

Maximiliane Freymann

Validierungsmethode UTAPAM: Einsatz und Nutzen im User-Centered Design Prozess

1. Einleitung

Validierungen sind im Rahmen eines User-Centered Design Prozesses eine wichtige Methode, um zu überprüfen, ob die erstellten Lösungen die Anforderungen der Nutzer erfüllen und gebrauchstauglich sind.

Laut DIN EN ISO 9241-11 setzt sich Gebrauchstauglichkeit (Usability) aus Effektivität, Effizienz und Zufriedenheit zusammen. Effektivität meint dabei, inwieweit Nutzer ihre Ziele mithilfe des Systems erreichen können. Mit Effizienz ist der Aufwand gemeint, den die Nutzer dabei betreiben müssen. Die Zufriedenheit bezieht sich auf die Einstellung der Nutzer gegenüber der Software.

Eine Schwierigkeit im Zusammenhang mit diesen drei Usability-Dimensionen ist, dass ihre Beurteilung erst nach Fertigstellung des Systems möglich ist. Während des Entwicklungsprozesses ist ihre unmittelbare Beurteilung kaum möglich. Die Zielerreichung, der damit verbundene Aufwand und die Zufriedenheit der Nutzer kann meist erst dann abschließend beurteilt werden, wenn das vollständige System vorliegt und im realen Nutzungskontext eingesetzt wird.

Die Sicherstellung einer hohen Usability während des nutzerzentrierten Entwicklungsprozesses findet in der Regel anhand von Teilzielen oder der indirekten Bewertung der Effizienz und Effektivität statt. Beispielsweise können die erhobenen Nutzeranforderungen auf Vollständigkeit und Konsistenz überprüft werden, um somit eine effektive Nutzung des Systems sicherzustellen.

Im nutzerzentrierten Entwicklungsprozess sollten Annahmen über Nutzerziele, -aufgaben und -anforderungen immer wieder überprüft werden. Durch entwicklungsbegleitende Validierungen

ist es möglich, erhobene Anforderungen, entwickelte Konzepte und erstellte Lösungen auf ihre Richtigkeit und Vollständigkeit zu überprüfen und somit die effiziente, effektive und zufrieden stellende Nutzung sicherzustellen.

Die Integration von nutzerzentrierten Validierungsaktivitäten in einen Softwareentwicklungsprozess stellt nach wie vor eine Herausforderung dar. In gängigen Softwareentwicklungsprozessen sind Validierungen generell integriert. So sind Validierungen beispielsweise im V-Modell (Boehm 1979), im Spiralmodell (Boehm 1984), in der agilen (Cockburn 2003) und evolutionären Entwicklung (Gilb 1981) und im Extreme Programming (Beck et al. 2000) ein fester Bestandteil. Sie zielen in der Regel aber auf funktionelles Prüfen der Software (z. B. Fehlerfreiheit, Robustheit, Vollständigkeit der Funktionalität etc.).

Die Aufgabe des Usability Engineerings ist es, für nutzerzentrierte Validierungsaktivitäten ein angemessenes Instrumentarium bereitzustellen, welches

in den Softwareentwicklungsprozess integriert werden kann.

2. Validierungsmethoden

Es existieren verschiedene Validierungsmethoden, die in unterschiedlichen Phasen der Entwicklung mit User-Centered Design angewendet werden können. Bei einer frühen Validierung können Anforderungen und Ziele untersucht werden, die formative und prospektive Validierung findet entwicklungsbegleitend statt und kann schon auf Teile des späteren Systems angewendet werden. Eine summativ Validierung findet abschließend statt und ist auf das Gesamtsystem fokussiert. Diese summativ Validierung kann z. B. mit Hilfe von Usability Tests, Interviews und Befragungen von Nutzern geschehen.

Innerhalb der Aktivitäten der Produktentwicklung gibt es verschiedene Arbeitsergebnisse, die validiert werden können. Die drei Aktivitäten sind Analy-

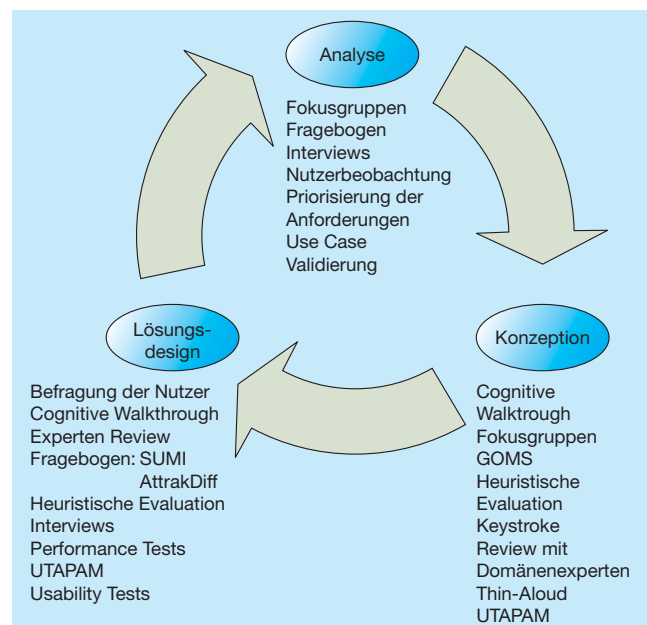


Bild 1: Aktivitäten und Methoden

se, Konzeption und Lösungsdesign. Zusätzlich können die Arbeitsergebnisse, die während der Durchführung der Aktivitäten entstehen, hinsichtlich der Usability-Dimensionen der DIN EN ISO 9241-11 (Effektivität, Effizienz und Zufriedenheit) untersucht werden.

Während der ersten Aktivität werden die Anforderungen spezifiziert, dabei wird zwischen funktionalen und nicht-funktionalen Anforderungen unterschieden. Diese Anforderungen werden z. B. in Use Cases festgehalten, die auf Vollständigkeit, Konsistenz und Korrektheit überprüft werden können. Die beiden Usability-Dimensionen Effektivität und Effizienz der Anforderungen beziehen sich in diesem Kontext auf die Frage, ob die erhobenen Anforderungen alles abdecken, so dass Nutzer ihre Ziele mit der Software mit möglichst geringem Aufwand erreichen können. Methoden, um diese Aspekte zu überprüfen, sind Beobachtungen der Nutzer, Interviews und Fokusgruppen. Diejenigen Anforderungen, die besondere Auswirkung auf die Zufriedenheit der Nutzer mit dem System haben, können mit Hilfe einer Priorisierung der Anforderungen, Fragebogen und weiteren Interviews identifiziert werden.

Nach der Konzeptionsphase können entstandene Konzepte auf Vollständigkeit, Konsistenz und Korrektheit überprüft werden.

Auf diese Weise kann sichergestellt werden, dass alle relevanten Anforderungen erfasst und korrekt umgesetzt wurden.

Die Effektivität kann mittels Cognitive Walkthroughs, Heuristischer Evaluation und Experten Reviews mit Domänenexperten geprüft werden. Um die Effizienz der Konzepte zu validieren, existieren Methoden wie GOMS (Goals, Operators, Methods and Selection Rules; Card et al. 1983), das Keystroke-Level Modell (Card et al. 1983) und die User Task Performance Assessment Method, kurz UTAPAM (Komischke 2004). Hierbei werden verschiedene Performanzparameter wie Mausklicks, ASCII-Eingaben, etc. gemessen. Durch Validierung der Konzepte mit Fokusgruppen, Think-Aloud, und Befragungen der Nutzer kann herausgefunden werden, wie zufrieden zukünftige Nutzer mit den bisher erstellten Konzepten sind.

Während der Designphase werden die Lösungen erstellt. Zur Validierung der Lösungen in Bezug auf ihre Effektivität kön-

nen klassische Usability-Tests eingesetzt werden, Heuristische Evaluation, Cognitive Walkthrough und Experten Reviews.

Mit Hilfe von Usability Tests, Performance Tests und UTAPAM kann die Effizienz des Systems getestet werden. Darüber, wie zufrieden die Nutzer mit der Lösung letztendlich sind, geben Fragebogen wie SUMI (Software Usability Measurement Inventory; Sauro, Kindlund 2005), SUS (System Usability Scale) und AttrakDiff (Hassenzahl et al., 2003) Aufschluss.

3. User Task Performance Assessment Method (UTAPAM)

Die Methode kann auf zwei Weisen eingesetzt werden. Zum einen zur Validierung von Konzepten zu unterschiedlichen Zeitpunkten im Entwicklungsprozess und zum Abgleich mit Zielvorgaben und objektiven Kriterien, die vor Beginn der Entwicklung erstellt wurden. Zum anderen kann diese Methode für den direkten Vergleich zweier Produktversionen benutzt werden, um herauszufinden, wie diese sich bezüglich der fünf Effizienzparameter unterscheiden. Das Besondere dieser Methode liegt darin, dass sie zu jedem Zeitpunkt der Entwicklung zur Überprüfung der Effizienz eingesetzt werden kann, ohne dass repräsentative Nutzer nötig sind.

Die User Task Performance Assessment Method wurde von Komischke entwickelt (Komischke 2004). Sie basiert auf der Annahme, dass die Performanz der Nutzer beim Durchführen einer Aufgabe quantitativ mittels fünf verschiedener Parameter festgehalten werden kann: Mausklicks, Scrolling, ASCII Eingaben, Screen Flips und Lesemenge (Komischke 2004). Diese Parameter alleine haben keine Aussage, im Zusammenhang können sie jedoch Aufschlüsse über die Effizienz beim Benutzen des Systems liefern, insbesondere beim Vergleich mit Effizienzparametern, die mit anderen Systemen beim Durchführen einer ähnlichen Aufgabe entstehen.

Der Parameter „Clicks“ steht bei UTAPAM für die Anzahl der Mausklicks, die zum Durchführen einer Aufgabe benötigt werden, „ASCII“ für die Tastatureingaben, „Screenflips“ bezeichnen Kontextwechsel auf dem aktuellen Screen. „Reading Load“, die Lesemenge, wird in

static und dynamic unterteilt. Lesemenge bezeichnet in diesem Kontext die Anzahl an sinnvollen Einheiten („chunks“), die dem Nutzer präsentiert werden. Mit „Static Reading Load“ sind diejenigen Einheiten gemeint, die sich beim Durchführen einer Aufgabe nicht verändern, z. B. Titel oder Kontrollbuttons. „Dynamic Reading Load“ sind Chunks, die sich ändern können oder nur manchmal sichtbar sind. Ein Beispiel für dynamische Leseinheiten wären Suchergebnisse oder Fehlermeldungen.

Die User Task Performance Assessment Methode ist zu jedem Zeitpunkt der Entwicklung anwendbar, kann also als Hilfsmittel bei der formativen oder summativen Evaluation dienen. Ein weiterer Vorteil dieser Methode ist, dass keine repräsentativen Nutzer erforderlich sind, die reale Aufgaben mit einem funktionierenden System durchführen. Zur Durchführung reichen schon Prototypen, Screenshots oder Konzeptdarstellungen aus. Ein Nachteil ist jedoch, dass nur quantitative Werte erfasst werden, qualitative Ergebnisse müssen mit Hilfe anderer Methoden wie beispielsweise Fragebogen zusätzlich gemessen werden.

Die User Task Performance Assessment Method wurde in einer Fallstudie zum Vergleich zweier Produktversionen eines radiologischen Informationssystems eingesetzt, um herauszufinden, inwieweit sich die beiden Produktversionen quantitativ bezüglich der Effizienz unterscheiden.

4. Web Order Entry und Referring Physician Workplace

Bei den beiden Softwareprodukten Web Order Entry (WOE) und Referring Physician Workplace (RPW) handelt es sich um ein webbasiertes System, das für Ärzte entwickelt wurde, die damit Untersuchungsanforderungen an eine Radiologie-Abteilung oder -Praxis schicken können. Anwender sind sowohl niedergelassene Ärzte als auch Ärzte in Krankenhäusern, die Patienten zu radiologischen Untersuchungen überweisen. Beide Softwareprodukte wurden bei Siemens Medical Solutions im Bereich Health Care Systems – Image Management entwickelt.

Web Order Entry, die ältere Produktversion, wurde ohne User-Centered De-

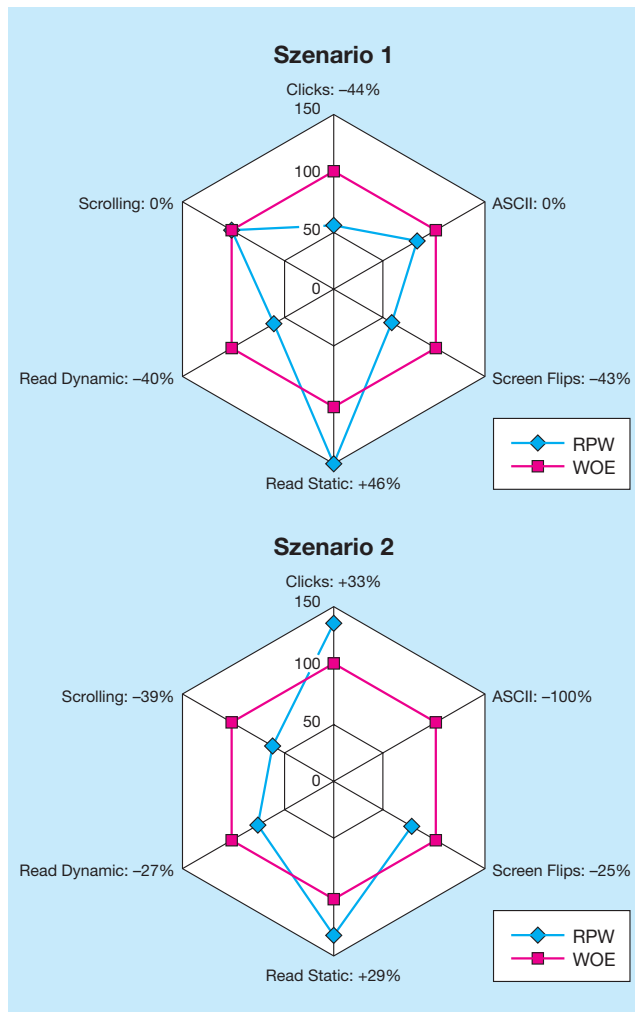


Bild 2: Ergebnisse

kumentiert. Die Ergebnisse wurden zunächst nach Szenario und Produktversion getrennt aufgezeichnet und später einander gegenübergestellt.

Die unterstützende Software dabei war Camtasia Studio (<http://www.techsmith.com/camtasia.asp>), mit deren Hilfe der Bildschirm aufgenommen werden kann. Zusätzlich können Mauswege und -klicks graphisch dargestellt werden.

4.2 Ergebnisse

Die Ergebnisse können in einem Radardiagramm dargestellt werden, um für jedes der beiden Systeme auf einen Blick die Unterschiede zwischen den gemessenen Parametern erkennen zu lassen. Dabei stellt ein System die Basislinie dar, indem die Werte normiert werden. Alle Werte für dieses System sind nun auf 100% gesetzt, die Werte für das andere System stehen dann in Relation dazu. Dadurch, dass die Werte des einen Systems die Basislinie darstellen, können schnell Verbesserungen bei den einzelnen Effizienzparametern festgestellt werden.

Bei der Durchführung der Fallstudie mit den beiden Softwareprodukten Web Order Entry und Referring Physician Workplace wurden die Ergebnisse der älteren Produktversion (Web Order Entry, WOE) als Basislinie gewählt. Im Diagramm sind die Werte von Web Order Entry daher bei beiden Szenarien bei genau 100%.

Für das erste Szenario ergaben sich bei der konkreten Durchführung von UTAPAM folgende Werte (Bild 2): Mit dem neuen Produkt Referring Physician Workplace (RPW) benötigte der Nutzer dazu 44% weniger Mausklicks, 43% weniger Screenflips, es gab 40% weniger dynamische Lesemenge und 46% mehr statische Lesemenge.

Die Anzahl der Tastatureingaben waren bei beiden Produkten gleich. Daraus lässt sich schließen, dass der Nutzer weniger Informationen suchen muss, da mehr Informationen auf einem Screen sind, wodurch die geringere Anzahl an Screen Flips und die höhere Menge an statischen Leseeinheiten zustande kommt. Dadurch wird zusätzlich die Navigation erleichtert. Insbesondere für unerfahrenen Nutzer sind wenige Kontextwechsel, wie sie Screen Flips darstellen und weniger dynamische Lesemenge von Vorteil, da dieses sich positiv auf die Erlernbarkeit des Systems auswirkt. Beim zweiten Szenario ergaben sich folgende

sign entwickelt. Referring Physician Workplace hingegen entstand im Kontext eines User-Centered Design Prozesses.

In einer Fallstudie wurden die beiden Softwareprodukte gegenübergestellt. Dabei sollte untersucht werden, wie gut die neue Produktversion im direkten Vergleich mit der alten abschneidet. Durch quantitative Messungen mit der User Task Performance Assessment Methode sollte gezeigt werden, welche Auswirkungen User-Centered Design auf ein Produkt hat, mit Hilfe von UTAPAM wurde dabei die Effizienz der beiden Produktversionen gemessen. Der Zufriedenheitsgrad der Nutzer wurde mit Fragebogen untersucht. Dabei kam heraus, dass die Software Referring Physician Workplace von den Nutzern subjektiv besser bewertet wurde.

4.1 Vorgehen

Für die Untersuchung wurden aus einem Use Case zwei Testszenarien entwickelt. Dieser Use Case beschreibt die Aktivitä-

ten des überweisenden Arztes während der Sprechstunde mit dem Patienten. Die Szenarien wurden mit beiden Softwareprodukten durchgeführt. Bei jedem Szenario wurden die Performanzparameter der UTAPAM gezählt (Clicks, ASCII, Screen Flips, statische und dynamische Lesemenge, Scrolling) und schließlich die Ergebnisse der beiden Produkte miteinander verglichen.

Beim ersten Szenario sollte sich der Nutzer zunächst einloggen, seinen nächsten Patienten im System aufrufen und sich einen Überblick über diesen verschaffen. Das zweite Szenario beinhaltete die Anforderung einer Untersuchung für einen bestimmten Patienten. Diese Untersuchung sollte innerhalb eines festgelegten Zeitrahmens stattfinden.

Beide Szenarien wurden von der Autorin mit Hilfe von Prototypen beider Produktversionen durchgeführt. Die Performanzparameter von UTAPAM wurden dabei gleichzeitig teilweise elektronisch gemessen und später in einer Tabelle do-

Werte (in Bild 2 im unteren Teil dargestellt): die neue Produktversion benötigte 100 % weniger ASCII Eingaben, 25 % weniger Screen Flips, 27 % weniger dynamische Lesemenge und 39 % weniger Scrolling. Die statische Lesemenge war um 29 % größer und der Nutzer benötigte 33 % mehr Clicks zum Durchführen der Aufgabe.

Die mittels UTAPAM gemessenen Werte beim zweiten Szenario ähneln also denen, die beim ersten Szenario gemessen wurden. Auch hier liegen bis auf die Anzahl der Mausklicks die Werte bei der neuen Produktversion Referring Physician Workplace unter denen von Web Order Entry. Augenscheinlich ist die neue Produktversion beim Vergleich der benötigten Mausklicks dennoch weniger effizient als die alte Produktversion. Bei einer genaueren Untersuchung stellte sich jedoch heraus, dass die zusätzlichen Klicks aufgrund von Pflichtfeldern zustande kamen, die in der alten Produktversion nicht vorhanden waren. Diese Pflichtfelder waren jedoch extrem wichtig für die Patientensicherheit und damit eine wichtige Verbesserung, da relevante Informationen eingetragen werden mussten.

In diesem Falle war die Effizienz der neuen Produktversion zwar geringer, jedoch liegt die Vermutung nahe, dass die Effektivität dadurch erhöht wurde, da nun zusätzliche Eingaben gemacht werden müssen, die der Patientensicherheit dienen, in der alten Produktversion jedoch nicht beachtet wurden.

5. Schlussfolgerung

Die Methode ist geeignet, schon während der Entwicklung eines Produktes Vergleiche mit älteren Versionen zu ziehen oder auch gegen vorher definierte Ziele zu validieren. Auf diese Weise können konkrete Verbesserungsmaßnahmen abgeleitet werden, mit denen gezielt die Effizienz gesteigert werden kann. Wichtig an dieser Stelle ist aber auch, Ziele und Anforderungen bezüglich der Effektivität nicht zu vernachlässigen.

UTAPAM liefert ausschließlich quantitative Ergebnisse. Um weitere qualitative Hinweise auf die Gebrauchstauglichkeit eines Produktes zu bekommen, müssen weitere Schritte erfolgen, beispielsweise in Form einer Nutzerbefragung. Im Zusammenhang mit weiteren Validierungs-

methoden, die auch Aspekte der Effektivität und Zufriedenheit berücksichtigen, kann eine umfassende Bewertung der Nutzungsqualität erfolgen.

Vorteile der Methode UTAPAM liegen zum einen darin, dass sie leicht handhabbar ist und mit wenig Aufwand durchgeführt werden kann. Es ist möglich, bereits während des Entwicklungsprozesses erstellte Konzepte hinsichtlich ihrer Effizienz zu überprüfen. Es ist nicht notwendig, diese Tests mit Nutzern durchzuführen. Auch Usability-Experten können z. B. anhand von Szenarien eine Bewertung durchführen. Neben der Verbesserung der Effizienz eines Softwareproduktes können die Ergebnisse, die mit UTAPAM erzielt werden, auch genutzt werden, um die Vorteile eines nutzerzentrierten Vorgehens zu demonstrieren.

Im Gegensatz zu subjektiven Bewertungen oder vagen qualitativen Aussagen liefert UTAPAM quantitative Daten, die auf eine plakative und leicht verständliche Art und Weise präsentiert werden können. Im Falle der vorliegenden Untersuchung konnten die Ergebnisse darüber hinaus auch im Rahmen von Kundenpräsentationen genutzt werden, um zu verdeutlichen, worin die Verbesserungen der neuen Produktversion im Vergleich zur Vorgängerversion bestehen.

6. Ausblick

Durch die Studie konnte das Potential von UTAPAM als formative und prospektive Validierungsmethode gezeigt werden. Zukünftige Studien könnten sich damit auseinandersetzen, wie Ergebnisse, die mit UTAPAM erzielt werden, konkret in einem Entwicklungsprozess genutzt werden können. Insbesondere das Zusammenspiel mit weiteren Methoden, die auch Aspekte der Effektivität und Zufriedenheit erfassen, sollte umfassender betrachtet und untersucht werden. Anhand der Ergebnisse können einerseits Maßnahmen abgeleitet werden, eine andere Möglichkeit zur Verwendung der UTAPAM-Ergebnisse ist das Marketing. Die Methode liefert quantitative Ergebnisse, mit denen ein System eingeschätzt werden kann. Diese Ergebnisse können konkrete Verbesserungspotentiale eines Produktes aufzeigen und Aufschluss darüber geben, wo das System im Vergleich mit anderen Systemen steht.

Danksagung

Ich bedanke mich bei Siemens Medical HS IM und besonders bei Susanne Laumann dafür, dass mir dort die Durchführung meiner Untersuchung ermöglicht wurde. Tobias Komischke möchte ich dafür danken, dass ich die Methode UTAPAM verwenden durfte. Natalie, Dirk, Karsten und Lennart danke ich für ihre Geduld und die großartige Unterstützung bei der Erstellung dieses Beitrags.

Literatur

- Balzert, H.: Lehrbuch der Software-Technik: Software-Entwicklung. Heidelberg: Spektrum Akademischer Verlag, 1996.
- Beck, K.; Cunningham, W.; Jeffries, R.: eX-treme Programming. <http://www.extremeprogramming.org/>
- Card, S.; Moran, T.; Newell, A.: The Psychology of Human-Computer Interaction. Hillsdale, NJ: Erlbaum, 1983.
- Cockburn, A.: Agile Software-Entwicklung. Bonn: mitp-Verlag, 2003.
- Constantine, L.; Lockwood, L.: Software for Use: A Practical Guide to the Models and Methods of Usage-Centered Design. Reading, MA: Addison-Wesley Longman, Inc., 1999.
- DIN EN ISO 9241-11: Anforderungen an die Gebrauchstauglichkeit, (1999).
- Gilb, T.: Evolutionary Development. ACM Sigsoft, Software Engineering Notes 6/Nr. 2 (1981) 17.
- Hassenzahl, M.; Burmester, M.; Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Mensch & Computer 2003 Interaktion in Bewegung (J. Ziegler; G. Szwillus Eds.) (pp. 187–196). Stuttgart: B. G. Teubner, 2003.
- Komischke, T.: Usability Engineering in the Area of Computerized Physician Order Entry. Human Factor Issues in Human-Computer Interaction 6 (2004).
- Sauro, J.; Kindlund, E.: Using a single usability metric (SUM) to compare the Usability of competing products. In: Proceedings of the HCI International 2005.



Maximiliane Freymann, 25, studiert seit 2000 Informatik mit Nebenfach Medienwissenschaften an der Universität Paderborn. Seit 2001 arbeitet sie als studentische Hilfskraft im Bereich Usability im C-LAB Paderborn. Dort schrieb sie im vergangenen Jahr eine Studienarbeit über User-Centered Design und die Auswirkungen auf die Produktentwicklung anhand eines Vergleichs zweier Produktversionen. Im Jahr 2003 machte ein Praktikum bei Siemens Medical Solutions in Malvern, USA und war dort im UI-Team tätig. Ihre Interessensgebiete liegen im Bereich Usability, User-Centered Design und der Erstellung von UI Konzepten.
E-Mail: Maximiliane.Freymann@c-lab.de