

Web 系统维护中的逆向工程研究

王爽¹, 华庆一²

(1. 西北大学 信息学院 计算机科学系, 陕西 西安 710069;

2. 中国科学院 计算机科学国家重点实验室, 北京 100080)

摘要:分析了软件维护当中的问题,提出了一种针对 Web 系统的逆向工程方法。该方法以 Web 站点为输入分析页面的源码,从页面链接和交互进行逆向,构建出站点的部署和任务模型,从而直观地表示出对维护人员有帮助的系统信息,同时介绍了笔者自行开发的逆向工程辅助工具。通过该方法,维护人员能够直接得到系统信息,而不必去分析源代码,克服了缺乏设计文档的困难。通过一个 Web 站点的逆向实例,演示了这种方法。

关键词:逆向工程;可扩展标记语言;数据抽取;任务模型

中图分类号:TP311.53

文献标识码:A

文章编号:1673-629X(2008)01-0129-03

Research on Reverse Engineering in Maintenance of Web System

WANG Shuang¹, HUA Qing-yi²

(1. Department of Computer Science, Institute of Information Science, Northwest University, Xi'an 710069, China;

2. State Key Laboratory of Computer Sci., Chinese Academy of Sci., Beijing 100080, China)

Abstract: Analyses the problems in the maintenance of software, presents the approach of reverse engineering from Web system. This approach uses Web as input to analyse the source code of pages, reversing the link and interaction to construct the deployment and task model, thus it can express useful information clearly for maintainer, and provide a reverse engineering tool developed by the author. Using this method, maintainer can get the information of system without analyzing the source code, and get over the difficulty of lack for design documents. Finally, gives an example of reverse engineering for a Web site to illustrate this method.

Key words: reverse engineering; XML; data extraction; task model

0 引言

随着软件规模和复杂度的日益提高,软件维护已成为软件生存期中费用最高、难度最大的一个阶段^[1]。由于维护人员相对于原系统开发人员对系统的理解程度不同、系统设计文档的缺乏等原因,维护性开发工作变的非常困难。当维护人员接到一个 Web 系统的维护性项目时,需从页面的操作去了解系统的大体功能和页面的部署情况。这时如缺乏相关的文档,就需通过试用系统 release 版本的方法来熟悉系统,必要时甚至还要去分析页面的静态和动态代码,这给维护工作带来了不便。而逆向工程对现有系统进行分析,从中抽取和恢复设计信息,识别出系统的组成及相互的关

系,生成对系统结构的描述文档,给维护工作提供了极大的方便。在 Web 系统中利用逆向工程方法,一方面分析页面交互构建页面的任务模型,另一方面分析页面的链接,得到 Web 站点的部署信息。这样就可直观快速地了解系统,为理解系统的其他模块节约了时间。

1 软件维护的现状

软件维护^[2]是一项复杂的活动,软件维护是指软件产品在交付给用户投入使用后,所有与软件修改有关的活动。在维护过程中^[3],要正确有效地修改程序,必须充分地理解程序,如程序的功能和目标,程序的结构信息,数据源的构造以及对程序的操作。Web 系统在维护阶段是要经常更新的。更新的规模有大有小,更新的内容可以是页面的内容,也可以是系统的结构框架。由于这些变化的大量存在,加之有些系统规模庞大,给程序理解带来了很大的困难。由此,提出了对 Web 系统的逆向研究,分析现有系统,在比源代码高的层次上创建程序的某种表示^[4]。

收稿日期:2007-08-03

基金项目:国家“863”高技术发展研究计划(2006AA01Z328);中科院计算机科学国家重点实验室开放课题基金(SYSKF0704)

作者简介:王爽(1983-),男,河南开封人,硕士研究生,研究方向为人机交互;华庆一,教授,博士,研究方向为人机交互的理论和新技术、三维可视化技术及分布式交互系统。

2 逆向工程的基本概念

软件逆向工程是软件工程的一个重要分支。对软件逆向工程最好的定义是:对现有系统的分析过程,识别出系统各组成部分及其相互间的关系,并将系统以其它形式来表现^[5]。

在传统的软件开发周期模型中,首先是问题定义、可行性分析以及需求分析,然后进行设计,最后进行编码、测试,最终进入维护阶段^[6]。这是一个从高层抽象和逻辑实现转变为物理实现的过程。而逆向工程的过程相反,它的目的在于对现有系统得到高层次的设计信息。抽象层次越高,开发人员越容易理解^[2,7]。

3 Web 系统逆向工程解决方案

在 Web 系统维护中主要关注对页面的分析,因而在逆向工程时主要从站点的任务和部署的提取两方面来研究。

3.1 任务模型的逆向

在 Web 系统中有大量的交互技术。对这些交互用构建任务模型的方法来描述。任务模型是描述系统要完成用户目标所涉及的一系列动作,同时还描述各动作之间的相互关系。在任务模型中,一般按功能抽象程度将任务进行逻辑划分并组成层次结构^[8,9]。用任务树来表示任务模型。在任务树中,采用层次化结构,低层结点表示用户与界面进行交互操作且不能再分解的任务,称为基本任务,用“○”表示。由多个交互组成的任务称为复杂任务,用“□”表示。图 1 描述了任务模型结点的含义。

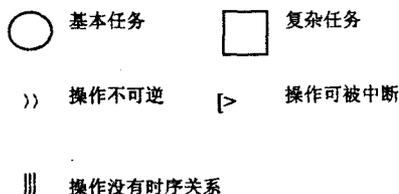


图 1 任务模型的结点

页面中主要有以下几种交互:

* 按钮(button):代码里是通过“<INPUT type = “button”>”联系任务的,通过它就可以分析提交信息的任务。

* 单选按钮(radio):单选按钮的选择是一个基本任务,用“select”+ 单选按钮名。

* 文本框(text area):文本框的任务模型可以用“select”+ 文本框名表示。

* 多选框(checkbox):多选框是对多个对象的交互,其任务模型可以再细分为多个子任务。

3.2 站点部署的逆向

对于站点部署的研究和分析,主要是分析页面中的链接信息。通过自行设计的逆向工程工具分析其静

态页面的源码从而得到这些页面的部署图^[10]。该工具运行环境为在 Windows 平台,处理的对象是静态页面的源代码。

要提取页面的链接信息,必须在代码中找到字符串“href”,该字符串后的文件名就是要找的连接信息,其后缀一般为“html”或“htm”。笔者设计的这种逆向工程工具以 Web 站点为输入,通过文件的路径和所要查找的字符串“href”查找字符串后第一个完整的文件名。

比如一个简单的 Web 网站包含如下文件:首页: index.html;关于我们: about.html;公司介绍: gsjs.html;产品简介: cpjj.html;技术支持: jszc.html;联系我们: lxwm.html。它们相互之间都有链接,那么这些页面逆向出来的部署信息如图 2 所示。

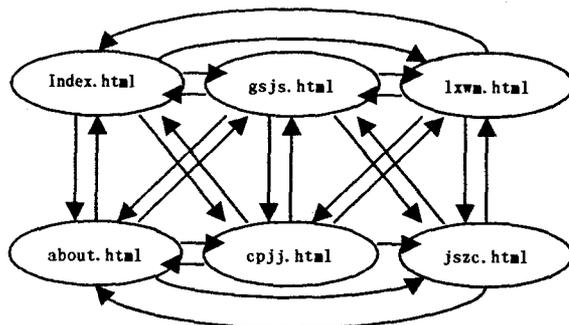


图 2 页面部署图

3.3 逆向工程的一个简单实例

根据以上论述,用一个简单的 Web 页面模拟在 Web 系统中页面的逆向。图 3 是所要逆向的页面。

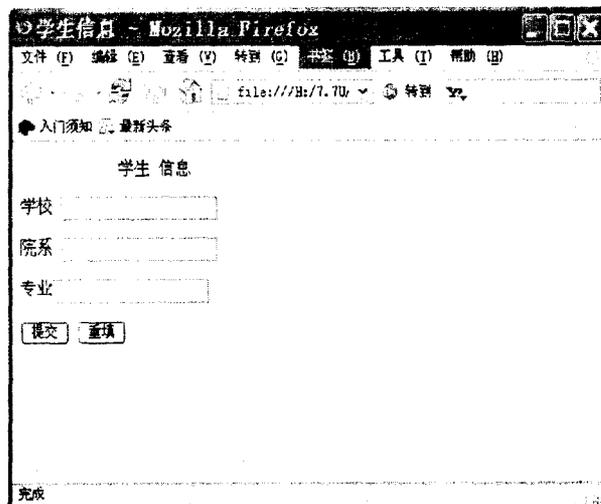


图 3 实例页面

以下是该页面的核心代码:

```
<title>学生信息</title>
<form method = “post” name = “form1” action = “saveinfo. htm”>
<!-- “saveinfo” 将表单信息存于 info. txt 中 -->
```

```

<p> 学生 信息 </p>
<p>学校 <input name="text" type="text"></p>
<p>院系 <input name="text2" type="text"></p>
<p>专业 <input name="text3" type="text"></p>
<p>
  <input type="submit" name="Submit" value="提交">
  <input type="submit" name="Submit2" value="重填">
</p>
</form>

```

图 4 是根据该页面的表单代码逆向得到的任务模型。在该页面的代码中, input 为输入任务。表单中选择按钮 submit 和 reset 分别表示提交信息(send form)和重填(reset)两个任务。

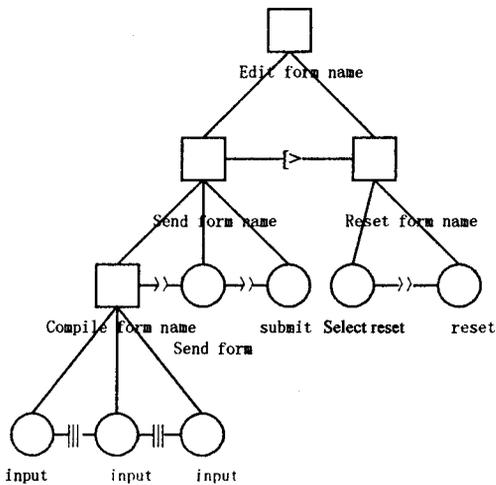


图 4 任务模型

笔者对 Web 系统的逆向工程解决方案还存在一些问题。如该方案对站点部署的逆向还只限于静态网站,没有涉及动态网站的逆向。这些都还要在以后的工作中做深入研究。

4 总 结

对 Web 系统维护中的逆向工程技术进行了分析,对软件维护当中的问题,逆向工程基本方法进行了简要介绍。最后用一个 Web 页面逆向的例子提出了针对 Web 系统逆向工程的方法。由于 web 系统应用的越来越广泛,技术越来越复杂,因此也越来越有必要对 Web 系统逆向工程进行更深一层次研究和发展。

参考文献:

- [1] Bennet K H. The Software Maintenance of large Software Systems: Management, Methods and Tools[M]. [s. l.]: Elsevier Science Publishing Co., Inc., 1990: 88-98.
- [2] 普瑞斯曼. 软件工程 - 实践者的开发方法[M]. 第 5 版. 黄柏素, 梅 宏译. 北京: 机械工业出版社, 1999.
- [3] Biggerstaf T J. Design Recovery for Maintenance and Reuse [J]. IEEE Computer, 1989, 22(7): 36-49.
- [4] 张 亮, 赵文耕, 彭 鑫, 等. 一种支持时序图恢复的逆向工程方法[J]. 南京大学学报, 2005, 41(10): 386-387.
- [5] Elliot C, James C. Reverse Engineering and Design Recovery: a Taxonomy[J]. IEEE Software, 1990, 7(1): 13-17.
- [6] 刘宗田. 软件维护与逆向工程评述[J]. 计算机软件与应用, 1995(1): 1-8.
- [7] Jamesrunbaugh, Javobson J, Booch G. UML 参考手册[M]. 第 2 版. UMLCHINA 译. 北京: 机械工业出版社, 2005.
- [8] Merlo E. Reverse Engineering[J]. IEEE, 1993, 41(3): 171-173, 176-178.
- [9] Milner R. Communication and Concurrency[J]. IEEE Software, 1989, 37(4): 268-280.
- [10] Estiévenart F, Francois A, Henrard J. A tool supported method to extract data and schema from web sites[C]//Proc of the 5th IEEE Int'l Workshop on Web Site Evolution. New York: IEEE Press, 2003: 3-11.

(上接第 128 页)

- [5] Stricker M A, Orengo M. Similarity of color images[C]//In: Proc of SPIE: Storage and Retrieval for Image and Video Databases III. San Jose, CA: [s. n.], 1995: 381-392.
- [6] Hsu W, Chua T S, Pung H K. An integrated color - spatial approach to content - based image retrieval[C]// In: Proc ACM Multimedia '95 Conference. San Francisco: [s. n.], 1995: 305-313.
- [7] Smith J R, Chang S F. Tools and techniques for color image retrieval[C]//In: Proc of SPIE: Storage and Retrieval for Image and Video Database. San Jose, CA: [s. n.], 1996: 426-437.
- [8] 章毓晋. 图像理解与计算机视觉[M]. 北京: 清华大学出版社, 2000.
- [9] Mehre B M, Anhalii M K, Lee W F. Shape Measures for

- Content Based Image Retrieval: a Comparison[J]. Information Processing and Management, 1997, 33(3): 319-337.
- [10] ZHANG Lei, LIN Fu-zong, ZHANG Bo. Support vector machine learning for image retrieval[EB/OL]. International Conference on Image Processing. (2001-08-16)[2006-12-15]. <http://research.microsoft.com/users/leizhang/Paper/ICIP01.pdf>.
- [11] HONG Peng-yu, TIAN Qi, HUANG T S. In corporate support vector machines to content - based image retrieval with relevant feedback[EB/OL]. Proceedings of International Conference on Image Processing. (2000-03-10)[2006-12-15]. <http://www.ifp.uiuc.edu/~qitian/e-paper/icip00svm/icip00-svm.pdf>.