

基于数据驱动的工作流引擎的设计与实现

陈义松, 汪 芸

(东南大学 计算机科学与工程学院, 江苏 南京 211189)

摘 要: 工作流是计算机化的业务流程. 工作流引擎为工作流流程提供运行环境, 在流程的执行过程中起着关键性作用. 传统的工作流技术在处理复杂多变的流程时, 缺乏良好的建模以及适应机制. 本文提出一种基于数据驱动的工作流运行方式, 实现了活动与活动之间的松散耦合, 并以此为基础设计并实现支持这种数据驱动方式的工作流引擎. 该引擎在应对复杂多变的业务流程时具有较强的处理能力, 并能够实现流程的自动建模.

关键词: 工作流; 数据驱动; 工作流引擎; 订阅/发布

中图分类号: TP311

文献标识码: A

文章编号: 1000-7180(2012)11-0138-03

Design and Implementation of Data-Driven Based Workflow Engine

CHEN Yi-song, WANG Yun

(School of Computer Science and Engineering, Southeast University, Nanjing 211189, China)

Abstract: Workflow is a computerized business process. Workflow engine provides the runtime environment for workflow processes, playing a key role in the implementation of processes. The traditional workflow technology reflects a number of deficiencies in dealing with complicated and changing process. This paper proposes a data-driven based mode to achieve loose coupling between activities, and designs and implements workflow engine with the purpose of supporting this kind of data-driven mode. The engine copes with the complicated business processes with strong processing capabilities and can automatically mode the processes.

Key words: workflow; data-driven; workflow engine; publish/subscribe

1 引言

为适应现代企业业务流程的新特性, 很多学者开始致力于新的工作流技术的研究. 文献[1]在一定程度上可以支持流程的动态变化, 然而在此基础上的工作流仍是面向过程的控制驱动, 活动的执行顺序在流程建立时就已经确定, 活动之间具有较强的依赖性, 在面临复杂的结构变化时需要大量修改流程定义. 文献[2]可以在流程运行时出错或需要进行修改时进行部分替换, 但在映射时要求流程部分相似, 在一些情况下并不现实.

有学者提出以数据驱动工作流运行的概念. 文献[3]提出在流程运行前, 设计人员根据对流程的理解程度, 对流程部分定义. 流程运行时, 根据需要对

流程进行必要的补充, 随着对流程理解的逐渐深入, 完成整个流程的定义. 但这种方式要求流程开发人员在相应的领域具备一定专业能力.

文献[4]将 FreeFlow 中的活动执行的状态分为用户态和系统态, 用户态表示需要用户补充数据, 系统态表示流程可以自动运行. 这种方式虽然可以对数据的融合以及整个流程的构造做到有的放矢, 但是与以往的工作流驱动方式相比, 开发周期过长, 并且需要设计人员对相应的领域要有比较深入的了解, 这在很多情况下也并不现实.

在寻求数据驱动方式的领域, 较有影响的是 case handing^[5]. 该方法将工作流的执行方式从“应该做什么”转变为“可以做什么”. 通过利用能够得到的所有信息而不是部分信息驱动工作流的运行. 但

收稿日期: 2012-05-05; 修回日期: 2012-06-09

基金项目: 国家“八六三”高科技计划项目(2001AA040502)

这种方式并未显式支持复杂的对象和数据结构. 另一个对数据驱动工作流的研究产生较大影响的理论是 COREPRO^[6]. 该方式借鉴了面向对象程序设计中对象的概念, 并将其引入 workflow 技术中. 但是并没有对流程运行过程中出现问题提出有效的解决方法.

本文基于数据驱动的理念, 提出一种新的数据驱动工作流运行方式. 设计并实现了支持该方式下流程运行的引擎, 以提高工作流的灵活性和健壮性. 本文首先综述相关的研究状况; 然后提出以松耦合方式实现的工作流, 随后详细地论述了以数据驱动实现松耦合工作流引擎的设计与实现; 最后对文中提出的工作流引擎做一评价.

2 问题描述

实现数据驱动的工作流引擎要解决两个问题:

(1) 流程设计

控制驱动的工作流中, 将流程表示为如图 1 所示的有向图. 流程从 A 活动开始执行, 按照预先定义的顺序执行活动, 直至完成 F 活动. 这对于结构稳定的流程具有较高的执行效率, 然而对需要经常修改的流程来说则体现出明显不足: 对流程的修改可能非常繁琐, 甚至导致整个流程的重建. 随着计算机应用领域对工作流灵活性的要求越来越高, 这种传统的方式已不能适应当前需求, 造成了工作流的研究瓶颈, 甚至在一定程度上限制了 workflow 技术的发展. 因此, 需要引进新的设计方式实现流程的灵活性, 松耦合机制能够满足这一要求.

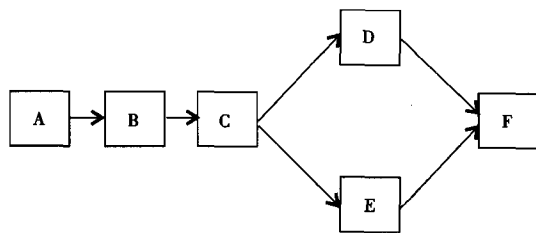


图 1 控制驱动下的流程结构

(2) 流程执行

传统的控制驱动的工作流的控制信息和数据信息都包含在整个流程中, 流程的各个部分都能访问到流程的控制信息和数据信息, 流程仅需根据流程的控制信息决定流程的运行方向. 而在本文提出的数据驱动方式下的工作流程中, 活动与活动之间是松耦合的关系, 所以必须要有相应的机制负责活动的调度, 以达到执行流程的目的.

3 松耦合工作流引擎设计与实现

为提高流程的灵活性, 必须改变现有流程的设计; 设计并实现支持松耦合流程执行的工作流引擎.

3.1 数据驱动流程的设计

以松耦合实现的数据驱动机制将触发活动执行的条件参数化作为输入, 根据输入判别活动的执行条件是否满足. 活动执行完后, 修改自身参数作为输出, 该输出同时也是其他活动的输入. 为使活动能够及时响应数据变化, 本文采用订阅/发布机制: 当活动的订阅内容更新时, 由订阅/发布系统通知引擎发送消息给该活动, 该活动继而判别当前输入是否满足运行条件, 若满足, 则激活该活动. 当前活动并不需要知晓由什么活动触发了它, 只需接收从引擎传输过来的数据. 此时活动与活动之间是一个松耦合的关系, 并没有像控制驱动方式下的显式控制. 在此方式下, 需要对流程重新设计. 如图 2 所示, 实线表示每个活动向订阅/发布系统订阅数据, 作为活动的输入. 活动完成之后更新其数据作为输出, 更新之后, 订阅/发布系统通知工作流引擎调用相应的活动, 如图中虚线表示. 由于当活动 B 获悉所需参数已更新并经自身判别之后满足执行条件, 则开始执行, 直至到达终止活动.

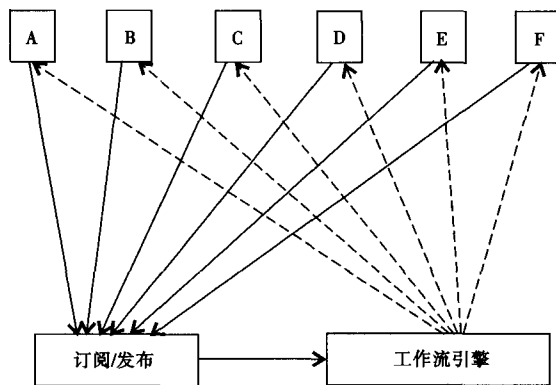


图 2 数据驱动下的流程结构

工作流流程定义包含了流程执行需要的所有信息, 包括流程的开始、结束状态, 流程的活动构成、执行规则、需要用户干涉的任务, 相关数据定义以及需要调用的应用等, 活动通过执行规则完成流程运行.

3.2 松耦合工作流引擎的设计

工作流引擎为流程的执行提供软件执行环境, 是流程执行的关键, 它直接关系到流程是否能够顺利执行以及流程的执行效率. 工作流引擎的主要功能包括: 流程的解释, 控制、管理流程实例的运行以及提供流程与外界的交互功能.

(1) 数据预处理层

负责解释整个流程并对相关数据进行处理,是流程执行前的必要准备.由于本文提出的这种以数据驱动建立的流程并没有显式的业务流程控制图,流程执行之前无法直观地观察到业务的执行拓扑序列,也就无法在流程建立后直接判别流程是否可行.所以该层还负责通过订阅\发表可以形成活动的拓扑图用以判断流程是否可以执行.

(2) 调度层

活动执行完成之后,由调度层选择下一个要执行的活动.流程在数据预处理层解析完成之后,由该层将每个活动的订阅数据传输给订阅/发布系统,对订阅/发布系统进行初始化.当前活动完成操作之后将自身的输出数据传送给订阅/发布系统,订阅/发布系统查询到数据更新后,通知引擎找到订阅这些数据的活动并分析活动的执行条件是否满足,若满足,则将这些活动提交给执行层;不满足,则不进行任何操作,这种情况下,订阅/发布系统如果在规定的时间内仍然没有接收到有效的更新数据,则终止该流程的执行.其调度算法如图 3 所示.

```
FindStartActivity(); // 获得初始活动
Init(初始活动); // 初始化初始活动
doJob(初始活动)
{
    excute(初始活动);
    if(活动为终止活动)
        end();
    updata(); // 更新数据
    publish(); // 寻找订阅更新数据的下一个活动的 ID 号
    CallNext() // 寻找下一个活动
    {
        for(活动 ID 表不为空)
        {
            根据活动的 ID 号查找活动 act;
            init(); // 初始化找到的活动 act
            if (act 执行条件满足)
                doJob(act);
        }
    }
}
```

图 3 活动调度算法

(3) 执行层

该层是整个工作流引擎系统的核心,负责控制活动的执行.其主要功能有:①根据活动定义完成活动操作;②控制不同权限的用户实行其权限内的操作;③调用系统应用与客户应用;④维护工作流相关

信息;⑤维护流程执行日志;

(4) 控制管理层

对于一些需要人工干预的活动,由该层负责角色的注册与退出.客户可以通过该接口实现对流程的干预,包括提前执行、挂起、激活、删除、暂停、终止等操作.

本文提出的流程执行方式注重数据在激活、调用活动中的作用,相应的引擎操作方式也从传统的被动接受活动的执行任务转移至主动寻求那些执行条件满足的活动.引擎通过对数据的管理,并在活动执行条件满足时激活活动.在这样的方式下,每个活动都独立地存在,它们之间通过参数的间接影响以实现活动之间的松散耦合.

3.3 松耦合工作流引擎的实现

当一组活动交给工作流引擎后,由引擎的解释模块将流程解析为工作流引擎可识别的表示方式.流程解析模块负责形成本流程的输入输出表(用于检测流程)、活动列表、流程相关信息(包括角色控制等工作流执行必需的所有信息)、活动项(活动内部操作)、活动需要调用的应用程序以及完成订阅发布系统的初始化.

初始化完成之后,引擎中负责检测流程的可行性,查看流程中是否存在循环以及是否可以到达预设的终止活动.执行模块还需维护流程执行日志,并随时更新已执行活动表、未执行活动表以及正在执行的活动表,通过这些表可以获知流程的执行状况.在执行模块对活动进行操作的同时,用户可以通过控制模块对流程的执行进行查询或必要的控制,以参与到流程的运行过程中.

4 结束语

本文基于数据驱动工作流的基本思想,设计和实现了数据驱动的工作流引擎.由数据驱动的理念,提出了设计流程的新方式,该方式降低了活动与活动之间的依赖程度,实现了活动与活动之间的松散耦合,提高了流程的灵活性以及健壮性,有利于流程的自动构建.并在此基础上设计并实现了相应的工作流引擎系统,验证了这本文提出的设想.提高工作流引擎的健壮性、适应性以及可扩展性将是我们今后研究的方向.

参考文献:

- [1] 罗海滨,范玉顺,吴澄. 工作流技术综述[J]. 软件学报. 2000. (下转第 144 页)

参考文献:

- [1] 胡丽芳,关欣,邓勇.一种三角模糊数型多属性群决策方法[J]. 控制与决策, 2011,12;(26):1878-1880.
- [2] 刘姗姗.基于三角模糊数的价值工程评标方法研究[D]. 天津大学, 2008.
- [3] 龚艳冰.方案偏好已知的三角模糊数型多属性决策方法[J]. 控制与决策, 2012,27(2):281-284.
- [4] 陈晓红,阳熹.一种基于三角模糊数的多属性群决策方法[J]. 系统工程与电子技术, 2008,30(2):279-282.
- [5] Zhou-jing Wang, Edward Y Qian. A vague-set-based fuzzy multi-objective decision making model for bidding purchase[J]. Journal of Zhejiang University SCIENCE A, 2007,8(4):644-650.
- [6] Jue Wang, San-Yang Liu, Jie Zhang, et al. On the Parameterized OWA Operators for Fuzzy MCDM Based on Vague Set Theory[J]. Fuzzy Optimization and Decision Making, 2006,5(1):5-22.
- [7] 汪金霞.基于梯形模糊数多属性群决策方法研究[D]. 合肥工业大学, 2009.
- [8] 周亚.多属性决策中的 TOPSIS 法研究[D]. 武汉理工

大学, 2009.

- [9] Zhang L, Li R H. Designing of classifiers based on immune principles and fuzzy rules[J]. Information Sciences, 2008(178): 1836-1847.

作者简介:



郑世明 男,(1981—),博士,工程师.研究方向为数据挖掘,作战仿真.

高志年 男,(1957—),教授.博士生导师,研究方向为计算机仿真,军事运筹.

张 璐 男,(1982—),硕士,工程师.研究方向为通信服务性能测试.

吴从晖 女,(1975—),硕士,副教授.研究方向为作战模拟.

韦 伟 男,(1965—),博士,教授.研究方向为作战模拟.

(上接第 140 页)

- [2] Manfred Reichert, Stefanie Rinderle, Ulrich Kreher, Peter Dadam. Adaptive Process Management with AD-EPT2[C]// Proc Int Conf on Data & Knowledge Engineering. Tokyo, Japan: IEEE, 2005.
- [3] Pesic M, Schonenberg H, W M Pvander Aalst. DECLARE: full support for loosely-structured processes [C]// proceedings of the 11th IEEE International Enterprise Distributed Object Computing Conference, Washington, DC, USA: IEEE, 2007: 287-298.
- [4] Cutler R, Davis L. Robust real-time periodic motion detection, analysis, and application[J]. IEEE Trans Pattern Analysis and Machine Intelligence, 2000.

- [5] W M P van der Aalst, Mathias Weske, Dolf Grunbauer. Case Handling: a new paradigm for business process support[J]. Data & Knowledge Engineering, 2005, (53): 129-162.

- [6] Dominic Müller, Manfred Reichert, Joachim Herbst, et al. Data-driven modeling and coordination of large process structures[J]. Computer Science, 2007.

作者简介:

陈义松 男,(1986—),硕士.研究方向为柔性工作流技术.

汪 芸 女,(1967—),博士生导师.研究方向为分布式计算、中间件技术和容错理论.