

Web 应用测试的通用编解码器研究

柳永坡¹ 吴 际¹ 刘霜梅²

(北京航空航天大学计算机学院 北京 100191)¹ (青岛农业大学理学与信息科学学院 青岛 266109)²

摘 要 编解码器是 TTCN-3 测试系统的重要组成部分,用来对 TTCN-3 测试数据和被测软件的输入/输出数据进行双向转换。面向 Web 应用测试系统的通用编解码器的研究,就是要把 TTCN-3 测试数据转化成 URL 请求,并对网页进行分析,抽取测试关心的信息,将其转化成 TTCN-3 测试数据。针对 Web 应用页面数据量大、干扰信息多的特点,提出了多层解析组装的实现方案。实践表明,该方案可行,可大大减少编解码开发工作量。

关键词 TTCN-3,测试系统,Web 应用测试,编解码器,多层解析

中图分类号 TP311.5 **文献标识码** A

Research of Generic Codec for Web Application Testing

LIU Yong-po¹ WU Ji¹ LIU Shuang-mei²

(School of Computer Science and Engineering, Beihang University, Beijing 100191, China)¹

(Science and Information College, Qingdao Agricultural University, Qingdao 266109, China)²

Abstract Codec is a necessary part of TTCN-3 test system which transfers TTCN-3 template data to SUT input data, and SUT output data to TTCN-3 template data. This paper presented the design of generic codec for testing Web application. The generic codec needs to transform TTCN-3 template into URL request string, and extract Web page content to transform into TTCN-3 template. Conscious of the Web pages with large volume of data and noise data, this paper presented the multi-level parsing and composition strategy. Experiments show that the strategy works well, and could save much codec development effort.

Keywords TTCN-3, Test system, Web application testing, Codec, Multi-level parsing

TTCN-3 是由 ETSI 制定和推动的测试规范和测试实现标准。它是一种描述能力丰富的基于黑盒的测试描述规范,能够应用于多种形式的分布式系统的测试规约描述^[1]。随着 TTCN-3 测试技术的不断发展,它已经被越来越多地应用到各种测试领域,基于 TTCN-3 的测试系统开发已经具备了和软件开发相似的特征^[2]。但随着测试系统规模的增长以及测试人员的变更,对于庞大的测试系统的管理和维护,已经变得越来越困难。因此对基于 TTCN-3 的测试系统进行逆向工程,可以帮助测试人员从更高层次把握测试系统的设计,同时可以检验测试设计和测试实现之间的一致性,对于测试系统的维护、扩展以及评估都有重大的意义和重要的价值。

TTCN-3 由于具有丰富的数据描述手段,因此适用于测试基于数据交互的软件或系统。针对 TTCN-3 的领域适用性,作者曾开展以 Web 应用软件和桌面应用软件为被测对象的测试研究。通过对 Web 应用软件进行负载、功能和 API 测试,对桌面应用软件进行 GUI、功能和 API 测试,获得了关于测试开发时间、测试代码规模等一系列实验数据^[3]。在对数据进行有效的分析之后,从语言规范、被测系统和测试开发人

员 3 个方面对 TTCN-3 的适用性进行了分析与评估。研究表明,无论是 TTCN-3 标准自身,还是其支持工具,都更加适合测试以数据交互为核心的 Web 应用,此结论也是本文的理论基础。但是在面向 Web 应用的测试系统构造方面,编解码的开发工作量非常大,不同网页需要的编解码规则也存在较大的差异。因此本文对于 Web 应用软件测试系统中通用编解码器设计与实现的研究具有重要的实用价值和前瞻性。

1 TTCN-3 测试系统

TTCN-3 测试系统是一个可以与被测软件进行交互从而完成指定测试行为的系统,包含 6 个实体和 2 个接口标准,每个实体都对应测试系统的一个功能实现。这些实体负责管理测试执行、环境适配、结果收集,以及完成与被测系统的通信、实现外部函数、处理计时器操作等,如图 1 所示。

(1) TTCN-3 执行部件(TTCN-3 Executable, TE),可执行测试代码;

(2) 系统适配器(System Adapter, SA),通过抽象测试系统接口来适配 TTCN-3 测试系统与被测系统的通信;

到稿日期:2012-10-29 返修日期:2013-01-13 本文受国家高技术研究发展计划(863 计划)项目(2009AA01Z145)资助。

柳永坡(1971—),男,博士,副教授,CCF 高级会员,主要研究方向为数据挖掘、软件测试、Web 应用性能分析与诊断,E-mail:Liuyongpo@bjut.edu.cn;吴 际(1974—),男,博士,副教授,主要研究方向为嵌入式软件建模与验证、软件测试与可靠性;刘霜梅(1978—),女,硕士,主要研究方向为计算机应用、软件工程、数据库理论与技术、计算机图形学。

(3)平台适配器(Platform Adapter,PA),使 TTCN-3 测试系统适应具体的执行平台,并且实现了外部函数以及操作定时器的相关方法;

(4)编解码器(Code/Decode,CD),转换被测系统和 SA 之间交互的数据,使其具有接收方能够处理的格式;

(5)测试管理部件(Test Management, TM),提供用户接口,管理 TTCN-3 测试系统;

(6)构件控制部件(Component Handling,CH),负责管理分布式测试构件;

(7)TTCN-3 控制接口(TTCN-3 Control Interface,TCI),为测试系统中的 TM、CH 和 CD 实体提供面向特定被测系统的标准适配机制^[4]。

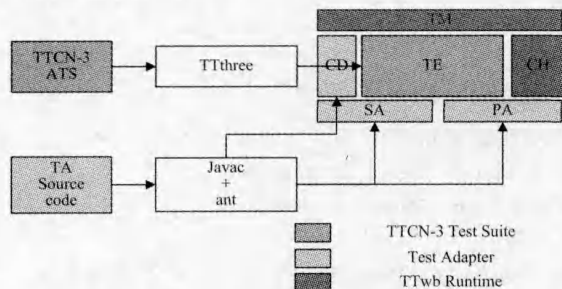


图1 TTCN-3 测试系统实体实现图

其中 TA 集成了 SA、PA 和 CD 3 个实体,完成测试系统与 被测系统以及测试系统与执行平台之间的适配;CD 实体 虽然被包含入 TA 实体,但实际上它只是在执行时被 TA 引 用,而在开发中则是作为单独实体来实现的,是整个开发工 作的核心;基于 TTCN-3 核心语言的抽象测试集(Abstract Test Suite, ATS)经由 TTworkbench 中的 TTCN-3 编译器 (TTthree)编译就成为了 TE 实体。

2 基于 TTCN-3 的 Web 应用测试

基于 B/S 架构的 Web 应用系统和传统的软件系统不同, 它是由相互协作而又彼此独立的软件子系统构成的,各个子 系统之间通过通信协议进行通信,而子系统内部则通过消息 进行通信^[5]。从体系结构上讲,Web 应用软件系统由多层软 件结构构成,一般包括客户端浏览器、服务器端应用服务子 系统和数据库服务子系统,如图 2 所示。

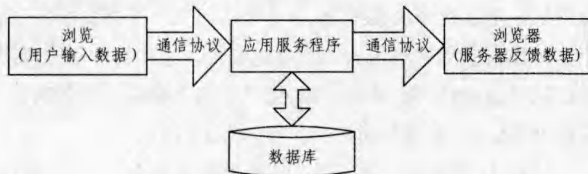


图2 Web 应用软件系统体系结构

由于 TTCN-3 语言所特有的动态配置、同步异步通信机 制等强大功能,使得测试人员在从测试设计到测试实现阶段 很难保证彼此之间的一致性;其次,面对庞大的测试系统的管 理和维护,也需要有工具辅助。这些问题必将随着 TTCN-3 测试语言的日益成熟而受到人们更多的关注。传统的逆向工 程通过分析程序源代码、文档和其他可用的资源,从中抽象出

设计信息、功能描述和需求。它的主要目的首先是重现考察 对象所体现的原有设计知识,其次是能发现一些在设计中没有 有明显表示出的、存在于设计人员头脑中的设计知识。随着 TTCN-3 在测试描述方面所展现出的巨大优点,它已经被应 用到了各种软件测试领域,同时有更多的遗留系统需要被管 理和维护。因此对基于 TTCN-3 的测试系统进行逆向分析, 可以帮助测试人员和维护人员从更高层次上把握测试系统的 设计,同时可以检验测试设计和测试实现之间的一致性,这些 工作对于测试系统的维护、扩展以及评估都有重大的意义。

针对上述 Web 应用系统,根据 TTCN-3 测试系统结构, 测试开发所要实现的 3 个部分(TE、TA 和 CD)应分别完成以 下功能:

- TE:定义测试数据和测试行为,其中测试数据包括向 被测系统输入的激励数据和期望自被测系统输出的响应数 据;
- TA:辅助实现测试系统与 被测系统的交互行为(适 配),完成向被测系统发送激励以及从被测系统接收响应的功 能;
- CD:转换数据编码格式,使得激励数据和响应数据可 以分别被被测系统和测试系统接受。

测试执行中三者的相互关系如图 3 所示。

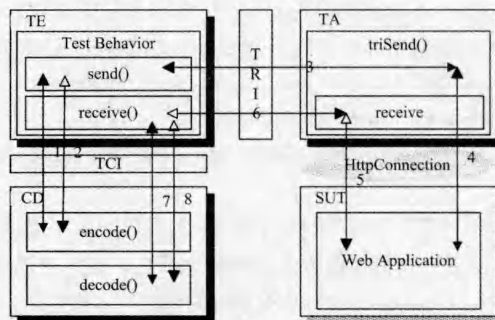


图3 测试系统执行流程图

1. 当测试系统执行到 TE 中的 send 方法时,首先调用 CD 的 encode 方法,对 send 方法的参数进行编码;
2. CD 对传入的数据(原始激励数据)进行编码,并将编码 后的数据回送给 TE;
3. 测试系统调用 TA 的 triSend 方法,向被测系统发送经 过编码的激励数据^[6];
4. TA 利用 HTTP 协议将激励数据发送给被测系统;
5. 被测系统对激励数据进行处理后,利用 HTTP 协议将 响应数据传回给 TA(TA 实体虽然实现了接收响应数据的功 能,但并不存在名为 receive 的方法);
6. TA 将原始响应数据回送给 TE;
7. 测试系统调用 CD 的 decode 方法将原始响应数据解码 为测试系统可以理解的格式;
8. CD 将解码后的响应数据回送给 TE,进行数据比对。

3 编解码器(CD)实现方案

研究结果表明,TTCN-3 适合于 Web 应用测试,但同时 也存在问题。由于 Web 应用系统的所有页面不尽相同,因此

需要为每个页面设计相应的编解码方案。因此一般地,在面向 Web 应用的 TTCN-3 测试系统开发中,编写 CD 代码就成为最复杂、最困难的工作。然而,尽管页面内容不同,但页面脚本的结构特点是一致的,因此页面脚本解析具有一般性,这个特点使得 CD 实体可以具有一定的领域通用性。以此为基础,提供面向 Web 应用领域的通用编解码器能够有效提高测试开发工作的效率,同时也可以增强 TTCN-3 在 Web 应用测试领域的适用性。

如前所述,对于 TTCN-3 测试系统,TA 接收到的 SUT 响应是完整的 Web 页面,从数据格式的角度讲即是 HTML 代码字符串,其中隐含着测试人员关注的信息。CD 实体针对 TE 中的数据定义解析 Web 页面,抽取数据并格式转换。不同于测试传统的基于客户/服务器(Client/Server, C/S)结构的 Web 应用,在测试 B/S 结构 Web 应用活动中,测试人员关注的对象不仅包括数据元素,即应用服务程序对输入数据的响应结果,还包括页面元素,即 Web 页面的展现形式,比如表单、文本框等。因此,本文将 TE 中定义的期望接收数据抽象成一种层次结构,从而可以更有效地描述目标数据, Pulei Xiong 等人在 Web 内容测试中也进行了相关研究^[7],如图 4 所示。由于 Frame、Form、Select、Input 和 Hyperlink 是 Web 页面的常用元素,因此将它们抽取出来归为 WebElement,而其他特殊的或页面特有的元素则归为 OtherElement,再结合描述具体数据的数据元素(DataElement),这样的结构化方法可以提高期望接收数据定义的通用性,具体定义见如下代码。

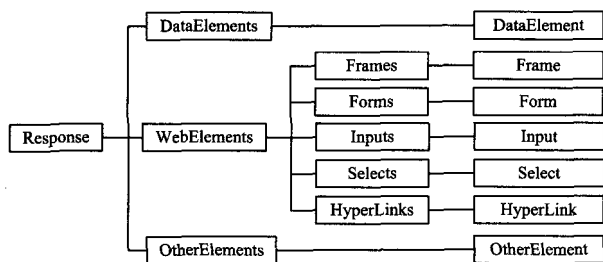


图 4 期望数据层次结构

```

module ResponseType {
  type set DataElement {
    charstring DataField,
    charstring DataValue
  }
  type set of DataElement DataElements;
  type set Frame {
    charstring FrameName,
    charstring FrameSrc
  }
  type set of Frame Frames;
  type set Form {
    charstring FormName,
    charstring FormAction,
    charstring FormMethod
  }
  type set of Form Forms;
  type set Input {

```

```

  charstring InputName,
  charstring InputType
}
type set of Input Inputs;
type set Select {
  charstring SelectName,
  charstring SelectId optional,
  charstring SelectClass optional
}
type set of Select Selects;
type set Hyperlink {
  charstring Href,
  charstring Target optional
}
type set of Hyperlink Hyperlinks;
type set WebElements {
  Frames frames,
  Forms forms,
  Inputs inputs,
  Selects selects,
  Hyperlinks hyperlinks
}
type set OtherElement {
  charstring OEField,
  charstring OEValue
}
type set of OtherElement OtherElements;
type set Response {
  DataElements dataElements,
  WebElements webElements,
  OtherElements otherElements optional
}
}

```

在上面的代码中,模块(module)是 TTCN-3 代码的顶层单元,测试数据、配置、行为都在其中定义。集合(set)是一种无序的结构化类型,类似 C 语言中的 struct 类型,而 set of 是对所包含元素都是 set 类型的一种集合化描述。代码中的 DataElement、Frame、Form、Input、Select、Hyperlink、OtherElement 分别定义了图 4 中末端节点的 7 种基本元素,它们包含各自属性的名/值对;DataElements、Frames、Forms、Inputs、Selects、Hyperlinks、OtherElements 是对应各基本元素的集合;WebElements 是 Web 元素的集合;Response 包含数据元素、Web 元素和其他元素 3 种集合,对应整个页面。代码中的 optional 保留字表示该项是可选项,在实际数据中可能不存在。

4 目标数据转换

本文研究的重点是通用性,而通用性意味着一定程度的抽象。CD 实体是一个完整的工作单元,有特定输入、处理对应输入的方法以及相应的输出。这其中,输入是由被测系统决定的,随着被测系统的变化而变化,因而对应输出也是不同

的。但从另一个角度讲,尽管 CD 实体的输入,即原始数据,其内容纷繁复杂,但作为 Web 页面,其内容仍旧具有一定的共同点,这决定了通用编解码器的通用性应该体现在数据的处理方法上。

被解析的 Web 页面结构复杂,信息量大,干扰信息多,若按照期望接收数据直接从原始响应数据中抽取,效率将非常低。针对上述特点,本文采取两层解析的策略将问题分解,通过连续两次解析得到目标数据,再经组装完成数据格式的转换,如图 5 所示。

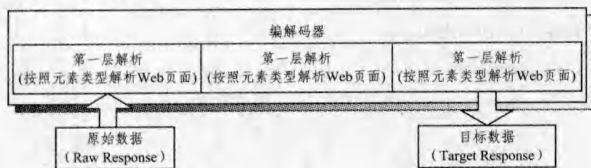


图 5 两层解析策略

根据不同被测系统来定制相应的解析规则作为指导,从而使该方法可以处理不同的原始数据。XML 语言是一种规范的标记语言,并且具有标准的结构化数据描述方法,所以本文采用 XML 语言描述解析规则,并将描述的规则以文件形式输入到每一层解析功能模块中。因此,CD 的数据流可以表示为图 6。应用两层解析方法时,针对不同被测系统,只需修改解析规则描述文件,就可以实现对不同目标的解析。事实上,两层解析和组装共同实现了 CD 实体的 decode 方法。第一层解析的作用是过滤不相关数据,即将原始数据按照前述的元素类型进行解析,形成对应数据元素、页面元素和其他元素的数据对象集,本文采用了开源的第三方软件或类库(如 HTML Parser、JTidy 等)来实现该层次的解析功能;第二层解析的作用是从各数据对象集中提取目标数据应包含的具体数据元素;组装是将第二层解析输出的元素组合成具有与期望接收数据相同格式的目标数据,这部分功能是通过实现 TCI 标准接口中数据转换和数据包装等相关方法来完成的。以上一节代码中定义的期望接收数据为例,第一层解析后应得到 7 个数据对象集,分别对应 Data 元素、Frame 元素、Form 元素、Input 元素、Select 元素、Hyperlink 元素以及其他元素;第二层解析抽取期望接收数据中的具体元素,为组装提供素材;组装后形成的即为目标数据。

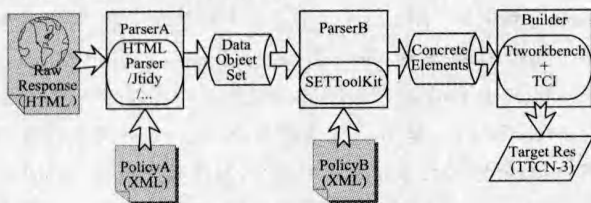


图 6 CD 实体工作流程图

当然,本文的设计也存在一个问题,即利用两层解析方法时,解析一个页面要求输入两个解析规则描述文件。如果被测系统是一个大型的 Web 站点,那么解析规则描述文件的数量将比较可观。因此,有必要将解析规则进行一定程度的抽象,形成具有模型性质的预定义解析规则,使其也具有一定的

通用性,从而减少解析规则描述文件的数量,进一步增强编解码器的通用性。

结束语 TTCN-3 作为 ETSI 的测试标准规范,以其平台无关性和通用性,正在越来越多的领域中得到应用,目前国内外都在大力研究和推广。从 TTCN-3 规范自身看,无论是其核心语言,还是其接口设计,都为其在各个领域的测试应用打下了良好的基础;另一方面,就目前 TTCN-3 支持工具而言,比如基于 Java 语言的 Ttworkbench、基于 C++ 语言的 Exhaustif/TTCN 等,都为 TTCN-3 的推广提供了大力的支持;而目前欠缺的是比较成熟的 TTCN-3 测试系统开发方法。本文以此为出发点,在 TTCN-3 领域适用性研究结论的基础上,选取 Web 应用软件为测试对象,提出了一种面向 Web 应用软件的测试编解码器(CD)开发方法。此方法针对 Web 应用软件数据量大、干扰信息多的特点,采用多层解析策略对原始数据进行处理,最终获得目标数据。

本研究旨在提出一个具有一定通用性的基于 TTCN-3 的 Web 应用软件测试系统实现方案,从而有助于增强 TTCN-3 规范在 Web 应用软件测试领域的适用性。目前,此方法的设计工作已经完成,下一步将进行具体功能的实现。此外,本方法目前关注的是数据的有无和正误,对于 Web 页面元素的位置分布还没有进行考虑,今后的研究将重点关注这个方面。

参考文献

- [1] ETSI. European Standard (ES) 201 873-6 V3. 1. 1 (2005-06); Methods for Testing and Specification(MTS); The Testing and Test Control Notation version 3; Part 6; TTCN-3 Control Interface (TCI). European Telecommunications Standards Institute (ETSI), Sophia-Antipolis, France, also published as ITU-T. Recommendation Z. 140, 2005
- [2] ETSI. European Standard (ES) 201 873-5 V3. 1. 1 (2005-06); Methods for Testing and Specification(MTS); The Testing and Test Control Notation version 3; Part 5; TTCN-3 Runtime Interface(TRI). European Telecommunications Standards Institute(ETSI), Sophia-Antipolis, France, also published as ITU-T. Recommendation Z. 140, 2005
- [3] Probert R L, Stepien B, Xiong P. Formal Testing of Web Content using TTCN-3[C]// TTCN-3 User Conference 2005. Sophia-Antipolis, France, 2005
- [4] 路晓丽. Web 应用软件的测试技术研究[D]. 西安:西北大学, 2006
- [5] Schieferdecker I, Pietsch S, Vassiliou-Gioles T. Systematic Testing of Internet Protocols- First Experiences in Using TTCN-3 for SIP[C]// 5th Africom Conference on Communication Systems. Cape Town, South Africa, May 2001
- [6] 高翔, 蒋凡. TTCN-3 研究综述[J]. 计算机工程与科学, 2004, 126(16)
- [7] 刘小勇, 蒋凡. 面向对象的 TTCN-3 测试系统的实现[J]. 计算机工程, 2006, 32(10)