

一种面向可理解性的 可复用软件开发方法*

董韫美 李开德

(中国科学院软件研究所)

AN UNDERSTANDABILITY ORIENTED METHOD FOR REUSABLE SOFTWARE DEVELOPMENT

Dong Yunmei and Li Kaide

(Institute of Software, Academia Sinica)

ABSTRACT

A programming method for supporting software understanding and reusing is presented in this paper, and a prototype system URS-1 to implement this method is also outlined.

摘 要

本文介绍一种支持软件理解和复用的程序设计方法, 以及实现该方法的实验性系统 URS-1.

§ 1. 引 言

可理解性是指软件的结构、功能、及其细节能容易地被该软件作者以外的其他人所理解。可复用性是指软件或其构件能容易地被用来构造其他新的软件系统。现今的大多数软件基本上不具备这些品质, 然而对于任何软件来说, 这两种属性是和可靠性、效率同等重要的, 应该成为新一代软件所必须具备的性质。如果真正达到了这些目标, 所谓“软件危

* 1989年5月7日收到。本项目得到国家自然科学基金委员会、机械电子工业部、国家科学技术委员会高技术计划的资助。

机”、“软件复杂性障碍”就将成为过去。

许多表面上不同的研究课题或技术开发，如软件工程、软件说明技术、面向对象的开发方法等，实质上也有意无意地以软件复用及理解为目标。我们最初是研究软件复用技术，不久就发现软件的可复用性强烈地依赖于其可理解性，因而把这两个问题连同起来考虑。在本文中我们将介绍自己解决问题的方法，以及据此进行的实验工作。

关于软件复用技术，已有不少文献报导^[1-4]。按照[5]中的分类办法，本文的方法可归入基于构件合成的复用一类。我们将有一个可复用软件库，在程序设计过程中当需要某种功能时，首先在库中寻找。需要而又找不到时，才设计新的软构件。一旦一个新构件已被建立并且经过检验，它将被存放在库中并从此可以被复用。

软件的可理解性依赖于完备而结构合理的文档，特别是中文文档，因此我们将专门讨论在软件文档中使用中文的问题。可理解性还依赖于软件本身的内部结构是否自然而合理，Knuth^[9]提出的文化程序设计使得可以非常自然地写软件，我们参照其思想，也开发了自己的系统。

以下我们将概述这种支持软件理解和复用的设计方法，然后概述实现这一方法的实验性系统 URS-1 的设计思想、总体功能和结构，最后叙述试用情况及今后的方向。

§ 2. 可复用软件设计过程

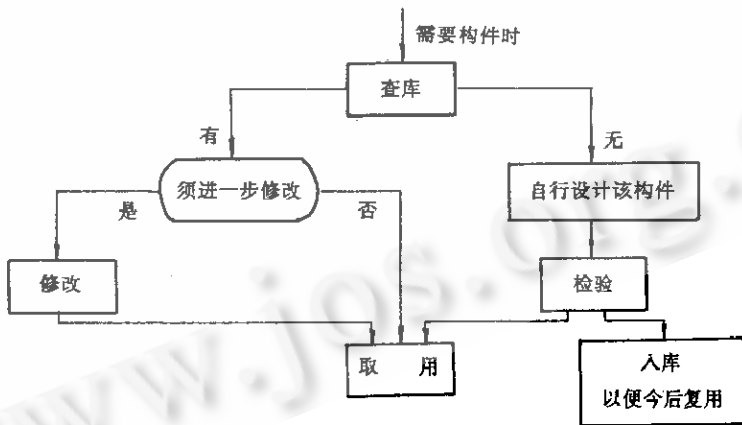


图 1 复用设计方式

我们将按照软件生存期的术语分阶段叙述。

2.1 需求分析

本阶段的结果是需求说明。它表述用户的需求，使用中文、英文、图表、可能还有数学记号。这是系统功能和用法的一种非形式描述，通常由用户或者系统分析员写出。它的作者将可以用一个专门的编辑器来帮助写作。此外，写作时应先在可复用软件库中查找，看有无类似的说明可资利用。

2.2 功能分析

本阶段的结果是功能说明，它精确地描述系统应该提供的功能，可能用形式化的记法并掺以自然语言。实际上，可以把需求说明的一个副本改写，适当地划分并且对每个部分进一步具体化。需求说明应该是它的一个组成部分。同样，在写功能说明时应该先查可复用软件库。

2.3 详细设计

仅仅需要设计那些真正新的功能。详细设计可以按照文化程序设计的方法进行，功能说明被扩展以补充如何实现各功能的细节。其结果是一系列文件，其中包含了系统的全部信息。

2.4 编码

产生可执行代码和可读文档的工作将由文化程序设计系统 CDS^[7]完成。

我们并不象 Knuth 那样强调文档排版系统的自由使用，而是代之以由系统自动生成高可读性文档，足以作为正式保存之用。

2.5 维护

我们没有建议有关测试和排错的任何特殊手段，而是采用现成的软件工具。然而重要的是，维护是在高层次即功能说明一级上进行，对系统功能和实现的任何修改都要回到功能说明一级。这使得维护变得非常容易，并且软件系统的可靠性也得到保证。

2.6 软件的复用和理解

一旦一个系统经过充分考验，它的设计和实现将来可能会被复用。于是应该把作为系统设计方案的需求说明和功能说明存放在软件库中，作为独立功能的某些部分也应该改写成可复用软件放在库中。软件的理解由完备而具有高可读性的文档及顺应程序员思考方式的软件结构所保证。

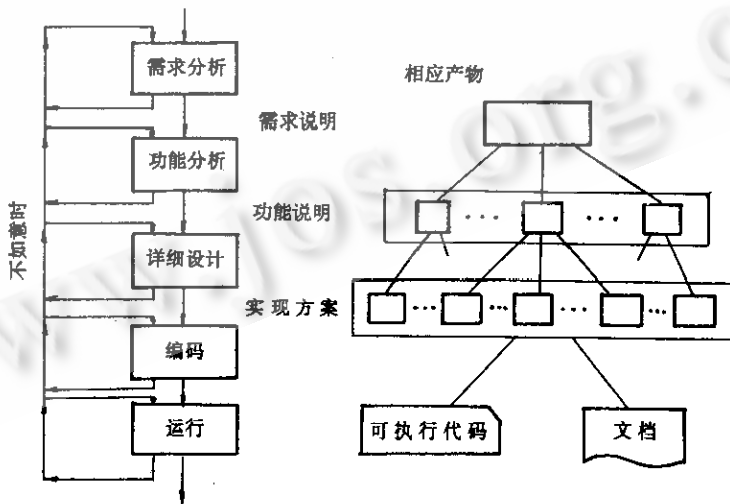


图2 可复用软件设计过程

§ 3. 程序设计系统 URS-1

根据以上的思想，我们研制了一个实验性系统 URS-1(Understandable and Reusable

software design System-1), 它可以在 80286 或 68000 为 CPU 的微型计算机上工作。

URS-1 包括三个子系统:

SBMS, 软件库管理系统^[6];

CDS, 文化程序设计系统^[7];

PLED, 面向个人的中西文屏幕编辑系统^[8];

每个子系统都可以作为一项工具独立地用于程序设计过程中的某一阶段。

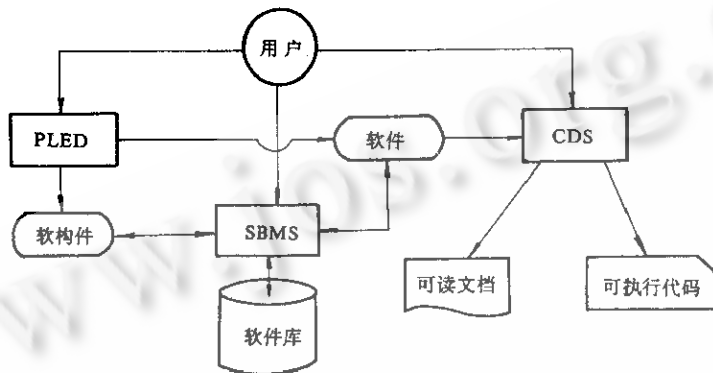


图 3 URS-1 的构成和作用

以下将叙述各子系统的设计思想及总体功能。

3.1 软件库管理系统 SBMS

SBMS 管理若干个软件库, 库中存放可复用软构件。每个软件库都有一些由其设计者定义的关键词。每个关键词管辖一批构件, 关键词贴切地刻划了这批构件的共同性质。SBMS 提供手段来浏览关键词表和在其中进行查找, 从而用户可以很方便地找到所需要的构件类。每个构件都有一个不超过 100 个字符的简短描述称为构件定义。同样地, SBMS 提供手段来浏览和查找构件定义表以及构件的详细说明。因而用户可以进一步很方便地找到所需要的构件, 并且抽取其实质部分以供复用。

每个软构件可以有四个组成部分如下

- 说明: 它是什么, 做什么事, 如何使用它, 以及构件作者想要说明的任何其他事情。这一部分是专门供理解构件之用的, 目前仅仅是面向用户的, 但是将来也可以是面向系统的。
- 属性: 构件的某些管理信息, 如构件作者、建立日期、版本号码、算法出处、所用的语言、效率、可靠性, 等等。这一部分随 SBMS 功能的增强将发挥更大的作用。
- 代码: 某一程序设计语言写的一段源程序。可以是过程、函数、数据说明、若干语句、或者甚至是注释。它将被 SBMS 抽出作为复用的实质部分。为达到较大的灵活性, SBMS 对用来写代码的语言没有加任何限制, 但是为了使用 CDS 系统, 应该用 C 语言来写代码。
- 环境: 本部分和代码部分合并之后将提供一个可执行的程序, 用以展示代码的行为, 因而它也是服务于理解构件的。为构件提供环境并非容易的事, 所以本部分

也可以缺省。

SBMS 提供一个友好的用户界面。在 SBMS 中有两个操作菜单，一个供系统管理员，一个供普通用户使用。差别是管理员可封锁、开锁、增加、修改、删除软件库中的任何构件，而普通用户则不能。然而，用户可以拥有私人的软件库，其他人不能访问该库，用户对其私有的库具有管理员的地位。一旦库主决定与其他人共享该库时，管理员可以把它合并到公共的软件库中。

3.2 文化程序设计系统 CDS

Donald Knuth 提出过一种新的程序设计方法学称为“文化程序设计”(Literate Programming)^[9]，其出发点是：人们在构造一个程序时，应该把注意的重点从教计算机做什么转移到向人解释打算让计算机做什么上来，换言之即把重点转移到软件的可理解性上来。他并且实现了一个称为 WEB 的系统，以支持这种新方法学。

WEB 系统由两个语言组成，一个是他自己发明的排版语言 TEX，另外一个程序是程序设计语言 PASCAL。WEB 程序是两种语言的混合物，其中 PASCAL 部分可以由一个子系统 TANGLE 抽出，并且提交 PASCAL 编译，得到可执行代码。另一个子系统 WEAVE 将加工 WEB 程序以得到 TEX 的输入，并用来产生可读的文档。特别值得注意的是，WEB 容许各种不同的程序设计风格。可以是自顶向下、由底向上、甚至是意识流式的。总之，它允许程序员按照自己思考的顺序来写程序。WEB 已被用来出版某些大型软件杰作如 TEX^[10]METAFONT^[11]以及 WEB^[12]本身，质量极为精美。

文化程序设计主要是着眼于软件的可理解性，然而我们相信，这种工作方式本身也完全适于使用可复用软构件来构造新软件系统，所以决定开发自己的兼顾可理解性和可复用性的文化程序设计系统 CDS^[7]。

CDS 中的程序设计语言是 C，排版系统是自己设计的 SP。CDS 的程序由若干段组成，段具有非常自由的形式。段可以有自已的名字。段可以包含一个注释部分以便其作者提供面向读者的信息。段可以包含宏定义以提供更强的表达能力。段的实质部分是 C 语言的程序，但是非常灵活，可以包含宏调用或者引用其他段。段的模块性很强，并且非常灵活。段应该就是被复用的单位，也就是 SBMS 中的软构件。

CDS 的功能包括：

- 从 CDS 源程序中抽取 C 语言程序，以供 C 编译器进一步加工得到可执行代码；
- 加工 CDS 源程序得到高可读性的程序文档。

所产生的文档包括以下内容：

- 程序段的目录；
- 程序模块名字的索引表；
- 交叉引用索引表，即 C 程序中非保留字的名字的定义性出现和应用性出现部位；
- 源程序结构信息，反映模块之间的关系，模块与程序段之间的关系；
- 考虑到美观与结构清晰要求的 C 语言程序和注释。

排版系统 SP 的功能包括普通的文书排版，并特别考虑了程序排版的需要，不用现成的排版系统而自行设计 SP 的理由是：所要求的文档排版系统除了处理中西文之外还要能

处理程序，并能在 8086 和 68000 序列的芯片上工作，最重要的是我们必须掌握系统的源程序，以便作可能的修改和扩充，并作为将来发展的基础。在这些约束条件之下，除了自己开发之外，别无其他选择。

象 WEB 中那样使用两种语言，对于程序员来说是一种沉重的负担，特别是 TEX 那样的内容极其丰富而复杂的命令式排版语言。因而在 CDS 中 SP 对程序员可以是隐蔽的。即程序员可直接用 SP 排版命令插入源程序以控制版式，也可以不用它而由系统完全自动决定版式。与 WEB 比较，程序员的负担减轻了，但是排版格式的灵活性也有所损失。

3.3 在软件文档中使用中文

对于软件工程师来说，表达其设计意图的最自然方式是使用本国语言。大多数软件读者大约也希望有本国语言写的文档。对于英语或者其他使用罗马字母表的语言来说，利用机器产生文档没有任何困难。然而对于中国的软件工程师所惯用的汉语，情形就不同了。

3.3.1 面向写作人员个人的中文输入方法

汉语使用很大的字符集合，迄今已经发明了数百种汉字输入方案，其中大多数是为了中文打字员高速汉字录入用，于是通常需要一个艰苦的学习过程和记忆若干不自然的规则。

我们认为软件工程师是一类很特别的中文打字员。他们是写作人员，自主地打他们自己想要写的东西，而不是被动地打别人写出来的东西。中文打字不是他们的专业，他们不在意打字的速度，也难以要求他们记忆不自然的规则，他们更重视的是简单而自然的输入规则。实际上当他们打字的时候，经常要停下来思考，以寻找表达他们思想的最好方式。已经观察到每一个人都有自己的写作习惯，只从一个相当受限的词汇表中选词用字。作者们并且受到自己写作主题的限制。

基于以上的想法，我们提出一种中文输入方法，适于包括软件工程师在内的写作人员使用。这种输入方法的特点是：用户可以建立若干个自己专用的词典，在其中包含了他所惯用的字和词，这些字词是作为“意义”而通过其关键字检索得到的。关键字是由用户按照自己的记忆习惯来定义的 ASCII 字符串。用户输入中文时，只需要打关键字，系统将据此检索出相应字词作为输入的对象。一个关键字允许有多个意义，系统将全部提供出来供用户选择。用户在输入过程中可以不断地扩充他的词典。这样，虽然在一开始时用户的输入速度可能较慢，随着个人专用词典的扩充和用户对这种输入方式的逐渐适应，将会获得越来越高的输入速度。

虽然每个人只为自己建立词典，但是看来有一个词典大家都会用到。按照字词频率统计材料^[13]，我们建立了一个公共词典并向所有用户推荐。

公共词典中收有近千个常用字，累计字频为 89.12%，近五千个常用双字词，累计词频为 84%。这些字词的编码规则很容易记忆，那就是汉语拼音字头缩写。例如，“程序”一词的汉语拼音是 Chengxu，它的关键字就是 cx。按照这一原则，“抽象”一词的关键字也是 cx。但如前所述，这并不导致问题。我们建议，用户首先浏览公共词典并且熟悉一下它的编码规则。如果不喜欢这种规则，用户可以按照自己的意愿来改造公共词典，可以使用地方方言、特殊符号或甚至英语来作为关键字。

3.3.2 中西文屏幕编辑系统 PLED

这个输入方法已经在—个中西文屏幕编辑系统 PLED 中实现^[1]，其功能包括：

- 屏幕编辑：功能和用法与 UNIX 系统中的屏幕编辑器 VI 兼容；
- 图表输入：可用任意字符绘制图表；
- 中文输入：面向用户个人用词习惯的专用词典、高频字词标准词典、在编辑的同时扩充及修改词典、CCDOS 及其他中西文终端所支持的输入方法；
- 词典管理：词典的创建、选择、修改、翻阅、复制、删除；
- 友好的用户界面：入门介绍、联机 HELP，菜单驱动。

§ 4. 结束语

URS-1 的三个子系统已经经过多次改进甚至重新设计，现行版本的程序量大约相当于 25,000 行 C 源程序。它们现在 DUA83/80 和 IBMPC/AT 上工作，作为独立的软件工具已经被使用了一段时间，其结果是令人鼓舞的。PLED 被研究室的同事广泛地用来撰写中文论文和报告。CDS 和 SBMS 被用于它们自身的开发和改进(以自展方式)。

但是 URS-1 作为一个集成系统，还需要进行更多实验，尤其是需要积累更多的可复用软构件(这个任务由用户来完成，更为合适)，才能得到比较确切的结论。特别是当以下的条件被满足时，软件复用才是现实可能的：

- 待复用的构件相当复杂，复用是值得考虑的；
- 已经拥有丰富的可复用构件资源；
- 用于寻找可供复用的构件的代价、理解它的代价、修改它使得完全符合使用要求的代价，其总和少于重新开发它的代价；
- 待复用的构件的正确性没有问题，时间、空间效率是可接受的。

今后的工作：看来建立某种帮助理解已有软件，并从中提取可复用构件的工具，对于积累软件库，将会是有用的。软构件的参量化功能应该加强，以便使较少的软构件能适用于较多的复用情形。虽然当前起作用的还是较低层次(代码级)的构件复用，但从获取更大效益出发，高层次上(知识、经验)的复用应该作为今后研究工作的努力方向。

致谢：研究生张旭波、曾云峰、蔡德明完成了子系统的全部程序设计工作，刘骏参加了系统设计的讨论，谨致谢意。

参 考 文 献

- [1] IEEE Transactions on Software Engineering, Vol. SE-10, No.5, Sept. 1984.
- [2] P. Freeman (ed.), Software Reusability, IEEE Computer Society Press, 1987.
- [3] IEEE Software, Jan. 1987.
- [4] IEEE Software, July 1987.
- [5] Ted J. Biggerstaff & Alan J. Perlis, Foreword, IEEE Transactions on Software Engineering, Vol. SE-10, No.5, pp. 474-477, Sept. 1984.
- [6] 蔡德明，可复用软件库管理系统 SBMS，1988 年度智能计算机系统学术和工作会议，北京。
- [7] 曾云峰，文化程序设计系统 CDS，同上。

- [8] 张旭波, 面向个人的词汇输入方式中西文屏幕编辑器 PLED, 同上。
- [9] Donald E. Knuth, Literate Programming, The Computer Journal, Vol. 27, No. 2, pp. 97-111. 1984.
- [10] Donald E. Knuth, Computers & Typesetting, Vol. B, TEX: The Program, Reading, Mass. Addison-Wesley 1986.
- [11] Donald E. Knuth, Computers & Typesetting, Vol. D, METAFONT: The Program, ibid.
- [12] Donald E. Knuth, The WEB System of Structured Documentation, Stanford Computer Science Report 980, Stanford University, 1983.
- [13] 刘源等, 社会科学·自然科学综合词频统计表。

《软件学报》创刊征稿征订通知

为了适应我国计算机软件不断发展的需要和广大读者的要求, 并受中国计算机学会软件专业委员会的委托, 决定由中国科学院软件研究所主办《软件学报》。该刊已经国家科委批准, 将于一九九零年初创刊, 暂定季刊, 国内外公开发行, 全国各地邮局订阅, 欢迎科研部门研究人员、大专院校的师生以及广大计算机界科技人员投稿。

办刊方针:

加强计算机软件这一领域的学术交流, 为我国计算机软件事业的发展和国民经济建设服务, 重点刊登我国计算机科学基础理论、软件理论、软件技术和软件工程方面的重要研究成果, 及时反映我国计算机软件的学术水平。

刊登内容:

数理逻辑、自动机及其在计算机科学中的应用, 可计算理论、网论, 算法设计与分析, 思维科学, 计算语言学, 形式语言和语义理论, 程序设计逻辑, 软件开发形式化技术, 软件自动生成, 人工智能, 专家系统, 知识工程, 定理机器证明, 自动翻译, 计算机辅助软件工程, 程序设计语言, 操作系统, 数据库, 计算机网络, 分布式系统, CAD/CAM/CAT/CAE, 图形、图像、汉字、语言等信息处理, 计算机辅助教学, 软件新技术的应用开发。

读者对象: 广大的计算机科学研究人员、工程技术人员、大专院校的教师及高年级学生、研究生等。

订阅办法: 全国当地邮局订阅(过期期刊可到编辑部购置, 售完为止, 需要邮寄者外加邮费10%)。

本刊主编: 许孔时

通讯处: 北京 8718 信箱《软件学报》编辑部。

地址: 北京海淀区中关村科学院南路 6 号中国科学院软件研究所。

邮政编码: 100080 **电报:** 5547 **电话:** 2562563 **联系人:** 鞠玉兰

国内统一刊号: CN11-2560

国内邮发代号: 82-367

单价: 2.40 元