

Vom kritischen Ereignis zum Nutzungsproblem: Die qualitative Analyse in diagnostischen Usability Tests

Marc Hassenzahl

Technische Universität Darmstadt
Institut für Psychologie
Steubenplatz 12, 64293 Darmstadt
hassenzahl@psychologie.tu-darmstadt.de

Frauke Seewald

Technische Universität Darmstadt
Institut für Psychologie
Steubenplatz 12, 64293 Darmstadt
frauke.seewald@gmx.de

Zusammenfassung

Um ein interaktives System auf Gebrauchstauglichkeit (Usability) zu überprüfen, können unterschiedliche Methoden eingesetzt werden. Eine bereits recht verbreitete Methode ist der usability test. Im Rahmen eines solchen Tests fall

en insbesondere qualitative Daten, sogenannte kritische Ereignisse an, die dann zu Nutzungsproblemen verdichtet werden müssen. Der vorliegende Text stellt eine Vorgehensweise zur Analyse solcher Daten vor.

Keywords

usability test, Auswertungsverfahren, qualitative Daten

1.0 Einleitung

Die Gebrauchstauglichkeit (usability) interaktiver Produkte wird als zunehmend wichtige Qualität wahrgenommen. Dies stellt Hersteller vor die Aufgabe, diese angemessen zu sichern. Eine übliche Methode dazu ist der usability test. Hierbei werden potentielle Benutzer eines Produkts um die Bearbeitung verschiedener, möglichst realistischer Aufgaben gebeten (z.B. das Tätigen eines Anrufs beim Test eines Mobiltelefons). Je nach Anwendungsart können sowohl quantitative (Bearbeitungszeit, Grad der Zielerreichung) als auch qualitative (Nutzungsprobleme) Informationen über ein komplexes, interaktives Produkt erhoben werden. Die spezifische Anwendungsart ist abhängig von der jeweiligen Zielsetzung des Tests und den verfügbaren Ressourcen (Zeit, Geld).

In der Praxis werden usability tests häufig mit dem Ziel eingesetzt, ein sich in der Entwicklung befindendes Produkt (Prototypen, funktionale Entwürfe) weiter zu verbessern (vgl. 4). Dazu werden konkrete, gestaltungsorientierte Verbesserungsvorschläge benötigt. Der wichtigste Schritt eines usability tests ist daher die Diagnose von

Nutzungsproblemen aufgrund qualitativer, gestaltungsorientierter Informationen. Diese werden hauptsächlich aus einer Kombination von Beobachtung und verbalen Protokollen (dem sogenannten »Lauten Denken«) gewonnen².

Die Anwendungsart eines Tests wird weiterhin durch die vorhandenen Ressourcen bestimmt. Wahrgenommener Zeit- und Budgetdruck führten Anfang der 90er Jahre zur sogenannten discount usability⁶. Man wollte sich vom Wissenschaftlichen abgrenzen und einfachere, möglicherweise weniger zuverlässigere, aber dafür »praxistauglichere« Methoden entwickeln. In den letzten Jahren lässt sich jedoch ein entgegengesetzter Trend beobachten. Dazu gehört es auch, die Aufmerksamkeit stärker auf die Qualitätssicherung und Standardisierung des usability testing zu richten (vgl. Hamborg in diesem Band). Ansätze zur Standardisierung von usability tests bietet beispielsweise das NCIST/ANSI Common Industry Format (CIF). Es definiert die Art und Weise, mit der Testergebnisse berichtet werden sollen. Dabei orientiert es sich jedoch an den

Empfehlungen der ISO 9241-11, welche qualitative Daten kaum berücksichtigt. Das macht CIF für diagnostische, d.h., stark gestaltungsorientierte usability tests nur bedingt tauglich. Eine Standardisierung ist aber vor allem für den diagnostischen Ansatz und der damit verbundenen Interpretation qualitativer Daten von großer Bedeutung. Die vorliegende Arbeit möchte dazu einen Beitrag leisten, indem sie die Auswertung der anfallenden qualitativen Daten (verbale Protokolle kombiniert mit Beobachtungen) näher beleuchtet (siehe auch Hamborg & Hoemske in diesem Band). Es wird eine Vorgehensweise beschrieben, mit der man ausgehend von den erhobenen Daten Schritt für Schritt zu Ergebnissen und Empfehlungen gelangen kann.

2.0 Modell zur Diagnose von Nutzungsproblemen

Um konkrete Verbesserungsvorschläge geben zu können, müssen zunächst Nutzungsprobleme erkannt werden. In der Praxis werden Nutzungsprobleme jedoch unterschiedlich definiert. Hassenzahl und

Burmester 4 haben in diesem Zusammenhang ein Modell zur Diagnose von Nutzungsproblemen diskutiert, das einen mehrstufigen Prozess beschreibt. Der Prozess beginnt mit den Erfahrungen, die ein Evaluator während einer Testsituation macht. Erfahrung meint hier die Summe von Ereignissen und Aktivitäten, die in einer Situation auftreten. Der Fokus des Evaluators liegt auf jenen beobachtbaren Aktivitäten, die Probleme mit der Benutzung eines Produktes erkennen lassen (z.B. suboptimales Verhalten, Kommentare, Gestik, Mimik). Solche Ereignisse werden als kritische Ereignisse bezeichnet.

Nach der Sammlung solcher kritischer Ereignisse folgt der Schritt der Konstruktion. Die Ereignisse werden miteinander verglichen und Kategorien zugeordnet. Ziel ist die Strukturierung und Reduktion. Die Konstruktion ist ein subjektiver Vorgang und stellt damit nur ein mögliches Ergebnis dar. Andere Evaluatoren könnten zu anderen Konstruktionen kommen.

Den letzten Schritt stellt die Interpretation dar. Der Evaluator versucht, jede Gruppe von Ereignissen in ihrer Gesamtheit zu deuten und Erklärungen für ihr Auftreten zu finden. Von der ursprünglichen Beobachtung soll zum Kern bzw. zur Ursache vorgedrungen werden. Erst durch diesen Schritt wird das eigentliche Nutzungsproblem beschrieben. Die beobachtbaren Ereignisse stellen lediglich das Rohmaterial für die Nutzungsprobleme dar.

3.0 Systematische Konstruktion und Interpretation der Nutzungsprobleme

Das beschriebene Modell zur Diagnose von Nutzungsproblemen hilft, den Prozess der Analyse (vom kritischen

Ereignis zum Verbesserungsvorschlag) in Teilschritte zu zerlegen. Diese Teilschritte sind:

- Erfahrung (Datenerhebung): Identifikation kritischer Ereignisse durch Beobachtung und Befragung (siehe 3)
- Konstruktion (Datenauswertung): Gruppierung und Klassifizierung der Ereignisse
- Interpretation: Unterstellung einer gemeinsamen Ursache für eine Ereignisgruppe (Nutzungsproblem).

Zusätzlich erfolgt in der Regel

- die Priorisierung der Nutzungsprobleme und
- die Erarbeitung von Verbesserungsvorschlägen.

Der usability test stellt für den Evaluator eine Möglichkeit dar, kritische Ereignisse zu sammeln. Ein kritisches Ereignis ist in den meisten Fällen ein Stocken oder Feststecken bei der Aufgabenbearbeitung. Es kann aber auch schon bei der Handlungsplanung auftreten, wenn Teilnehmer erst gar nicht Teilziele und angemessene Handlungen zum Erreichen des Aufgabenziels generieren können. Ein kritisches Ereignis besteht also mindestens aus einem Ziel und einer »Aktivität« des Teilnehmers, die nicht im Einklang mit dem Ziel steht. Dies kann dann eine Handlung oder eine verbale/non-verbale Äußerung sein. Zusätzlich können auch direkte Konsequenzen notiert werden. Ein kritisches Ereignis setzt sich also aus folgenden Elementen zusammen: Ziel, beobachtete Handlung (statt zielführender Handlung), Kommentar und Folge. Tabelle 1 zeigt einige Beispiele für kritische Ereignisse.

Tabelle 1: Beispiele für kritische Ereignisse

Beobachtete Ereignisse	
1	Teilnehmer möchte neue Daten eingeben [Ziel]. Tn bemerkt: »Ich weiß nicht, wie ich neue Daten eingeben soll, da keine leere Zeile vorhanden ist.« [Kommentar]
2	Teilnehmer versucht, eine Funktion in der Menüleiste auszuwählen [Ziel]. Tn ist irritiert [Kommentar nonverbal] und kommt nicht weiter [Folge].
3	Teilnehmer versucht, die Diagrammansicht zu beenden [Ziel]. Tn ist unsicher und bemerkt »Muss ich dazu jetzt auf »Fenster schließen« gehen oder schließt sich dann das komplette Programm?« [Kommentar]
4	Teilnehmer möchte ein Dokument ausdrucken [Ziel]. Er wählt alle Funktionen in der Menüleiste aus [Handlung] und bemerkt: »Ich finde keinen Befehl zum Drucken« [Kommentar]. Tn ist verärgert [Kommentar nonverbal] und bricht die Aufgabe ab [Folge].

Wichtig ist, dass kritische Ereignisse nur Beobachtungen und keine Interpretationen enthalten. Der Evaluator muss seine Diagnose von der eigentlichen Beobachtung trennen, um vorschnelle Schlussfolgerungen und Verzerrungen beim Erheben zu minimieren. Die fehlende Trennung zwischen der eigentlichen Beobachtung und der Interpretation ist einer der häufigsten Fehler beim usability testen.

Kritische Ereignisse können nur bei der Bearbeitung von Aufgaben beobachtet werden. Das Erstellen dieser Aufgaben ist nicht ganz einfach. Sie müssen realistisch sein, d.h., auf das Wissen der Teilnehmer zurückgreifen. Sie dürfen nicht zu einfach, müssen aber lösbar sein. Es dürfen nur Ziele, aber keine notwendigen Handlungen vorgegeben werden. Jede Aufgabe enthält nur ein klar verständliches Ziel, so dass der Teilnehmer selbst erkennen kann, ob und wann das Ziel erfolgreich erfüllt wurde. Es sollten möglichst viele verschiedene Bereiche des

Produkts berührt werden, mindestens aber die Kernbereiche. Wenn möglich sollte sich die Auswahl der Aufgaben auf Ergebnisse der Anforderungsanalyse beziehen. Der Evaluator muss sich bewusst sein, dass die Ergebnisse des Tests entscheidend von den Aufgaben abhängen (vgl. 1). Komplexe Aufgaben (z.B. mit vielen Teilzielen) erhöhen die Wahrscheinlichkeit für kritische Ereignisse, während einfache Aufgaben sie vermindert.

Je nach Aufgabenkomplexität kann somit dasselbe Produkt als mehr oder weniger gebrauchstauglich erscheinen. Aufgaben können ausgehend von dem eigentlichen Ziel in weitere Teilziele gegliedert werden. Zum besseren Erkennen von kritischen Ereignissen kann eine Teilzielliste angefertigt werden. Sie zeigt, welche Schritte zum Erreichen des Aufgabenziels erfüllt werden müssen. Dabei darf ein Teilziel nicht mit einer Handlung verwechselt werden. Ein Teilziel ist beispielsweise »Antragsdaten abspeichern«. Ob dies nun mit dem Menüeintrag oder STRG-S geschieht, ist zunächst irrelevant. In der Praxis hat es sich bewährt, Aufgaben in realistische Szenarien einzubetten. Tabelle 2 zeigt ein Aufgabenbeispiel.

Tabelle 2:
Beispiel für eine Aufgabe und ihre Teilziele

Benutzer
Szenario mit konkretem Aufgabenziel Für die Erstellung des anfallenden Jahresberichts bittet Sie ihr Chef einige Daten in einem Diagramm anschaulich darzustellen. Erstellen Sie ein Balkendiagramm mit den vorliegenden Werten aus Ihrer Excel-Tabelle. Die Balken sollten als 3-D-Grafik und in blauer Farbe angezeigt werden. Beschriften Sie die Achsen entsprechend und geben Sie der Grafik eine passende Überschrift.
Evaluator (zusätzlich)
Teilzielliste: 1 Diagrammassistenten aufrufen 2 Diagrammtyp »Balken« auswählen. 3 Diagrammuntertyp »3-D Balken« auswählen. 4 ...

Kritische Ereignisse können entweder während des Testdurchlaufs direkt protokolliert oder mit einer Videokamera aufgezeichnet werden. Im industriellen Kontext ist aus Zeit- und Kostengründen die direkte Protokollierung gängiger. Ein dafür ideales »Test-team« besteht aus einem Moderator, der den Teilnehmer instruiert und sich um den reibungslosen Ablauf des Tests bemüht, sowie einem Protokollanten, der sich ganz auf das Identifizieren kritischer Ereignisse konzentriert. Es ist allerdings auch möglich, dass eine Person alleine beide Aufgaben übernimmt. Dies ist sehr fordernd und benötigt entsprechende Erfahrung.

Die beobachteten kritischen Ereignisse sollten direkt nach Abschluss eines jeden Testdurchlaufs (wenn die Eindrücke noch »frisch« sind) zur weiteren Bearbeitung festgehalten werden. Dafür eignet sich ein Tabellen-Kalkulationsprogramm. So können die Informationen schnell und einfach nach unterschiedlichen Kriterien sortiert werden. Neben dem eigentlichen kritischen Ereignis sollte die Teilnehmernummer, die

Aufgabennummer, wenn möglich das betreffende Teilziel und der betroffenen Produktteil (Maske, Bedienelemente etc.) notiert werden.

Auf Basis der identifizierten kritischen Ereignisse kann nun die Diagnose von Nutzungsproblemen erfolgen. Hierfür müssen kritische Ereignisse zu Gruppen zusammengefasst und eine gemeinsame Ursache identifiziert werden. Ein Nutzungsproblem besteht aus einer zusammenfassenden Beschreibung der ihm zugrunde liegenden kritischen Ereignisse, einer Ursachenzuschreibung und einem Verbesserungsvorschlag. Das diagnostizierte Nutzungsproblem geht somit über die einzelnen kritischen Ereignisse hinaus, ist aber noch so spezifisch, dass Verbesserungsvorschläge abgeleitet werden können. Erst durch den Schritt der Interpretation lassen sich Ergebnisse auch auf nicht direkt getestete Teile des Produkts übertragen (siehe Kasten 1 für ein Beispiel).

Kasten 1:

Beispiel für ein Nutzungsproblem
Zwei kritische Ereignisse:
Teilnehmerin will neue Daten eingeben. Sie bemerkt »Ich weiß nicht, wie ich neue Daten eingeben soll, es gibt keine leere Zeile«.
Teilnehmer hat Werte in die Zeile einer Tabelle eingegeben. Er drückt wiederholt die Tabulatortaste, um eine neue Zeile zu erstellen. Er bemerkt »Wie soll denn das gehen?«

Die kritischen Ereignisse beziehen sich auf unterschiedliche Teilziele bzw. Aufgaben. Die zugrunde liegende Ursache ist jedoch gleich: Es ist nicht offensichtlich, wie eine neue Zeile erstellt werden kann.

Betrachtet man jedes Ereignis für sich, könnte man zu unterschiedlichen Lösungen kommen. Die Lösung für Beispiel 1 könnte lauten, eine Schaltfläche »Neue Zeile anlegen« einzufügen. In Beispiel 2 wäre es sinnvoll, der Tabulatortaste die allgemeinübliche Funktion der Erstellung einer neuen Zeile zuzuweisen. Die eigentliche Ursache wird jedoch erst durch das Zusammenfassen und Interpretieren identifiziert. Dadurch lassen sich einheitliche Verbesserungsvorschläge geben, die auch auf andere Bereiche des Produkts übertragen werden können.

Ein Nutzungsproblem auf Basis der beiden Eingangs beschriebenen Ereignisse würde wie folgt aussehen:

Teilnehmer wollen Elemente in eine neue Zeile eintragen Ziel]. Es wird keine direkt ersichtliche Möglichkeit geboten, dies zu tun Ursache]. Es sollte immer eine leere Zeile vorhanden sein. Lösung]

Das Zusammenfassen der kritischen Ereignisse ist u.U. schwierig und bedarf einiger Übung. Die folgenden Dinge helfen bei dieser Aufgabe:

- Von abstrakteren Prinzipien ausgehen: Um eine Krankheit zu diagnostizieren muss ein Arzt auch abstrakteres Wissen über mögliche Symptome und Krankheiten haben. Dem Usability Professional bieten z.B. die Dialogprinzipien der ISO 9241-10 (z.B. Selbstbeschreibungsfähigkeit bei dem Beispiel in Kasten 1) oder Nielsen's Heuristiken 7 einen Ausgangspunkt.
- Aufgabenweise zusammenfassen: Um die Zahl der zu gruppierenden Ereignisse zu reduzieren, ist es hilfreich, aufgabenweise vorzugehen. Das bedeutet, es werden zunächst nur die Ereignisse einer bestimmten Aufgabe über alle Teilnehmer gruppiert. Der Vorteil liegt darin, dass die kritischen Ereignisse einer Aufgabe mit einiger Wahrscheinlichkeit bereits ähnlich sind, da sie sich auf dieselben Ziele beziehen.
- Über Aufgaben generalisieren: Nach der Gruppierung pro Aufgabe ist es sinnvoll, über Aufgaben hinweg zu generalisieren. Dabei ist es einfacher, bei Aufgaben mit ähnlichen Teilzielen (wenn vorhanden) zu beginnen.

- Vorläufig bleiben, mehrfach sichten: Die Ereignisgruppen sollten möglichst lange vorläufig bleiben. Sie sollten immer wieder umsortiert und revidiert werden, bis die ideale Gruppierung gefunden ist. Ein zu frühes Festlegen auf eine Ursache kann dazu führen, dass nicht in die Diagnose passende Ereignisse abgewertet werden.
- Diagnose als sozialer Prozess: Eine Alternative zur Verwaltung im Tabellenkalkulationsprogramm stellen Karten dar. Auf jeder Karte wird schon beim Erheben ein Ereignis notiert. Diese Karten können dann einfach sortiert und gruppiert werden. Dies kann auch im Rahmen eines Workshops gemacht werden. So wird die Diagnose zum sozialen Prozess.

Die durch den usability test aufgedeckten Probleme (Beschreibung, Ursache, Verbesserungsmöglichkeiten) werden den Entwicklern als Ergebnis zurückgemeldet. Meist ist es so, dass nicht alle Probleme bearbeitet werden können und eine Priorisierung vorgenommen werden muss. Dies kann zunächst auf der Basis eines Urteils des Usability Professionals geschehen (siehe z.B. 8, »severity ratings«). Ein etwas objektiveres Kriterium ist die »Auftrittsstabilität« eines Problems 5. Als Indikator gilt die Zahl unterschiedlicher Teilnehmer, die kritische Ereignisse zu einem Problem beigesteuert haben. Je mehr Teilnehmer kritische Ereignisse beisteuern, desto ernster und dringender ist das Problem. Ursachen für Probleme, die nur auf kritischen Ereignissen einer Person beruhen, liegen oft in dieser Person. Sie entstehen beispielweise bei mangelndem Aufgabenwissen.

Für das Erstellen von Verbesserungsvorschlägen werden recht unterschiedliche Kompetenzen benötigt. Ein guter Diagnostiker (Usability Professional) ist nicht unbedingt auch ein guter Gestalter und vice versa. Dies sollte dem Professional klar sein. Die beste Diagnose kann durch schlechte Umgestaltungsvorschläge in Frage gestellt werden. Gleichzeitig ist es offensichtlich, dass die Auftraggeber oft nur wenig Interesse an den eigentlichen Nutzungsproblemen haben. Vielmehr interessieren sie die notwendigen Veränderungen. Dieses Dilemma kann gemildert werden, indem die Ergebnispräsentation in eine Art Workshop umgewandelt wird. In diesem Workshop können Probleme erläutert werden, um im Dialog mit den Entwicklern Lösungsansätze zu erarbeiten.

4.0 Schluss

Das Diagnostizieren von Nutzungsproblemen ist ein komplexer Prozess. Gesammelte Beobachtungen müssen schrittweise zusammengefasst und eingegrenzt werden, um den eigentlichen »Kern« – das Problem – benennen zu können. Die Schwierigkeit bei der Umsetzung dieses Vorgehens besteht darin, zwischen einem kritischen Ereignis und dem eigentlichen Nutzungsproblem zu unterscheiden. Während der Testdurchführung muss der Evaluator eine rein beobachtende Haltung einnehmen; er oder sie muss sich immer wieder kontrollieren, um die gemachten Beobachtungen nicht gleichzeitig zu interpretieren. Die Wahrnehmung muss sich objektiv auf die Handlungen und Kommentare der Teilnehmer konzentrieren. Durch die Trennung der drei Schritte, Beobachtung, Konstruktion und Interpretation, kann ver-

hindert werden, dass die Ergebnisse durch Voreingenommenheit und selektive Wahrnehmung verzerrt werden.

Das hier Vorgestellte ist sicher noch weit von einer detaillierten Verfahrensweise entfernt. Trotzdem hoffen wir damit, zumindest die Aufmerksamkeit auf Probleme und Möglichkeiten der Analyse qualitativer Daten aus einem usability test gelenkt zu haben.

5.0 Literatur

- 1 Cordes, E. R.: Task-selection bias: A case for user-defined tasks. In: International Journal of Human-Computer Interaction 13 4 (2002), 411-420.
- 2 Ebling, M. R. & John, B. E.: On the contributions of different empirical data in usability testing. Proceedings of the conference on designing interactive systems (DIS 2000): processes, practices, methods, and techniques. 2000: 289-296.
- 3 Gediga, G. & Hamborg, K.-C.: Ergonomische Evaluation von Software: Methoden und Modelle im Software-Entwicklungsprozess. In: Zeitschrift für Psychologie 210 1 (2002), 40-57.
- 4 Hassenzahl, M. & Burmester, M.: Zur Diagnose von Nutzungsproblemen: Praktikable Ansätze aus der qualitativen Forschungspraxis. In: K.-P. Timpe, H.-P. Willumeit, und H. Kolrep (Hg.): Bewertung von Mensch-Maschine-Systemen. 3. Berliner Werkstatt Mensch-Maschine Systeme. Düsseldorf, 2000: VDI Verlag. 171-184.
- 5 Hassenzahl, M., Prümper, J. & Sailer, U.: Die Priorisierung von Problemhinweisen in der software-ergonomischen Qualitätssicherung. In: R. Liskowsky, B. M. Velichkovsky, und W. Wüschmann (Hg.): Software-Ergonomie '97. Usability Engineering: Integration von Mensch-Computer-Interaktion und Software-Entwicklung. Stuttgart, 1997: B.G. Teubner. 191-201.
- 6 Nielsen, J.: Usability engineering at a discount. In: G. Salvendy und M. J. Smith (Hg.): Designing and Using Human-Computer Interfaces and Knowledge Based Systems. Amsterdam, 1989: Elsevier Science Publishers. 394-401.

7 Nielsen, J.: Heuristic Evaluation. In: J. Nielsen und R. L. Mack (Hg.): Usability Inspection Methods. New York, 1993: Wiley. 25-62.

8 J. Nielsen: Usability Engineering. Boston, San Diego, 1993: Academic Press.

»Es ist erlaubt digitale und Kopien in Papierform des ganzen Papers oder Teilen davon für den persönlichen Gebrauch oder zur Verwendung in Lehrveranstaltungen zu erstellen. Der Verkauf oder gewerbliche Vertrieb ist untersagt. Rückfragen sind zu stellen an den Vorstand des GC-UPA e.V. (Postfach 80 06 46, 70506 Stuttgart).
Proceedings of the 2nd annual GC-UPA Track
Paderborn, September 2004
© 2004 German Chapter of the UPA e.V.«





