连接成一个长的字符串 T ;
- 5) return T .

软构件的查询转换成查询字符串 T 与软构件后缀排序构件库 SD 的匹配为题, 即转换为子串查找问题, 模式字符串 T 在主串 SD 中是否出现, 以及出现的位置问题.

软构件的查询算法有如下几个主要步骤:

输入为软构件查询树字符串 T , 软构件库形成的后缀排列字符串 SD , 而输出为符合查询条件的软构件集合 R ;

- 1) $R = \varnothing$;
- 2) 根据字符串 T 中符号“#”出现次数, 计算查询树的路径数 m ;
- 3) for SD 中的每一棵软构件树对应的字符串 S ;
- 4) 采用字符串匹配算法, 计算 T 在 S 中出现的情况, 以判断软构件是否符合查询条件;
- 5) if 没有找到符合距离阈值的子字符串, 则退出本次循环;
- 6) else 对于在 T 中出现未在 S 中出现的字符(树的节点), 根据概念距离图计算插入代价; 对于未在 T 中出现而在 S 出现的字符(节点), 计算节点删除代价; 对于在无法通过删除或修改操作而成为阈

值范围内的节点,计算修改代价,并将这 3 种代价之和作为匹配近似度的参考值;

- 7) 统计查询树字符串 T 与构件树字符串 S 匹配路径数目 k ;
- 8) if($k=m$),将该查询字符串对应构件树加入结果集 R ;
- 9) 退出 for 循环;
- 10) 返回结果集 R (根据匹配代价从小到大排序).

假设查询树有 k 条路径,构件树中有 k 条路径与查询树字符串子串匹配,则说明满足查询要求,该方法描述了松弛匹配的过程.

2.4 构件的匹配

基于上述检索方法中得到的构件不一定能够满足复用者对组装的要求,为了进一步提高查准率,特安排第二次匹配.在此,采用构件的规约语法匹配方法进行精确匹配.

构件规约语法描述的重要组成部分是构件接口,包括构件对外提供的服务描述和对外需求的服务描述两部分组成,构件规约的语法匹配特指构件接口的语法匹配.接口的服务的参数类型列表和返回值类型部分是语法匹配的重要信息.然后,根据服务的参数类型和返回值类型的语义相似程度来判断构件的规约语法匹配的高低等级,按照此匹配等级重新排列构件结果集给复用者.

2.5 软构件检索过程的实例分析

通过一个实例来说明如何在检索过程中利用剖面检索和规约语法匹配技术进行构件检索.假定用户需要查询一个“管理用户”构件,软构件检索有 3 个主要过程.

1) 用户填写剖面描述信息.使用环境为 OS,windows DB,oracle,所属领域为 MIS,件功能为 user management,表示方法为 C++,抽象层次为 design.

2) 系统自动将用户的输入信息转化为构件剖面查询树(图 4),并将查询树与构件库中的构件描述树的后缀排列匹配,将查询结果返回给用户.按字典顺序排好序的生成字符串为:component-AA-MIS# component-AE-DB-oracle# component-AE-OS-windows# component-FUN-user management# component-LAY-design# component-REP-C++.

3) 规约语法匹配查询阶段.主要提供源代码级构件更精确的查找.用户输入软构件的语法描述信息,系统将用户输入的规约信息经语法分析器转换成规范化的语法信息,与软构件库中的候选构件的规范化语法信息进行语法匹配算法匹配,得出候选匹配结果集,用户输入的语法信息如图 5 所示.

以上给出的构件检索与匹配模型及算法分析,为精确查找到满足组装需要的构件提供了可能.

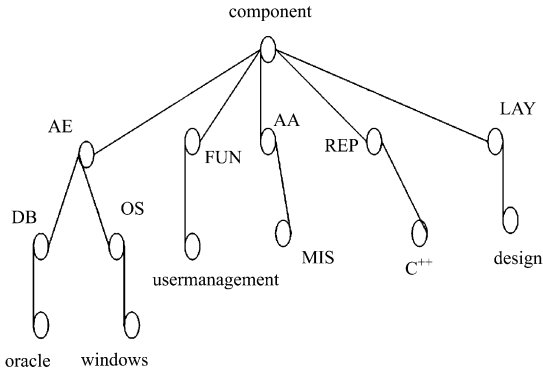


图 4 系统生成的构件查询树

Fig. 4 Component query tree of the system generated



图 5 用户输入规约语法查询信息

Fig. 5 User's input specification syntax information

3 结 论

文中分析了当前研究的各种软构件检索方法,详细研究当前应用比较成功的基于构件刻面描述的检索方式,并设计一种基于字符串匹配的软构件刻面检索算法. 在考虑刻面检索没有支持组装这一需求的同时,提出在检索过程中插入基于构件规约的语法匹配过程,以提供较准确的构件组装信息及提高复用效率.

由于刻面描述方法,不具有对构件行为方面的描述,对于软构件复用者的组装需求无法很好满足,因此,文中采用以刻面分类为主,语法规约描述相结合的方法对构件进行分类. 在同一系统中支持多种分类模式和检索方法的组合使用,以较好地满足构件检索的查全率和查准率的要求

参考文献:

[1] FRALES W B,POLE T P. Proteus: A software reuse library system that supports multiples representations methods[J]. ACM SIGER Forum,1990,24(3):43-55.

[2] 邓勇. 基于本体与刻面相结合的构件检索方法研究[D]. 武汉:华中师范大学,2007:14-16.

[3] 韩忠明. XML 数据查询与信息检索系统[M]. 北京:中国水利水电出版社,2010:3-4.

[4] SHASHA D,TSONG J,WANG L. Exact and approximate algorithm for unordered tree matching[J]. IEEE Trans on Systems, Man and Cybemetrics,1994,24(4):668-678.

[5] WANG Zhong-jie,ZHANG De-chen,XU Xiao-fei. A component retrieval method based on feature tree matching[J]. International Journal of Information Technology,2006,12(8):60-72.

[6] 姚全珠,刘波. 一种高效的基于刻面分类的构件检索算法研究[J]. 计算机工程与应用,2010,46(2):118-120.

[7] 张弛,吴健,胡正国. 基于规约匹配的软件组件组装[J]. 西北工业大学学报,2005,23(5):623-627.

Research on Component Retrieval Based on Facet and Specification Description

CHEN Du-ying, LIU Shao-tao

(College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China)

Abstract: By analysis of components retrieval methods based on facet classification and component specification syntax. We find that the faceted classification and retrieval methods focus on the components' static characteristics, without considering the component behavior description. While component specification syntax description focuses on emphasize the aspect of component behavior description of the statutes of the components which is independent of the semantics of each function in the component description. We design a software component facets retrieval algorithm based on the string matching, and considering the facets retrieval didn't support for the demand of the assembly. In the retrieval process, insert specification syntax matching process based on component specification, to provide accurate information of the component composition and improve the efficiency of reuse.

Keywords: component retrieval; facet classification; specification description; string matching

(责任编辑: 钱筠 英文审校: 吴逢铁)