

# 数据库检索系统可用性评价指标与实证研究

胡晓青 张建勇

(中国科学院国家科学图书馆, 北京 100190)

**[摘要]** 通过分析数据库检索系统与一般网站间的可用性特征差异, 在传统可用性评价理论的基础上, 提出了适合于数据库检索系统的新型可用性评价指标集: 可理解性、可操作性、信息获取度和信息辨别度。同时采用总结性测试方法对中国科学引文数据库(CSCD)进行可用性测试, 验证可用性评价指标集的适用性。最后分析了测试结果以及产生问题的原因, 提出进一步完善系统可用性的建议。

**[关键词]** 数据库检索系统; 可用性评价指标; 可用性测试; 实例分析

**[分类号]** TB18

## Usability Evaluation Indicators and Empirical Study of Database Retrieval System

Hu Xiaoqing Zhang Jianyong

(National Science Library, Chinese Academy of Sciences, Beijing, 100190)

**[Abstract]** By analyzing the differences in the characteristics of usability between the database retrieval system and general websites, a new set of evaluation indicator of usability was put forward on the basis of the traditional evaluation theory of usability, including: intelligibility; operability; accessing to information; identify information. At the same time, the author tested the usability of Chinese Science Citation Database (CSCD) with the method of summative testing and verified the applicability of the new set of evaluation indicator of usability. Finally, the suggestions of the further improving the usability of database retrieval system were put forward after analyzing testing results and the source of problems.

**[Keywords]** Database Retrieval System; Usability Evaluation Indicators; Usability Testing; Case Study

## 1 前言

可用性是评估用户界面容易使用程度的一种属性, 也是网站生存的必备条件之一。友好的用户界面将大大提高系统的利用频率和使用效率。ISO FDIS 9241-11 把可用性定义为“产品在特定使用情境下被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”。<sup>[1]</sup>也就是说, 一个易于使用的网站可使用户在短时间内清晰了解网页内容、顺利完成界面操作、增加对网站的主观满意度和接受程度。衡量站点的可用性就需要进行可用性测试。根据ISO13407 对测试方法的说明: 在设计中可以邀请用户对即将发布或已经发布的产品以及设计原型进行评估, 并通过评估数据的分析进行迭代式设计直至达到可用性目标。<sup>[2]</sup>目前, 业界普遍采用的网站设计方式是“User-Centered Design (以用户为中心的设计)”, <sup>[3]</sup>可用性测试就是其重要的组成部分之一, 以用户为本的设计过程本身就应该包括对性能和偏好进行评价

的一系列测试。

对于可用性概念的涵盖范围研究者们曾给出过多种解释。1998年，Hartson 提出可用性包含两层含义：有用性和易用性。有用性是指产品能否实现一系列的功能。易用性是指用户与界面的交互效率、易学性以及用户的满意度。虽然 Hartson 对这一概念的界定比较全面，但对于用户的操作过程缺乏进一步的分析。而 Nielsen 在其著作《可用性工程》中所提出的观点则弥补了这一缺陷，并成为网站可用性测试的重要依据。他认为可用性的定义包含以下五个要素<sup>[4]</sup>：

- 易学性：初次接触这个设计时，用户完成基本任务的难易程度。包括：界面设计是否符合用户习惯，系统操作帮助和提示功能设计是否完善等。

- 交互效率：用户完成操作任务的速度。包括：界面导航的设置，用户完成操作任务的成功率、所需时间和点击量。

- 可记忆性：在一段时间没有使用之后再次使用该设计，用户重新熟练操作的难易程度。

- 出错频率：用户操作错误的数量，错误的严重性，错误的易恢复性。

- 用户主观满意度：用户对设计是否满意。

Nielsen 指出产品在每个要素上都达到了很好的水平，才具有高可用性。同时，网站可用性的优劣也是由这五个要素决定的。

## 2 数据库检索系统的可用性评价指标与测试方法

目前，许多数字图书馆都为自己的特色资源建立了各类型的数据库检索系统。在不断实现数字化、网络化资源共享的同时，提高用户使用图书馆电子资源的频率和效率也应是图书馆积极进取的目标。在建设检索系统时，如果单纯去追求先进的自动化技术，而不从可用性的角度去考虑用户在实际使用上的需求、方便和效率，用户对系统的满意度不一定会随着先进技术的提高而提高，甚至反而会适得其反，达不到系统设计预期的效果。因此，用户在可用性方面的满意度是检索系统存在的核心价值之一。

尽管网站可用性评价的方法和指标为数据库检索系统的评价提供了丰富的借鉴与启发，但是检索系统自身所具有的特点决定了它不能够照搬一般网站的可用性评价模式。数据库检索系统的可用性评价应在传统可用性评价理论的基础上建立起一套自己的评价体系，以弥补原有评价指标在测试结果分析上的不适应以及评价角度的不全面。

### 2.1 数据库检索系统与一般网站间的可用性特征差异

对于任何开放的系统和网站来说，吸引更多的用户是它们获得发展的重要手段，甚至是目标之一。这也使检索系统界面设计的某些方面需要用网站的可用性要求进行规范，如：良好的视觉感受，强大的系统导航功能，用户使用效率和系统操作的记忆性等。但是数据库检索系统与网站间存在的差异也是可用性测试、评价中所不能忽视的问题，因为这些问题将直接影响到测试指标的设置和测试结果的分析。本文从用户群、用户检索习惯、操作任务、信息辨别、信息获取、界面设计等方面比较了数据库检索系统和一般网站间的可用性特征差异，见表 1。

角度	数据库检索系统	一般网站
用户群	用户群相对固定，用户的信息素养较高，具有一定相关专业的知识。	用户群分散，信息素养层次不一。
用户检索习惯	大部分用户的检索经验主要来自于网络搜索引擎，数据库检索需要用户具有相关的操作知识。	用户的操作经验来自于一些较知名的门户网站，在操作知识方面一般对用户不作要求。
操作任务	用户操作目的性明确，重复操作次数多。	用户操作行为随意性大。
信息辨别	检索系统信息准确度较高，会根据用户的检索条件显示相应的检索结果。	所获得的信息内容较为分散，准确度不高，一般通过分类法组织信息，用户不用输入具体的检索条件。
信息获取	资源下载是用户获取信息的主要手段。	网页浏览是用户获取信息的主要手段
界面设计	界面设计专业性强，检索界面模式较为固定，多为符合用户使用习惯的设计，界面用语专业化。	界面设计强调特色化、新颖性以吸引更多网民的浏览。在界面使用习惯上追求大众化。

表 1: 数据库检索系统与一般网站间的特征差异

## 2.2 数据库检索系统可用性评价指标的建立

数据库检索系统和一般网站间可用性特征差异表明：传统网站可用性测试的五项指标并不能完全满足检索系统测试的需求，需要建立新的评价指标以提高检索系统可用性评价的准确度。本文基于数据库检索系统的可用性特征，从用户检索理解、具体操作、信息获取和信息辨别等 4 个角度出发建立起可理解性、可操作性、信息获取度、信息辨别度 4 项评价指标。由于原有可用性测试中易学性指标的功能已被可理解性和可操作性所覆盖，可记忆性需要进行迭代测试，所以在本次测试中将忽略这两个指标。因此，测试将采用可理解性、可操作性、信息获取度、信息辨别度、出错频率、交互效率和用户主观满意度等七个指标。前六个指标的数据来源于仪器记录和测试员观察，用户主观满意度指标则通过用户满意度调查问卷所收集到的定量数据统计得出。

### 2.2.1 可理解性

对于非图情专业的用户来说，使用某些专业性很强的检索系统是十分困难的。因为他们可能不理解检索系统中某些功能性名词的含义，如：开放链接、耦合度等。如果没有相应的提示或帮助，用户只能通过尝试性的点击来了解相应功能，这将使他们容易在检索系统中迷失方向，从而大大降低用户的检索效率。因此，提高用户对相关功能性名词理解的准确度是数据库检索系统可用性设计的主要任务之一。

### 2.2.2 可操作性

可操作性是指用户可顺利完成信息检索所需的各项功能的操作。由于大部分用户的检索经验来自于网络搜索引擎的使用，对于专业数据库检索知识储备的不足将使他们对检索结果

的深层次查询、排序和筛选等方面的操作感到困难。同时由于用户会不断更换检索条件以优化检索结果，所以在数据库检索中重复操作的机率会比一般网站大，因此，可操作性的优劣将直接影响到用户的检索效率。如何让用户轻松使用检索系统中的各项功能，是系统界面设计人员需要着重解决的问题。

### 2.2.3 信息辨别度

帮助用户从众多的检索结果中快速发现与自己检索条件相关的数据记录是提高检索系统使用效率的关键。在某种程度上可以减少用户进行检索结果筛选、二次检索和重新检索的机率。除了运用检索技术提高检索结果的准确度之外，许多数据库检索系统都在系统界面的设计上为此做出了许多尝试，如：根据用户的检索条件在检索结果中突出显示某些符合用户要求的关键词，使用户能够对资源的准确性快速做出判断，从而减少用户重复操作的次数，提高系统的使用效率。

### 2.2.4 信息获取度

与一般网站不同，在数据库检索系统中用户需要通过下载数据的方式获取更为丰富的信息。检索系统下载功能的设计是用户评价系统满意度的关键因素，系统应保证用户能轻松找到各类资源下载的连接并获取到相关数据。因此，在设计系统界面时应着重考虑用户信息获取功能的易用性问题。

## 2.3 数据库检索系统的可用性测试方法

测试采用总结性测试方法，即：邀请一定数量的真实和潜在用户使用该系统完成一些典型任务，并在任务完成后填写调查问卷。由测试员收集诸如任务成功率、用户主观满意度等方面的定量数据，并通过任务完成中的观察和测试之后与用户的交流获得定性数据，进而通过对这些数据的统计分析获知该系统的总体可用性状况。测试要求用户采用发声思维法完成任务<sup>[5]</sup>，发声思维法强调在操作过程中让用户积极思考并说出自己的思维过程。

根据数据库检索系统测试指标集：可理解性、可操作性、信息获取度、信息辨别度、出错频率、交互效率、用户主观满意度等七个指标，测试员采集用户在测试过程中产生的定量数据和定性数据。定量数据来自于对用户操作状况的标准测量，包括：用户完成任务的成功率、所需时间、点击量和发生操作错误的次数；用户在操作中使用系统导航、帮助和提示功能的次数以及通过用户填写调查问卷所得到的定量数据。定性数据是指在系统操作时因用户情感因素、心理因素、文化背景和个人经验所造成的不确定的行为表现，这些需通过测试员的观察进行记录。它包括：用户在完成任务过程中所表现出的困惑、不解、迟疑、焦急；用户解决操作错误的方式；用户完成任务所使用的路径以及用户调查问卷中的选择和开放性问题。

## 2.4 测试用户的选择

根据 Tom Landauer 关于测试人员数量和测试效果之间相关度的公式： $N(1-(1-L)^n)$ （假设一个可用性测试的测试人员数量为  $n$ ， $N$  为所有可用性测试发现的问题总数， $L$  是单个测试人员的问题发现率），可证明：在可承受的测试成本范围内，使用 5 名测试用户，即能发现 85% 的可用性问题，达到很好的测试效果。<sup>[6]</sup>

## 2.5 用户满意度调查问卷的计算方法

为了避免可用性本身的模糊性和用户体验概念的不确定性，用户满意度调查问卷将采用模糊综合评价和层次分析法（AHP）对收集到的数据进行处理。<sup>[7]</sup>基本步骤是：确立评价指标体系；设定评价集并确立评价因素的隶属度函数；确定评价因素的权重；进行模糊综合评判，用层次分析法计算整个系统的评价得分。

问卷设计了用户评分、选择和开放性三种问题类型。选择和开放性用于收集定性数据，而用户评分问题则通过 7 点标度问卷的形式收集定量数据。它将评价标准分为 5 个等级{很差，较差，一般，较好，很好}，阈值分别为：1，2，3.5，5.5，6.5，7。所涉及到的评价内容包括：操作绩效、信息获取过程质量和界面质量，它们的权重分别为 0.490，0.312，0.198。

按照模糊综合评价方法，计算评价内容和评价标准的隶属程度。计算后两者的模糊关系矩阵 R 如下：

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{bmatrix}$$

式中  $r_{ij}$  表示评价内容属于评价标准的隶属程度。然后可以根据下面的公式计算综合评价：

$$B = A \circ R = (b_1, b_2, b_3 \dots b_n) \quad (A \text{ 表示权重指标})$$

$$b_j = \sum_{i=1}^n a_i r_{ij} = \min \left\{ 1, \sum_{i=1}^n a_i r_{ij} \right\}$$

当对多层内容进行评价时，同层子目标的评语集  $B_{i,j}$ （ $i$  表示第  $i$  层， $j$  表示同一层中第  $j$  个子目标）又可以构成新的模糊矩阵  $R_i$ 。这样就可以迭代计算出整个系统的评价得分。

## 3 CSCD 系统可用性测试

中国科学引文数据库（以下简称：CSCD，<http://www.sciencechina.cn>）<sup>[8]</sup>2008 年 8 月全新改版，除了在系统结构、数据质量和检索功能等方面有所突破之外，对检索界面也进行了全新的设计。本文以 CSCD 为测试对象，选择新构建的可用性指标集测试该系统，对测试结果在可用性方面存在的缺陷进行分析，并面对面访谈用户了解其对该系统的使用意见。为进一步提高 CSCD 系统的可用性提供依据，节约后期完善设计的成本。

### 3.1 测试工作的前期准备

采用选择代表性用户的原则<sup>[9]</sup>，本次测试将邀请 5 名不同专业的研究生参与测试。虽然他们都是初次使用 CSCD 系统，但都拥有一定的数据库检索经验和相关知识，并且在测试前，允许他们对系统进行 5 分钟的试用，这样既可保证测试结果的公正，也可与其他检索系统进行可用性对比。

根据 CSCD 系统的特点，主要测试系统的五个功能模块，包括：首页、简单检索界面、高级检索界面、检索结果界面和详细信息界面。在每个功能模块中再设计若干任务，任务覆盖系统所有的功能。为了保证测试结果可反映出共性问题，每一位测试用户所需完成的任务

都是大体相同的，但具体到检索条件（如：作者、刊名、题名等内容）可能会根据用户专业背景的不同有所区别。

### 3.2 CSCD系统可用性测试结果及问题分析

采用模糊综合评价和层次分析法（AHP）对通过5位用户进行测试所得到的定量数据进行分析、计算，用户满意度为5.63，属于较好水平。这说明用户对整个系统比较满意，但还存在着一些亟待改进的问题。同时，通过观察所得到的定性数据表明：5位用户平均任务操作成功率为90.12%，平均操作错误量为8.4次，80%的用户表示今后会使用该系统并将其推荐给自己的同学。

通过总结5位用户在测试过程中的行为规律可发现一些容易被系统界面设计人员所忽略的问题。统计这些问题发生的频率，分析其产生的原因，提出相应的修改建议，可大幅提高用户对于系统的使用效率。

#### 3.2.1 理解性问题

在系统操作过程中用户对链接含义的不理解严重影响其使用效率，主要表现在不理解专业名词的含义和链接功能定义不明确这两个方面。

##### 3.2.1.1 专业名词理解难度大

对于图书馆自建的检索系统来说，以专业名词设置的链接大多不能被非图情专业的用户所理解。在本次测试中，所有用户均不理解“开放链接”的含义，其中有3位用户不理解何为“引证文献”。由于“开放链接”与全文下载功能密切相关，因此所有参与测试的用户均未下载到全文，从而大大影响了系统的信息获取度和用户满意度。通过测试还发现用户几乎没有在使用系统前阅读帮助文档的习惯，所以仅在“帮助”中解释专业名词的含义是远远不够的，更换通俗易懂的词语或在链接上添加相关词汇含义的提示才是解决该问题的根本途径。

##### 3.2.1.2 链接功能定义不明确

在详细信息界面中，共有3个名为“下载”且图标相同的链接，然而其功能却各不相同。其中两个为参考文献目次的下载，另一个则用于下载该篇文章的详细信息。测试中，由于链接定义不明确，用户大多不能准确分辨出这些链接的功能，只能通过尝试性的点击进行操作，甚至有4名用户把它们当作了全文下载链接，这样无形中延长了操作时间和下载的准确度。同样，名为“E-mail”的链接也存在类似的问题。所以，功能不同的链接应在名称和图标上予以区别，以防用户混淆。

#### 3.2.2 操作性问题

测试中在系统的可操作性方面主要发现了两个问题：一是操作方式不符合用户习惯；二是有些链接不易被用户发现。

##### 3.2.2.1 操作方式不符合用户习惯

检索结果筛选功能是提高用户检索效率，减少重复检索次数的主要手段。测试发现：虽然“检索结果分布”的筛选更为直观且不需要用户输入任何检索条件，但与二次检索功能相比，大部分用户则更为喜欢后者。通过观察用户行为并与他们交流后发现：用户并不熟悉也不习惯使用前者进行筛选，而且在观看分布结果时，点击一次“更多”只能多显示出三条结

果，这将给用户的查找带来很大不便。高级检索功能也存在类似的情况，用户反映（尤其是非计算机类专业的用户）：他们并不习惯生成高级检索式这样的检索方式。因此，在系统设计前应广泛调研用户的检索习惯，以减少用户使用系统的障碍。

### 3.2.2.2 链接设置

CSCD系统链接设置的问题主要有：重点链接设置不突出；链接设置位置不准确；链接设置隐蔽等。如：所有测试用户都没有点击“开放链接”的原因除了不理解其含义外，更重要的因素是其设置不突出，甚至没有吸引到用户的尝试性点击。获取参考文献详细信息的链接设置在了该篇文献的来源期刊上，让用户误以为其是查找该期刊的链接。检索结果界面的“被引频次排序”链接没有经过任何特殊设计，如果没有光标移过，用户将很难发现这个链接。在详细信息界面中，使用“相关文献”功能和直接点击查询词所得到的检索结果是一致的，但是通过测试发现有4名用户更加偏爱直接点击查询词进行检索。其实对于使用复合检索条件的用户来说“相关文献”功能更为符合他们的需求，但用户反映其设计过于隐蔽，很难被发现。诸如此类的链接设置问题在很大程度上影响了系统功能的被使用率。因此，如果没有合理的链接设计，再先进的检索功能也是无法发挥其预期作用的。

### 3.2.3 信息辨别问题

提高检索结果辨别度是提高检索系统使用效率的关键。测试发现：用户习惯于通过直接点击检索结果来了解文章的详细信息，检索结果中应突出显示与用户检索条件相关的内容。目前，CSCD系统还不具备这两项功能，相应的界面修改将会提高用户对检索结果进行人工筛选的效率。

## 4 结语

CSCD系统的可用性测试表明：用户针对系统提出的所有问题均在这七个指标的涵盖范围之内。因此，数据库检索系统的可用性测试指标是全面、完整的。只有在这七个指标都表现良好的情况下，检索系统才具有较高的可用性。对于系统设计人员来说，在前期设计中注重遵循这七个指标所描述的原则，将会大大提高系统原型的可用性，减少后期设计成本。

对于图书馆自建的数据库检索系统来说，提高系统可用性应注重以下几个方面：用专业名词命名链接要慎重，对于易混淆的链接要进行明确地设置；系统操作设计简洁，符合用户使用习惯；突出重点功能链接的设置，设计上尽量多参考知名的搜索引擎和数据库检索系统；重视检索结果的呈现和筛选功能的设计；保证用户轻松下载获取检索结果。可用性的提高可增加用户对图书馆检索系统使用的忠诚度，增强图书馆自建系统的对外竞争力，进而使图书馆的馆藏资源得到充分的利用。

[1] ISO 9241-11 Human- centered design processes for interactive systems[S].International Standards Organization, 1999.

[2] ISO 13407 Human- centered design processes for interactive systems[S].International Standards Organization, 1999.

[3] Vredenburg K, Isensee S, Righi C. User - centered design: an integrated approach [M].New Jersey: Prentice Hall, 2001.

[4] 可用性简介[EB/OL].[2008-09-03].[http://www.upachina.org/userfriendly2006/news/02\\_4.pdf](http://www.upachina.org/userfriendly2006/news/02_4.pdf)

[5] 甘利人.可用性测试方法在 IA 研究中的应用[J].情报理论与实践,2004,04:415-417,443

[6] Jakob Nielsen. Usability Testing With 5 Users[EB/OL].[2008-09-15].

---

<http://www.useit.com/alertbox/20000319.html>

<sup>[7]</sup> 周荣刚.IT 产品用户体验质量的模糊综合评价研究[J].计算机工程与应用,2007,43(31):102-105

<sup>[8]</sup> ScienceChina 中国科学文献服务系统  
[EB/OL].[2008-09-15].[http://sdb.cSDL.ac.cn/index\\_more1.jsp](http://sdb.cSDL.ac.cn/index_more1.jsp)

<sup>[9]</sup> Jakob Nielsen. Usability 101: Definition and Fundamentals - What, Why, How[EB/OL].[2008-09-15].<http://www.useit.com/alertbox/20030825.html>