

Share your knowledge: Usability von Wissensmanagementsystemen

Heinz Holling
Philipp Alexander Freund
Jörg-Tobias Kuhn
Psychologisches Institut IV
Westfälische Wilhelms-Universität
Münster

Martin Salaschek
Christian Gawlista
Westfälische Wilhelms-Universität
Münster

Meinald Thielsch
Psychologisches Institut I
Westfälische Wilhelms-Universität
Münster

Abstract

Die Organisation und Verfügbarkeit von Informationen stellt eine zentrale Ressource moderner Organisationen dar. Unter Wissensmanagement lässt sich dabei eine Managementstrategie verstehen, die den zielorientierten und optimalen Umgang mit Informationen verspricht. Besonders onlinegestützte Wissensmanagementsysteme erfreuen sich wachsender Beliebtheit in vielen Arbeits- und Forschungsfeldern.

Die Benutzbarkeit bereits existierender Wissensmanagementsysteme fällt sehr

unterschiedlich aus: Viele gute Ideen auf Entwicklerseite werden bedauerlicherweise häufig nicht bedienerfreundlich realisiert. Neben vorgeschalteten Anforderungsanalysen zur Bestimmung der nutzerseitigen Bedürfnisse sind hier vor allem Studien zur Evaluation der Usability zentral bei der Systemoptimierung. Nur so kann erreicht werden, dass auch breit definierte Zielgruppen problemlos mit solchen Systemen arbeiten können.

Keywords

Usability-Evaluation, Wissensmanagement, Wissensmanagement-Plattformen

1.0 Einleitung

Die in diesem und dem letzten Jahrhundert entstandene Wissensexplosion macht eine neue Betrachtung des Wissens als Ressource, als Wettbewerbsvorteil und als Produkt notwendig. Der heutigen Flut an immer spezieller und ausdifferenzierter werdendem Wissen und Erfahrungen muss mit probaten Mitteln – nicht nur in der Forschung, sondern auch in der freien Wirtschaft – begegnet werden. Aus diesem Grund werden das systematische Handling, die Aufbereitung und die gezielte Bereitstellung von Wissen seit einigen Jahren besonders thematisiert. Das Schlagwort „Wissensmanagement“ begegnet uns seit einigen Jahren in fast allen Teilen unserer Gesellschaft. Nicht nur in Forschung und Lehre ist die systematische Bereitstellung und Zugänglichkeit von Wissen von besonderer Bedeutung, sondern auch in Unternehmensberatungen, Wirtschaftsprüfungsunternehmen, Anwaltskanzleien und anderen Dienst-

leistungsbranchen wird der effektive und effiziente Umgang mit Wissen bereits als Wettbewerbsvorteil betrachtet und wahrgenommen. „As you deal with knowledge professionals and the huge amount of information out there, you have to understand which information is important, get people to share it, and find ways to make it easy and understandable to use“ (DiMattia & Oder, 1997, S.33).

In dieser Publikation werden drei Beiträge zusammengefasst, die sich des Problems einer auf Wissensmanagementsysteme zugeschnittenen Usability-Evaluation annehmen. Der erste Beitrag gibt einen Überblick über derzeit eingesetzte Verfahren zur Usability-Evaluation in diesem Bereich. Im zweiten Beitrag werden exemplarisch ausgewählte prototypische Vertreter solcher Verfahren in der Anwendung auf ein konkretes Wissensmanagementsystem vorgestellt und beurteilt. Der abschließende Beitrag unter-

sucht den Einfluss unterschiedlich intensiven Nutzungsverhaltens bei einem Wissensmanagementsystem in der universitären Lehre auf die wahrgenommene Usability.

2.0 Derzeit eingesetzte Verfahren zur Usability-Evaluation

Zur Evaluation der Usability von Software existiert bereits eine Vielzahl etablierter Verfahren. Viele von ihnen bieten sich auch an, wenn es um die Evaluation von Websites geht. Bei der Usability-Evaluation von Wissensmanagement-Plattformen im Internet müssen einige Besonderheiten berücksichtigt werden. Die zentralen Ziele – Bereitstellen und Erhalten von Informationen – so „benutzbar“ wie möglich zu gestalten wird erschwert durch die Einschränkungen eines Web-Interfaces. Zudem ist die Zielgruppe hinsichtlich Vorwissen und Erwartungen oft sehr heterogen.

Im Folgenden wird ein Überblick über verschiedene Instrumente zur Software-Evaluation gegeben.

2.1 Prädiktive Evaluationsmethoden

Prädiktive Evaluationsmethoden sind gut geeignet um Schwächen bei der Benutzbarkeit der Software aufzuzeigen. Beurteiler sind meist Experten, die einen User simulieren.

2.1.1 Walkthrough-Methoden

Beim klassischen Walkthrough werden dem Beurteiler Bildschirmansichten präsentiert, die man zum Ausführen einer Funktion benötigt. Die Teilnehmer bewerten aus User-Sicht, wie das Programm zu bedienen wäre und geben Bewertungen anhand festgelegter Kriterien ab.

Cognitive Walkthroughs, bei denen der Fokus auf die kognitiven Prozesse des Nutzers beim Bearbeiten einer Aufgabe gelegt wird, sind ein beliebtes Mittel um Nutzungsprobleme und deren Ursachen bereits in einer frühen Phase der Produktentwicklung aufzudecken. Da Walkthrough-Methoden sehr zeitaufwändig sind, findet die Evaluation häufig in Gruppen statt um die Effektivität zu steigern.

2.1.2 Expertenurteile und Heuristische Evaluation

Bei Expertenurteilen wird eine Software von Spezialisten bewertet, die nicht in die Entwicklung involviert sind. Das freie, das strukturierte und das methodengeleitete Expertenurteil unterscheiden sich in der Vorgabenstrenge bezüglich bestimmter Auswertungs- und Bewertungskriterien.

Die Heuristische Evaluation ist ein Sonderfall des strukturierten Expertenurteils, bei dem der Bewerter sich an einer Reihe von Heuristiken orientiert. Nielsen (1994) schlägt folgende Regeln zur Beachtung vor:

- Simple and natural dialogue
- Speak the user's language
- Minimize user memory load
- Be consistent
- Provide feedback
- Provide clearly marked exits
- Provide shortcuts
- Good error messages
- Prevent errors

Sowohl die Cognitive Walkthrough-Methode als auch Expertenurteile erfordern gute Kenntnisse über Nutzer und Aufgabenanforderungen.

2.2 Deskriptive Evaluationsmethoden

Deskriptive Evaluationsmethoden beschreiben den Zustand der Software möglichst objektiv, reliabel und valide aus Sicht des Users.

2.2.1 Verhaltensbasierte Evaluationsmethoden

Hier interagiert ein Benutzer mit einer lauffähigen Version des Systems. Dabei werden beispielsweise seine Aktionen aufgezeichnet (z. B. Mouse-Tracking und Tastatureingaben, Videoaufzeichnung), und/oder er wird zu lautem Denken aufgefordert um bewusste Kognitionen und Emotionen zu erfassen.

2.2.2 Meinungsbasierte Evaluationsmethoden

Meinungsbasierte Befragungen umfassen Interviews und Fragebögen. Sie sind in Planung, Durchführung und Auswertung insgesamt in der Regel erheblich ökonomischer als Beobachtungen.

Interviews werden üblicherweise als Nachbefragung oder in Bezug auf sogenannte critical incidents durchgeführt. Standardisierte Verfahren sind rar, jedoch gibt Nielsen (2001) einige Hinweise zur Durchführung. Mit der halbstandardisierten Repertory-

Gridtechnik lassen sich mehrere (Papier-)Prototypen in einzelnen Aspekten miteinander vergleichen und zugleich differenzieren um erwünschte Programmfunktionen oder -eigenschaften zu generieren. Die Frage nach Validität und Reliabilität der Ergebnisse der Technik ist weitgehend ungeklärt.

Fragebogenmethoden werden in der Spezifikations- und Designphase, hauptsächlich aber beim Redesign eingesetzt. Fragebogenuntersuchungen lassen sich günstig durchführen und können zumeist auch computergestützt dargeboten werden.

Während ad hoc entworfene Fragebögen, die direkt für das Testobjekt entworfen wurden, direkte Hinweise zur Usability und ggf. zum Redesign einzelner Komponenten geben, leiden sie üblicherweise unter niedriger Reliabilität, und aufgrund mangelnder Standardisierung lassen sich die Ergebnisse nicht mit anderen Systemen vergleichen. Folglich lässt sich auch nur schwerlich abschätzen, wann ein Redesign einer Komponente empfehlenswert oder gar dringend erforderlich ist.

Im folgenden Abschnitt werden hingegen vier etablierte Verfahren diskutiert, die sich insbesondere in der Redesign-Phase einer Software anbieten.

2.3 Vergleich ausgewählter Verfahren

2.3.1 ISONORM

ISONORM (Prümper & Anft, 1993) basiert auf der ISO-Norm 9241-10 (Grundsätze der Dialoggestaltung). ISONORM besteht aus 7 Skalen, die den Usability-Kriterien der ISO 9241 entsprechen (Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Steuerbarkeit, Erwartungskonformität, Fehlertoleranz, Individualisierbarkeit, Lernförderlichkeit). Jede Skala enthält 5 Items, die auf einer siebenstufigen Ratingskala beantwortet wird. Die ermittelten Reliabi-

litäten von 0.59 bis 0.89 sind nur teilweise als gut anzusehen. Allerdings verlangen die Autoren, dass ein Beurteiler das Programm sehr gut kennen sollte (tägliche Nutzung über Monate hinweg). Somit eignet sich das Verfahren nicht zur Evaluation von Prototypen.

2.3.2 IsoMetrics

IsoMetrics (zusammenfassend siehe Gediga & Hamborg, 1999) orientiert sich ebenfalls an Teil 10 der ISO 9241 und enthält ebenfalls sieben Skalen zu den in der Norm formulierten Kriterien. Mit 75 fünffach gestuften Items ist der Fragebogen recht umfangreich. Eine für die formative Evaluation gedachte Langversion enthält zusätzlich ein Feld zur Antwortbegründung nach jeder Frage sowie eine Einschätzung der Wichtigkeit dieses Aspekts. Den Autoren zufolge eignet sich IsoMetrics auch für Software-Prototypen und für Evaluationen, bei denen der Proband die Software erst beim Testen kennen lernt. Die ermittelten Reliabilitäten betragen 0.69 bis 0.86.

2.3.3 SUMI und WAMMI

Das SUMI (Software Usability Measurement Inventory) (Kirakowski, 1995) umfasst 5 Skalen mit insgesamt 50 Items, die die subjektiv wahrgenommene Usability erfassen. Außerdem gibt es eine „Global“-Skala, die die 25 am stärksten auf einem allgemeinen Usability-Faktor ladenden Items enthält. Zwar ist auch das SUMI an ISO 9241 angelehnt, die einzelnen Dimensionen (Affect, Control, Efficiency, Helpfulness und Learnability) sind jedoch eigenständig entwickelt und psychometrisch abgesichert worden. Durch die Skala *Affect* grenzt sich das SUMI dabei von anderen Instrumenten ab (z. B.: „I would not like to use this software every day“). Das SUMI ist sehr gut standardisiert, und es existieren Normwerte zur leichteren Interpretation der Ergebnisse. Das *WAMMI* (Website Analysis and

Measurement Inventory; Kirakowski et al., 1998) ist ein weit verbreitetes Instrument zur Erfassung der Usability von Websites und stellt eine spezifische Weiterentwicklung des SUMI dar. In der Kurzversion enthält es 5 Skalen mit insgesamt 20 Items, die in der Benennung den SUMI-Skalen ähneln. Einzig die Skala *Affect* heißt hier *Attractiveness*. Die einzelnen Dimensionen wurden ohne Überprüfung der Faktorenstruktur aus dem SUMI übernommen, daher kann nicht beurteilt werden, ob und welche der Items in einer Faktorenanalyse als zusammengehörig erscheinen würden. Die Gesamtreliabilität beträgt 0.96 (Cronbachs Alpha) in der Version 1.1.

2.3.4 QUIS

Das QUIS (Questionnaire for User Interface Satisfaction) (Shneiderman & Plaisant, 2005), das mittlerweile in Version 7 vorliegt, fragt nach der subjektiven Zufriedenheit mit der Software im Allgemeinen sowie in Bezug auf spezielle Aspekte. Die Langversion enthält insgesamt zwölf Skalen, u.a. zu den Aspekten Nutzerzufriedenheit, Bildschirmgestaltung, Erlernbarkeit, Leistungen der Software wie Geschwindigkeit und Zuverlässigkeit, sowie zu Handbüchern und Online-Hilfen. Verschiedene Themenblöcke können für verschiedene Produkte benutzt werden um den Fragebogen an die Software anzupassen. Chin et al. (1988) berichten über gute Reliabilitätswerte des QUIS-Gesamtscores (Cronbachs Alpha = 0.94), machen jedoch keine Angaben zu Reliabilitätswerten der einzelnen Skalen. Sie merken außerdem an, dass bisher keine Ergebnisse zur Konstruktvalidität vorliegen. Für die neueste Version des QUIS existieren noch keine psychometrischen Untersuchungen.

3.0 Einsatz ausgewählter Verfahren bei HERBIE

Die Wissensmanagementplattform HERBIE (*Hybrid Education and Research Base for Information Exchange*) wurde an der Westfälischen Wilhelms-Universität, Münster entwickelt (mehr Informationen unter <http://www.herbiegroup.de>). Sie stellt eine online basierte Wissensmanagement Plattform zur verteilten Forschung dar. Somit bietet sie die Möglichkeit, räumlich verteilten Arbeitsgruppen Informationen bereitzustellen, auszutauschen, zu erarbeiten und zu diskutieren.

Eine einfache Ordnerstruktur gliedert die vorhandenen Dokumente den Wünschen der Anwender entsprechend. Durch die Beschreibung des Kontextes der Dokumente, das Versehen mit Autoren und verschiedenen Versionshinweisen sowie die Charakterisierung der Inhalte anhand von Taxonomien und Ontologien wird das zielsetzungsgerechte Bereitstellen und das einfache Auffinden von Dokumenten ermöglicht. Zudem stehen eine Volltext- und eine Kontextsuche zum schnellen Auffinden von Informationen zur Verfügung. Abbildung 1 zeigt einen Screenshot von HERBIE.

Für die Zusammenarbeit der Nutzergruppen besteht die Möglichkeit der Nutzung verschiedener synchroner und asynchroner Kommunikationskanäle. So können die Dokumente über objektspezifische Foren oder Chats durch die Nutzer diskutiert und kommentiert werden. Ein allgemeiner Nachrichtenbereich ermöglicht die zeitgesteuerte, nicht an Objekte gebundene Publikation von Hinweisen und Nachrichten.

Ein differenziertes Rechtssystem auf Objektebene ermöglicht die Verwaltung der Mitglieder, deren Gruppierung und der Zugänglichkeit von Dokumenten. Des Weiteren können die Mitgliederprofile durch Erfassung der Interessen und Fachgebiete, durch Bereitstel-

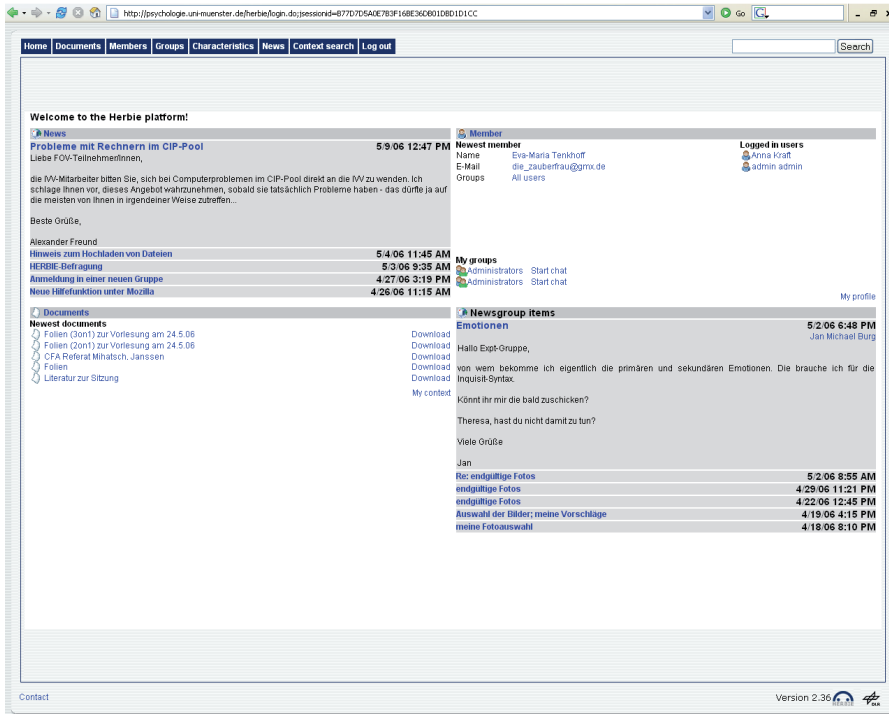


Abbildung 1: HERBIE-Screenshot

lung von persönlichen Homepages und durch die Personalisierung mit Fotos bereichert werden.

Eine solche Wissensmanagement-Plattform steht und fällt mit der erfolgreichen Benutzung durch die Anwender. Da die Zielgruppe, die eine solche Plattform nutzen soll, relativ heterogen ist und eine große Varianz in Bezug auf die Fähigkeiten der Anwender im Umgang mit solchen Plattformen zu erwarten ist, muss ein großes Augenmerk bereits während der Entwicklung auf die Benutzerfreundlichkeit gelegt werden. Denn der Nutzen einer solchen Plattform besteht, neben der Möglichkeit räumlich getrennt zusammen zu arbeiten, vor allem auch darin, die Effizienz und Effektivität im Umgang mit Informationen zu steigern. Dieser Nutzen kann aber sehr schnell durch vom Anwender empfundene Barrieren und Schwächen in der Bedienbarkeit geschmälert werden.

Zudem konnte von Igbaria, Schiffmann & Wieckowski (1994) gezeigt werden, dass auch Spaß bei der Arbeit mit Softwareprodukten zu erhöhter Akzeptanz und Zufriedenheit führen kann. „Joy of use“ ist also auch in diesem Bereich ein nicht zu vernachlässigender Faktor, der bei der Konzeption und Entwicklung von Software, nicht nur im Unterhaltungsbereich, Beachtung finden sollte (zur Übersicht siehe Norman, 2004). Für eine online basierte Software wie HERBIE ebenfalls von großer Bedeutung, ist die optimale Erfahrung des „Flow-Erlebnisses“ wie es Novak, Hoffman und Yung (1999) in den Vordergrund stellen. Um dieses Erlebnis im Umgang mit der Software nicht zu verhindern, gilt es, störende Bugs und Unzulänglichkeiten in der Anwendung der Software zu identifizieren und zu beseitigen.

Zu diesem Zweck wurde HERBIE in der Version V2 im Sommersemester 2005 in zwei Seminaren an der Westfälischen Wilhelms-Universität, Müns-

ter, testweise eingesetzt und evaluiert. Zur Anwendung kamen der IsoMetrics-Fragebogen (Gediga & Hamborg, 1999) und ein Usability-Test.

3.1 Die angewandten Methoden

Zum Einsatz kam der IsoMetrics-Fragebogen in der Kurzversion (siehe Abschnitt 2.3.2). Um auch den Verlauf des subjektiven Eindrucks der Seminarteilnehmer zu erheben, wurden die Teilnehmer zu drei Zeitpunkten im Semester befragt, da damit zu rechnen war, dass die Beurteilungen zu Beginn schlechter, mit etwas Erfahrung im Umgang mit der Software jedoch besser werden würden. Zudem lag nahe, dass es Schwächen in der Bedienbarkeit geben würde, die erst bei einer längerfristigen Auseinandersetzung mit den Möglichkeiten der Software zu Tage treten.

Unterschiedliche Evaluationsverfahren liefern häufig unterschiedliche Ergebnisse, deren Schnittmenge nicht groß ausfällt (Molich et al., 2004). Daher wurden zusätzlich Usability-Tests durchgeführt, um eine möglichst umfassende Sicht des Zustandes der Benutzbarkeit der Software zu erreichen. Außerdem ermöglicht dieses Vorgehen einen Vergleich der mit den beiden Verfahren erzielten Ergebnisse.

An dem Test nahmen fünf Personen aus den beiden Seminaren teil, die bereits mit der Plattform gearbeitet haben, sowie sieben Personen, denen die Plattform gänzlich unbekannt war. Die Probanden waren angehalten, einige für die Benutzer typische Aufgaben auf der Plattform durchzuführen und ihre handlungsbegleitenden Kognitionen während der Arbeit, nach der Methode des lauten Denkens (Carroll & Mack, 1984; Nielsen, 1993), zu äußern. Während dieser perikationalen Methode der Verhaltensbasierten Evaluation wurden die Äußerungen der Probanden, sowie deren Vorgehen auf dem Bildschirm, aufgezeichnet.

Diese „[...] Vorgehensweise hat den Vorteil, dass zusätzlich zu den Äußerungen der Probanden Kontextinformationen registriert werden, die ggf. zum Verständnis der Äußerungen und deren Auswertung herangezogen werden können“ (Gediga & Hamborg, 2002). Vor der Bearbeitung der eigentlichen Aufgaben gab es vier Übungsaufgaben, die normale Tätigkeiten am Computer und im Internet simulieren sollten, um den Probanden die Möglichkeit zu geben, sich „an die aus dem parallelen Handeln und Denken erwachsene Doppelbelastung“ (Gediga & Hamborg, 2002) zu gewöhnen. Des Weiteren dienten diese Aufgaben auch dazu, den Katalog mit Leitfragen, der den Probanden vorgelegt wurde, kennen zu lernen und zu verinnerlichen. Die Leitfragen, anhand derer die Versuchsteilnehmer eventuelle Probleme identifizieren sollten, waren an die von Nielsen und Molich (1990) verfassten zehn Heuristiken zur Usability angelehnt, und bezogen auf die oben genannte Software formuliert. Die zwölf zu bearbeitenden Aufgaben bestanden aus den für Nutzer der Plattform alltäglichen und typischen Aufgaben wie z. B. „Finden Sie den Beitrag ‚...‘ von ‚...‘ oder „Antworten Sie in dem Forum auf den Beitrag ‚...‘“.

Die aus den beiden Verfahren gewonnenen Ergebnisse sollen in den nächsten Entwicklungsstufen der Software einfließen, um die Bedienbarkeit weiter an die Bedürfnisse der Nutzer anzupassen und somit eine dauerhaft erhöhte Akzeptanz, Effizienz und Effektivität im Umgang mit HERBIE zu gewährleisten.

Die bisherige Betrachtung der Ergebnisse bestätigt die Annahme, dass verschiedene Verfahren unterschiedliche Schwächen einer Software zu Tage fördern. Die Ergebnisse des Usability-Tests zeigten tendenziell eher Barrieren in der Handhabung auf, die während des Umgangs auftreten, jedoch durch etwas

Ausprobieren überwunden werden können und somit in der retrospektiven Beurteilung durch den Benutzer in einem Fragebogen häufig nicht mehr angemerkt werden. Die Ergebnisse des IsoMetrics Fragebogen spiegeln eher gröbere Schwächen der Software und deren subjektive Wahrnehmung wider. Gerade die Kombination der Ergebnisse scheint eine deutliche Verbesserung der Software zu ermöglichen.

Auf Basis der bisherigen Nutzungsdaten und den individuellen Kommentaren innerhalb des IsoMetrics-Fragebogens lässt sich erkennen, dass besonders der objektspezifische Diskurs der online verfügbaren Dokumente von großer Bedeutung für die Nutzer zu sein scheint. Dieser wurde von den Pilot-Anwendern verstärkt genutzt und in dem Fragebogen kommentiert. Die Möglichkeit in Form eines Forums über bestimmte Inhalte diskutieren zu können und diese weiterzuentwickeln war zentraler Bestandteil der Seminare und wurde mehrheitlich positiv bewertet. Viele der von den Anwendern abgegebenen Kommentare befassten sich mit der Gestaltung dieser Funktion und der Möglichkeit diese um weitere Features zu ergänzen. „Im Forum sollte immer das Datum des letzten Beitrags und nicht des ersten sichtbar sein, so lässt sich auf einen Blick erkennen, ob neue Diskussionsbeiträge erstellt worden sind.“ war so zum Beispiel die Äußerung eines Nutzers. Die Übersichtlichkeit und Überschaubarkeit des Forums scheint in diesem Zusammenhang noch als verbesserungswürdig wahrgenommen zu werden. In diesem Kontext wurde auch deutlich, dass die Wiki-Funktion hohe Akzeptanz und Verwendung fand. Vielfach wurde eine bessere Anbindung der Wiki-Funktion zu den in den Foren erarbeiteten Informationen erwünscht.

4.0 Einfluss von Nutzungsintensität

Um den Einfluss der Nutzungsintensität zu untersuchen, wurde HERBIE in verschiedenen universitären Lehrveranstaltungen im Fachbereich Psychologie und Sportwissenschaften an der Westfälischen Wilhelms-Universität Münster im Verlauf des Wintersemesters 2005/2006 eingesetzt. Insgesamt wurde das System in vier Seminarveranstaltungen und zwei Vorlesungen genutzt. Auf Grund von Unterschieden in der organisatorischen Ausrichtung (Vorlesungen vs. Seminare) und der damit verbundenen erwarteten unterschiedlichen Nutzungsintensität und Nutzungsfunktionalität von HERBIE wurden die Veranstaltungen in drei Gruppen zusammengefasst. In Gruppe 1 sollte die Plattform als reiner Dokumentenverteiler genutzt werden. Zu dieser Gruppe gehörten die beiden Vorlesungen und zwei reguläre Seminare.

Weiterführende Programmmerkmale sollten auf einer kontinuierlichen Basis von den Nutzern in Gruppe 2 in Anspruch genommen werden. Es sollten also hier nicht hauptsächlich nur Dokumente verwaltet werden, sondern darüber hinaus wurde auch eine aktivere Kommunikation zwischen Dozenten und Seminarteilnehmern beziehungsweise unter den Seminarteilnehmern selbst erwartet. Eine der beiden Seminarveranstaltungen wurde auf Grund ihrer konzeptuellen Anlage (Propädeutikum einer zweisemestrigen Veranstaltung) dieser zweiten Gruppe zugerechnet.

In der dritten Gruppe schließlich sollten die Möglichkeiten von HERBIE von den Mitgliedern am intensivsten genutzt werden. Diesen Kriterien entsprach die Konzeptualisierung des Praxissemesters eines Projektseminars, in dem die gesamte Kommunikation über HERBIE abgewickelt werden sollte.

4.1 Stichprobenbeschreibung

Da die Teilnehmer der Untersuchung nicht randomisiert den einzelnen Veranstaltungen – und somit Nutzungsbedingungen – zugeordnet werden konnten, liegt ein quasiexperimentelles Untersuchungsdesign vor. Neben der nicht zufälligen Aufteilung der Teilnehmer auf die einzelnen Gruppen wurde der between-Subjects Faktor „Grad der Nutzungsintensität“ von den Dozenten durch den Veranstaltungscharakter und die damit verbundenen Anforderungen an die Studierenden festgelegt.

Die Evaluation wurde online durchgeführt; insgesamt beteiligten sich an ihr $N = 57$ Teilnehmer, was einer Quote von ca. 17 % der Gesamtstichprobe von 336 Veranstaltungsteilnehmern entspricht. Für Onlineumfragen sind derartig geringe Rücklaufquoten durchaus typisch (siehe z. B. Welker, 2001), und auch vor dem Hintergrund einer allgemein zu verzeichnenden Abnahme der Umfragebereitschaft überrascht die beobachtete Rücklaufquote nicht (Tourangeau, 2004).

4.2 Ergebnisse

Eine Inspektion der Ergebnisse zeigte einen sehr homogenen Verlauf der Mittelwerte über die sieben IsoMetrics-Skalen für alle Veranstaltungen auf. Es gab jedoch eine bedeutsame Ausnahme: Die Mittelwerte in der der Gruppe 3 zugeordneten Veranstaltung lagen deutlich unter denen der anderen Gruppen. Da der Profilverlauf der Mittelwerte der übrigen fünf Gruppen nahezu kongruent ausfiel, wurden diese Mittelwerte aggregiert. Abbildung 2 plottet diese aggregierten Mittelwerte gegen die Werte der Gruppe 3. Die Konsequenz dieses Vorgehens besteht natürlich darin, dass die ursprünglich vorgenommene Dreiteilung der Gruppen mit Rücksicht auf die angestrebte Nutzungsintensität zu Gunsten einer Zweiteilung in intensi-

ve und einfache (in erster Linie als Dokumentenmanager) Nutzung geändert wurde. In Abbildung 2 werden die Mittelwerte der aggregierten Gruppen 1 und 2 durch die weißen Marker gekennzeichnet, die der Gruppe 3 durch die schwarz ausgefüllten. Die in Abhängigkeit von der Nutzungsintensität wahrgenommenen Unterschiede in der Usability des Softwaresystems werden auf diese Weise klar herausgestellt.

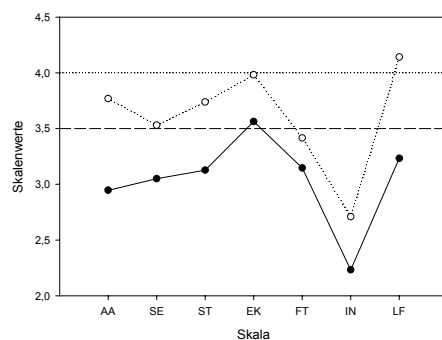


Abbildung 2: ISOMetrics-Mittelwerte

In Abbildung 2 zeigen zwei Referenzlinien die Werte 3,5 und 4,0 als unteren beziehungsweise oberen Standard an, der für das Programmsystem HERBIE erreicht werden sollte. Mit Bezug auf andere Wissensmanagementsysteme (Blackboard, Stud.IP und WebCT) kann der Wert 3,5 als durchschnittlich, der Wert 4,0 bereits als deutlich besser angesehen werden (Ollermann, Gruber & Hamborg, 2003).

Die Mittelwerte der Gruppe 3 liegen in nahezu allen IsoMetrics-Skalen signifikant unter denen der hier gewählten Vergleichsgruppe. Der Faktor Gruppenzugehörigkeit erreichte in einer MANOVA höchste statistische Signifikanz [$F(7, 49) = 4.92; p = .00$]. Es sind vor allem die Skalen Aufgabenangemessenheit, Steuerbarkeit und Lernförderlichkeit, in denen die größten Unterschiede in der beurteilten Usability auftreten.

Von einigen Teilnehmern wurden nach dem Ausfüllen des IsoMetrics-Fragebogens auch offene Anmerkungen zu dem von ihnen genutzten HERBIE-System gemacht. Diese Anmerkungen fielen insgesamt relativ uniform aus: Bemängelt wurden größtenteils (selten auftretende) technische Probleme, wie beispielsweise temporäre Unerreichbarkeit des Servers, in erster Linie jedoch Probleme bei der Nutzerregistrierung und -verwaltung. Einige konkrete Verbesserungsvorschläge wurden auch unterbreitet.

5.0 Ausblick

Die Evaluationsergebnisse sollen dazu dienen, unmittelbar zu behobende Schwachstellen im System zu identifizieren und Verbesserungen zu implementieren.

Bei dem Vergleich von „normaler“ bzw. limitierter und intensiver Nutzung fallen einige Probleme ins Auge. HERBIE erscheint für die intensive Nutzung in seiner in dieser Studie untersuchten Form zu statisch und umständlich und nicht zu 100 % für die Verwaltung großer Informationsmengen geeignet. Diese Interpretation wird auch durch die Erfahrungsberichte einzelner Teilnehmer belegt, die beklagten, dass ein hoher Informationsinput auch unbedingt eine erhöhte Zugriffsfrequenz mit dem Programm erfordert, mit dem Resultat, dass man nach längerer Abwesenheit von der Plattform zunächst einmal alle mit Dokumenten und sonstigen Informationen gefüllten Hierarchieebenen des Systems überprüfen muss, um festzustellen, was sich in der Zwischenzeit ereignet hat. Die Nutzung des Systems sollte jedoch in erster Linie eine Arbeitersparnis bewirken und *nicht* für mehr organisatorischen Aufwand sorgen. Dieses Ziel konnte zumindest in den Veranstaltungen mit einer normalen Nutzungsintensität erreicht werden.

Die Nutzung von HERBIE in der universitären Lehre stellt somit eine viel versprechende Perspektive auf dem Weg zu einer erfolgreichen Anreicherung traditioneller Lehrkonzepte dar. Die bisherigen Ergebnisse zeigen, dass hier bereits Erfolge erzielt werden konnten, die sich in letzter Instanz auch unabhängig vom unterrichteten Fach einstellen sollten (bisher liegen in erster Linie Untersuchungen zu Lehrveranstaltungen im Fach Psychologie vor). Das System soll auf der Basis dieser und folgender Evaluationsstudien kontinuierlich weiterentwickelt und optimiert werden.

6.0 Referenzen

Carroll, J.M. & Mack, R. (1984): Learning to use a word processor: By doing, by thinking and by knowing. In: J. Thomas & M. Schneider (Eds.), *Human Factors in Computing Systems* (pp. 13-52). Norwood: Ablex

Chin, J. P.; Diehl, V. A.; Norman, K. L. (1988): Development of an instrument measuring user satisfaction of the human-computer interface. In: *Proceedings of SIGCHI '88* (pp. 213-218). New York: ACM/SIGCHI

DiMattia, S.; Oder, N. (1997): Knowledge Management: Hope, Hype or Harbinger?; In: *Library Journal*, 1997, September, S33-35.

Gediga, G.; Hamborg, K.-C. (1997): Heuristische Evaluation und IsoMetrics: Ein Vergleich. In: R. Liskowsky, B. M. Velichkovsky; W. Wünschmann (Hrsg.): *Software Ergonomie '97, Usability Engineering: Integration von Mensch-Computer-Interaktion und Software-Entwicklung* (S. 145-155). Stuttgart: Teubner

Gediga, G., Hamborg, K.-C. (1999): IsoMetrics: Ein Verfahren zur Evaluation von Software nach ISO 9241-10. In: H. Holling & G. Gediga (Hrsg.), *Evaluationsforschung* (S. 195-234). Göttingen: Hogrefe.

Gediga, G., Hamborg, K.-C., & Düntsch, I. (1999): The isometrics usability inventory: An operationalization of iso 9241-10 supporting summative and formative evaluation of software systems. *Behaviour & Information Technology*, 18, 151 - 164.

Gediga, G., Hamborg, K.-C. (2002): Evaluation in der Software-Ergonomie. *Zeitschrift für Psychologie*, 210, S.40-57, Göttingen: Hogrefe.

Igbaria, M.; Schiffman, S.J.; Wieckowski, T.J. (1994): The respective roles of perceived usefulness and perceived fun in the acceptance of microcomputer technology. *Behaviour & Information Technology*, 13, 349-361.

Kirakowski, J (1996): The Software Usability Measurement Inventory: background and usage. In: P. W. Jordan; B. Thomas, B. A. Weerdmeester; I. L. McClelland (Eds.): *Us*

ability Evaluation in Industry (pp. 169-177). London: Taylor & Francis.

Kirakowski, J.; Cierlik, B. (1998): Measuring the usability of web sites. In: *Human Factors and Ergonomics Society 42nd Annual Meeting*. Santa Monica, CA: HFES.

Molich, R., Ede, M. R., Kaasgaard, K., & Karyukin, B. (2004): Comparative usability evaluation. *Behaviour & Information Technology*, 23, 65 - 74.

Nielsen, J. (1993): *Usability Engineering*. Boston: AP Professional

Nielsen, J. (1994): Heuristic Evaluation. In: J. Nielsen; R. L. Mack (Eds.), *Usability Inspection Methods* (pp. 25-62). New York: J. Wiley & Sons

Nielsen, J. & Molich, R. (1990): Heuristic Evaluation of User Interfaces. In: J.C. Chew & J. Whiteside (Eds.), *Chi'90 Conference Proceedings, Empowering People* (pp. 249-256). New York: ACM Press.

Nielsen, J. (2001): The 3 Cs of Critical Web Use: Collect, Compare, Choose. Jakob Nielsen's Alertbox. www.useit.com/alertbox

Norman, D. A. (2004). *Emotional design: Why we love (or hate) everyday things*. New York, USA: Basic Books.

Novak, T.P.; Hoffman, D.L.; Yung, Y.F. (2000): Measuring the Customer Experience in Online Environments: a Structural Modelling Approach. *Marketing Science*, 19, 22-44.

Ollermann, F., Gruber, C. & Hamborg, K.-C. (2003): *Software-ergonomische Evaluation der Kursmanagementsysteme Blackboard, Stud.IP und WebCT*. Universität Osnabrück: Unveröffentlichter Arbeitsbericht.

Prümper, J.; Anft, M. (1993): Die Evaluation von Software auf Grundlage des Entwurfs zur internationalen Ergonomie-Norm 9241 Teil 10 als Beitrag zur partizipativen Systemgestaltung - Ein Fallbeispiel. In K.-H. Rüdiger (Ed.), *Software-Ergonomie '93* (S. 145-156). Stuttgart: Teubner.

Shneiderman, B., & Plaisant, C. (2005): *Designing the user interface* (4th ed.). Boston: Addison-Wesley.

Tourangeau, R. (2004): Survey research and societal change. *Annual Review of Psychology*, 55, 775-801.

Welker, M. (2001): E-mail surveys: Non-response figures reflected. In: U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of internet science* (pp. 231-238). Lengerich: Pabst.

»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 4th annual GC UPA Track Gelsenkirchen, September 2006* © 2006 German Chapter of the UPA e.V.«



