

软件可用性与美国工业可用性标准报告

Software Usability and the US Industry USability Reporting(IUSR)

李 佳 刘正捷 李冠宇 段建丽

(大连海事大学计算机学院欧盟可用性中国支持中心 大连116026)

Abstract In the software product development, a special attention should be paid to the usability of the product while realizing its functionality. Usability pertains to the user's view of the quality of a product, which is the synonym of the quality in use. It is one of the critical factors to the success of the product. In order to improve the usability of software products, we must first make usability visible. Thus a kind of objective, uniform and quantitative measurements for usability is in great need. The Common Industry Format for usability testing is a standardized test for software usability, which is also called CIF test for short. It is the main contribution of the IUSR project initiated by the US National Institute of Standards and Technology (NIST). The CIF Test is based on a carefully designed evaluation procedure and an appropriate metrics. It enables suppliers and purchasers of software products measure the usability of the products in an objective and quantitative way, therefore promote the emphasis on products' usability. After giving the concept of software usability, this paper describes the IUSR project and its main contribution-Common Industrial Format for usability testing. The authors' experiences in using the standard are discussed finally.

Keywords Software, Usability, NIST, IUSR, CIF, Effectiveness, Efficiency, Satisfaction, Context of use

一、前言

近几十年来, IT 产业发展十分迅速。然而在 IT 产品特别是软件产品的开发中一直存在的一个问题是在重视技术创新的同时往往对人的因素考虑不足, 这导致多数产品存在着程度不同的可用性问题。例如根据一项统计^[1], 在英国, 一般商用计算机用户只有30%-40%的有效生产率; 80%的软件缺陷是由于不能有效满足用户需求, 而其中60%是由于存在着可用性质量问题; 70%的英国公司对他们的网站设计不够满意。这说明在软件产业的发展中, 软件的可用性质量是一个十分重要的问题。规范地说, 可用性是指产品为特定用户用于特定用途时所具有的有效性、效率和用户主观满意度。具有良好可用性质量的软件产品可以减少用户的操作错误, 减少人员培训和系统维护的费用, 提高用户的生产效率和满意程度, 从而提高产品的市场竞争力。目前造成许多企业忽视产品可用性质量的一个原因是缺乏一个客观、统一和定量的衡量标准。这是一个必须要解决的问题。这些年来学术界和工业界都为

此进行了不懈的努力, 并正在取得积极的进展。本文以下部分首先介绍可用性与可用性工程的概念, 然后重点介绍美国 IUSR 计划及其所发展的 CIF 可用性测试, 最后是作者在实际使用 CIF 测试中的初步体会。

二、可用性与可用性工程

1. 可用性(Usability)的含义

提起可用性有些人马上会想到用户界面, 其实可用性并不仅仅是用户界面的问题。通俗地说, 产品可用性包含了五方面的含义^[2]:

- 1) 易学: 应该容易学习, 使用户能尽快掌握其使用方法;
- 2) 高效: 其使用应能给用户带来较高的工作效率;
- 3) 好记: 使用方法应该容易记忆, 当用户在一段时间没有使用该产品后, 再使用时不需要重新学习就能很快使用;
- 4) 少错: 应能减少用户的操作错误, 一旦发生错误用户也可以容易地纠正, 并且应杜绝致命性错误的产生;

李 佳 硕士生, 研究方向为可用性工程。刘正捷 教授, 研究方向为人机交互、可用性工程和软件过程改进。李冠宇 副教授, 研究方向为信息系统工程和可用性工程。段建丽 硕士生, 研究方向为可用性工程。

5)满意:用户在使用时可以有愉快、满意的感觉。

在有关的国际标准如 ISO9241-11、ISO14598-1、ISO9126-1中对可用性质量作了范围更加广泛而确切的定义。其中,在 ISO9241-11中对可用性是这样定义的:产品为特定用户用于特定用途时所具有的有效性、效率和用户主观满意度。这里给出了三个指标,即有效性、效率和满意度。有效性是指用户完成特定任务和达到特定目标时所具有的正确和完整程度。效率是指用户所完成任务与所使用资源(如时间)之间的比率。满意度是一个主观标准,体现用户在使用产品过程中具有的满意程度。在这里,可用性质量的概念与通常的方便使用的概念是不相同的,它所体现的是满足用户需求这样一个广义的目标,因此也被称为使用质量(quality in use)。在 ISO14598-1中就采用了使用质量这一定义,它把使用质量解释为从用户角度感受到的产品质量,即以使用产品的有效结果来衡量的质量,而不是产品本身所固有的属性。只有某一产品能使用户有效地达到使用目的、充分地利用时间等资源并感到满意,才能说这一产品符合了用户的需求,即具有好的使用质量。这种使用质量的概念与可用性质量是一致的。

2. 可用性工程(Usability Engineering)

自八十年代以来,人们对产品可用性质量的重视促进了可用性工程这一概念的出现,并相应地形成了一个在学术界和工业界均十分热门的领域^[2]。可用性工程泛指以提高和评价产品可用性质量为目的的一系列过程、方法、技术和标准。可用性工程近十几年来在发达国家的 IT 工业界得到了广泛重视和使用,比如在微软、IBM 等许多企业都有十几年实际运用的历史,并设有几百人规模的专门的产品可用性部门。为了在工业界的普遍应用,国际标准组织 ISO 已经和正在制订有关的国际标准如 ISO9214、ISO13407等。欧盟以及美、日等国政府正在实施有关的推广计划,这正在使可用性工程的运用成为今后软件及 IT 工业界的一个发展趋势。可用性工程既包括在软件开发生命周期中为提高可用性所运用的一系列技术方法(如用户为中心的设计方法),也包括面向原型或最终软件产品所进行的可用性评价和测试,其中,可用性测试为衡量和改进可用性质量提供了必要的基础,在可用性工程领域一直备受关注,然而可用性质量评价不易定量、难以统一的特点,也使其一直是一个难点。针对这一问题,可用性工程领域多年来做了许多尝试和努力,并正在取得显著的进展,在这其中,美国的 IUSR 计划是一个重要的里程碑。

三、IUSR 计划

1. IUSR 计划的内容

在1997年10月,美国国家标准技术局(NIST)顺应来自各方面的普遍需求,联合自愿参加的主要软件产品供应商和采购商发起了工业可用性报告计划(the Industry USability Reporting--IUSR)^[3]。NIST 隶属于美国商业部,是推动制定工业标准的政府机构。参与此项计划的美国主要软件供应商有 IBM、Microsoft、HP、Sun、Oracle、Compaq 等,大的软件采购商有如 Boeing、Northwest Mutual Life 和 State Farm Insurance 等。许多企业在软件生产和使用中由于可用性问题遭受了难以估量的损失,如生产成本和周期的增加、产品竞争力的降低、企业内部和外部培训和技术支持的费用、用户操作错误造成的损失、产量和工作效率的降低等。针对这些问题,IUSR 计划致力于提高软件产品可用性质量的可见性,这主要是通过对软件产品的可用性质量进行标准测试并提供测试结果的标准报告格式来实现的。这项计划最终要达到三个目的:

- 1)鼓励软件供应商和采购商更好地理解用户的可用性需求和任务;
- 2)建立一个通用可用性质量报告规范,使可用性信息更加规范以利于相互沟通和共享;
- 3)进行先导试验来确定 IUSR 可用性报告规范的使用效果和软件采购中的使用价值。

IUSR 计划所开发的可用性质量测试结果的报告规范被称为通用工业规范(Common Industry Format--CIF),目前为1.1版。CIF 报告所提供的是采购商在评价产品可用性质量时所需的最基本的信息,在对具体产品使用 CIF 报告的时候可以选择增加内容。CIF 报告的基本内容包括产品说明、测试目的、测试对象、测试任务、测试环境、测试的方法和过程、指标体系和数据采集及数据分析结果等。CIF 是对测试方法和测试结果进行详细说明的标准格式。CIF 报告对可用性测试起到了指导和规范的作用,实际上给出了一种软件可用性测试的标准。

2. IUSR 与其它标准的关系

目前,IUSR 计划是一项独立的工作,不依赖于任何其它标准组织,但为协调工作中的重复和交叉,IUSR 的成员与 ISO、IEC、IEEE 等标准组织的代表有充分的沟通。IUSR 计划所制定的 CIF 报告符合 ISO9241-11和 ISO13407国际标准,可以在这方面作为有关国际标准的补充。

四、CIF 可用性测试

目前对可用性质量结果的报告可以有多种形

式,因而产品采购方对报告本身的可比性、可靠性、有效性存在质疑,通用标准格式的报告将有助于解决这个问题。客观的、定量的、可重复进行的 CIF 报告使采购方更容易评价测试、解释测试结果,因而将大大方便软件供应方和采购方对产品可用性质量的评价和对比。

1. 测试类型

CIF 测试是对最终软件产品可用性质量的衡量,因而它不同于软件开发过程中所进行的可用性测试,后者是为了发现问题,收集用户反馈信息以改进设计。采购方需要知道的是在特定的环境下,特定类型的用户使用某软件产品完成特定任务的效果。这不仅仅是界面简洁友好的问题,它体现了产品在何种程度上提供了适当的功能,以及是否具有可靠性和便捷性等质量,所以,CIF 测试侧重的是提供同类产品之间可比的可用性质量的数据,是一种黑箱方式的定量测试。

2. 测试相关人员

CIF 测试可以由几方面的可用性专业人员来进行,软件供应方可以设计并进行 CIF 测试,为采购方提供 CIF 报告;采购方也可以进行 CIF 测试来验证某个测试报告;还可以由独立的第三方可用性专业人员来进行,CIF 报告主要供两类人员使用:

1) 第一类是采购方的管理和决策人员,他们根据测试结果来决定是否购买某种软件产品。

2) 另一类是可用性专业人员,他们对可用性测试的技术价值和产品的可用性进行评价,或者重复进行可用性测试以对报告结果加以验证。

3. 测试计划

可用性是产品为特定的用户用于特定的用途时所具有的有效性、效率和用户主观满意度,因此只有让测试对象在特定的环境下完成特定的任务才能得到可靠的可用性质量测试结果。这里所谓的特定是指应尽可能地接近实际环境和工作任务。这需要事先有一个详细设计的测试计划,给出测试过程和适当的指标体系。在测试计划中首先必须给出产品描述,说明要测试产品的版本和组成部分等基本信息。其次要分析产品的使用环境(context of use)^[4],并在此基础上设计和确定测试环境。

4. 测试环境

产品的可用性不仅取决于产品本身的特点而且受产品使用环境的影响。产品的使用环境包括以下几方面:

1) 用户的特征:这包括用户的知识背景、技能、经验、学历、年龄和体力等内容,有时还可以将用户分类,如按不同经验、不同职务或不同能力来分组。

2) 使用产品的工作目标:产品对用户来说的主要

用途,以及所有可能影响使用质量的工作特点如工作时间长短、强度等。

3) 其它社会物理环境:如计算环境(包括计算机网络状况),周围的温度、湿度环境,法律环境,社会文化习惯环境等。

在详细分析使用环境的基础上,设计出测试环境并在测试计划中明确定义,应使测试环境具有使用环境的典型特征:

1) 测试对象特征:说明测试对象的知识背景、技能、学历、年龄和体力等内容。

2) 测试任务:测试任务应符合产品的主要用途并覆盖其主要功能,详细规定任务要求和执行顺序。

3) 测试的社会物理环境:具体说明所使用的物理设备、软件和其它所需设备以及相关的物理和社会环境特征。

5. 衡量可用性质量的指标体系

CIF 测试的客观统一性体现在它使用了一套明确的指标体系来定量地衡量可用性质量,这个指标体系通常包括三方面的指标:

1) 有效性指标:有效实现的工作目标与预期工作目标的比值。通常包括任务完成率、出错率、寻求帮助的次数等具体指标。其中完成率又包括:当任务不可分时,有效完成任务的测试对象所占比重;当任务可分时,每个测试对象所完成的工作与全部工作的比率。

2) 效率指标:使用单位资源所完成任务(这里的资源通常指时间),如单位时间的工作量。

有效性和效率指标都是通过黑箱方式测试的,它们不考虑所用技术和达到目的的方式,因而,这种测试结果可有效地用于同类产品的对比。

3) 满意度指标:描述测试对象使用产品的主观满意程度,可使用标准的问卷进行调查,如 SUMI(Software Usability Measurement Inventory)、SUS(System Usability Scale)等问卷。

6. 测试过程和数据收集

在测试过程中首先要向各位测试对象介绍测试的内容和目的,然后,让测试对象熟悉一下测试的物理环境。值得注意的是在进行测试的过程中,为了使测试环境符合使用环境,除了用户在正常使用环境中所能得到的帮助外不能向测试对象提供任何额外帮助。

接着正式开始测试,由测试管理员介绍产品及测试任务和完成时间。测试过程中管理员记录测试数据并计算他们完成任务的以下各项指标:平均完成率、平均效率、平均完成时间等。

最后,询问测试对象对于这个产品是否有其他的意见和建议。除了衡量有效性、效率和满意度等指标外,还应记录下测试对象遇到的具体问题。若条件允许

应使用低干扰的测试工具,如果使用摄像机拍摄下测试的过程,将更有利于对细节的分析,并且可以使测试对象更加放松自然,从而得到更准确的测试结果。

7. 数据的分析和解释

对所记录的测试数据经过计算整理后可得出用来定量衡量有效性、效率和满意度这三方面指标的数据,并可以用图表等方式表现出来,使结果更加直观。

供应方和采购方双方都应有其可用性专业人员对测试结果进行解释说明,并对所做的解释负责。对于供应方进行的测试,供应方应给予所有的采购方同样的权利来获取这些信息,而不应使某些采购方享有任何特权。

五、CIF 测试报告格式

作为标准测试结果的记录和说明,CIF 报告采用统一规范的格式^[4]。

在报告的封面要标明本报告为 CIF 报告,并写明测试产品的名称、版本、测试人员、测试时间、报告时间、报告人及报告人的联系方式等内容。

正文的第一部分是概述,这部分主要针对的读者为产品采购方的决策人员,他们往往并不关心具体的技术细节,而注重产品描述和测试结果,所以概述部分主要是产品描述、测试目的和测试结果等内容。

除第一部分外的其它部分针对的读者主要是可用性专业人员,因此这些部分提供的信息必须足够详细以便从技术角度判断测试的可信性,并可据以重复测试。

第二部分是测试方法,介绍测试对象的人数(至少为8人)和特点,测试环境的情况如测试任务、测试设施、测试的计算机环境,以及测试工具和测试管理人员等。

第三部分是测试过程,详细说明测试的具体过程和步骤、对测试对象的指导,测试任务说明和使用的指标体系。

最后一部分是测试数据、计算及分析结果,包括数据计算标准、计算公式和计算结果,这是报告的主要技术性部分,应以数据表格的形式表示出来。

我们在对 CIF 可用性测试方法进行研究的過程中,选取了一些软件产品进行 CIF 测试的试用,其中包括一种电子日记本产品,从中获得了一些体会。

电子日记本的用户对象十分广泛,可以是任何有基本电脑操作基础的用户,尤其适用于有记日记习惯并经常使用电脑的青年人。

六、使用 CIF 的体会

我们在对电子日记本进行试用时,事先拟定了详

细的测试计划书。根据产品的主要功能及用途设计了测试任务:测试对象使用该产品建立新的用户记录并完成一份指定的日记。

在测试过程中,首先向各位测试对象介绍测试的内容和目的。然后,让测试对象熟悉测试的物理环境,让其填写一份关于其一般情况的表格,包括测试对象的专业、年龄、使用 Windows 的经验和对使用电子日记本的态度等内容。其间着重向测试对象申明测试的目的是对产品的可用性质量进行评价,而不是对用户的能力和素质进行任何评价,测试结果的记录将采取匿名方式。

在正式测试时,由测试管理员 A 向所有测试对象介绍电子日记本产品,测试管理员 B 设置测试的初始计算机状态。之后对测试对象逐个进行测试。管理员 A 告知测试对象每项任务的内容和最长完成时间,要求测试对象在完成每一项任务后示意管理员,并且告知他们在完成任务的过程中将不提供任务帮助。管理员现场记录每项任务的完成时间、帮助次数和正确率等测试数据,并计算他们完成任务的各项指标:平均完成率、平均效率、平均完成时间等。在每个测试对象完成最后一项任务后,让其填写一份 SUMI 问卷。最后,征询测试对象对于该产品的其他意见和建议。

这次测试所使用的指标体系如下。有效性指标包括以下各项:用户完成率--正确完成任务的用户比率。平均完成率--用户完整、正确完成某项任务的平均程度。协助--是否提供了额外帮助。

效率指标包括以下各项:完成时间--用户正确完成每项任务的平均时间。

满意度评测使用 SUMI 问卷。

以下为此次测试的部分数据结果。

数据分析结果汇总:

	独立完成率%	协助完成率%	完成时间(分钟)	帮助次数
平均值	77.25%	95.375%	37'55"98(37.93)	5.75
标准差	10.02%	8.35%	6.98	1.09
标准偏差	5.01%	4.18%	3.49	0.54
最小值	63.5%	97.5%	29'24"00(29.4)	4
最大值	90%	91%	48'40"51(48.67)	7

这一 CIF 测试的结果和报告基本上能对产品的可用性质量给予一个客观定量的评价,可以用来对同类产品进行比较,在测试中也的确发现产品的某些方面存在着可用性问题,在 CIF 报告的使用过程中,我们总结了这样几点体会:

1) 可用性测试的核心任务是测试的设计包括任务的设计和测试对象的选择。其中测试任务的设计既要

具有代表性又要尽可能全面。测试对象的选择也要有足够的代表性,使其反映不同类型用户的分布。

2)测试环境的安排要尽量使测试对象轻松自然。在有的测试中,由于条件所限,测试管理人员对测试数据进行现场人工记录,而未使用低干扰性的工具如摄像机等,这可能会使测试对象感到紧张,影响测试结果的客观性。

3)对比性测试可使可用性测试效果更加显著。孤立的一种产品的测试结果容易使人感到缺乏参照,不能形成清晰的概念。而同类产品的对比性测试会清楚体现出不同产品在可用性质量方面的优劣,使得测试结果更有意义。

4)CIF测试主要是对成品的可用性质量进行评价,因此测试采取无协助的方式,但在测试中仍发现一些可以改进的软件可用性问题。因此也可以基于这种方法由测试人员进行引导式的协助性测试来发现问题,帮助改进产品设计。

总结 目前,IUSR计划正处在CIF测试在工业界的两年试用阶段的第二年,美国国家标准技术局以及主要由工业界巨头组成的IUSR协作体正在积极推进它的完善和应用,欧盟的UsabilityNet计划以及日本的有关政府和民间机构也将CIF测试列为一项重要的推广内容。据最新消息,CIF测试的V2.0版正在

修订当中,将在近期内做为美国国家信息技术委员会标准(NCTIS)颁布,并进一步成为ISO国际标准。鉴于CIF测试是一项主要由IT工业界自发形成的标准,并且由IT工业发达的国家来推动,可以相信它在今后将为工业界所广泛采用,并带来产品可用性质量的普遍提高。这值得我国的学术界和工业界及早关注并进行相应的准备。作为欧盟可用性中国支持中心(Chinese Support Center for EU UsabilityNet),我们正在与美国国家标准技术局合作在国内工业界进行CIF测试的试点和推广工作。

致谢 本文工作得到了欧盟委员会第五框架研究开发计划和中国政府科学技术部国际科技合作计划的支持。

参考文献

- 1 Bevan N. Tutorial 11--Industry Standard Usability Tests. Interact'99 Conference, Edinburgh, 1999
- 2 Nilsen J. Usability Engineering. Academic Press, 1993
- 3 Blanchard H E. Standards for Usability Testing ACM SIGCHI Bulletin, 1998, 30(3)
- 4 Bevan N. Common Industry Format Usability Tests. In: Proc. Usability Professionals Association Conf. Scottsdale, Arizona, 1996. 5
- 5 Engler D R. The Design and Implementation of a Prototype Exokernel Operating System. [MA Thesis]. MIT, Oct. 1998
- 6 Engler D R. The Exokernel Operating System Architecture. [PhD Thesis]. MIT, Oct. 1998
- 7 Ford B, et al. Microkernel Meet Recursive Virtual Machines. In: Proceedings of the Second Symposium on Operating Systems Design and Implementation, Oct. 1996. 137~152
- 8 Custer H. Inside Windows NT. Microsoft Press, Redmond, WA, 1993 (见:程渝荣译,《Windows NT 技术内幕》,清华大学出版社,1993)
- 9 陈莉君,编著. Linux 操作系统内核分析. 人民邮电出版社, 2000
- 10 Tanenbaum A S, Woddhull A S. Operating Systems: Design and Implementation. Second Edition. Prentice Hall, Englewood Cliffs, NJ, 1997
- 11 Banerji A, Cohn D L. Protected Shared Libraries. A New Approach to Application-Extensible Operating Systems. [Technical Report TR-94-35]. Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, 1994
- 12 Seltzer M, et al. An Introduction to the Architecture of the VINO Kernel. In: VINO: The 1994 Fall Harvest. [Technical Report TR-34-94]. Harvard University, 1994
- 13 Cheriton D, Duda K. A Caching Model of Operating System Kernel Functionality. In: Proc. of the First Symposium on Operating Systems Design and Implementation, 1994. 179~193
- 14 Probert D, Bruno J. Building Fundamentally Extensible Application-Specific Operating Systems in SPACE. [Technical Report TRCS95-06]. Computer Science Department, UC Santa Barbara, Mar. 1995
- 15 Roscoe T. The Structure of a Multi-Service Operating System. [PhD Thesis]. Queens' College, University of Cambridge, Apr. 1995
- 16 Campbell R H, Tan See-Mong. μ Chioce: An Object-Oriented Multimedia Operating System. In: Proceedings of the Fifth Workshop on Hot Topics in Operating Systems, Orcas Island, Washington, May 1995
- 17 Veitch A C, Hutchinson N C. Kea---A Dynamically Extensible and Configurable Operating System Kernel. In: Proc. of the Third Conf. on Configurable Distributed Systems, 1996

(上接第121页)