

Prüfung von Software auf die Grundsätze ergonomischer Dialoggestaltung

Karl-Heinz Rödiger, Berlin und Ulrich Piepenburg, Hamburg

Zusammenfassung

Mit jeder DIN-Norm ist die Frage verknüpft, wie deren Einhaltung geprüft werden kann. In der DIN 66 234, Teil 8 "Grundsätze ergonomischer Dialoggestaltung" ist angeführt, daß es "derzeit noch nicht möglich (ist), die Erfüllung einzelner ... Leitsätze objektiv zu überprüfen, da geeignete Überprüfungsverfahren noch nicht bekannt sind". Ziel dieses Beitrags ist es, Mindestanforderungen für die Konformitätsprüfung vorzustellen. Zu deren Begründung sind Methoden zur Evaluierung in der Literatur ebenso untersucht worden wie die Praxis anerkannter Softwareprüfstellen.

1 Einleitung

Im Rahmen des Landesprogramms "Mensch und Technik - Sozialverträgliche Technikgestaltung" der Landesregierung Nordrhein-Westfalen erarbeiteten wir ein Gutachten, in dem untersucht wurde,

- a) welche Evaluationsverfahren für welche Gestaltungsdimensionen der Mensch-Rechner-Interaktion vorgeschlagen werden;
- b) wieweit die DIN 66 234, Teil 8 in die Prüfpraxis anerkannter Softwareprüfstellen integriert ist;
- c) welche kriterienabhängigen Mindestanforderungen an die Prüfung von Software auf Konformität mit dieser DIN-Norm benannt werden können (Piepenburg und Rödiger 1988).

Im Februar 1988 wurde die DIN-Norm 66 234, Teil 8 "Grundsätze ergonomischer Dialoggestaltung" veröffentlicht. Deren fünf Gestaltungsgrundsätze lauten Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Steuerbarkeit, Erwartungskonformität und Fehlerrobustheit.

Wie bei jeder DIN-Norm stellt sich auch bei dieser die Frage, wie sich deren Einhaltung prüfen läßt. Im Bereich technischer Normen ist dies relativ exakt zu beantworten; bei ergonomischen Normen finden technische Prüfverfahren jedoch ihre Grenzen. Die Frage nach Evaluierungsmöglichkeiten erhebt sich bei der DIN 66 234, Teil 8 noch entschiedener, da sie mit ihren Gestaltungsgrundsätzen Forderungen an Dialogsoftware erhebt, die in jedem Anwendungsfall der Operationalisierung bedürfen.

2 Methoden der Evaluation von Benutzungsfreundlichkeit

Augenblicklich werden Bemühungen verstärkt, präskriptiv aus einer theoretischen Analyse heraus Kriterien zu identifizieren, die Benutzungsfreundlichkeit garantieren, wenn sie bei der Systementwicklung berücksichtigt werden. Ausgehend von Aufgabenstellungen, die ausschließlich mit dem Arbeitsmittel Rechner bearbeitet werden sollen, wird in diesen Ansätzen versucht, die sich daraus ergebende kognitive Strukturierungsleistung von Benutzern abzubilden. Vorwiegend werden gerätebezogene Zielhierarchien und Abfolgen von Benutzeraktionen modelliert. Auf Basis solcher Modelle werden Vorhersagen getroffen, wie sich ein realer Benutzer bei einer vorgegebenen Aufgabe verhalten wird. Erstes und oft einziges Vorhersagekriterium sind die Ausführungszeiten.

2.1 Analytische Evaluation

Methoden der analytischen Evaluation sind u.a. GOMS- und das Key-stroke-Level-Model (Card, Moran and Newell 1983), die Cognitive Complexity Theory, CCT (Kieras and Polson 1985) und die Task-Action-Grammar, TAG (Green, Schiele and Payne 1986). Hier ist nicht der Raum, diese Methoden vorzustellen und zu bewerten. Zur ausführlicheren Diskussion verweisen wir auf unser Gutachten.

Eine Bewertung analytischer Modellierungen von Benutzerverhalten sollte entlang der Frage verlaufen, was zu welchem Zweck modelliert wurde. Zweck der Modellierung ist zumeist, Entwicklern von

Softwaresystemen Kriterien an die Hand zu geben, menschengerechte, oft gleichbedeutend mit effizient handhabbare, Systeme zu entwickeln. Um beurteilen zu können, ob dies mit den Modellen einzulösen ist, müssen zumindest drei Aspekte berücksichtigt werden: Die psychologische Basis der Modelle, die Angemessenheit der Darstellungsmittel und die empirischen Befunde über die Modellprüfung an der Realität.

In der Literatur werden diese Modelle oft zitiert und diskutiert, in der Praxis der Softwareentwicklung oder -prüfung spielen sie hingegen kaum eine Rolle. Bei der Einschätzung ihrer prinzipiellen Eignung für prospektive oder korrektive Gestaltungsmaßnahmen scheint uns mehr Skepsis angebracht, als sie dort zum Ausdruck kommt (vgl. Piepenburg und Rödiger 1988, S. 62 ff.).

2.2 Evaluationsschwerpunkte

Bei der Bewertung der Mensch-Rechner-Interaktion bezogen auf Benutzungsfreundlichkeit sind vier Einflußgrößen als Evaluationsschwerpunkte zu berücksichtigen: DV-System, Benutzer, die zu erledigenden Arbeitsaufgaben sowie die organisatorische Einbettung des DV-Systems. Entsprechend kann man die Literatur dazu ausdifferenzieren in Berichte über systemorientierte, personenorientierte, aufgabenorientierte und organisationsorientierte Evaluation. Allerdings erhält jeder Evaluationsschwerpunkt einen unterschiedlichen Stellenwert bei der Untersuchung der Kriterien von Benutzungsfreundlichkeit. Bei allen Evaluationsschwerpunkten bildet das Expertenurteil das am häufigsten angewendete Hilfsmittel, das jedoch den methodischen Nachteil einer nur bedingt nachvollziehbaren Urteilsfindung beinhaltet.

2.2.1 Systemorientierte Evaluation

Anhand systemorientierter Methoden wird die Eignung von Systemen oder Systemkomponenten bezogen auf Benutzungsfreundlichkeit untersucht. Folgende Methoden werden dazu benutzt:

- **Experimentelle "Walkthroughs"**: Personen werden aufgefordert, mit Hilfe eines Systems eine Aufgabe zu lösen bzw. sich in ein System einzuarbeiten. Begleitend wird der Handlungs- bzw. der Denk- und Planungsablauf protokolliert.
- **Gruppenvergleiche**: Versuchspersonen werden unterschiedlichen Systemausprägungen gruppenweise zugeordnet. Anhand vorgegebener Aufgaben wird die Performanz der Gruppen über die Kriterien Bearbeitungszeit, Fehlerhäufigkeit, Ergebnisqualität und Handlungsablauf verglichen. Für die Auswertung können inferenzstatistische Methoden herangezogen werden.

2.2.2 Personenorientierte Evaluation

Bei personenorientierten Evaluationen wird die Eignung von Dialogsystemen oder Systemkomponenten für unterschiedliche Benutzer oder Benutzererfordernisse hinterfragt. Hierzu werden folgende Methoden eingesetzt:

- **Gruppenvergleiche**: Einzelne Benutzungsfaktoren werden im experimentellen Design untersucht. Dazu werden Benutzergruppen, die im untersuchten Kriterium variieren, spezifiziert, die mit einem Dialogsystem Aufgaben bearbeiten; dabei werden Performanzmaße, meist Ausführungszeit, Ergebnisqualität oder Ausführungsverlauf erhoben. Anhand der Performanzmaße werden die Gruppen verglichen.
- **Arbeitsanalysen**: Untersuchungen komplexer psychologischer Konzepte zielen auf die Erfassung von Größen, wie sie in der arbeitspsychologischen Forschung konzeptionalisiert werden. Kriterien zur Bewertung sind Arbeitszufriedenheit, Handlungs- und Entscheidungsspielraum sowie Persönlichkeitsförderlichkeit bzw. Qualifizierungsmöglichkeiten.

2.2.3 Aufgabenorientierte Evaluation

In aufgabenorientierten Evaluationen ist die Eignung von Systemen oder Systemausprägungen bezogen auf das Erreichen eines vorgege-

benen Arbeitsergebnisses Untersuchungsgegenstand. Einziges methodisches Hilfsmittel ist dabei die **Arbeitsanalyse**, die vorwiegend in folgenden Formen eingesetzt wird:

- Entlang einer Standardaufgabe werden Dialogsysteme auf ihre Benutzungsfreundlichkeit geprüft;
- die Arbeitssituation und der Arbeitsablauf werden erfaßt;
- automatisierte Anteile der Arbeit werden kenntlich gemacht;
- mit subjektiven oder bedingungsbezogenen Verfahren werden Belastungs- und Beanspruchungsmomente sowie Arbeitszufriedenheit ermittelt.

2.2.4 Organisationsorientierte Evaluation

Organisationsorientierte Evaluationen zielen auf die Einpassung eines Systems in das arbeitsorganisatorische Umfeld; Untersuchungen dazu sind rar. Als methodische Hilfsmittel sind **Arbeits- und Kommunikationsanalysen** einsetzbar.

Zur Prüfung auf Benutzungsfreundlichkeit steht mithin eine begrenzte Auswahl methodischer Strategien zur Verfügung: der experimentelle Walkthrough, der Gruppenvergleich, die Arbeitsanalyse, die Kommunikationsanalyse und das Expertenurteil. In der folgenden Tabelle sind die Methoden den Evaluationsschwerpunkten zugeordnet dargestellt.

	Walk-through	Gruppenvergleich	Arbeitsanalyse	Kommunikationsanalyse	Expertenurteil
systemorientiert	X	X			X
personenorientiert		X	X		X
aufgabenorientiert	X		X		X
organisationsorientiert			X	X	X

Tab. 1

3 Einschätzung der Prüfpraxis von Softwareprüfstellen

Von einer einheitlichen Praxis kann - wenn überhaupt - nur bei den Prüfungen der Gütegemeinschaft Software (GGS) auf der Basis von DIN V 66 285 "Prüfgrundsätze" gesprochen werden. Zwar wird über folgende Maßnahmen versucht, eine einheitliche Prüfpraxis zum Gütesiegel der GGS sicherzustellen: einheitliche Vorgabe der Güte- und Prüfbestimmungen durch den Güteausschuß der GGS; zentrale Vergabe des Gütesiegels durch den Güteausschuß nach Überprüfung des Prüfberichts; jährliche Kontrolle der Prüfpraxis durch das Prüfinstitut, die GMD; ständiger Erfahrungsaustausch der Softwareprüfstellen untereinander. Es läßt sich jedoch nicht feststellen, ob die Prüfpraxis wirklich einheitlich ist, da das Prüfinventar, wie Checklisten, Fragebögen, rechnergestützte Analyseverfahren etc. zum Know-how der einzelnen Prüfstellen gehört und damit vor dem Zugriff Dritter geschützt wird.

So, wie sich die Prüfpraxis heute darstellt, ist sie nicht unter dem Aspekt der Uneinheitlichkeit zu kritisieren. Vielmehr ist der der Vornorm 66 285 zugrundeliegende Qualitätsbegriff von Software und der Umfang sowie die Art der Prüfung kritisch zu würdigen. Unsere Kritik an der auf DIN V 66 285 basierenden Praxis macht sich an folgenden Punkten fest:

- an dem für die Prüfung zentralen Qualitätskriterium der Übereinstimmung von Herstellerangaben mit dem Programm und der Dokumentation;
- am Ausschluß der Prüfung von Maßnahmen zum Datenschutz;
- an der Unentwickeltheit der Prüfkriterien zur Benutzungsfreundlichkeit.

Trotz der Bemühungen einiger Softwareprüfstellen, insbesondere den letzten Kritikpunkt durch Eigeninitiative zu überwinden - allen voran ist hier der TÜV Bayern zu nennen - ändert auch der Entwurf zu DIN 66 285 vom 30.9.1988 an diesem Zustand nichts.

4 Mindestanforderungen an die Prüfung auf Konformität mit DIN 66 234, Teil 8

Nachfolgend werden für jeden Gestaltungsgrundsatz der DIN-Norm Mindestanforderungen in Form von methodischen Rahmenbedingungen formuliert. Darüber hinaus sind im Gutachten Anforderungen an die Prüfer und für eine einheitliche Prüfpraxis aufgeführt.

4.1 Aufgabenangemessenheit

Zur Prüfung dieses Gestaltungsgrundsatzes müssen Dialogsysteme bezogen auf typische Arbeitsaufgaben eines Benutzers evaluiert werden. Außerdem verweist das Kriterium auf die Art der Einbettung eines Softwaresystems in eine Organisation.

Angemessenheit ist kein boolescher Wert, er stellt ein Kontinuum von Bewertungen dar, deren Zustandekommen von den Vorerfahrungen und den individuellen Wünschen von Benutzern abhängen. Ob ein Dialog einer Arbeitsaufgabe angemessen ist, sollte daher auch personenorientiert, in Abhängigkeit von unterschiedlichen Benutzerklassen geprüft werden. Ein Softwareprodukt sollte demnach wie folgt auf Aufgabenangemessenheit evaluiert werden:

	system-	personen-	aufgaben-	organisations-
			orientiert	
Aufgabenan-				
gemessenheit		X	X	X

Die Schnittmenge der zur Evaluation von Dialogsoftware auf Aufgabenangemessenheit notwendigen Schwerpunkte (siehe dazu auch Tab. 1) gibt den minimal zu leistenden Aufwand an. Das Expertenurteil nimmt bei allen Gestaltungsgrundsätzen eine Sonderstellung ein, da es bei jedem Evaluationsschwerpunkt einzusetzen ist; es wird daher nicht mehr gesondert erwähnt.

Mindestanforderung 1

Zur Prüfung von Software auf Konformität mit dem Gestaltungsgrundsatz Aufgabenangemessenheit sind mindestens Arbeitsanalysen durchzuführen.

4.2 Selbstbeschreibungsfähigkeit

Dieser Gestaltungsgrundsatz verweist zunächst auf Eigenschaften, die systemimmanent zu prüfen sind. Ob ein Dialogschritt, ein Bezeichner oder eine Erläuterung jedoch auch für Benutzer verständlich ist, ist wiederum in starkem Maße von deren Vorerfahrungen abhängig. Neben einer systemorientierten Evaluierung ist für diese Kategorie auch eine personenorientierte vonnöten.

	system-	personen-	aufgaben-	organisations-
			orientiert	
Selbstbeschreibungs-				
fähigkeit	X	X		

Zur Bestimmung des minimal zu treibenden methodischen Aufwands, um Dialogsysteme auf Selbstbeschreibungsfähigkeit zu prüfen, ergibt die Schnittmenge der Schwerpunkte die Mindestanforderung.

Mindestanforderung 2

Zur Prüfung von Software auf Konformität mit dem Gestaltungsgrundsatz Selbstbeschreibungsfähigkeit sind mindestens Gruppenvergleiche anzustellen.

4.3 Steuerbarkeit

Dieser Gestaltungsgrundsatz ist systemimmanent zu prüfen, da er jenseits individueller Wünsche objektive Eigenschaften eines Dialogsystems beschreibt. Wieweit ein Dialog subjektiv als steuerbar empfunden wird, läßt sich hingegen nur personenorientiert überprüfen. Steuerbarkeit hängt darüber hinaus von der Arbeitsaufgabe als dritter unabhängiger Variable ab. Wenn von einem Dialog ge-

fordert wird, er soll unterbrechbar sein, verweist dies auf die organisatorische Einbettung eines Systems in das DV-Umfeld.

	system-	personen-	aufgaben-	organisations-
			orientiert	
Steuerbarkeit	X	X	X	X

Die Schnittmenge dieser Evaluationsschwerpunkte ergibt das Expertenurteil als notwendiger Voraussetzung für die Konformitätsprüfung.

Mindestanforderung 3

Zur Prüfung von Software auf Konformität mit dem Gestaltungsgrundsatz Steuerbarkeit sind mindestens Expertenurteile erforderlich.

4.4 Erwartungskonformität

Die Forderung nach einheitlichem Dialogverhalten ist immanent zu prüfen. Dialogschrittweite, Systemreaktionen, Sprache sind Dinge, die auf innere Logik und Konsistenz geprüft werden können. Kommt jedoch als weitere Bezugsgröße die Erwartungshaltung von Benutzern und deren Schulungserfahrung hinzu, muß dieser Gestaltungsgrundsatz auch personenorientiert evaluiert werden.

	system-	personen-	aufgaben-	organisations-
			orientiert	
Erwartungskonformität	X	X		

Die Schnittmenge ergibt den minimal zu erbringenden Aufwand.

Mindestanforderung 4

Zur Prüfung von Software auf Konformität mit dem Gestaltungsgrundsatz Erwartungskonformität sind mindestens Gruppenvergleiche anzustellen.

4.5 Fehlerrobustheit

Die durch diesen Gestaltungsgrundsatz angesprochenen Systemeigenschaften lassen sich nahezu ausnahmslos systemorientiert evaluieren. Fehlerrobustheit bezieht sich aber auch auf Fragen nach dem minimalen Aufwand bei einer Korrektur bzw. nach der Verständlichkeit von Meldungen und Hinweisen. Diesen Problemen ist mittels eines personenorientierten Ansatzes nachzugehen.

	system-	personen-	aufgaben-	organisations-
			orientiert	
Fehlerrobustheit	X	X		

In der Übersicht erweist sich der Gruppenvergleich als mindestens zu erbringender methodischer Aufwand.

Mindestanforderung 5

Zur Prüfung von Software auf Konformität mit dem Gestaltungsgrundsatz Fehlerrobustheit sind mindestens Gruppenvergleiche anzustellen.

Es sei an dieser Stelle noch einmal betont, daß die den Mindestanforderungen zugrundeliegenden methodischen Schwerpunktsetzungen eine analytische Trennung von Bereichen wiedergeben, die in der Praxis eine Einheit bilden können. Dennoch sind wir der Auffassung, daß eine analytische Trennung sinnvoll und notwendig ist, da damit die Gestaltungsgrundsätze der DIN 66 234, Teil 8 prüfbar gemacht werden können.

5 Evaluation der Evaluation

Softwareprüfung setzt unter den heutigen Produktionsbedingungen von Softwareentwicklern am falschen Punkt an. Eine falsche Anforderungsermittlung in einer frühen Phase des Softwareentwicklungsprozesses kann Weichen stellen, die nicht mehr reversibel sind

und ein Produkt zu lebenslanger Nichtkonformität mit beliebigen Normen verdammen. Die Gestaltungsgrundsätze der DIN-Norm und die hier formulierten Mindestanforderungen sollten vielmehr dazu genutzt werden, Bewußtseinsprozesse unter Anwendern, Benutzern und Entwicklern für die Bedeutung ergonomischer Anforderungen in Gang zu setzen und den Fokus abzuwenden von der Prüfung fertiger Produkte hin zu einer entwicklungsbegleitenden Qualitätssicherung, die sich diese Mindestanforderungen zu eigen macht.

Fertiger Software kann man letztlich nur ein Prädikat zukommen oder nicht zukommen lassen; in der Entwicklung hingegen besteht noch die Chance, Einfluß auf das Produkt im Hinblick auf benutzungsfreundliche Gestaltung zu nehmen.

6 Literatur

- CARD, S.K., T.P. MORAN and A. NEWELL (1983). The psychology of human-computer interaction, Hillsdale, N.Y.
- Deutsches Institut für Normung e.V. (1988). DIN 66 234, Teil 8: Grundsätze ergonomischer Dialoggestaltung, Berlin
- Deutsches Institut für Normung e.V. (1985). DIN V 66 285: Anwendungssoftware - Prüfgrundsätze, Vornorm, Berlin
- GREEN, T.R.G., F.SCHIELE and S.J. PAYNE (1988). Formalizable models of user knowledge in human-computer interaction, in: T.R.G. Green et al. (eds.): Theory and outcomes in human-computer interaction, London
- KIERAS, D.E. and P.G. POLSON (1985). An approach to the formal analysis of user complexity, Int. Journ. Man-Machine Studies 22, pp. 365-394
- Piepenburg, U. und K.-H. Rödiger (1988). Mindestanforderungen an die Prüfung von Software auf Konformität nach DIN 66 234, Teil 8, Gutachten im Rahmen des Landesprogramms "Sozialverträgliche Technikgestaltung" des Landes NRW, erscheint in der Reihe Werkstattberichte "Mensch und Technik" des MAGS NRW, Düsseldorf, 182 S.

Dr.-Ing. Karl-Heinz Rödiger

Forschungsgruppe für Softwaretechnik und Software-Ergonomie
Mainzer Str. 18
1000 Berlin 31
Tel. 030/8 53 68 16

Dipl.-Psych. Ulrich Piepenburg

Buchenring 67
2000 Hamburg 67
Tel. 040/6 03 75 23