

· 技术与方法 ·

文章编号:1006-8309(2000)03-0054-04

诊查型用户界面可用性评价方法(IM)(上)

——简介与评价

吴昌旭,张侃

(中国科学院心理研究所,北京 100101)

TB18

TP311.5

摘要:在软件设计早期,运用诊查型方法(Inspection Methods, IM)对用户界面可用性进行及时的评价,有助于提高其可用性。文章回顾了近年来的用户界面可用性评价方法的7种主要的诊查型评价方法,对各个方法进行了简要介绍与评价,力图使界面设计与评价人员能够在实际中加以运用,并注意它们各自的局限。

关键词:人机界面;用户界面;可用性评价;诊查型方法 IM

中图分类号:TB18;TP31 **文献标识码:**A

1 引言

“用户界面的技师已成为软件质量的一个重要维度,它显著地影响着用户使用系统的满意感、操作绩效,乃至整个信息系统的效率与成功率”。^[1]“从市场角度看,具有友好的用户界面的软件具有很强的竞争力”。^[2]另一方面,“用户界面在系统中所占的比重越来越大,有的可达设计总量的60%~70%”。^[3]因此,如何评价和改进用户界面,已成为软件设计、评测人员与用户共同关心的问题。

1.1 用户界面的可用评价及其在我国现状

“可用性(usability)是用户界面最重要、最基本的目标”。^[3]可用性评价主要是对特定用户界面的易学性、易用性、操作绩效、操作满意感,以及用户操作的记忆负荷等因素进行评价,发现用户界面存在的可用性,进而重新改进原界面(还包括帮助系统、用户手册等)的过程(ISO9241-11^[5,6,7])。

可用性评价不仅能为用户减少学习时间,提高工作效率与满意度,还能帮助软件系统开发者突出产品特点,增加软件销售量,减少软件开发与支持费用。^[8]可用性评价可以使开发者获得较高的投入产出比率^[6]。

当前在我国,用户界面的可用性评价,一是通

过 α 测试或 β 测试,即直接发行软件测试版,收集用户反馈意见;二是采用一些直觉经验性的评价方法,如邀请一些用户使用某一系统观察用户行为或听取其意见等。^[10]但是,这些方法缺乏理论根据,而且都必须在软件系统已基本成型的基础上进行,还需要大量实际用户的参与。

1.2 诊查型评价方法及其优势

1.2.1 诊查型评价方法 综合目前各种用户界面可用性方法,^[5,8,9]它们主要可分为三大类:①诊查型评价法(Inspection Methods, IM);②以理论模型为基础的方法(theory-based evaluation methods);③以用户为基础的方法(可用性测试)(user-based evaluation methods)。其中,诊查型评价方法主要是在设计早期,在几乎不直接考察真实用户与界面交互的情况下,由评价人员凭借某些系统的评价方法来诊断、检查及预测将来用户与真实系统界面交互过程(包括学习、使用界面)中可能出现的问题,及时修改原设计。^[11]而后两大类方法,分别借助一定的数学心理模型(如GMOS),或较多真实用户与成型人机系统的实际交互来预测与发现可用性问题。

1.2.2 诊查型评价方法的优势 用户界面的设计方案与修改费用,随着软件开发的深入,呈现出下列关系(见图1)。因此,如果能在设计早期便

能运用 IM,对界面进行及时的修改,则可以大大降低设计与修改时间和费用。此外,与后两类方

法比较,IM 无需昂贵的试验设备与复杂的理论计算,是比较廉价和可靠的。^[12]

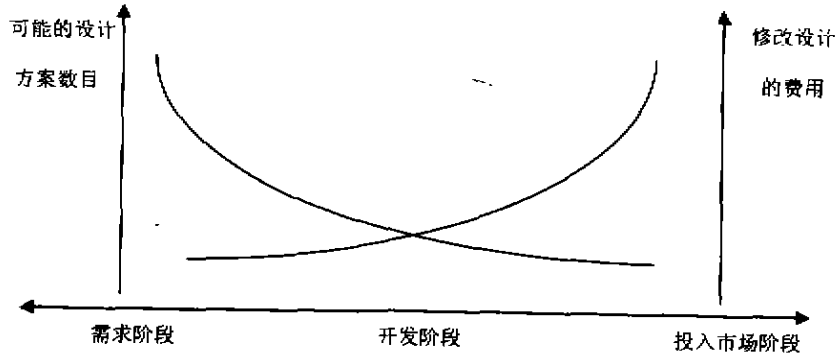


图1 设计方案、修改费用与软件开发阶段的关系^[8]

2 主要的诊断型方法及其评价

我们对现有 IM 做了以下分类:

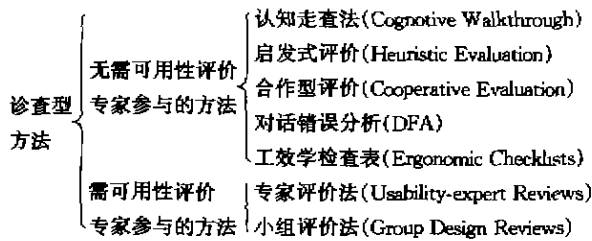


图2 诊断型方法的分类

2.1 认知走查法

该方法以“通过探索进行学习”理论及“现代问题解决理论”为基础^[11],集中评价界面的可用性的一个维度:易学性^[13]。

该方法首先确定任务的正确操作序列,设计设应的界面(可以用纸笔法模拟);然后评价者在这些界面雏形上按上述操作序列步步推进,同时对照下列4个环节,从用户的背景经验,任务目标出发,判断界面是否帮助用户顺利完成对界面的“探索”。从这4个环节出发,修改原界面(包括帮助系统与用户手册)。这4个环节是:①用户头脑中是否有完成任务的某一预测的目标?②用户是否注意到界面提供了达到这一目标的操作?③用户是否将这一界面提供的操作与所需完成的目标相联系?④如果正确操作得以执行,用户是否理解界面的反馈?

认知走查法的优点在于:它尤其适合在软件开发的初期应用,因为它能找出可用性问题的原因,因而能有的放矢地修改原界面;另外,该方法可为无认知科学知识的软件开发应用^[13,14],并且增加评价人数可以提高方法的使用效果。^[12]

但是,由于该方法只考虑界面的易学性,因而发现的可用性问题数量有限(Lewis 发现该方法的问题数量只有可用性测试方法发现问题数的40%^[14]);另外它不能提供整体分析信息。这一缺点可以通过再参照可用性评价原则等其它方法的组合运用加以弥补。

2.2 启发式评价法

这一方法由 Nielsen & Molich 首先提出^[6],主要是由多位评价者通过对照一些可用性评价原则来独立评价一个用户界面的可用性^[7]。

该方法让评价者考察界面的画面情景(Scene)(可用草图或快速编程工具建立功能模拟)两次,第一次集中考察其信息流(系统对用户的输出、用户对系统的输入),第二次考察对话元素(各种对话框、图标、界面用词等);之后,让评价者对照一套评价原则进行评价,这些原则在各种资料中均有详细论述^[3,5]。最后,对各评价者加以组合,讨论界面潜在的可用性问题及解决措施,并对问题的严重性按它们出现频繁程度等因素作一排序。

这一方法的优点在于:廉价与迅速。只需少量有经验的评价者便能发现大量可用性问题,并提出修改的建议,Neilsen 发现,使用5个评价者便能发现75%的可用性问题^[7]。

这一方法的缺点:①评价结果的有效性随着评价者经验减少而降低^[11];②方法中运用的评价原则过于笼统,实际操作对具体界面的评价时有一定难度。为此,我们建议:

①在评价时进一步运用工效学检查表(详见本文2.5),能够降低对评价者经验的要求;或者让缺少评价经验的人员进行一定的评价训练及交

流讨论^[11,12],或者适当增加评价者人数来提高评价水平^[12];②如果条件允许,可将启发式评价与可用性测试结合。

2.3 合作型评价

这一方法通过评价者与真实用户共同合作,来发现可用性问题和提出解决建议^[15]。整个可用性评价在一个轻松、非正式的气氛中进行^[5]。

首先选择极少量有代表性的用户与交互任务,让用户以评价合作者的身份参与评价。在评价过程中,评价者让用户在操作界面的同时进行大声思维(thinking aloud),评价者可以直接对用户的操作行为询问原因,用户也可以向评价者询问任务过程中遇到的问题。整个过程可进行录像,便于数据分析。

该方法的优点,一是有较高的生态效度,也可用于成型产品的改进^[16],二是用户及时参与初期设计,在一定程度上降低了评价者的经验要求。

该方法的缺点在于:由于只采用了极少量的真实用户,这些用户的经验水平直接限制了发现问题的数量与质量;另外,在设计初期,实现能与用户进行交互的界面原型可能有一定困难,但前者可以通过提高用户特性的典型性解决,后者可以通过 VB、DEPHI 等可视化编程工快速建立界

面原型加以克服。

2.4 对话错误分析法

该方法适用于只有少量交互步骤的简单用户界面^[17]。首先确定完成任务目标的操作序列,对每一步操作推论其发生错误的直接原因(主要是用户的外显行为)及其背后的主要原因(一般主要分析界面的设计问题),再通过估计错误的发生频率与可逆性,评估问题的严重性,找出相应的修改措施。

该方法比 PHEA、TAFEI^[18]等提出错误预测方法更为简便易用^[17];但它对错误的分类过于笼统,不利于错误原因的寻找^[18]。这一缺点可以通过运用大声思维等方法进行弥补;它的另一个缺点是,对错误的推断仍会在一定程度上受到评价人员的影响,这一点可以借鉴 PHEA 方法中的错误归类与分析方法加以解决。

2.5 工效学检查表

这一方法实际上是对上一些可用性评价原则的具体化陈述,形成由评价人员填写的界面可用性评价检查表(如现已有 Kappold & Flanagan, Smith & mosier, Woodson^[19])。例如,对界面的一致性的检查,如图 3。

检查项目	十分满意	较为满意	一般	较不满意	十分不满意
整个系统中表示相同功能的图标一致吗?		√			
整个系统的缩写	√				

综上所述,您对该软件系统的界面评价如何:

十分满意	较为满意	一般	较不满意	十分不满意
	√			

图 3 可用性评价检查表举例

这样,该方法一方面可以针对可用性的某一环节进行改进;另一方面,也可以汇总对软件系统界面可用性的各个方面,获得对界面可用性的整体评价。

它的最大优点就在于有较强的可操作性,并且项目能够覆盖整个界面可用性的各个方面。但是,它又缺乏一定的运用灵活性,Neville Stanto 等人称之为情景敏感性(Situation Sensitivity),某些项目只适用于某类界面,另一些并不适用于该类界面,因此只能根据具体某类界面编制特定的表,提高了可用性评价工作的复杂性。

2.6 专家评价法

这一方法类似于启发式评价,唯一的区别在于必须由专家(最好是兼有系统与用户经验的可用性专家)充当评价者的角色,它可以为专家提供一套指导原则,或不使用确定的指导原则^[12];但尚未有关于这两种方式评价的比较性研究。这一方法的优点类似启发式评价,专家的评价是高效并且实用。

2.7 小组评价法

小组评价法是由多个评价人员共同合作,共同同时对用户界面进行评价,以发现潜在的可用性问题;小组的组成通常是由系统开发者、人机专家(至少一人)等组成^[6]。小组评价法的一个具

体方法是多元走查法^[12]。即让预期的用户、产品开发者及人机专家组成小组,运用类似启发式评价的评价方法;由于小组成员的多元化,可以从多方面考察用户界面的可用性,提高评价效果。此外,它还可以用来评价界面是否达到了某些可用性标准^[14]。

小组评价法与专家评价法一样,能够比无需可用性专家的方法发现更多的可用性问题,但是由于实际应用中,一般公司难以得到这方面专家的帮助,特别是在我国(欧洲、亚洲地区人机专家介入软件公司的人数很少^[12]),因此这两种方法在实际推广中可能会受到阻力。

参 考 文 献

- [1] Mohamed Khalifa. Computer-assisted evaluation of interface designs[J]. The DATA BASE for Advances in Information Systems-Winter. 1998, 29(1):66-81.
- [2] 郑人杰. 实用软件工程[M]. 北京:清华大学出版社,1996.38-144.
- [3] 杨文龙. 软件工程[M]. 北京:电子工业出版社,1997.446,456-458.
- [4] Brooks J. Adding Value to Usability Testing in Usability Inspection Methods[M]. Neilsen J (eds), Wiley NY and Son, 1994. 258-265.
- [5] Rogers Y. Human Computer Interaction[M]. Addison-wesley Publishing company. Worryham, England, 1994. 603-708.
- [6] Karat J. User-centered Software Evaluation Methods in Handbook of Human Computer Interaction(2nd) [M]. Helander M. (eds), Elsevier Science B.V, 1997. 689-704.
- [7] Neilsen J. Usability Engineering [M]. Academic Press; Boston, 1993. 16-21, 71-267.
- [8] Ehrlich, Rohn. Cost-justification of Usability Engineering : a vendor's perspective[M]. in Cost-justifying Usability Randolph G. Bias (eds). Academic Press, 1994. 83-87.
- [9] Sweeny M. Evaluating user-computer interaction[J]. A Framework, Man-Machine studies 1993, 38: 689-711.
- [10] 电脑报评测实验室. 随心所欲说英语 II 评测报告 [N]. 电脑报, 2000-1 \ 10(68).
- [11] Neilsen J, Mark Heuristic. Evaluation in Usability Inspection Methods[M]. Neilsen J (eds), Wiley NY and Sons, 1994: 1-7, 36-61.
- [12] Robert A, Virzi. Usability Inspection Methods in Handbook of Human Computer Interaction(2nd) [M]. Helander M. (eds), Elsevier Science BV, 1997. 705-714.
- [13] Cathleen Wharton. The Cognitive Walkthrough Method: A Practitioners Guide in Usability Inspection Methods[M] Mack R, Neilsen J (eds), Wiley NY and Son, 1994. 105-140.
- [14] Clayton, Lewis. Cognitive Walkthrough in Handbook of Human Computer Interaction(2nd) [M]. Helander M. (eds), Elsevier Science BV, 1997. 17-732.
- [15] Wight Monk. A cost-effective evaluation method for use by designers[J]. Man-Machine Studies 1991: 35, 891-912.
- [16] Alan JD. Human Computer Interaction[M]. Prentice Hall Europe, 1998. 405-443.
- [17] Iain W, Connell. Error analysis of ticket vending machines: Comparing analytic and empirical data[J]. Ergonomics, 1998, 41(7): 927-961.
- [18] Baber C, Stanton NA. Human error identification techniques applied to public technology: predictions compared with observed use[J]. Applied Ergonomics, 1996, 27(2): 119-131.
- [19] Adelman L. Handbook for Evaluating knowledge-based systems [M]. Kluwer Academic Publisher, 1997. 230-267.

[收稿日期]2000-07-27