

“十二五”普通高等教育本科国家级规划教材

北京高等教育精品教材

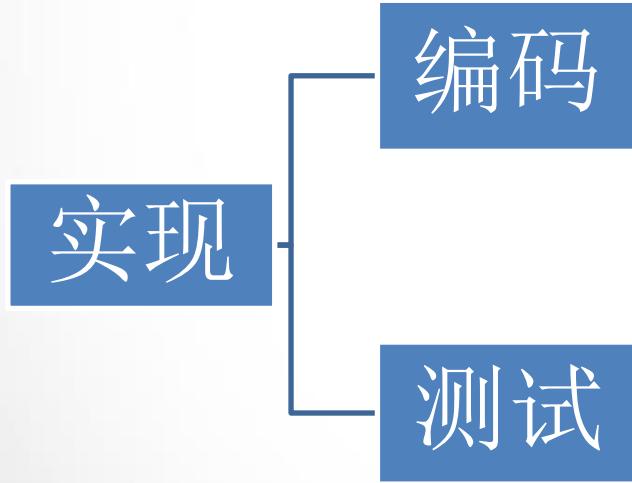
21世纪软件工程专业规划教材

软件工程导论（第6版）



第7章 实现

第7章 实现



编码就是把软件设计结果翻译成用某种程序设计语言书写的程序，是对设计的进一步具体化。

程序的质量主要取决于软件设计的质量。软件**测试**是保证软件质量的关键步骤，是对软件规格说明、设计和编码的最后复审。

主要内容

- 7. 1 编码
- 7. 2 软件测试基础
- 7. 3 单元测试
- 7. 4 集成测试
- 7. 5 确认测试
- 7. 6 白盒测试技术
- 7. 7 黑盒测试技术
- 7. 8 调试
- 7. 9 软件可靠性

主要内容



7.1 编码

7.2 软件测试基础

7.3 单元测试

7.4 集成测试

7.5 确认测试

7.6 白盒测试技术

7.7 黑盒测试技术

7.8 调试

7.9 软件可靠性

7.1 编码

7.1.1. 选择程序设计语言

程序设计语言是人和计算机通信的最基本的工具，会影响人的思维和解题方式，影响人和计算机通信的方式和质量，影响其他人阅读和理解程序的难易程度。

选择适宜的程序设计语言的原因：

- 根据设计去完成编码时，困难最少；
- 可以减少需要的程序测试量；
- 可以得到更容易阅读和更容易维护的程序。

7.1 编码

高级语言优于汇编语言：

- 汇编语言编码需要把软件设计翻译成机器操作的序列，既困难又容易出差错；
- 高级语言写程序比用汇编语言写程序生产率可以提高好几倍；
- 用高级语言写的程序容易阅读、容易测试、容易调试、容易维护。

7.1 编码

理想标准：

- 应该有理想的模块化机制，以及可读性好的控制结构和数据结构；
- 使编译程序能够尽可能多地发现程序中的错误；
- 应该有良好的独立编译机制。

实用标准：

- 系统用户的要求；
- 可以使用的编译程序；
- 可以得到的软件工具；
- 工程规模；
- 程序员的知识；
- 软件可移植性要求；
- 软件的应用领域。

7.1 编码

7.1.2 编码风格

源程序代码的逻辑简明清晰、易读易懂是好程序的一个重要标准，为了做到这一点，应该遵循下述规则。

1. 程序内部的文档

所谓程序内部的文档包括恰当的标识符、适当的注解和程序的视觉组织等。

- 标识符：含义鲜明的名字、缩写规则一致、为名字加注解；
- 注解：正确性，简要描述模块的功能、主要算法、接口特点、重要数据以及开发简史或解释包含这段代码的必要性；
- 视觉组织：适当的阶梯形式使程序的层次结构清晰明显。

7.1 编码

2. 数据说明

数据说明的原则：

- 数据说明的次序应该标准化；
- 当多个变量名在一个语句中说明时，应该按字母顺序排列这些变量；
- 如果设计时使用了一个复杂的数据结构，则应该用注解说明用程序设计语言实现这个数据结构的方法和特点。

7.1 编码

3. 语句构造

下述语句构造的原则有助于使语句简单明了：

- 不要为了节省空间而把多个语句写在同一行；
- 尽量避免复杂的条件测试；
- 尽量减少对“非”条件的测试；
- 避免大量使用循环嵌套和条件嵌套；
- 利用括号使逻辑表达式或算术表达式的运算次序清晰直观。

7.1 编码

4. 输入输出

在设计和编写程序时需考虑有关输入输出风格的规则：

- 对所有输入数据都进行检验；
- 检查输入项重要组合的合法性；
- 保持输入格式简单；
- 使用数据结束标记，不要要求用户指定数据的数目；
- 明确提示交互式输入的请求，详细说明可用的选择或边界数值；
- 程序设计语言对格式有严格要求时，应保持输入格式一致；
- 设计良好的输出报表；
- 给所有输出数据加标志。

7.1 编码

5. 效率

效率主要指处理机时间和存储器容量两个方面。

- 效率是性能要求，因此应该在需求分析阶段确定效率方面的要求；
- 效率是靠好设计来提高的；
- 程序的效率和程序的简单程度是一致的，不要牺牲程序的清晰性和可读性来不必要地提高效率。

7.1 编码

5. 效率

(1) 程序运行时间

写程序的风格会对程序的执行速度和存储器要求产生影响，应遵循的规则如下：

- 写程序之前先简化算术的和逻辑的表达式；
- 仔细研究嵌套的循环，以确定是否有语句可以从内层往外移；
- 尽量避免使用多维数组；
- 尽量避免使用指针和复杂的表；
- 使用执行时间短的算术运算；
- 不要混合作用不同的数据类型；
- 尽量使用整数运算和布尔表达式。

7.1 编码

5. 效率

(2) 存储器效率

- 在大型计算机中必须考虑操作系统页式调度的特点，一般说来，使用能保持功能域的结构化控制结构，是提高效率的好方法。
- 在微处理机中如果要求使用最少的存储单元，则应选用有紧缩存储器特性的编译程序，在非常必要时可以使用汇编语言。
- 提高执行效率的技术通常也能提高存储器效率。提高存储器效率的关键同样是“简单”。

7.1 编码

5. 效率

(3) 输入输出的效率

简单清晰是提高人机通信效率的关键。从写程序的角度看，却有些简单的原则可以提高输入输出的效率。

- 所有输入输出都应该有缓冲，以减少用于通信的额外开销；
- 对二级存储器(如磁盘)应选用最简单的访问方法；
- 二级存储器的输入输出应该以信息组为单位进行；
- 如果“超高效的”输入输出很难被人理解，则不应采用这种方法。

这些原则对于软件工程的设计和编码两个阶段都适用。

主要内容

7.1 编码

7.2 软件测试基础

7.3 单元测试

7.4 集成测试

7.5 确认测试

7.6 白盒测试技术

7.7 黑盒测试技术

7.8 调试

7.9 软件可靠性

7.2 软件测试基础

7.2.1 软件测试的目标

G. Myers给出的关于测试的一些规则如下：

- 测试是为了发现程序中的错误而执行程序的过程。
- 好的测试方案是极可能发现迄今为止尚未发现的错误的测试方案。
- 成功的测试是发现了至今为止尚未发现的错误的测试。

测试的正确定义是“为了发现程序中的错误而执行程序的过程”。应该认识到测试决不能证明程序是正确的。即使经过了最严格的测试之后，仍然可能还有没被发现的错误潜藏在程序中。另外，在综合测试阶段通常由其他人员组成测试小组来完成测试工作。

7.2 软件测试基础

7.2.2 软件测试准则

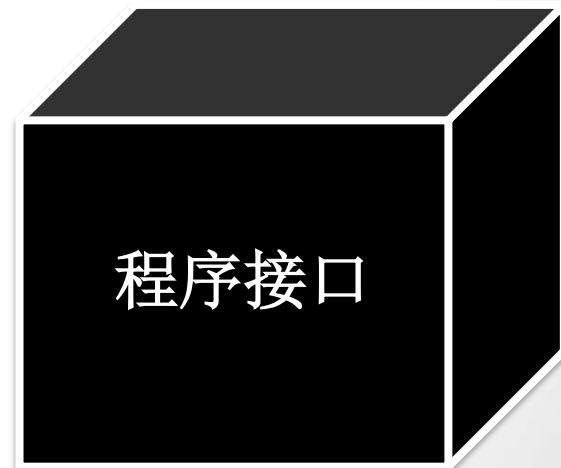
主要的软件测试准则如下：

- 所有测试都应该能追溯到用户需求；
- 应该远在测试开始之前就制定出测试计划；
- 把Pareto原理应用到软件测试中；
- 应该从“小规模”测试开始，并逐步进行“大规模”测试；
- 穷举测试是不可能的；
- 为了达到最佳的测试效果，应该由独立的第三方从事测试工作。

7.2 软件测试基础

7.2.3 测试方法

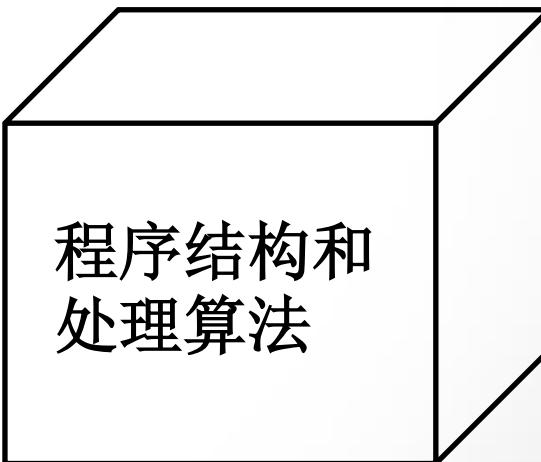
黑盒测试（又称功能测试）把程序看作一个黑盒子，完全不考虑程序的内部结构和处理过程。黑盒测试是在程序接口进行的测试，只检查程序功能是否能按照规格说明书的规定正常使用，程序是否能适当地接收输入数据并产生正确的输出信息，程序运行过程中能否保持外部信息（例如数据库或文件）的完整性。



7.2 软件测试基础

白盒测试（又称结构测试）

是把程序看成装在一个透明的白盒子里，测试者完全知道程序的结构和处理算法。这种方法按照程序内部的逻辑测试程序，检测程序中的主要执行通路是否都能按预定要求正确工作。



7.2 软件测试基础

7.2.4 测试步骤

根据第4条测试准则，测试过程也必须分步骤进行，后一个步骤在逻辑上是前一个步骤的继续。

大型软件系统通常由若干个子系统组成，每个子系统又由许多模块组成，因此，大型软件系统的测试过程基本上由**模块测试、子系统测试、系统测试、验收测试和平行运行**等五个步骤组成。

7.2 软件测试基础

1. 模块测试

在设计得好的软件系统中，每个模块完成一个清晰定义的子功能，而且这个子功能和同级其他模块的功能之间没有相互依赖关系。因此，有可能把每个模块作为一个单独的实体来测试，而且通常比较容易设计检验模块正确性的测试方案。

模块测试的目的是保证每个模块作为一个单元能正确运行，所以模块测试通常又称为**单元测试**。在这个测试步骤中所发现的往往是编码和详细设计的错误。

7.2 软件测试基础

2. 子系统测试

子系统测试是把经过单元测试的模块放在一起形成一个子系统来测试。模块相互间的协调和通信是这个测试过程中的主要问题，因此，这个步骤着重测试模块的接口。

3. 系统测试

系统测试是把经过测试的子系统装配成一个完整的系统来测试。在这个过程中不仅应该发现设计和编码的错误，还应该验证系统确实能提供需求说明书中指定的功能，而且系统的动态特性也符合预定要求。在这个测试步骤中发现的往往是软件设计中的错误，也可能发现需求说明中的错误。

子系统测试和系统测试，都兼有检测和组装两重含义，通常称为**集成测试**。

7.2 软件测试基础

4. 验收测试

验收测试把软件系统作为单一的实体进行测试，测试内容与系统测试基本类似，但是它是在**用户**积极参与下进行的，而且可能主要使用实际数据（系统将来要处理的信息）进行测试。

验收测试的目的是验证系统确实能够满足用户的需要，在这个测试步骤中发现的往往是系统需求说明书中的错误。验收测试也称为确认测试。

7.2 软件测试基础

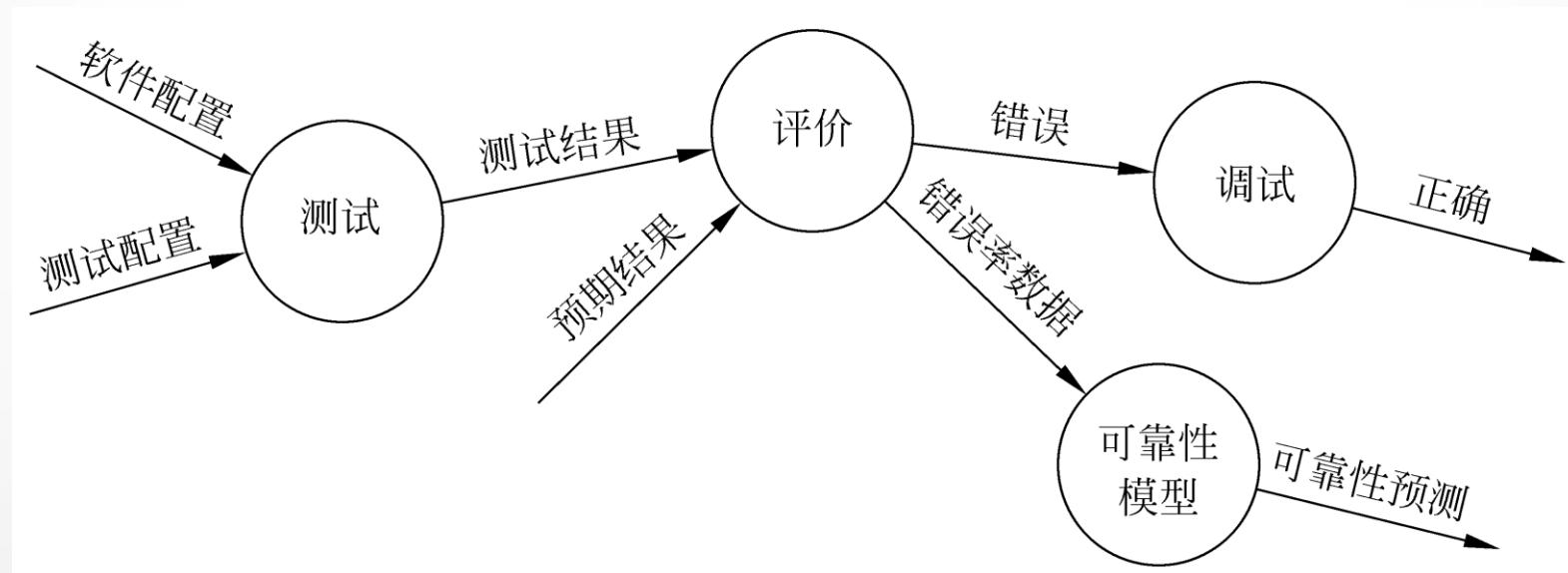
5. 平行运行

所谓**平行运行**就是同时运行新开发出来的系统和将被它取代的旧系统，以便比较新旧两个系统的处理结果。这样做的具体目的有如下几点。

- (1) 可以在准生产环境中运行新系统而又不冒风险。
- (2) 用户能有一段熟悉新系统的时间。
- (3) 可以验证用户指南和使用手册之类的文档。
- (4) 能够以准生产模式对新系统进行全负荷测试，可以用测试结果验证性能指标。

7.2 软件测试基础

7.2.5. 测试阶段的信息流



上图描绘了测试阶段的信息流，这个阶段的输入信息有两类：

- (1) 软件配置，包括需求说明书、设计说明书和源程序清单等；
- (2) 测试配置，包括测试计划和测试方案。

7.2 软件测试基础

测试方案不仅仅是测试时使用的输入数据(称为测试用例)，还应该包括每组输入数据预定要检验的功能，以及每组输入数据预期应该得到的正确输出。

测试配置是软件配置的一个子集，最终交出的软件配置应该包括上述测试配置以及测试的实际结果和调试的记录。

比较测试得出的实际结果和预期的结果，如果两者不一致则很可能是程序中有错误。

7.2 软件测试基础

如果经常出现要求修改设计的严重错误，那么软件的质量和可靠性是值得怀疑的，应该进一步仔细测试。

如果看起来软件功能完成得很正常，遇到的错误也很容易改正，则仍然应该考虑两种可能：(1)软件的可靠性是可以接受的；(2)所进行的测试尚不足以发现严重的错误。

如果经过测试，一个错误也没有被发现，则很可能是因为对测试配置思考不充分，以致不能暴露软件中潜藏的错误。

软件可靠性模型使用错误率数据估计将来出现错误的情况，并进而对软件可靠性进行预测。

主要内容

7.1 编码

7.2 软件测试基础

7.3 单元测试

7.4 集成测试

7.5 确认测试

7.6 白盒测试技术

7.7 黑盒测试技术

7.8 调试

7.9 软件可靠性

7.3 单元测试

- 单元测试集中检测软件设计的最小单元——模块。
- 单元测试和编码属于软件过程的同一个阶段。
- 在源程序代码通过编译程序的语法检查后，可以用详细设计描述作指南，对重要的执行通路进行测试，以便发现模块内部的错误。
- 可以应用人工测试和计算机测试这样两种不同类型的测试方法，完成单元测试工作。
- 单元测试主要使用白盒测试技术，而且对多个模块的测试可以并行地进行。

7.3 单元测试

7.3.1 测试重点

在单元测试期间着重从以下5个方面对模块进行测试。

1. 模块接口

对模块接口进行测试时主要检查以下几个方面：

- 参数的数目、次序、属性或单位系统与变元是否一致；
- 是否修改了只作输入用的变元；
- 全局变量的定义和用法在各个模块中是否一致。

7.3 单元测试

2. 局部数据结构

对于模块来说，局部数据结构是常见的错误来源。应该仔细设计测试方案，以便发现局部数据说明、初始化、默认值等方面的问题。

3. 重要的执行通路

由于通常不可能进行穷尽测试，因此，在单元测试期间选择最有代表性、最可能发现错误的执行通路进行测试是十分关键的。应该设计测试方案用来发现由于错误的计算、不正确的比较或不适当的控制流而造成的错误。

7.3 单元测试

4. 出错处理通路

好的设计应该能预见出现错误的条件，并且设置适当的处理错误的通路。不仅应该在程序中包含出错处理通路，而且应该认真测试这种通路。评价出错处理通路应该着重测试下述一些可能发生的错误。

- (1) 对错误的描述是难以理解的；
- (2) 记下的错误与实际遇到的错误不同；
- (3) 在对错误进行处理之前，错误条件已经引起系统干预；
- (4) 对错误的处理不正确；
- (5) 描述错误的信息不足以帮助确定造成错误的位置。

7.3 单元测试

5. 边界条件

- 边界测试是单元测试中最后的也可能是最重要的任务。
- 软件常常在它的边界上失效，例如，处理n元数组的第n个元素时，或做到i次循环中的第i次重复时，往往会发生错误。
- 使用刚好小于、刚好等于和刚好大于最大值或最小值的数据结构、控制量和数据值的测试方案，非常可能发现软件中的错误。

7.3 单元测试

7.3.2 代码审查

代码检查是指由审查小组正式对源程序进行人工测试。它是一种非常有效的程序验证技术，对于典型的程序来说，可以查出30%~70%的逻辑设计错误和编码错误。审查小组最好由下述4人组成。

- (1) 组长，应该是一个很有能力的程序员，而且没有直接参与这项工程；
- (2) 程序的设计者；
- (3) 程序的编写者；
- (4) 程序的测试者。

7.3 单元测试

在审查会上由程序的编写者解释他是怎样用程序代码实现设计的，通常是逐个语句地讲述程序的逻辑，小组其他成员仔细倾听他的讲解，并力图发现其中的错误。

审查会上需要对照程序设计常见错误清单，分析审查这个程序。当发现错误时由组长记录下来，审查会继续进行（**审查小组的任务是发现错误而不是改正错误**）。

审查会另外一种常见的进行方法，称为**预排**：由一个人扮演“测试者”，其他人扮演“计算机”。会前测试者准备好测试方案，会上由扮演计算机的成员模拟计算机执行被测试的程序。

7.3 单元测试

测试方案在代码审查中起一种促进思考引起讨论的作用。在大多数情况下，通过向程序员提出关于他的程序的逻辑和他编写程序时所做的假设的疑问，可以发现的错误比由测试方案直接发现的错误还多。

代码审查比计算机测试优越的是：一次审查会上可以发现许多错误；用计算机测试的方法发现错误之后，通常需要先改正这个错误才能继续测试，即：采用代码审查的方法可以减少系统验证的总工作量。

人工测试和计算机测试是互相补充，相辅相成的，缺少其中任何一种方法都会使查找错误的效率降低。

7.3 单元测试

7.3.3. 计算机测试

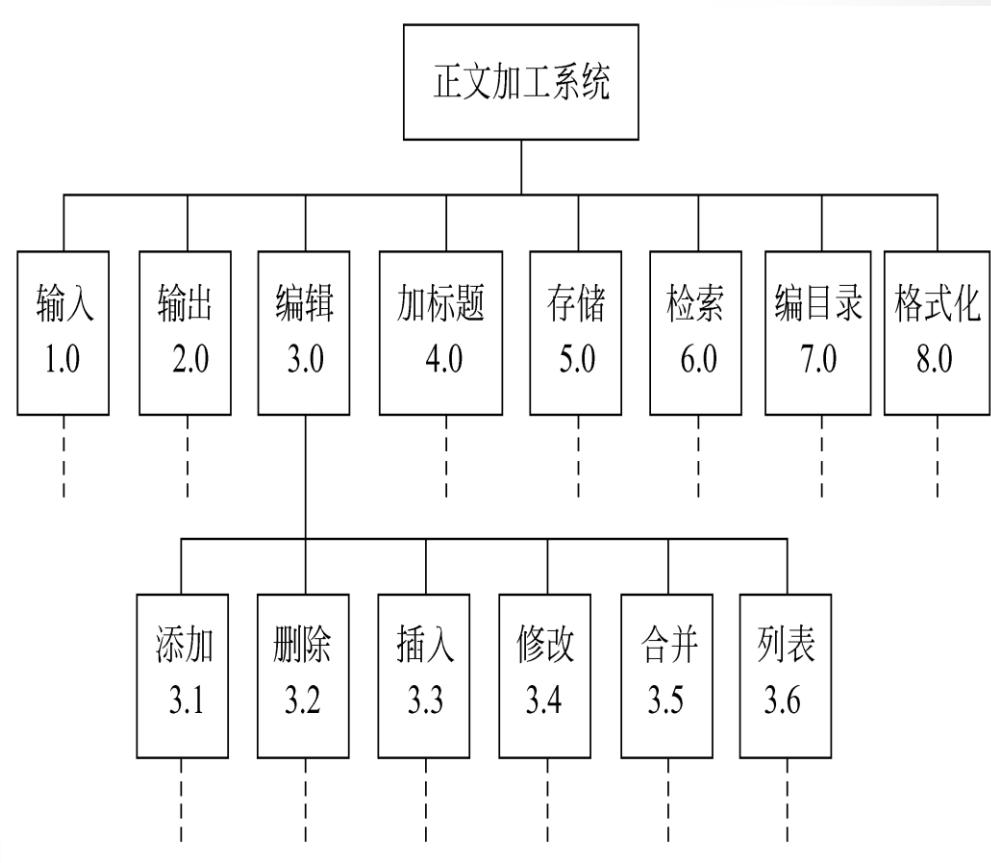
模块不是一个独立的程序，因此必须为每个单元测试开发驱动软件和(或)存根软件。

驱动程序是一个“主程序”，它接收测试数据，把这些数据传送给被测试的模块，并且印出有关的结果。

存根程序代替被测试的模块所调用的模块，它使用被它代替的模块的接口，可能做最少量的数据操作，印出对入口的检验或操作结果，并且把控制归还给调用它的模块。

7.3 单元测试

右图是一个正文加工系统的部分层次图，假定要测试编号为3.0的关键模块——正文编辑模块。正文编辑模块不是一个独立的程序，需要有一个测试驱动程序来调用它。这个驱动程序说明必要的变量，接收测试数据——字符串，设置正文编辑模块的编辑功能。并且需要有存根程序简化地模拟正文编辑模块的下层模块来完成具体的编辑功能。



7.3 单元测试

测试时，设置修改(CHANGE)和添加(APPEND)两种编辑功能，用控制变量CFUNCT标记要求的编辑功能，而且只用一个存根程序模拟正文编辑模块的所有下层模块。

TEST STUB(*存根程序*)

初始化；

输出信息“进入了正文编辑程序”；

输出“输入的控制信息是”CFUNCT；

输出缓冲区中的字符串；

IF CFUNCT=CHANGE

THEN

把缓冲区中第二个字改为***

ELSE

在缓冲区的尾部加???

END IF；

输出缓冲区中的新字符串；

END TEST STUB

TEST DRIVER(*驱动程序*)

说明长度为2500个字符的一个缓冲区；

把CFUNCT置为希望测试的状态；

输入字符串；

调用正文编辑块；

停止或再次初启；

END TEST DRIVER

主要内容

7.1 编码

7.2 软件测试基础

7.3 单元测试

7.4 集成测试

7.5 确认测试

7.6 白盒测试技术

7.7 黑盒测试技术

7.8 调试

7.9 软件可靠性

7.4 集成测试

集成测试是测试和组装软件的系统化技术。

由模块組裝成程序时有两种方法。一种方法是先分别測試每个模块，再把所有模块按设计要求放在一起结合成所要的程序，这种方法称为**非漸增式測試方法**；另一种方法是把下一个要測試的模块同已经測試好的那些模块结合起来进行測試，測試完以后再把下一个应该測試的模块结合起来測試。这种每次增加一个模块的方法称为**漸增式測試**，这种方法实际上同时完成单元測試和集成測試。

7.4 集成测试

非渐增式测试把所有模块放在一起，作为一个整体来测试。测试时会遇到许多的错误，改正错误非常困难，因为在庞大的程序中想要诊断定位一个错误非常困难，而且改正一个错误之后，马上又会遇到新的错误，这个过程会继续下去，没有尽头。

渐增式测试与“一步到位”的非渐增式测试相反，它把程序划分成小段来构造和测试，在这个过程中比较容易定位和改正错误；对接口可以进行更彻底的测试；可以使用系统化的测试方法。因此，目前在进行集成测试时普遍采用渐增式测试方法。

当使用渐增方式把模块结合到程序中去时，有**自顶向下**和**自底向上**两种集成策略。

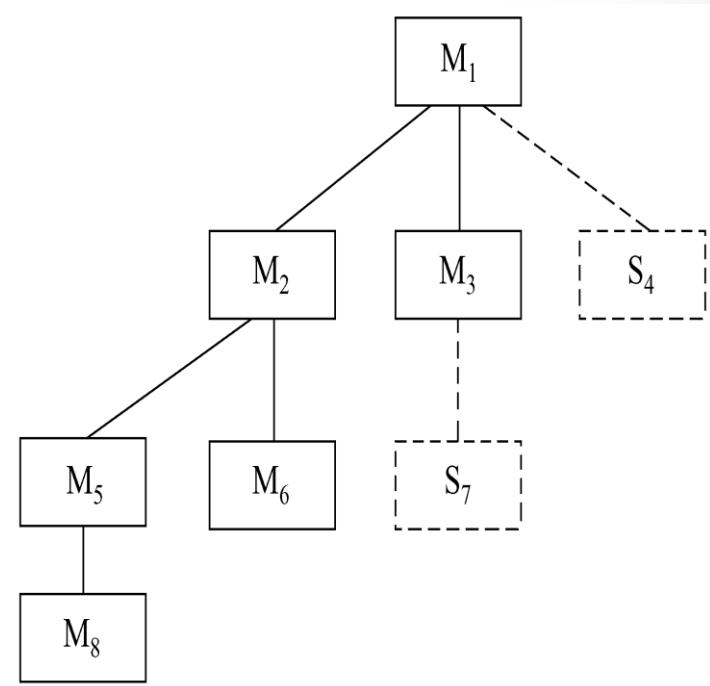
7.4 集成测试

7.4.1. 自顶向下集成

- **自顶向下集成方法**是从主控制模块开始，沿着程序的控制层次向下移动，逐渐把各个模块结合起来。在把附属于（及最终附属于）主控制模块的那些模块组装到程序结构中去时，或者使用深度优先的策略，或者使用宽度优先的策略。
- 深度优先的结合方法先组装在软件结构的一条主控制通路上的所有模块。选择一条主控制通路取决于应用的特点，并且有很大任意性。
- 宽度优先的结合方法是沿软件结构水平地移动，把处于同一个控制层次上的所有模块组装起来。

7.4 集成测试

如右图，使用深度优先的结合方法，选取左通路，首先结合模块M1\, M2和M5；其次，M8或M6(如果为了使M2具有适当功能需要M6)将被结合起来。然后构造中央的和右侧的控制通路。使用宽度优先的结合方法，首先结合模块M2\, M3和M4(代替存根程序S4)，然后结合下一个控制层次中的模块M5\, M6和M7；如此继续进行下去，直到所有模块都被结合起来为止。



7.4 集成测试

模块结合进软件结构的具体过程由下述4个步骤完成：

- ① 对主控制模块进行测试，测试时用存根程序代替所有直接附属于主控制模块的模块；
- ② 根据选定的结合策略（深度优先或宽度优先），每次用一个实际模块代换一个存根程序（新结合进来的模块往往又需要新的存根程序）；
- ③ 在结合进一个模块的同时进行测试；
- ④ 为了保证加入模块没有引进新的错误，可能需要进行回归测试（即全部或部分地重复以前做过的测试）。

从②开始不断地重复进行上述过程，直到构造起完整的软件结构为止。

7.4 集成测试

- **自顶向下的结合策略**能够在测试的早期对主要的控制或关键的抉择进行检验。在一个分解得好的软件结构中，关键的抉择位于层次系统的较上层，因此首先碰到。
- 如果选择**深度优先**的结合方法，可以在早期实现软件的一个完整的功能并且验证这个功能。
- 在自顶向下测试的初期，存根程序代替了低层次的模块，因此，在软件结构中没有重要的数据自下往上流。为了解决这个问题，测试人员有两种选择：①把许多测试推迟到用真实模块代替了存根程序以后再进行；②从层次系统的底部向上组装软件。

7.4 集成测试

7.4.2. 自底向上集成

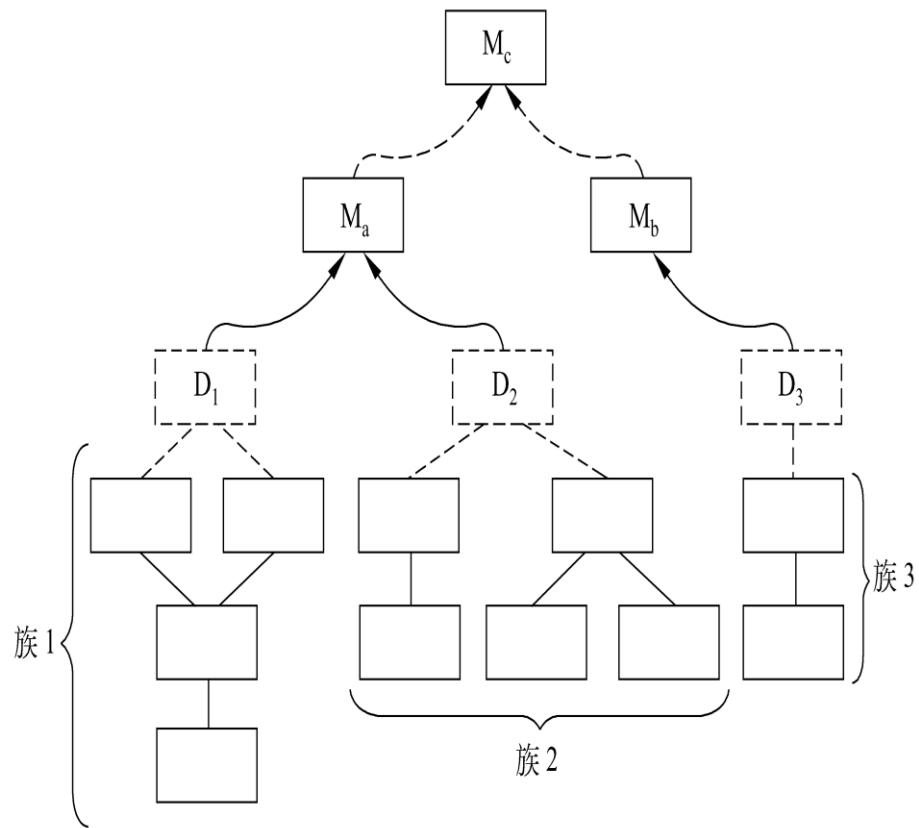
自底向上测试从“原子”模块(即在软件结构最低层的模块)开始组装和测试。因为是从底部向上结合模块，总能得到所需的下层模块处理功能，所以不需要存根程序。

用下述步骤可以实现自底向上的结合策略。

- ① 把低层模块组合成实现某个特定的软件子功能的族；
 - ② 写一个驱动程序(用于测试的控制程序)，协调测试数据的输入和输出；
 - ③ 对由模块组成的子功能族进行测试；
 - ④ 去掉驱动程序，沿软件结构自下向上移动，把子功能族组合起来形成更大的子功能族。
- 上述第②~④步实质上构成了一个循环。

7.4 集成测试

右图描绘了自底向上的结合过程。首先把模块组合成族1、族2和族3，使用驱动程序(图中用虚线方框表示)对每个子功能族进行测试。族1和族2中的模块附属于模块 M_a ，去掉驱动程序 D_1 和 D_2 ，把这两个族直接同 M_a 连接起来。类似地，在和模块 M_b 结合之前去掉族3的驱动程序 D_3 。最终 M_a 和 M_b 这两个模块都与模块 M_c 结合起来。随着结合向上移动，对测试驱动程序的需要减少了。



7.4 集成测试

7.4.3 不同集成测试策略的比较

- **自顶向下测试方法的主要优点**是不需要测试驱动程序，能够在测试阶段的早期实现并验证系统的主要功能，而且能在早期发现上层模块的接口错误。
- **自顶向下测试方法的主要缺点**是需要存根程序，可能遇到与此相联系的测试困难，低层关键模块中的错误发现较晚，而且用这种方法在早期不能充分展开人力。
- **自底向上测试方法的优缺点与上述自顶向下测试方法的优缺点刚好相反。**

7.4 集成测试

一般说来，纯粹自顶向下或纯粹自底向上的策略可能都不实用，人们在实践中创造出许多混合策略。

(1) **改进的自顶向下测试方法**。基本上使用自顶向下的测试方法，但是在早期使用自底向上的方法测试软件中的少数关键模块。一般的自顶向下方法所具有的优点在这种方法中也都有，而且能在测试的早期发现关键模块中的错误；但是，它的缺点也比自顶向下方法多一条，即测试关键模块时需要驱动程序。

(2) **混合法**。对软件结构中较上层使用的自顶向下方法与对软件结构中较下层使用的自底向上方法相结合。这种方法兼有两种方法的优点和缺点，当被测试的软件中关键模块比较多时，这种混合法可能是最好的折衷方法。

7. 4 集成测试

7. 4. 4. 回归测试

- 在集成测试过程中，每当一个新模块结合起来时，程序就发生了变化：建立了新的数据流路径，可能出现了新的I/O操作，激活了新的控制逻辑。在集成测试的范畴中，**回归测试**是指重新执行已经做过的测试的某个子集，以保证上述这些变化没有带来非预期的副作用。
- **回归测试**就是用于保证由于调试或其他原因引起的变化，不会导致非预期的软件行为或额外错误的测试活动。
- **回归测试**可以通过人工地进行，也可以使用自动化的捕获回放工具自动进行。利用捕获回放工具，软件工程师能够捕获测试用例和实际运行结果，然后可以回放（即重新执行测试用例），并且比较软件变化前后所得到的运行结果。

7.4 集成测试

回归测试集（已执行过的测试用例的子集）包括下述3类不同的测试用例。

- (1) 检测软件全部功能的代表性测试用例。
- (2) 专门针对可能受修改影响的软件功能的附加测试。
- (3) 针对被修改过的软件成分的测试。

在集成测试过程中，回归测试用例的数量可能变得非常大。因此，应该把回归测试集设计成只包括可以检测程序每个主要功能中的一类或多类错误的那样一些测试用例。

主要内容

7.1 编码

7.2 软件测试基础

7.3 单元测试

7.4 集成测试

7.5 确认测试

7.6 白盒测试技术

7.7 黑盒测试技术

7.8 调试

7.9 软件可靠性

7.5 确认测试

- **确认测试**也称为验收测试，它的目标是**验证**软件的有效性。
- 通常，**验证**指的是保证软件正确地实现了某个特定要求的一系列活动；**确认**指的是为了保证软件确实满足了用户需求而进行的一系列活动。
- **软件有效性**的一个简单定义是：如果软件的功能和性能如同用户所合理期待的那样，软件就是有效的。
- 需求分析阶段产生的软件需求规格说明书，准确地描述了用户对软件的合理期望，因此是软件有效性的标准，也是进行确认测试的基础。

7.5 确认测试

7.5.1. 确认测试的范围

确认测试必须有用户积极参与，或以用户为主进行。用户应该参与设计测试方案，使用用户界面输入测试数据并且分析评价测试的输出结果。

确认测试通常使用黑盒测试法。应该仔细设计测试计划和测试过程，测试计划包括要进行的测试的种类及进度安排，测试过程规定了用来检测软件是否与需求一致的测试方案。

通过测试和调试要保证软件能满足所有功能要求，能达到每个性能要求，文档资料是准确而完整的，此外，还应该保证软件能满足其他预定的要求（例如安全性、可移植性、兼容性和可维护性等）。

确认测试有下述两种可能的结果：

- (1) 功能和性能与用户要求一致，软件是可以接受的。
- (2) 功能和性能与用户要求有差距。

7.5 确认测试

7.5.2 软件配置复查

软件配置复查是确认测试的一个重要内容。复查的目的是保证软件配置的所有成分都齐全，质量符合要求，文档与程序完全一致，具有完成软件维护所必须的细节，而且已经编好目录。

除了按合同规定的内容和要求，由人工审查软件配置之外，在确认测试过程中还应该严格遵循用户指南及其他操作程序，以便检验这些使用手册的完整性和正确性。必须仔细记录发现的遗漏或错误，并且适当地补充和改正。

7.5 确认测试

7.5.3 Alpha和Beta测试

- 如果一个软件是为许多客户开发的（例如，向大众公开出售的盒装软件产品），那么绝大多数软件开发商都使用被称为**Alpha测试**和**Beta测试**的过程，来发现那些看起来只有最终用户才能发现的错误。
- **Alpha测试**由用户在开发者的场所进行，并且在开发者对用户的“指导”下进行测试。开发者负责记录发现的错误和使用中遇到的问题。
- **Alpha测试**是在受控的环境中进行的。
- **Beta测试**由软件的最终用户们在一个或多个客户场所进行。与Alpha测试不同，开发者通常不在Beta测试的现场。
- **Beta测试**是软件在开发者不能控制的环境中的“真实”应用。

主要内容

- 7. 1 编码
- 7. 2 软件测试基础
- 7. 3 单元测试
- 7. 4 集成测试
- 7. 5 确认测试
-  7. 6 白盒测试技术
- 7. 7 黑盒测试技术
- 7. 8 调试
- 7. 9 软件可靠性

7.6 白盒测试技术

- 通常把测试数据和预期的输出结果称为测试用例。
- 不同的测试数据发现程序错误的能力差别很大，为了提高测试效率降低测试成本，应该选用高效的测试数据。因为不可能进行穷尽的测试，所以选用少量“最有效的”测试数据，做到尽可能完备的测试就更重要了。
- 设计测试方案的基本目标是，确定一组最可能发现某个错误或某类错误的测试数据。已经研究出许多设计测试数据的技术，这些技术各有优缺点；同一种技术在不同的应用场合效果可能相差很大，因此，通常需要联合使用多种设计测试数据的技术。

7.6 白盒测试技术

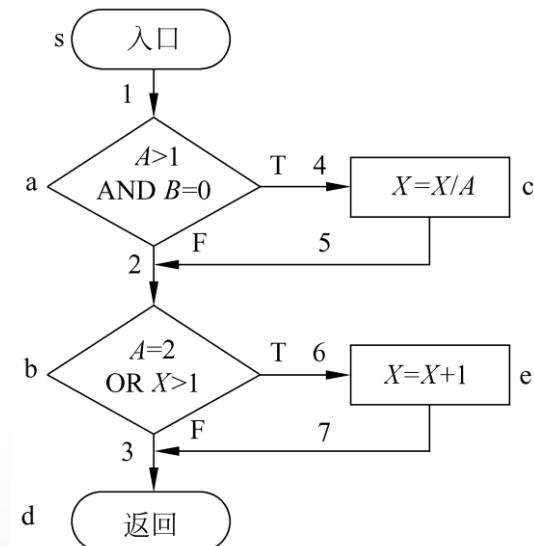
7.6.1. 逻辑覆盖

逻辑覆盖是对一系列测试过程的总称，这组测试过程逐渐进行越来越完整的通路测试。

1. 语句覆盖

语句覆盖的含义是，选择足够多的测试数据，使被测程序中每个语句至少执行一次。

右图为被测试模块的流程图为了使每个语句都执行一次，程序的执行路径应该是sacbed，为此只需要输入下面的测试数据（实际上X可以是任意实数）：A=2，B=0，X=4。



7.6 白盒测试技术

1. 语句覆盖

语句覆盖对程序的逻辑覆盖很少，在上面例子中两个判定条件都只测试了条件为真的情况，如果条件为假时处理有错误，显然不能发现。

语句覆盖只关心判定表达式的值，而没有分别测试判定表达式中每个条件取不同值时的情况。为了执行sacbed路径，以测试每个语句，只需两个判定表达式 $(A>1) \text{ AND } (B=0)$ 和 $(A=2) \text{ OR } (X>1)$ 都取真值，因此使用上述一组测试数据就够了。但是，如果程序中把第一个判定表达式中的逻辑运算符AND错写成OR，或把第二个判定表达式中的条件 $X>1$ 误写成 $X<1$ ，使用上面的测试数据并不能查出这些错误。

综上所述，可以看出语句覆盖是很弱的逻辑覆盖标准。

7.6 白盒测试技术

2. 判定覆盖

判定覆盖又叫分支覆盖，它的含义是，不仅每个语句必须至少执行一次，而且每个判定的每种可能的结果都应该至少执行一次，也就是每个判定的每个分支都至少执行一次。

对于上述例子来说，能够分别覆盖路径sacbed和sabd的两组测试数据，或者可以分别覆盖路径sacbd和sabed的两组测试数据，都满足判定覆盖标准。例如，以下两组测试数据就可做到判定覆盖：

- ① A=3, B=0, X=3 (覆盖sacbd)
- ② A=2, B=1, X=1 (覆盖sabed)

判定覆盖比语句覆盖强，但是对程序逻辑的覆盖程度仍然不高，例如，上面的测试数据只覆盖了程序全部路径的一半。

7.6 白盒测试技术

3. 条件覆盖

条件覆盖的含义是，不仅每个语句至少执行一次，而且使判定表达式中的每个条件都取到各种可能的结果。

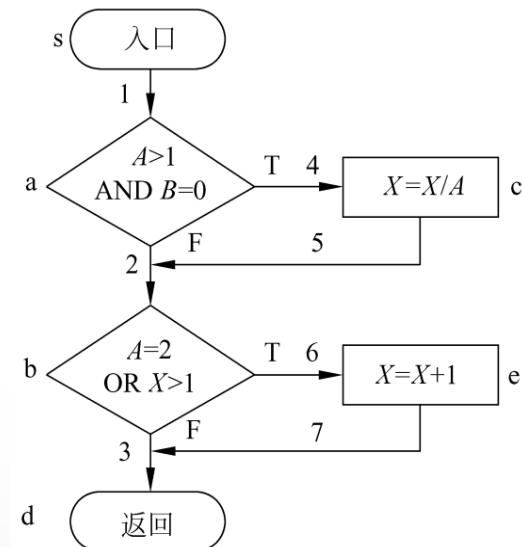
上例中共有两个判定表达式，每个表达式中有两个条件，为了做到条件覆盖，应该选取测试数据满足下面的要求。

在a点有下述各种结果出现：

$A > 1, A \leq 1, B = 0, B \neq 0;$

在b点有下述各种结果出现：

$A = 2, A \neq 2, X > 1, X \leq 1;$



7.6 白盒测试技术

3. 条件覆盖

只需要使用下面两组测试数据就可以达到上述覆盖标准：

- ① $A=2, B=0, X=4$ (满足 $A>1, B=0, A=2$ 和 $X>1$ ，执行路径sacbed)
- ② $A=1, B=1, X=1$ (满足 $A\leqslant 1, B\neq 0, A\neq 2$ 和 $X\leqslant 1$ ，执行路径sabd)

条件覆盖通常比判定覆盖强，但满足条件覆盖的测试数据不一定满足判定覆盖。例如，上面两组测试数据也同时满足判定覆盖标准。但是，如果使用下面两组测试数据，则只满足条件覆盖标准并不满足判定覆盖标准(第二个判定表达式的值总为真)：

- ① $A=2, B=0, X=1$ (满足 $A>1, B=0, A=2$ 和 $X\leqslant 1$ ，执行路径sacbed)
- ② $A=1, B=1, X=2$ (满足 $A\leqslant 1, B\neq 0, A\neq 2$ 和 $X>1$ ，执行路径sabed)

7.6 白盒测试技术

4. 判定/条件覆盖

判定/条件覆盖是一种能同时满足判定覆盖和条件覆盖的逻辑覆盖，它的含义是，选取足够的测试数据，使得判定表达式中的每个条件都取到各种可能的值，而且每个判定表达式也都取到各种可能的结果。

对于上例而言，下述两组测试数据满足判定/条件覆盖标准：

- ① A=2, B=0, X=4
- ② A=1, B=1, X=1

但是，这两组测试数据也就是为了满足条件覆盖标准最初选取的两组数据，因此，有时判定/条件覆盖也并不比条件覆盖更强。

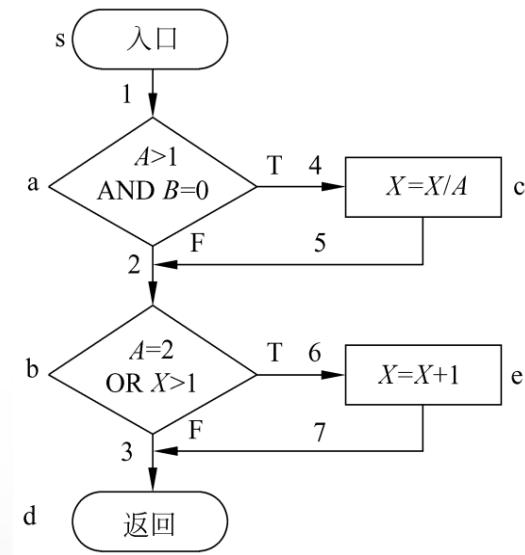
7.6 白盒测试技术

5. 条件组合覆盖

条件组合覆盖是更强的逻辑覆盖标准，它要求选取足够的测试数据，使得每个判定表达式中条件的各种可能组合都至少出现一次。

对于上例，共有8种可能的条件组合，它们分别是：

- | | |
|-----------------------|--------------------------|
| (1) $A > 1, B = 0$ | (2) $A > 1, B \neq 0$ |
| (3) $A \leq 1, B = 0$ | (4) $A \leq 1, B \neq 0$ |
| (5) $A = 2, X > 1$ | (6) $A = 2, X \leq 1$ |
| (7) $A \neq 2, X > 1$ | (8) $A \neq 2, X \leq 1$ |



7.6 白盒测试技术

5. 条件组合覆盖

下面的4组测试数据使上面列出的8种条件组合每种至少出现一次：

- ① A=2, B=0, X=4 (针对 (1) 和 (5) , 执行路径sacbed)
- ② A=2, B=1, X=1 (针对 (2) 和 (6) , 执行路径sabed)
- ③ A=1, B=0, X=2 (针对 (3) 和 (7) , 执行路径sabed)
- ④ A=1, B=1, X=1 (针对 (4) 和 (8) , 执行路径sabd)

显然，满足条件组合覆盖标准的测试数据，也一定满足判定覆盖、条件覆盖和判定/条件覆盖标准。因此，条件组合覆盖是前述几种覆盖标准中最强的。但是，满足条件组合覆盖标准的测试数据并不一定能使程序中的每条路径都执行到，例如，上述4组测试数据都没有测试到路径sacbd。

7.6 白盒测试技术

6. 点覆盖

从对程序路径的覆盖程度分析，能够提出下述一些主要的逻辑覆盖标准。

图论中**点覆盖**的定义如下：如果连通图G的子图G'是连通的，而且包含G的所有结点，则称G'是G的点覆盖。

在第6.5节中已经讲述了从程序流程图导出流图的方法。在正常情况下流图是连通的有向图。满足点覆盖标准要求选取足够多的测试数据，使得程序执行路径至少经过流图的每个结点一次，由于流图的每个结点与一条或多条语句相对应，显然，点覆盖标准和语句覆盖标准是相同的。

7.6 白盒测试技术

7. 边覆盖和路径覆盖

图论中**边覆盖**的定义是：如果连通图G的子图G”是连通的，而且包含G的所有边，则称G”是G的边覆盖。为了满足边覆盖的测试标准，要求选取足够多测试数据，使得程序执行路径至少经过流图中每条边一次。通常边覆盖和判定覆盖是一致的。

路径覆盖的含义是，选取足够多测试数据，使程序的每条可能路径都至少执行一次（如果程序图中有环，则要求每个环至少经过一次）。

7.6 白盒测试技术

7.6.2 控制结构测试

基本路径测试是Tom McCabe提出的一种白盒测试技术。使用基本路径测试设计测试用例时，首先计算程序的环形复杂度，并用该复杂度为指南定义执行路径的基本集合，从该基本集合导出的测试用例可以保证程序中的每条语句至少执行一次，而且每个条件在执行时都将分别取真、假两种值。

使用基本路径测试技术设计测试用例的步骤如下。

① 根据过程设计结果画出相应的流图。

例如，为了用基本路径测试技术测试下列的用PDL描述的求平均值过程，首先画出下图所示的流图。注意，为了正确地画出流图，这里把被映射为流图结点的PDL语句编了序号。

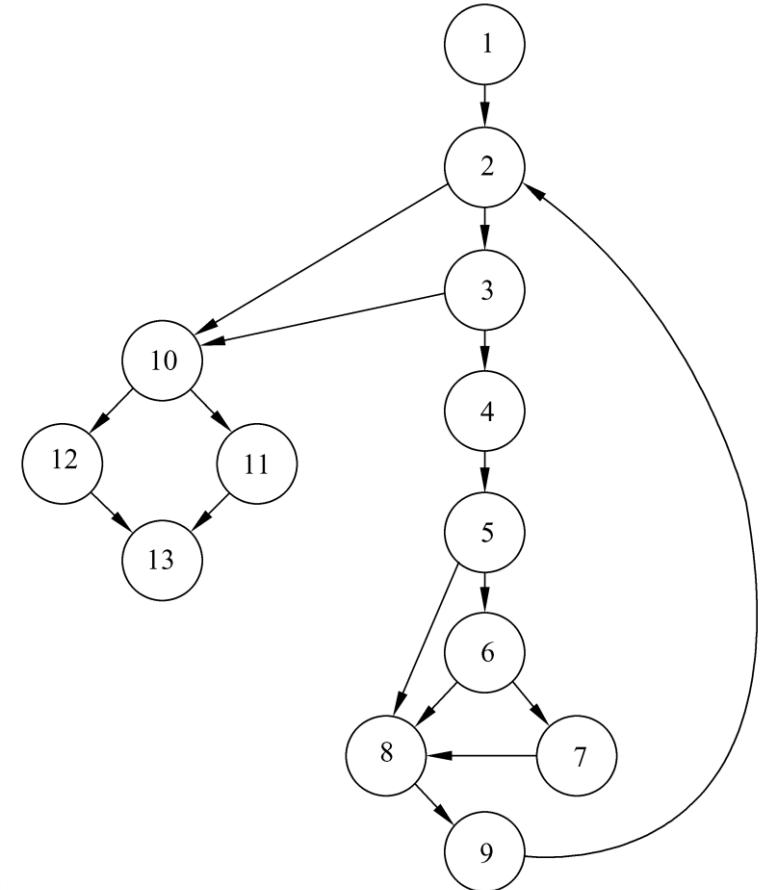
7.6 白盒测试技术

1. 基本路径测试

```

1:   i=1;
      total.input=total.valid=0;
      sum=0;
2:   DO WHILE value[i] <> -999
3:       AND total.input<100
4:       increment total.input by1;
5:       IF value[i]>=minimum
6:           AND value[i]<=maximum
7:           THEN increment total.valid by 1;
              sum=sum+value[i];
8:       ENDIF
              increment i by 1;
9:   ENDDO
10:  IF total.valid>0
11:      THEN average=sum/total.valid;
12:  ELSE average=-999;
13:  ENDIF
END average

```



7.6 白盒测试技术

1. 基本路径测试

② 计算流图的环形复杂度。

环形复杂度定量度量程序的逻辑复杂性。使用第6.5.1小节讲述的3种方法之一计算环形复杂度。经计算，流图的环形复杂度为6。

③ 确定线性独立路径的基本集合。

独立路径是指至少引入程序的一个新处理语句集合或一个新条件的路径，即独立路径至少包含一条在定义该路径之前不曾用过的边。

程序的环形复杂度决定了程序中独立路径的数量，而且这个数是确保程序中所有语句至少被执行一次所需的测试数量的上界。

上述程序的环形复杂度为6，因此共有6条独立路径。

路径1: 1-2-10-11-13

路径2: 1-2-10-12-13

路径3: 1-2-3-10-11-13

路径4: 1-2-3-4-5-8-9-2-...

路径5: 1-2-3-4-5-6-8-9-2-...

路径6: 1-2-3-4-5-6-7-8-9-2-...

7.6 白盒测试技术

1. 基本路径测试

④ 设计可强制执行基本集合中每条路径的测试用例。

应该选取测试数据使得在测试每条路径时都适当地设置好各个判定结点的条件。测试第③步得出的基本集合的测试用例如下。

路径1的测试用例：

value [k] =有效输入值，其中 $k < i$ (i 的定义在下面)

value [i] =-999, 其中 $2 \leq i \leq 100$

预期结果：基于k的正确平均值和总数

注意，路径1无法独立测试，必须作为路径4或5或6的一部分来测试。

路径2的测试用例：

value [1] =-999

预期结果： average=-999, 其他都保持初始值

7.6 白盒测试技术

1. 基本路径测试

路径3的测试用例：

试图处理101个或更多个值

前100个数值应该是有效输入值

预期结果：前100个数的平均值，总数为100

注意，路径3无法独立测试，必须作为路径4或5或6的一部分来测试。

路径4的测试用例：

value [i] =有效输入值，其中 $i < 100$

value [k] < minimum, 其中 $k < i$

预期结果：基于k的正确平均值和总数

路径5的测试用例：

value [i] =有效输入值，其中 $i < 100$

value [k] > maximum, 其中 $k < i$

预期结果：基于k的正确平均值和总数

7.6 白盒测试技术

1. 基本路径测试

路径6的测试用例：

value [i] =有效输入值，其中 $i < 100$

预期结果：正确的平均值和总数

在测试过程中，执行每个测试用例并把实际输出结果与预期结果相比较。一旦执行完所有测试用例，就可以确保程序中所有语句都至少被执行了一次，而且每个条件都分别取过true值和false值。

注意，某些独立路径（例如，本例中的路径1和路径3）不能以独立的方式测试，例如，为了执行本例中的路径1，需要满足条件`total.valid>0`。在这种情况下，这些路径必须作为另一个路径的一部分来测试。

7.6 白盒测试技术

2. 条件测试

用**条件测试技术**设计出的测试用例，能够检查程序模块中包含的逻辑条件。一个简单条件是一个布尔变量或一个关系表达式，在布尔变量或关系表达式之前还可能有一个NOT (\neg) 算符。关系表达式的形式如下：

$$E1 < \text{关系算符} > E2$$

其中，E1和E2是算术表达式，而<关系算符>是下列算符之一： $<$ ， \leqslant ， $=$ ， \neq ， $>$ 或 \geqslant 。布尔算符有OR(|)，AND(&)和NOT(\neg)。不包含关系表达式的条件称为布尔表达式。

因此，条件成分的类型包括布尔算符、布尔变量、布尔括弧（括住简单条件或复合条件）、关系算符及算术表达式。

如果条件不正确，则至少条件的一个成分不正确。因此，条件错误的类型有：布尔算符错、布尔变量错、布尔括弧错、关系算符错、算术表达式错。

7.6 白盒测试技术

2. 条件测试

条件测试方法着重测试程序中的每个条件。条件测试策略有两个优点：①容易度量条件的测试覆盖率；②程序内条件的测试覆盖率可指导附加测试的设计。

条件测试的目的不仅是检测程序条件中的错误，而且是检测程序中的其他错误。如果程序P的测试集能有效地检测P中条件的错误，则它很可能也可以有效地检测P中的其他错误。

在分支测试、域测试等条件测试技术的基础上，K. C. Tai提出了一种被称为**BRO (branch and relational operator) 测试的条件测试策略**。如果在条件中所有布尔变量和关系算符都只出现一次而且没有公共变量，则BRO测试保证能发现该条件中的分支错和关系算符错。

7.6 白盒测试技术

2. 条件测试

BRO测试利用条件C的条件约束来设计测试用例。包含n个简单条件的条件C的条件约束定义为 (D_1, D_2, \dots, D_n) ，其中 $D_i (0 < i \leq n)$ 表示条件C中第i个简单条件的输出约束。如果在条件C的一次执行过程中，C中每个简单条件的输出都满足D中对应的约束，则称C的这次执行覆盖了C的条件约束D。

对于布尔变量B来说，B的输出约束指出，B必须是真(t)或假(f)。类似地，对于关系表达式来说，用符号>，=和<指定表达式的输出约束。

作为第一个例子，考虑下列条件：C1: B1 & B2

其中，B1和B2是布尔变量。C1的条件约束形式为 (D_1, D_2) ，其中D1和D2中的每一个都是t或f。值 (t, f) 是C1的一个条件约束，并由使B1值为真B2值为假的测试所覆盖。BRO测试策略要求，约束集 $\{(t, t), (f, t), (t, f)\}$ 被C1的执行所覆盖。如果C1因布尔算符错误而不正确，则至少上述约束集中的一个约束将迫使C1失败。

7.6 白盒测试技术

2. 条件测试

作为第二个例子，考虑下列条件

C2: B1 & (E3=E4)

其中，B1是布尔变量，E3和E4是算术表达式。C2的条件约束形式为(D1, D2)，其中D1是t或f，D2是>, =或<。除了C2的第二个简单条件是关系表达式之外，C2和C1相同，因此，可以通过修改C1的约束集{(t, t), (f, t), (t, f)}得出C2的约束集。

注意，对于(E3=E4)来说，t意味=，而f意味着<或>，因此，分别用(t, =)和(f, =)替换(t, t)和(f, t)，并用(t, <)和(t, >)替换(t, f)，就得到C2的约束集{(t, =), (f, =), (t, <), (t, >)}。覆盖上述条件约束集的测试，保证可以发现C2中布尔算符和关系算符的错误。

7.6 白盒测试技术

2. 条件测试

作为第三个例子，考虑下列条件

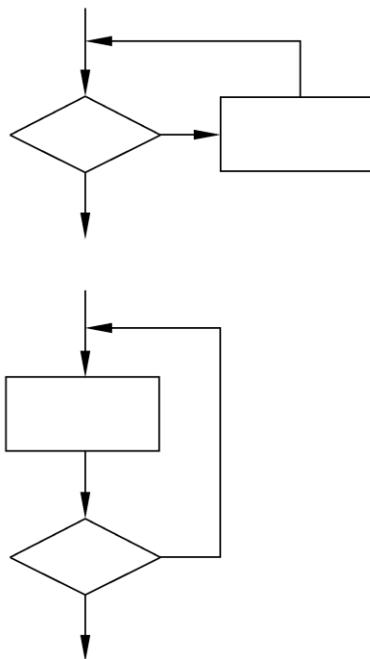
C3: $(E1 > E2) \& (E3 = E4)$

其中， $E1$ 、 $E2$ 、 $E3$ 和 $E4$ 是算术表达式。C3的条件约束形式为 $(D1, D2)$ ，而 $D1$ 和 $D2$ 的每一个都是 $>$ ， $=$ 或 $<$ 。除了C3的第一个简单条件是关系表达式之外，C3和C2相同，因此，可以通过修改C2的约束集得到C3的约束集，结果为： $\{(>, =), (=, =), (<, =), (>, <), (>, >)\}$ 覆盖上述条件约束集的测试，保证可以发现C3中关系算符的错误。

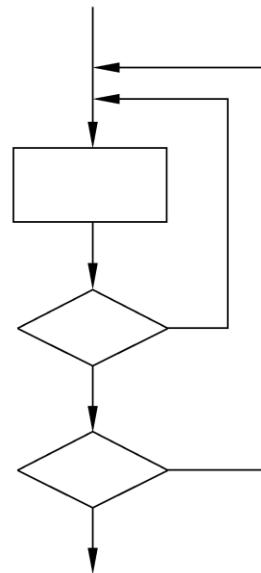
7.6 白盒测试技术

3. 循环测试

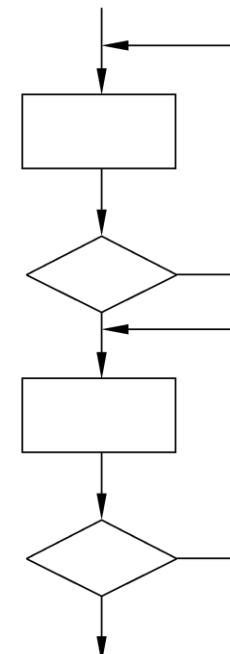
循环测试是一种白盒测试技术，它专注于测试循环结构的有效性。在结构化的程序中通常只有3种循环，即简单循环、串接循环和嵌套循环。



(a) 简单循环



(b) 嵌套循环



(c) 串接循环

7.6 白盒测试技术

3. 循环测试

(1) 简单循环

应该使用下列测试集来测试简单循环，其中n是允许通过循环的最大次数。

- 跳过循环。
- 只通过循环一次。
- 通过循环两次。
- 通过循环m次，其中 $m < n - 1$ 。
- 通过循环 $n - 1, n, n + 1$ 次。

7.6 白盒测试技术

3. 循环测试

(2) 嵌套循环

如果把简单循环的测试方法直接应用到嵌套循环，测试数就会随嵌套层数的增加按几何级数增长，B. Beizer提出了一种能减少测试数的方法。跳过循环。

- 从最内层循环开始测试，把所有其他循环都设置为最小值。
- 对最内层循环使用简单循环测试方法，而使外层循环的迭代参数（例如，循环计数器）取最小值，并为越界值或非法值增加一些额外的测试。
- 由内向外，对下一个循环进行测试，但保持所有其他外层循环为最小值，其他嵌套循环为“典型”值。
- 继续进行下去，直到测试完所有循环。

7.6 白盒测试技术

3. 循环测试

(3) 串接循环

如果串接循环的各个循环都彼此独立，则可以使用前述的测试简单循环的方法来测试串接循环。但是，如果两个循环串接，而且第一个循环的循环计数器值是第二个循环的初始值，则这两个循环并不是独立的。当循环不独立时，建议使用测试嵌套循环的方法来测试串接循环。

主要内容

7.1 编码

7.2 软件测试基础

7.3 单元测试

7.4 集成测试

7.5 确认测试

7.6 白盒测试技术

7.7 黑盒测试技术

7.8 调试

7.9 软件可靠性

7.7 黑盒测试技术

黑盒测试着重测试软件功能。黑盒测试并不能取代白盒测试，它是与白盒测试互补的测试方法，它很可能发现白盒测试不易发现的其他类型的错误。

黑盒测试力图发现下述类型的错误：

- (1) 功能不正确或遗漏了功能；
- (2) 界面错误；
- (3) 数据结构错误或外部数据库访问错误；
- (4) 性能错误；
- (5) 初始化和终止错误。

7.7 黑盒测试技术

白盒测试在测试过程的早期阶段进行，而黑盒测试主要用于测试过程的后期。设计黑盒测试方案时，应该考虑下述问题。

- (1) 怎样测试功能的有效性？
- (2) 哪些类型的输入可构成好测试用例？
- (3) 系统是否对特定的输入值特别敏感？
- (4) 怎样划定数据类的边界？
- (5) 系统能够承受什么样的数据率和数据量？
- (6) 数据的特定组合将对系统运行产生什么影响？

应用黑盒测试技术，能设计出满足下述标准的测试用例集。

- (1) 所设计出的测试用例能够减少为达到合理测试所需要设计的测试用例的总数。
- (2) 所设计出的测试用例能够告诉人们，是否存在某些类型的错误，而不是仅仅指出与特定测试相关的错误是否存在。

7.7 黑盒测试技术

7.7.1. 等价划分

等价划分把程序的输入域划分成若干个数据类，据此导出测试用例。等价划分法力图设计出能发现若干类程序错误的测试用例，从而减少必须设计的测试用例的数目。

如果把所有可能的输入数据(有效的和无效的)划分成若干个等价类，则可以合理地做出下述假定：每类中的一个典型值在测试中的作用与这一类中所有其他值的作用相同。因此，可以从每个等价类中只取一组数据作为测试数据。这样选取的测试数据最有代表性，最可能发现程序中的错误。

使用等价划分法设计测试方案首先需要划分输入数据的等价类，为此需要研究程序的功能说明，从而确定输入数据的有效等价类和无效等价类。

7.7 黑盒测试技术

划分等价类需要经验，下述的启发式规则可能有助于等价类划分。

- (1) 如果规定了输入值的范围，则可划分出一个有效的等价类(输入值在此范围内)，两个无效的等价类(输入值小于最小值或大于最大值)。
- (2) 如果规定了输入数据的个数，则类似地也可以划分出一个有效的等价类和两个无效的等价类。
- (3) 如果规定了输入数据的一组值，而且程序对不同输入值做不同处理，则每个允许的输入值是一个有效的等价类，此外还有一个无效的等价类(任一个不允许的输入值)。

7.7 黑盒测试技术

(4) 如果规定了输入数据必须遵循的规则，则可以划分出一个有效等价类(符合规则)和若干个无效等价类(从各种不同角度违反规则)。

(5) 如果规定了输入数据为整型，则可以划分出正整数、零和负整数3个有效类。

(6) 如果程序的处理对象是表格，则应该使用空表，以及含一项或多项的表。

7.7 黑盒测试技术

划分出等价类以后，根据等价类设计测试方案时主要使用下面两个步骤。

(1) 设计一个新的测试方案以尽可能多地覆盖尚未被覆盖的有效等价类，重复这一步骤直到所有有效等价类都被覆盖为止。

(2) 设计一个新的测试方案，使它覆盖一个而且只覆盖一个尚未被覆盖的无效等价类，重复这一步骤直到所有无效等价类都被覆盖为止。

注意，通常程序发现一类错误后就不再检查是否还有其他错误，因此，应该使每个测试方案只覆盖一个无效的等价类。

7.7 黑盒测试技术

假设有一个把数字串转变成整数的函数。运行程序的计算机字长16位，用二进制补码表示整数。这个函数是用Pascal语言编写的，它的说明如下：

```
function strtoint (dstr:shortstr):integer;
```

函数的参数类型是shortstr，它的说明是：

```
type shortstr=array [1..6] of char;
```

被处理的数字串是右对齐的，也就是说，如果数字串比6个字符短，则在它的左边补空格。如果数字串是负的，则负号和最高位数字紧相邻(负号在最高位数字左边一位)。

考虑到Pascal编译程序固有的检错功能，测试时不需要使用长度不等于6的数组做实在参数，更不需要使用任何非字符数组类型的实在参数。

7.7 黑盒测试技术

分析这个程序的规格说明，可以划分出如下等价类。

- **有效输入的等价类有**

- (1) 1~6个数字字符组成的数字串(最高位数字不是零)。
- (2) 最高位数字是零的数字串。
- (3) 最高位数字左邻是负号的数字串。

- **无效输入的等价类有**

- (1) 空字符串(全是空格)。
- (2) 左部填充的字符既不是零也不是空格。
- (3) 最高位数字右面由数字和空格混合组成。
- (4) 最高位数字右面由数字和其他字符混合组成。
- (5) 负号与最高位数字之间有空格。

7.7 黑盒测试技术

- 合法输出的等价类有

- (1) 在计算机能表示的最小负整数和零之间的负整数。
- (2) 零。
- (3) 在零和计算机能表示的最大正整数之间的正整数。

- 非法输出的等价类有

- (1) 比计算机能表示的最小负整数还小的负整数。
- (2) 比计算机能表示的最大正整数还大的正整数。

因为所用的计算机字长16位，用二进制补码表示整数，所以能表示的最小负整数是-32 768，能表示的最大正整数是32 767。

7.7 黑盒测试技术

根据划分出的等价类，可以设计出下述测试方案如下：

编号	描述	输入	预期输出
1	1~6个数字组成的数字串，输出是合法的正整数	‘1’	1
2	最高位数字是零的数字串，输出是合法的正整数	‘000001’	1
3	负号与最高位数字紧相邻，输出合法的负整数	‘-00001’	-1
4	最高位数字是零，输出也是零	‘000000’	0
5	太小的负整数	‘-47561’	错误——无效输入

7.7 黑盒测试技术

编号	描述	输入	预期输出
6	太大的正整数	‘132767’	错误——无效输入
7	空字符串	‘ ’	错误——没有数字
8	字符串左部字符既不是零也不是空格	‘×××××1’	错误——填充错
9	最高位数字后面有空格	‘12’	错误——无效输入
10	最高位数字后面有其他字符	‘1××2’	错误——无效输入
11	负号和最高位数字之间有空格	‘-12’	错误——负号位置错

7.7 黑盒测试技术

7.7.2. 边界值分析

经验表明，处理边界情况时程序最容易发生错误。例如，许多程序错误出现在下标、纯量、数据结构和循环等等的边界附近。因此，设计使程序运行在边界情况附近的测试方案，暴露出程序错误的可能性更大一些。

使用**边界值分析方法**设计测试方案首先应该确定边界情况，通常输入等价类和输出等价类的边界。选取的测试数据应该刚好等于、刚刚小于和刚刚大于边界值。

通常设计测试方案时总是联合使用等价划分和边界值分析两种技术。

7.7 黑盒测试技术

为了测试前述的把数字串转变成整数的程序，除了上一小节已经用等价划分法设计出的测试方案外，还应该用边界值分析法再补充下述测试方案。

编号	描述	输入	预期输出
1	使输出刚好等于最小的负整数	‘-32768’	-32768
2	使输出刚好等于最大的正整数	‘32767’	32767
3	使输出刚刚小于最小的负整数	‘-32769’	错误——无效输入
4	使输出刚刚大于最大的正整数	‘32768’	错误——无效输入

根据边界值分析方法的要求，应该分别使用长度为0, 1和6的数字串作为测试数据。

7.7 黑盒测试技术

7.7.3. 错误推测

错误推测法在很大程度上靠直觉和经验进行。它的基本想法是列举出程序中可能有的错误和容易发生错误的特殊情况，并且根据它们选择测试方案。

应该仔细分析程序规格说明书，注意找出其中遗漏或省略的部分，以便设计相应的测试方案，检测程序员对这些部分的处理是否正确。

经验表明，在一段程序中已经发现的错误数目往往和尚未发现的错误数成正比。例如，在IBM OS/370操作系统中，用户发现的全部错误的47%只与该系统4%的模块有关。因此，在进一步测试时要着重测试那些已发现了较多错误的程序段。

7.7 黑盒测试技术

等价划分法和边界值分析法都只孤立地考虑各个输入数据的测试功效，而没有考虑多个输入数据的组合效应，可能会遗漏了输入数据易于出错的组合情况。

选择输入组合的一个有效途径是利用判定表或判定树为工具，列出输入数据各种组合与程序应作的动作(及相应的输出结果)之间的对应关系，然后为判定表的每一列至少设计一个测试用例。

选择输入组合的另一个有效途径是把计算机测试和人工检查代码结合起来。

主要内容

7.1 编码

7.2 软件测试基础

7.3 单元测试

7.4 集成测试

7.5 确认测试

7.6 白盒测试技术

7.7 黑盒测试技术

7.8 调试

7.9 软件可靠性

7.8 调试

- **调试**（也称为纠错）作为成功测试的后果出现，即调试是在测试发现错误之后排除错误的过程。
- 软件错误的外部表现和它的内在原因之间可能并没有明显的联系。**调试**就是把症状和原因联系起来的尚未被人深入认识的智力过程。

7.8.1. 调试过程

- 调试不是测试。
- 调试过程从执行一个测试用例开始，评估测试结果，如果发现实际结果与预期结果不一致，则这种不一致就是一个症状，它表明在软件中存在着隐藏的问题。调试过程试图找出产生症状的原因，以便改正错误。

7.8 调试

调试过程总会有以下两种结果之一：①找到了问题的原因并把问题改正和排除掉了；②没找出问题的原因。在后一种情况下，调试人员可以猜想一个原因，并设计测试用例来验证这个假设，重复此过程直至找到原因并改正了错误。

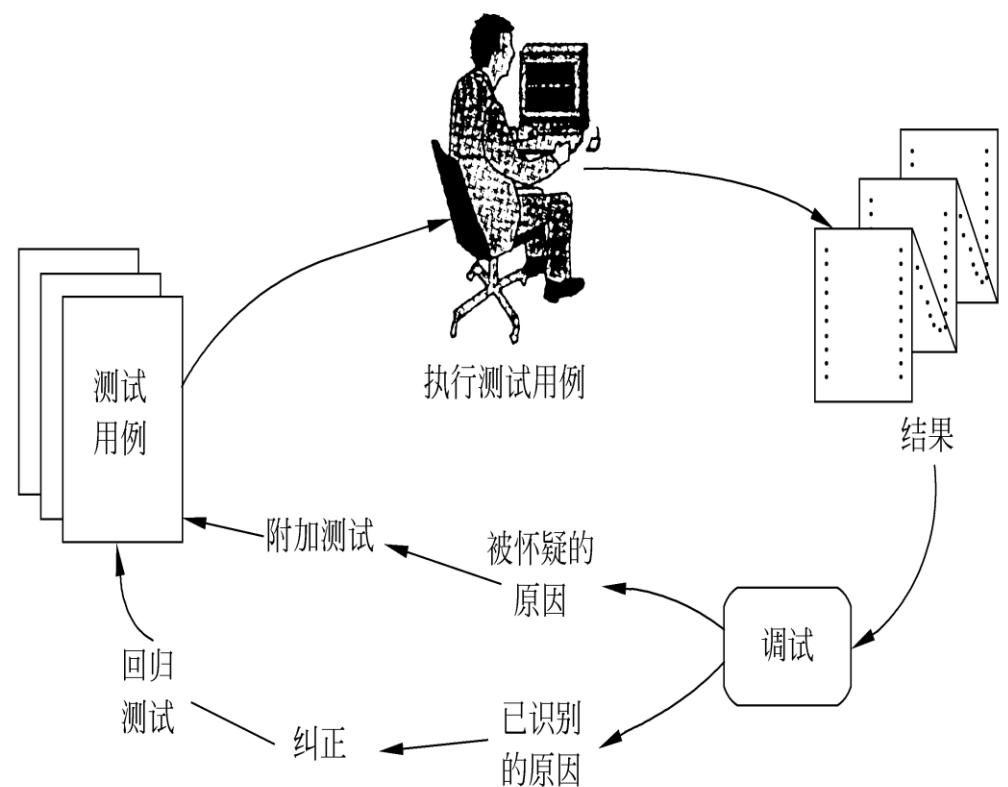


图 调试过程

7.8 调试

调试工作如此困难，软件错误的下述特征也是相当重要的原因。

(1) 症状和产生症状的原因可能在程序中相距甚远，也就是说，症状可能出现在程序的一个部分，而实际的原因可能在与之相距很远的另一部分。紧耦合的程序结构更加剧了这种情况。

(2) 当改正了另一个错误之后，症状可能暂时消失了。
(3) 症状可能实际上并不是由错误引起的（例如，舍入误差）。
(4) 症状可能是由不易跟踪的人为错误引起的。
(5) 症状可能是由定时问题而不是由处理问题引起的。
(6) 可能很难重新产生完全一样的输入条件（例如，输入顺序不确定的实时应用系统）。

(7) 症状可能时有时无，这种情况在硬件和软件紧密地耦合在一起的嵌入式系统中特别常见。

(8) 症状可能是由分布在许多任务中的原因引起的，这些任务运行在不同的处理机上。

7.8 调试

7.8.2. 调试途径

1. 蛮干法

- **蛮干法**可能是寻找软件错误原因的最低效的方法。仅当所有其他方法都失败了的情况下，才应该使用这种方法。
- **蛮干法**按照“让计算机自己寻找错误”的策略，这种方法印出内存的内容，激活对运行过程的跟踪，并在程序中到处都写上WRITE（输出）语句，希望在这样生成的信息海洋的某个地方发现错误原因的线索。
- 在更多情况下这样做只会浪费时间和精力。在使用任何一种调试方法之前，必须首先进行周密的思考，必须有明确的目的，应该尽量减少无关信息的数量。

7.8 调试

2. 回溯法

- 回溯是一种相当常用的调试方法，当调试小程序时这种方法是有效的。具体做法：从发现症状的地方开始，人工沿程序的控制流往回追踪分析源程序代码，直到找出错误原因为止。
- 随着程序规模的扩大，应该回溯的路径数目变得越来越大，回溯法不适用于这种规模的程序。

3. 原因排错法

对分查找法、归纳法和演绎法都属于原因排除法。

7.8 调试

对分查找法的基本思路是，如果已经知道每个变量在程序内若干个关键点的正确值，则可以用赋值语句或输入语句在程序中点附近“注入”这些变量的正确值，然后运行程序并检查所得到的输出。

归纳法是从个别现象推断出一般性结论的思维方法。使用这种方法调试程序时，首先把和错误有关的数据组织起来进行分析，以便发现可能的错误原因。然后导出对错误原因的一个或多个假设，并利用已有的数据来证明或排除这些假设。

演绎法从一般原理或前提出发，经过排除和精化的过程推导出结论。采用这种方法调试程序时，首先设想出所有可能的出错原因，然后试图用测试来排除每一个假设的原因。

主要内容

7.1 编码

7.2 软件测试基础

7.3 单元测试

7.4 集成测试

7.5 确认测试

7.6 白盒测试技术

7.7 黑盒测试技术

7.8 调试

7.9 软件可靠性



7.9 软件可靠性

7.9.1 基本概念

软件可靠性是程序在给定的时间间隔内，按照规格说明书的规定成功地运行的概率。软件可靠性随着给定的时间间隔的加大而减少。

一般说来，对于任何其故障是可以修复的系统，都应该同时使用可靠性和可用性衡量它的优劣程度。

软件可用性是程序在给定的时间点，按照规格说明书的规定，成功地运行的概率。

可靠性和可用性之间的主要差别是，可靠性意味着在0到 t 这段时间间隔内系统没有失效，而可用性只意味着在时刻 t ，系统是正常运行的。

7.9 软件可靠性

如果在一段时间内，软件系统故障停机时间分别为 t_{d1}, t_{d2}, \dots ，正常运行时间为 t_{u1}, t_{u2}, \dots ，则系统的稳态可用性为：

$$A_{ss} = \frac{T_{up}}{T_{up} + T_{down}}$$

其中， $T_{up} = \sum t_{ui}$, $T_{down} = \sum t_{di}$

如果引入系统平均无故障时间MTTF和平均维修时间MTTR的概念，则上式变为：

$$A_{ss} = \frac{MTTF}{MTTF + MTTR}$$

平均维修时间MTTR是修复一个故障平均需要的时间，它取决于维护人员的技术水平和对系统的熟悉程度，也和系统的可维护性有重要关系。平均无故障时间MTTF是系统按规格说明书规定成功地运行的平均时间，它主要取决于系统中潜伏的错误的数目。

7.9 软件可靠性

7.9.2 估算平均无故障时间的方法

1. 符号

在估算MTTF的过程中使用下述符号表示有关的数量。

E_T ——测试之前程序中错误总数；

I_T ——程序长度(机器指令总数)；

τ ——测试(包括调试)时间；

$E_d(\tau)$ ——在0至 τ 期间发现的错误数；

$E_c(\tau)$ ——在0至 τ 期间改正的错误数。

2. 基本假定

(1) 在类似的程序中，单位长度里的错误数ET/IT近似为常数。美国的一些统计数字表明，通常

$$0.5 \times 10^{-2} \leq ET/IT \leq 2 \times 10^{-2}$$

7.9 软件可靠性

(2) 失效率正比于软件中剩余的(潜藏的)错误数, 而平均无故障时间MTTF与剩余的错误数成反比。

(3) 假设发现的每一个错误都立即正确地改正了(即调试过程没有引入新的错误)。因此, $E_c(\tau)=E_d(\tau)$ 。剩余的错误数为 $E_r(\tau)=E_T - E_c(\tau)$, 单位长度程序中剩余的错误数为 $\varepsilon_r(\tau)=E_T/I_T - E_c(\tau)/I_T$ 。

3. 估算平均无故障时间

经验表明, 平均无故障时间与单位长度程序中剩余的错误数成反比, 即:

$$MTTF = \frac{1}{K(ET/IT - EC(\tau)/IT)}$$

其中, K为常数, 它的值应该根据经验选取。美国的一些统计数字表明, K的典型值是200。

7.9 软件可靠性

4. 符号

(1) 植入错误法

在测试之前由专人在程序中随机地植入一些错误，测试之后，根据测试小组发现的错误中原有的和植入的两种错误的比例，来估计程序中原有错误的总数 E_T 。

假设人为地植入的错误数为 N_s ，经过一段时间的测试之后发现 n_s 个植入的错误，此外还发现了 n 个原有的错误。如果可以认为测试方案发现植入错误和发现原有错误的能力相同，则能够估计出程序中原有错误的总数为 $\hat{N} = \frac{n}{n_s} N_s$

其中， \hat{N} 即是错误总数ET的估计值。

7.9 软件可靠性

4. 符号

(2) 分别测试法

为了随机地给一部分错误加标记，**分别测试法**使用两个测试员（或测试小组），彼此独立地测试同一个程序的两个副本，把其中一个测试员发现的错误作为有标记的错误。具体做法是，在测试过程的早期阶段，由测试员甲和测试员乙分别测试同一个程序的两个副本，由另一名分析员分析他们的测试结果。用 τ 表示测试时间，假设

- $\tau=0$ 时错误总数为 B_0 ；
- $\tau=\tau_1$ 时测试员甲发现的错误数为 B_1 ；
- $\tau=\tau_1$ 时测试员乙发现的错误数为 B_2 ；
- $\tau=\tau_1$ 时两个测试员发现的相同错误数为 b_c 。

7.9 软件可靠性

4. 符号

(2) 分别测试法

如果认为测试员甲发现的错误是有标记的，即程序中有标记的错误总数为 B_1 ，则测试员乙发现的 B_2 个错误中有 b_c 个是有标记的。假定测试员乙发现有标记错误和发现无标记错误的概率相同，则可以估计出测试前程序中的错误总数为

$$\hat{B}_0 = \frac{B_2}{b_c} B_1$$

使用分别测试法，在测试阶段的早期，每隔一段时间分析员分析两名测试员的测试结果，并且使用上面的公式计算 \hat{B}_0 。如果几次估算的结果相差不多，则可用 \hat{B}_0 的平均值作为 E_T 的估计值。

本章小结

1. 实现包括编码和测试两个阶段。
2. 高级程序设计语言较汇编语言有很多优点。
3. 通常软件测试至少分为单元测试、集成测试和验收测试3个基本阶段。
4. 软件测试不仅仅指利用计算机进行的测试，还包括人工进行的测试(例如，代码审查)。
5. 白盒测试和黑盒测试是软件测试的两类基本方法，设计白盒测试方案的技术主要有，逻辑覆盖和控制结构测试；设计黑盒测试方案的技术主要有，等价划分、边界值分析和错误推测。
6. 及时改正测试过程中发现的软件错误就是调试的任务。
7. 程序中潜藏的错误的数目，直接决定了软件的可靠性。通过测试可以估算出程序中剩余的错误数。

本章结束