

一种 Web 应用测试结果分析比较技术^{*}

许 蕾^{1,2} 徐宝文^{1,2,3}

(东南大学计算机科学与工程系 南京 210096)¹ (江苏省软件质量研究所 南京 210096)²

(武汉大学软件工程国家重点实验室 武汉 430072)³

摘要 合理分析、正确比较 Web 应用的测试结果能够有效评判测试工作的效果和效率。在现有工作的基础上,我们提出一种新的测试结果分析比较方法,利用语义标注和 XML 描述技术实现 Web 页面中数据与显示信息的分离,从而方便测试结果与预期结果的直接比较;并进一步引入反馈控制机制,把测试结果反馈给 Web 应用本身,根据 Web 应用的实际状况来指导其发展和演化,使得系统质量得到进一步改善。

关键词 Web 应用,测试结果,分析比较,反馈控制

An Analyzing and Comparing Technique for Testing Results of Web Applications

XU Lei^{1,2} XU Bao-Wen^{1,2,3}

(Department of Computer Science & Engineering, Southeast University, Nanjing 210096)¹

(Jiangsu Institute of Software Quality, Nanjing 210096)²

(State Key Laboratory of Software Engineering, Wuhan University, Wuhan 430072)³

Abstract Abstract Reasonable analysis and corrective compare for the testing results of Web applications can effectively judge the effect and efficiency of the testing. Therefore, based on the previous work, we propose a new method for testing results analysis and comparison, which uses the semantic label and XML description technique to realize the information separation between data and display in the Web pages, so as to directly compare the testing results and the expected results; Furthermore, we introduce the feedback control mechanism into the development and evolvement of Web applications, so as to further improve the system quality.

Keywords Web application, Testing result, Analysis and comparison, Feedback control

1 引言

Web 应用从最初的静态文本显示发展到目前的个性化服务,在动态、交互等功能上取得了很大突破,吸引了越来越多的用户,并在服务质量上也有了很大改善。在增强 Web 应用功能的同时,也增加了系统复杂性,因而相应的测试、维护工作亟待加强。Web 应用测试是保障其质量的一项重要措施和方法,得到人们越来越多的重视,并取得一些重要研究成果^[3~5,7]。我们在 Web 应用的测试方法和技术、测试用例生成选择策略以及相关测试工具原型的研究开发等相关方面已进行了一定的尝试和探索^[8~11]。

执行完 Web 应用测试后,为了评价 Web 应用的质量和测试的效率,需要根据预期输出结果对实际测试结果进行深入分析比较,以确定是否通过此项测试;此外,由于 Web 应用本身处于不断的发展变动中,还需要对故障问题进行修正和维护,以进一步改善系统质量。本文将探讨如何实现高效的 Web 应用测试结果分析比较,在第 2 部分提出一种新的测试结果分析比较方法,利用语义标注和 XML 描述技术实现 Web 页面中数据与显示信息的分离,从而方便测试结果与预期结果的直接比较;结合实际情况,第 3 部分引入反馈控制机制,把测试结果反馈给 Web 应用本身,使得系统质量得到进一步改善;最后是对我们工作的总结。

2 测试结果分析比较方法研究

通过比较分析测试用例中的预期测试结果和实际测试结果,可以发现 Web 应用中存在的一些问题和故障,从而判断 Web 应用是否通过此项测试,即:如果实际结果与期望输出相吻合,那么表明 Web 应用通过了该项测试,能够保证功能上的正确性或者性能上的可靠性;如果两者存在着差异,则需要错误的排查、定位和修复,并补充新的测试用例进行再测试(即回归测试)。因而,测试结果的分析比较将直接关系到何时终止测试。

目前,关于如何进行 Web 应用测试结果分析比较的研究并不多见,这是由于 Web 应用测试缺乏必要的规范,对测试结果的分析比较没有严格的过程要求,另外 Web 应用测试主要由测试人员人工完成,对测试结果的评判也是由人工审查而实现,整个过程显得较为随意。为了适应 Web 应用测试内容庞杂、任务繁多的特性,提高测试效率,研究人员已经在工具开发、测试自动化等方面开展了一定的工作,相应也要尽量实现自动化的测试结果分析比较。

2.1 存在的问题

在自动执行测试结果的分析比较方面,目前的困难主要在于预期结果和实际结果的表述不够精确。目前,用于描述 Web 应用的主要还是缺乏语义的 HTML 语言,只是定义元

^{*}本研究得到国家自然科学基金(60425206,9412003,60503033)、国家重点基础研究发展规划 973 资助项目(2002CB312000)、江苏省高新技术研究项目(BG2005032)、武汉大学软件工程国家重点实验室开放基金资助。许 蕾 博士,讲师,研究方向:Web 应用分析测试。徐宝文 博士,教授,研究方向:程序设计语言、软件工程、并行与网络软件等。

素在浏览器中的显示格式,而显示的实际内容和客户端的配置状况密切相关,从而难以进行页面元素的自动识别,给测试结果的自动判断带来了很大麻烦。另外,一些元素的显示和其周围元素提供的语义环境密切相关,而计算机通常难以自动识别这些语义环境,因而对服务器返回结果的正确性进行自动判断是非常困难的。现有一些方法能够部分实现测试结果的自动判断,如在整个页面的 HTML 文本中查找预期的字符,如果查找到,就说明测试结果正确;而如果出现“Error”、“Not Found”等字符时,就说明结果错误。不过,这种方法只是进行了简单的字符查找匹配,没有结合语义信息进行分析,误判的概率非常大,另外当输出结果包含在多媒体元素(如图片)中时,由于无法找到对应文本,该方法就会完全失效。

2.2 解决办法

针对上述 Web 应用特性,我们的主要研究工作就是将 Web 页面中的数据与显示分离,以方便相关数据的提取和比较。由于难以直接从 HTML 页面中提取所需信息以直接判断输出页面是否正确,因而需要采用适当的工具并按照一定的策略进行,具体过程示意图 1。

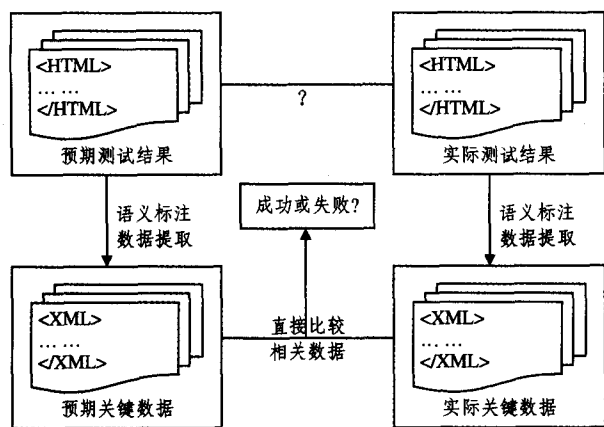


图 1 Web 应用测试结果分析比较过程示意

首先,在预期测试结果的准备方面,需要事先多次预运行测试用例,以获得较全面的各种动态返回结果,并按照功能的差别将返回页面划分、合并成不同的种类,同时还需要通过元素识别技术去除页面显示内容并提取页面数据内容,比较重要的信息有:页面 URL 地址、文件名、变量名及其取值等,然后引入 XML 语言来描述预期测试结果的相关结构和数据;

其次,对实际测试结果,同样通过元素识别技术从返回页面上提取关键信息,即将返回页面上的数据元素与显示元素分离,并选择与测试目的相对应的内容,也采用 XML 来描述;

再次,直接比较两个 XML 文档,可以通过自动识别查找对对应项中描述的数据,并通过对这些数据的比较以决定是否通过测试。在比较时需要按照一定的次序进行,即根据数据内容的重要程度先比较 URL 地址和文件名等数据信息,如果有差异,才需要继续比较变量、取值等其它数据信息,且比较时只要查看数据的本质内容,不用拘泥于显示方式的差别。

通过这样的方式进行 Web 应用测试结果的分析比较,就能够避免 2.1 节中的问题,并且由于我们直接比较相关数据信息,在结果的精度和整个过程的效率上都将会有明显改善。

在上述过程中,需要通过一定的方式在可能引起混淆的

页面元素上标注相应的语义信息,这就牵涉到语义标注^[2,6]的工作。为此,首先需要建立对应的本体,用以表示一个领域内的概念以及概念之间的关系(如 Web 应用测试本体或者页面元素本体),并给出确定性的描述;然后在一些工具的辅助下对页面上相关的内容进行语义标注,以使其具备自解释的能力,从而方便页面元素的提取以及对应内容的比较。

3 Web 应用测试结果的影响和反馈

根据测试结果的具体情况,如果 Web 应用中存在故障,则需要对故障的查找和修复工作,并进行必要的回归测试。同时,由于 Web 应用本身处于不断的发展变动中,这些通过测试发现的问题也能够对 Web 应用的发展、演化起到很好的指导和推动作用。

可以应用控制论^[1,13]的理论和方法,将测试结果反馈给 Web 应用本身,从而明确 Web 应用中存在的问题以及需要改进优化的地方,增强系统的功能并改善系统的可靠性、可用性、安全性等。具体的工作包括故障修复、经验提取以及系统优化等。

3.1 根据测试结果排查故障

根据测试得到的结果,可能会存在实际结果与期望结果不同的情况,这说明在 Web 应用中或者测试环节上出现了问题。在 Web 应用中,故障可能存在于客户端脚本、服务器端功能模块、甚至网络传输过程中,并且出错提示信息非常含糊,或者需要通过与期望输出对比后才知道是否出错。因此,如何查找、定位故障就显得非常重要,需要一些策略和技巧。

通常的方法就是首先记录好当前的相关信息,包括出错提示、系统状态以及环境信息等,这样可以为排查故障提供必要的线索,并且可以通过追踪重演的方式来重现错误发生的情景;然后根据线索以及测试人员的经验等因素查找可能的故障原因,并根据需要补充测试用例进行测试,以逐渐缩小可能故障原因的范围,直至最终查出原因。由此可见,测试结果为故障修复提供了第一手的现场资料和线索支持,测试结果对故障修复起到非常重要的作用。

3.2 测试经验总结

在大量实践的基础上,通过对测试结果的统计分析和经验总结,还可以提取出一些规律性的东西,如常见错误分类、设计模式归并、度量指标参照等,并需要不断地总结和更新。

就常见错误分类而言,一般包括错误名称、症状以及产生原因等,如“链接不可达”这一错误,其对应的症状是返回出错提示页面,显示“HTTP 404 Error”或者“File Not Found”等,其原因在于链接的指向有误。在实际测试时,如果可以借鉴这些常见错误分类,将会为测试人员查找和排除故障提供宝贵的经验,并起到积极的指导作用。

进一步,结合测试结果以及用户使用 Web 应用的反馈信息,可以确定符合用户实际需求的设计模式,如 Web 应用的结构框架安排、内容元素的层次组织等,从而能够安排好 Web 应用的结构和内容,方便用户的浏览和使用。具体的设计模式包含使用框架要尽量避免嵌套、到达目标页面的链接跳转次数要尽可能少、图片等多媒体元素在页面上的比重不可过多等。

同时,根据现有 Web 应用的实际运行状况,对其度量指标的取值具有很好的参照意义,事实上,Web 应用度量指标的参考值正是在总结了很多实际应用的基础上获得的。因此,不断地总结更新实际 Web 应用的度量取值,能够使得对

应的参考值具有良好的时效性。

3.3 Web 应用的优化和演化

在上述实践经验的有效指导下,我们能够对 Web 应用进行系统优化^[12],并有利于维护和演化等方面工作的开展,具体体现在以下几个方面:

首先是通过常见错误分类来方便故障的查找和排除,为测试人员提供非常有价值的经验和参考,有效提高测试人员的业务水平,进而改善测试工作的效率;

其次是在设计模式的指导下,Web 应用的逻辑结构更趋于合理、内容组织更加有层次且专业规范,能够更好地适应用户的需求,并且对整个系统性能的改善也起到积极的作用;

再次,通过计算 Web 应用度量指标的实际取值以及对参考值的及时更新,不仅能够明确该 Web 应用的质量水平和测试效率,还可以为其它 Web 应用的度量发挥参照作用。

小结 如何有效地处理 Web 应用测试结果,目前所受到的关注度还不是太高。但作为测试流程中的最后一步,Web 应用测试结果的分析比较还是必不可少和非常重要的。因为只有通过测试结果评判后,才能够决定测试是否可以终止;另外,通过对测试结果的统计分析,还可以得到一些故障排查的线索、设计测试中的经验等,从而有利于系统的不断进化。

本文主要研究如何对测试结果进行分析比较、如何利用测试结果反馈作用于 Web 应用以取得良好的发展演化效果。主要利用语义标注和 XML 描述技术实现 Web 页面中数据与显示信息的分离,从而方便测试结果与预期结果的比较;将测试结果反馈给 Web 应用本身,以便于故障修复并提取一些操作经验,然后应用到整个系统的优化工作中去,以指导其发展和演化。

参考文献

1 Cai K Y. Optimal software testing and adaptive software testing

in the context of software cybernetics. Information and Software Technology, 2002, 44: 841~855
 2 Handschuh S, Volz R, Staab S. Annotation for the deep Web. IEEE Intelligent Systems, 2003, 18(5): 42~48
 3 Kallepalli C, Tian J. Measuring and Modeling Usage and Reliability for Statistical Web Testing. IEEE Trans Software Engineering, 2001, 27(11): 1023~1036
 4 Liu C H. A Formal Object-Oriented Test Model for Testing Web Applications, 2002, Doctor Dissertation
 5 Ricca F, Tonella P. Web Testing: a Roadmap for the Empirical Research. In: Proc. of the 2005 7th IEEE Int. Symposium on Web Site Evolution (WSE'05), 2005. 63~70
 6 Soo Von-Wun, Lee Chen-Yu, Li Chung-Cheng, et al. Automated Semantic Annotation and Retrieval Based on Sharable Ontology and Case-based Learning Techniques. In: Proc. of the 2003 Joint Conference on Digital Libraries (JCDL'03), 2003. 61~72
 7 Warren P, Gaskell C, Boldyreff C. Preparing the ground for website metrics research. In: Proc. of the 3rd International Workshop on Web Site Evolution, 2001. 78~85
 8 Xu Lei, Xu Baowen. Research on the Analysis and Measurement for Testing Results of Web Applications. Second International Workshop on Web Computing in Cyberworlds (WCCW2005), 2005. 559~565
 9 Xu Lei, Xu Baowen, Nie Changhai, Huowang Chen, Hongji Yang. A Browser Compatibility Testing Method Based on Combinatorial Testing. Proc. of the Int. Conference on Web Engineering (ICWE), 2003. 310~313
 10 许蕾, 徐宝文. 用户行为获取方法在 Web 性能测试中的应用研究. 软件学报, 2003, 14 (增刊): 115~120
 11 许蕾, 徐宝文, 陈振强. Web 测试综述. 计算机科学, 2003, 30 (3): 100~104
 12 阳小华, 周龙钺. 基于用户访问模式的 WWW 浏览路径优化. 软件学报, 2001, 12(6): 846~850
 13 张卫丰, 徐宝文. 带反馈自适应 Web 搜索引擎研究. 计算机科学, 2004, 31(9): 3~8

(上接第 197 页)

别,并且保证每种类别有一定数量的实例个数,以满足 RBF 神经网络的训练需求。为了具有普适性,在对数据集进行分析的基础上,分别选定两组阈值参数进行试验,由于篇幅所限,此处只给出试验结果比较情况,如表 1 所示。

表 1 SCRBF 与 LSC 算法试验结果比较

算法	阈值参数 1	孤立点检测率	阈值参数 2	孤立点检测率
SCRBF	$k=100, \delta=0.10, \theta=0.03$	72.3%	$k=200, \delta=0.15, \theta=0.05$	45.5%
LSC	误差=0.03	68.4%	误差=0.05	43.9%

两组不同阈值参数的实验结果都表明,在误差率保持相等的情况下,SCRBF 算法的孤立点检测率都要高于 LSC 算法。而且,内在的隐单元简化机制使同样的问题 SCRBF 所需的网络结构相对较小,具有较低的网络复杂性。但本实验比较也存在不足之处,一是受限于实验用机的性能,所选取的训练集较小,SCRBF 在大训练集中的应用效果还不能完全确定;二是在本实验中,由于数据集限制,未出现噪声数据进入网络的情况,使 SCRBF 算法的一个重要的优越性未能体现出来。但这也从另一个侧面验证了采用统计聚类方法对神经网络进行初始化,可以有效地降低引进噪声样本点为隐单元的可能性。

结束语 本文提出的 SCRBF 孤立点检测算法创造性地将统计聚类方法引入孤立点检测中,既克服了 LSC 算法的泛化能力弱的缺陷,又通过合理的隐单元简化机制防止了传统 RBF 神经网络容易过拟合的缺点。对比实验表明,SCRBF 算法与 LSC 算法相比,在检测孤立点问题上具有很大的优越性。SCRBF 用于孤立点检测还有一个优点就是,可以在完成其他数据检测任务(如分类和聚类)的同时,完成孤立点检测的任务。由于结合了统计方法,使人工神经网络易产生的随机性和过拟合作用造成的不利影响减小,从而提高了孤立点检测的可靠性。

参考文献

1 Agyemang M. Local Sparsity Coefficient- Based Mining of Outliers. Windsor, Ontario, Canada; University of Windsor, 2002
 2 徐秉铨,张百灵,韦岗. 神经网络理论与应用. 广州:华南理工大学出版社,1994
 3 蒋建春,马恒太. 网络安全入侵检测:研究综述. 软件学报,2000, 11(11):1460~1466
 4 李俭川,秦国军,温熙森. 神经网络学习算法的过拟合问题及解决方法. 振动、测试及诊断,2002(4):260~265
 5 Axelsson S. The base-rate fallacy and its implications for the difficulty of intrusion detection. In: Tsudik, G, ed. Proceedings of the 6th Conference on Computer and Communication Security. New York: ACM Press, 1999. 1~7