

Evaluation webbasierter E-Learning-Systeme mittels implizit erfasster Benutzerinteraktionen

Ludger Schmidt

Lehrstuhl und Institut für Arbeitswissenschaft
RWTH Aachen
Bergdriesch 27
D-52062 Aachen
L.Schmidt@iaw.rwth-aachen.de

Abstract: Dieser Beitrag stellt eine Methode zur Evaluation webbasierter Systeme vor, bei der im Hintergrund aufgezeichnete Benutzerinteraktionen verwendet werden. Welche technischen Rahmenbedingungen sind bei der Aufbereitung und Interpretation dieser Daten zu berücksichtigen? Wie lassen sich diese implizit weitergegebenen Informationen zur Evaluation webbasierter E-Learning-Systeme nutzen? Diesen Fragen soll im Folgenden im Detail nachgegangen werden. Parallel zur schrittweisen Vorstellung der Methode werden die jeweiligen Ergebnisse der einzelnen Arbeitsschritte exemplarisch anhand einer konkreten E-Learning-Plattform veranschaulicht.

1 Einleitung

1.1 Exemplarischer Evaluationsgegenstand

Als exemplarischer Evaluationsgegenstand soll das webbasierte System INTEGRAL II dienen, das von 15 Hochschulpartnern zur mediengestützten Vermittlung von arbeitswissenschaftlichen Lerninhalten aufgebaut wurde. Die Entwicklung, Erprobung und Evaluation dieses E-Learning-Angebotes wurde vom BMBF im Förderprogramm „Neue Medien in der Bildung“ bis Ende 2003 unterstützt. Ein Teil der lokalen Studienangebote sollte im Rahmen dieses Kooperationsprojektes so miteinander vernetzt werden, dass thematisch abgegrenzte, von den einzelnen Partnern entwickelte Lernmodule standortübergreifend von allen Partnern über eine Internet-Lernplattform in der Lehre verwendet werden können (vgl. [Sc03], [SR03]). Über die gemeinsame Plattform steht ein Spektrum von 20 Modulen zur Verfügung, das unter der Adresse <http://www.integral2.iaw.rwth-aachen.de> nicht nur von den Projektpartnern, sondern auch von weiteren Interessenten eingesetzt werden kann. Neben diesen Lernmodulen stellt die Plattform zur webbasierten Kommunikation Funktionen wie Chat, Email, Forendiskussionen und Videoconferencing zur Verfügung. Dies kann ebenfalls von allen Beteiligten, also sowohl von den Lernenden, als auch den Dozenten, zwecks standortübergreifender Kommunikation genutzt werden.

Die arbeitswissenschaftliche Hochschullehre soll damit in ganz unterschiedlichen Studiengängen unterstützt werden, schwerpunktmäßig in den Ingenieur-, Geistes- und Wirtschaftswissenschaften, aber auch für die Bereiche Design, Architektur und Informatik. Der didaktische Nutzungskontext reicht von der multimedialen Unterstützung der klassischen Präsenzlehre bis hin zum selbstgesteuerten und ortsunabhängigen Lernen der Studierenden. Hinsichtlich der technischen Umsetzung wurde ein zentraler Webserver am Institut für Arbeitswissenschaft der RWTH Aachen aufgebaut, der die Lernplattform mit ihren Inhalten und Kommunikationsfunktionen bereitstellt. Ein detaillierter Überblick zur Lernplattform INTEGRAL II sowie zum Inhalt und Einsatz der einzelnen Lernmodule und Funktionen wird in [LS05] gegeben.

1.2 Untersuchungsfragestellungen

Losgelöst von diesem konkreten Beispielsystem lassen sich bei webbasierten E-Learning-Systemen verschiedene Evaluationsfragestellungen formulieren, die mit der hier vorgestellten Methode untersucht werden können. Betrachtet man einen Nutzungsvorgang, so können hinsichtlich der Evaluationsziele webbasierter Systeme folgende drei Bereiche differenziert werden.

Zunächst stellt sich die Frage, woher die Benutzer kommen und wie sie zum System gelangen. Dahinter steckt auch die zentrale Fragestellung, ob und wie die potentielle Zielgruppe des Systems erreicht wird und wie die tatsächliche Nutzergruppe aussieht. Analysieren kann man zum einen die geographische und institutionelle Herkunft der Benutzer, zum anderen lässt sich betrachten, über welche Querverweise auf anderen Webangeboten die Benutzer auf das System hingewiesen wurden. Ist ihnen die Webadresse bekannt und haben sie sie manuell in ihren Browser eingetippt oder ihre Bookmarks genutzt? Oder sind die Benutzer über Suchmaschinen zum System gelangt? Wenn dies der Fall ist, geben die dabei verwendeten Suchbegriffe näheren Aufschluss über den individuellen Zweck des Besuches.

Der zweite Bereich von Fragestellungen befasst sich damit, wie das Angebot in der Realität genutzt wird. Wie hoch ist der Anteil der tatsächlichen Benutzer gemessen an der potentiellen Zielgruppe? Welche Inhalte werden von welchen Benutzergruppen ausgewählt und wie lange setzen sie sich jeweils damit auseinander? Wird also das gesamte Spektrum genutzt oder selektiv nur ein Teil? Außerdem kann die Nutzungshäufigkeit, ihre Dauer und Intensität betrachtet werden, um daraus Rückschlüsse auf die Akzeptanz des Systems zu ziehen. Lassen sich typische Nutzungsmuster erkennen? Stimmt das Navigationsverhalten mit den geplanten Verläufen überein? Oder geben die Abweichungen dazu Anlass, das System hinsichtlich der Vernetzung der Seiten umzugestalten? Welche Funktionsbereiche werden wie stark frequentiert? Wie und wann wird beispielsweise eine Hilfefunktion in Anspruch genommen? Ist an diesen Stellen durch Umgestaltung eine bessere Benutzbarkeit zu erzielen? Für die Bewertung der ermittelten Kennzahlen kann zum einen ein relativer Vergleich zwischen Benutzern vorgenommen, zum anderen können aber auch externe Daten herangezogen werden, um z. B. einen Vergleich mit den Kennzahlen anderer E-Learning-Systeme anzustellen.

Die Evaluation der technischen Funktionalität zählt auch zu diesem zweiten Bereich, soll hier aber nicht weiter verfolgt werden.

Drittens kann man über die Betrachtung der Ausstiegsseiten aus dem System versuchen, die Frage zu beantworten, ob der Benutzer beim Ausstieg sein Ziel erreicht hat, also ob am Ende einer Benutzersequenz zum Beispiel der erfolgreiche Abschluss einer Lerneinheit steht oder der Benutzungsvorgang schon vorher abgebrochen wurde. Wie viele Teilschritte wurden zurückgelegt? War der Besuch einmalig oder kehrt der Benutzer erneut zum System zurück? Gibt es Zusammenhänge zwischen der Systemnutzung und anderen externen Daten der Benutzer?

2 Methodenanwendung und Ergebnisse

Um diese Fragen zu beantworten, soll im folgenden Teil des Beitrags beschrieben werden, wie mit Logfiles implizit erfasste Benutzerinteraktionen für die Evaluation webbasierter E-Learning-Systeme herangezogen werden können. Dazu werden sieben Anwendungsschritte und deren Ergebnisse vorgestellt.

2.1 Datenaufnahme

Die als datentechnische Grundlage genutzten Logfiles sind automatisch generierte Aufzeichnungen von Objktanforderungen und Aktivitäten der Benutzer, die ohne Beeinträchtigung der Interaktionen registriert werden. Im Bereich „Webserver“ werden solche Protokolle sogar standardmäßig angelegt, so dass die Verwendung von Logfiles hier besonders verbreitet ist, z. B. im Extended Logfile-Format (vgl. [HB05]). Objektabrufe durch die Webbrowser der einzelnen Benutzer werden dabei zentral auf dem Server erfasst. Oft werden ein längerer Zeitraum und viele Benutzer aggregiert betrachtet.

Da bestimmte Informationen nicht im Server-Logfile auftauchen (Navigation der Benutzer innerhalb einer Seite; Nutzung aktiver Inhalte wie Java-Applets; Softwareoperationen wie Ausdrucken und Fensteröffnen; Anfragen, die nicht an den Server weitergegeben werden, weil sie durch einen lokalen Zwischenspeicher oder einen zwischengeschalteten Proxy-Server bedient werden können), sind keine exakten Werte für die Zugriffe messbar, sondern nur näherungsweise Schätzungen, es sei denn, der Zugriff auf den lokalen Speicher oder Proxy-Server wird durch spezielle Seitengestaltung verhindert.

Unmittelbar im Logfile enthalten ist eine Angabe über die zuvor betrachtete Webseite, ggf. mit den dort übergebenen Parametern. Dieser Referrer-Eintrag ist sehr nützlich, da er u. U. unmittelbar verrät, von welcher Seite der Benutzer einem Hyperlink gefolgt ist, um zum betrachteten webbasierten System zu gelangen (z. B. Suchmaschine und dort eingegebene Begriffe).

Eine erste Nutzungskenngröße ergibt sich aus der Zeilenanzahl des Logfiles. Je Zeile wird ein einzelner Objektabruf protokolliert, den man auch als „Hit“ bezeichnet. Die Anzahl der Hits wurde zunächst auch als Maß für die Popularität eines Webangebotes betrachtet, allerdings ist diese Zahl zu Vergleichszwecken nicht besonders aussagekräftig, da sie unabhängig von einer erfolgreichen Übertragung an den Browser des Benutzers ist und jeder einzelne Objektabruf gezählt wird. Besteht eine abgerufene Seite z. B. aus einem Text und zehn dort verankerten Grafiken, so werden 11 Hits protokolliert. Daher haben sich stattdessen andere Standardmaße etabliert, wie z. B. die Anzahl der PageViews (auch als PageImpressions bezeichnet), die den Abruf einer vollständigen Seite beschreibt, ohne alle verankerten Elemente zusätzlich zu zählen. Dieses Maß muss aber zunächst aus dem Logfile abgeleitet werden und steht erst nach entsprechender Aufbereitung zur Verfügung.

2.2 Auflösung der erfassten IP-Adressen über den Domain Name Service

Nach der Datenaufnahme befinden sich im Webserver-Logfile die numerischen IP-Adressen der Rechner, mit denen auf das Webangebot zugegriffen wurde. Über den Domain Name Service (DNS) können diese IP-Adressen meist in den Rechnernamen aufgelöst werden, so dass z. B. aus der IP-Adresse „131.234.169.104“ der Name „cafe-bohne4.uni-paderborn.de“ wird. Daraus lässt sich in gewissem Maße auf die Herkunft der Anfrage rückschließen. Vermutlich handelt es sich bei diesem Beispiel um einen Rechner eines Internet-Cafes innerhalb der Uni Paderborn. Mit dem Ergebnis dieses Arbeitsschrittes ist bereits eine erste Beurteilung möglich, ob die intendierte Zielgruppe der webbasierten Plattform überhaupt erreicht wurde.

2.3 Vorfilterung

Einige aufgelöste IP-Adressen geben Hinweise auf Pseudo-Benutzer, z. B. automatische Suchmaschinen (Spider, Robots, Wanderer), E-Mail-Adressensammler oder Hyperlink-Checker, die sich im nächsten Auswertungsschritt ausfiltern lassen. Diese generelle Vorfilterung umfasst aber auch weitere Störfaktoren (z. B. Zugriffe auf das Webangebot mit Fehlermeldungen, Zugriffe auf einzelne Grafiken innerhalb von Seiten und auf Seitenformatierungsangaben). Danach lassen sich Aussagen über die Anzahl der betrachteten Seiten (PageViews) und über die anfordernden Rechner machen.

Eine konkrete Zuordnung zwischen diesen Rechnern und den Benutzern des webbasierten Systems ist nicht unproblematisch, da zum einen die Vergabe von IP-Adressen aufgrund eines beschränkten Adress-Kontingents beim Anmelden im Netzwerk insbesondere bei Dial-In- und WirelessLAN-Zugängen meist dynamisch erfolgt, d. h. eine IP-Adresse wird zu unterschiedlichen Zeitpunkten jeweils anderen Rechnern zugewiesen (bzw. dem gleichen Rechner bei mehreren Sitzungen unterschiedliche IP-Adressen zugewiesen). Zum anderen haben Rechner von Netzwerken, die durch eine Firewall vom Internet abgeschirmt sind, nach außen die gleiche IP-Adresse (Network Address Translation). Außerdem kann der gleiche Benutzer unterschiedliche Rechner verwenden (z. B. an der Uni und zu Hause) oder unterschiedliche Benutzer den gleichen Rechner (z. B. in einem Computer-Pool).

Die Konsequenz ist, dass allein über die IP-Adresse im Webserver-Logfile im Allgemeinen keine eindeutige Identifizierung von Benutzern möglich ist.

Eine Lösungsmöglichkeit dieses Problem besteht darin, neben dem Webserver-Logfile weitere Protokolle zu verwenden, z. B. über die Benutzeranmeldungen beim webbasierten System. Dies war beim E-Learning-System INTEGRAL II im Untersuchungszeitraum obligatorisch, da der offene Zugang ohne Anmeldung über einen freien Gastaccount gesperrt wurde. Ergänzend wurde auch das Protokoll der Neuanmeldungen verwendet (Benutzerdatenbank mit Registrierungsdaten).

2.4 Sessionizing

Das Logfile enthält einzelne Objektabrufe verschiedener Clients. Mit dem Sessionizing bezeichnet man das Zusammenfassen einer Folge von Abrufen eines Clients zu einem Visit. Ein Visit bezeichnet also einen zusammenhängenden Nutzungsvorgang durch einen Client.

Das Sessionizing kann durch Tracking der spezifischen Zusatzinformationen wie Browser und Betriebssystem und des Referrers erfolgen oder indem ein bestimmtes Zeitfenster betrachtet wird. Erfolgt innerhalb dieses Zeitfensters kein weiterer Zugriff, wird ein Visit als beendet angesehen, wobei üblicherweise eine Standardruhezeit von 10 Minuten zugrunde gelegt wird. Eine weitere Möglichkeit besteht in der Vergabe einer Session-ID für jeden Nutzungsvorgang, die an die Webadresse angehängt wird.

Mit den Kennzahlen „PageViews“, „Visits“ und „Zeit pro Visit“ sind nun periodische Zusammenfassungen zu zeitlichen Nutzungsstatistiken möglich. Ein Tagesverlauf zeigt die Nutzung während Arbeits- und Freizeitstunden oder, dass internationale Besucher nachts auftauchen. Beim Wochenverlauf wird die unterschiedliche Nutzung werktags und am Wochenende deutlich und der Jahresverlauf zeigt die Jahreszeiteinflüsse und langfristige Trends auf. In diesen zeitlichen Nutzungsstatistiken wird auch der Einfluss externen Faktoren sichtbar (Wetter, regionale Feiertage, Sport und andere Ereignisse, Marketing-Maßnahmen usw.).

Die Gesamtnutzung der E-Learning-Plattform in Abhängigkeit von der Zeit ist in Abbildung 1 dargestellt. Die Anzahl der registrierten Benutzer (Abbildung 1 oben) lag während der Projektlaufzeit in den letzten Monaten relativ konstant bei etwa 380. Nach einer Aktion zur Öffentlichkeitsarbeit Ende Dezember, bei der ca. 80 Institute angeschrieben wurden, sind Anfang Januar steigende Registrierungszahlen zu finden. Momentan sind zwar über 700 Nutzer registriert, aber möglicherweise sind damit nur einmalige Nutzungsvorgänge verbunden gewesen.

Ein wesentlich differenzierteres Bild für die Nutzungshäufigkeit des Systems liefert der Verlauf der Anzahl der Visits pro Woche wie er im Vergleich in Abbildung 1 unten aufgetragen ist. Hier wird der Nutzungsrückgang während der Weihnachtsferien deutlich und der sprunghafte Anstieg nach der oben beschriebenen Marketing-Maßnahme.

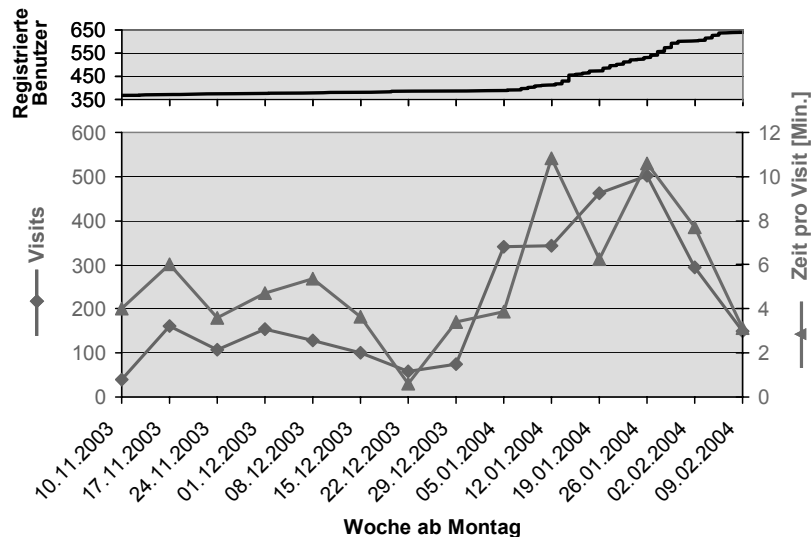


Abbildung 1: Nutzung des Gesamtsystems nach Wochen
(93-Tage-Zeitraum Anfang 2004, 2.916 Visits, 112.027 PageViews)

Betrachtet man zusätzlich die Zeit pro Visit als Maß für die Nutzungsintensität (vgl. [FG05]), so muss man allerdings konstatieren, dass diese zunächst bis Mitte Januar noch sehr niedrig war. Viele der neuen Benutzer haben sich also nur einen kurzen Eindruck verschafft, ehe sie die Lernplattform wieder verlassen haben. Die Steigerungen ab Mitte Januar werden weiter unten gesondert behandelt.

Die Nutzung des Gesamtsystems lässt sich auch in Abhängigkeit von der Uhrzeit betrachten. Beim Beispielsystem ist auffällig, dass die Nutzungshäufigkeit gemessen in Visits zwischen 15 und 16 Uhr ihr Maximum erreicht. Nimmt man die Nutzungsdauer pro Visit als Maß für die Intensität, so ist diese in den Abendstunden zwischen 21 und 0 Uhr deutlich größer als tagsüber.

Eine ähnlich inhomogene zeitliche Nutzungsverteilung der E-Learning-Plattform ergibt die Betrachtung nach Wochentagen. Montags bis donnerstags sind viele Visits zu verzeichnen, am Wochenende, insbesondere sonntags, ist mehr Zeit pro Visit zu messen, so dass man hier eine größere Nutzungsintensität feststellen kann.

2.5 Semantische Kodierung und Festlegen von Kategorien für die Auswertung

Wie oben beschrieben enthalten Logfiles Low-Level-Daten des Nutzungsverhaltens. Protokollierte Ereignisse wie Seitenabrufe und Interaktionen über Benutzungs-elemente wie ein Mausklick hatten jedoch in unterschiedlichen Bereichen möglicherweise jeweils eine andere Intention (vgl. [HR00]). Um die Frage nach der spezifischen Bedeutung der Benutzerinteraktionen besser beantworten zu können, werden die einzelnen Seiten inhaltlich semantischen Kategorien zugeordnet, deren Auswahl unmittelbar vom Evaluationsziel abhängt.

Setzt man sich z. B. die Modellierung und Analyse des Navigationsverhaltens über Zustands-Transitions-Diagramme oder Entscheidungsbäume zum Ziel, können Navigationssequenzen die Übergänge zwischen den Kategorien beschreiben (vgl. [HR00]). Hinter einer Sequenz könnte dann die Bedeutung liegen, dass sich ein Teil der Benutzer von einer Überblickskategorie A zum Detailbereich B bewegt, ein anderer Teil zur Unterkategorie C, dann gehen beide Gruppen ohne weitere Verzweigung jeweils tiefer in die folgenden Details usw. Mit Übergangswahrscheinlichkeiten versehen lassen sich so typische Verhaltensmuster identifizieren. In Untersuchungen werden daran z. B. bestimmte Vorgehensstrategien festgemacht (zielgerichtete Suche, zweckgebundenes Browsen oder exploratives Stöbern). Es zeigt sich dabei eine starke Abhängigkeit von den Gestaltungsmerkmalen des Angebotes (Charakter des Hypertextes, Hierarchietiefe, Navigationshilfen, Suchmöglichkeiten).

Untersucht man als semantische Kategorien verschiedene Funktionsbereiche, könnte man bei einer E-Learning-Plattform z. B. zwischen Login, rezeptivem Lernbereich, interaktiven Übungen, Hilfe, Kommunikationsbereich und Bookmarkfunktion unterscheiden. Eine Untersuchungsfragestellung wäre, ob sich damit Gruppen von Benutzern, z. B. verschiedene Lerntypen, identifizieren lassen.

Außerdem könnte sich eine Zuordnung auch am Inhalt orientieren. Als semantische Kategorien wurden bei der E-Learning-Plattform 17 einzelne thematische Lernmodule, zwei Querschnitts-Kategorien mit Methodensammlungen und eine Kategorie „Self-Assessment“ mit alten Prüfungsaufgaben verwendet. Nicht den Kategorien zuzuordnende Interaktionen (Forendiskussionen, Email, Videochat etc.) werden im nächsten Schritt von der Analyse ausgeschlossen. So lassen sich dann zusammen mit den Nutzungszeiten Hinweise auf die Interessen der Benutzer identifizieren.

Die Zuordnung der semantischen Kategorien lässt sich aufgrund der damit verbundenen Interpretation kaum automatisieren und muss „von Hand“ vorgenommen werden.

2.6 Zielfilterung

Nach dieser Zuordnung erfolgt im nächsten Schritt die Zielfilterung, die neben den Kategorien noch weitere Nebenbedingungen berücksichtigt.

Auf Basis einer Anforderungsanalyse wurden für die E-Learning-Anwendung INTEGRAL II vier didaktische Kernszenarien abgegrenzt (Präsentationshilfe für Dozenten in einer Präsenzveranstaltung, Aufgabenbearbeitung durch die Lernenden in einer Präsenzveranstaltung, orts- und zeitunabhängige Aufgabenbearbeitung durch die Lernenden sowie orts- und zeitunabhängige Bearbeitung komplexer Fallbeispiele). Das Szenario der orts- und zeitunabhängigen Aufgabenbearbeitung durch die Lernenden soll hier im Hinblick auf die E-Learning-gestützte Vertiefung des Vorlesungsstoffs und die gezielte Prüfungsvorbereitung untersucht werden.

Außerdem wurden anhand der Anforderungsanalyse fünf Benutzerrollen definiert (Lernende, Dozenten, Autoren, Administratoren und Gäste, die ohne Registrierung auf ein eingeschränktes Angebot zugreifen können).

Um hier die Zielgruppe der Lernenden herauszugreifen genügte die Sperrung des Gastaccounts und ein Untersuchungs-Zeitintervall am Ende der Projektlaufzeit jedoch nicht. Die Nutzerrollen ließen sich anhand der Protokolle nicht eindeutig zuordnen. Außerdem war die dezentral organisierte Nutzung an den einzelnen Standorten der Projektpartner zu unterschiedlich, so dass man auch nicht von kontrollierten Bedingungen sprechen konnte.

Daher erfolgte die Analyse mit der nicht am Projekt beteiligten Uni Paderborn als neutralem, nicht projektbeeinflussten Versuchsfeld. Hier gab es also auch keine Autoren und Betreuer der Lernmodule sowie keine Administratoren. Betrachtet wurde eine Grundlagenvorlesung „Arbeitswissenschaft“, die inhaltlich in etwa dem Spektrum der INTEGRAL-II-Lernmodule entspricht. Etwa 230 Studierende der Studiengänge Wirtschaftsingenieurwesen/Maschinenbau bzw. Elektrotechnik sowie Diplom Maschinenbau besuchten die Vorlesung, 201 Klausuranmeldungen lagen vor und schließlich gab es 173 Klausurteilnehmer. Es bestand eine prinzipielle Zugangsmöglichkeit für alle Studierenden über Ressourcen der Universität (WLAN, Computer-Pools, Dial-In-Zugänge). Diese kontrollierten Bedingungen ließen eine zielgenaue Filterung zu.

Es gibt einen klar abgegrenzten Nutzungszeitraum, der mit einer anfänglichen Info über das System an die Studierenden am 16.01.2004 beginnt. Zuvor war dieser Zielgruppe das webbasierte System unbekannt, und es gab keine registrierten Nutzer aus Paderborn. Im Hinblick auf das dritte Lernszenario erfolgte kein Einsatz von INTEGRAL II als Präsentationshilfe oder in einer Übung. Der Zeitraum endet mit der Klausur am 30.01.2004 nachmittags. Nach der Zielfilterung wird ab hier nur noch die Nutzung der Lernmodule durch diese Benutzergruppe im oben genannten Untersuchungszeitraum betrachtet (Abbildung 2).

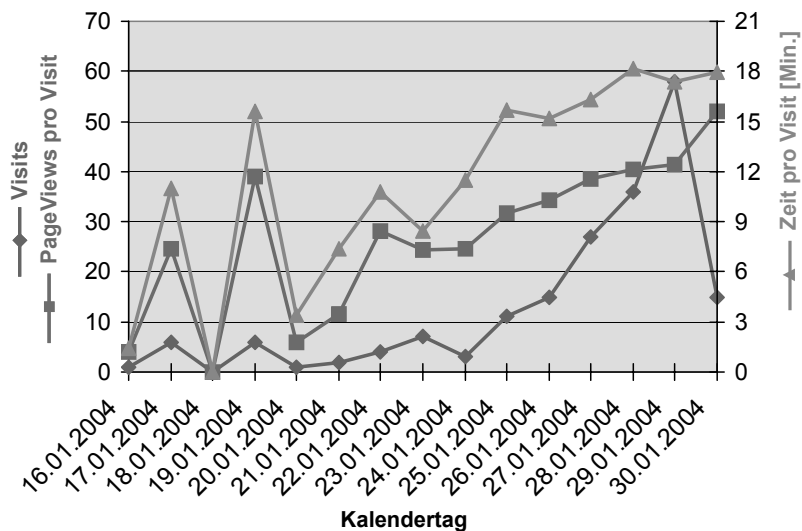


Abbildung 2: Nutzung der Lernmodule durch Klausurteilnehmer nach Kalendertag (15-Tage-Zeitraum, 76 Besucher, 192 Visits, 7.319 PageViews)

Man erkennt deutlich die Zunahme der Visits mit näherrückender Prüfung. Der Durchschnitt lag bei 12,8 Visits pro Tag. Während dieser 15 Tage konnten 76 unterschiedliche Besucher registriert werden, die durchschnittlich 2,5 Visits machten, 26 Besucher machten nur einen Visit. Die Anzahl der PageViews als Maß für die Nutzungsbreite schwankte stark und nahm auch tendenziell im Verlauf zu. Durchschnittlich fanden 38 PageViews pro Visit statt. Eine Seite wurde im Schnitt 25 Sekunden lang betrachtet. Die Anzahl der PageViews korrelierte stark mit der Zeit pro Visit, die ebenfalls in Abbildung 2 dargestellt ist. Zusammenfassend kann man also sagen, dass sowohl die Nutzungshäufigkeit als auch die Nutzungsintensität bis einen Tag vor der Klausur stark anstieg, dann brach die Anzahl der Visits wieder zusammen.

Differenziert man bei gleichem Untersuchungsfokus nach Uhrzeiten, so ergibt sich eine Aussage, zu welcher Tageszeit die Studierenden sich am meisten mit dem webbasierten System beschäftigt haben. Die Anzahl der Besuche war zwischen 10 und 20 Uhr am größten, aber auch nachts war noch eine intensive Auseinandersetzung mit dem System zu verzeichnen.

Von den 20 oben genannten Lernmodulen wurden 18 genutzt und sehr klausurzielorientiert betrug die Nutzungsdauer des Self-Assessments 85 % der Gesamtnutzungszeit. Die thematische Lerneinheit „Augmented Reality“ kam mit 5 h noch auf einen Anteil von 10 %, während sich die restlichen Lernmodule die verbleibenden 5 % teilten. Betrachtet man also die Interessenslage der Studierenden anhand der Verteilung der Nutzungszeit, so lässt sich der klare Fokus auf den Klausurselbsttest erkennen. Darüber hinaus schien Augmented Reality im Vergleich zu allen anderen thematischen Lernmodulen ein deutlicher Interessensschwerpunkt zu sein.

2.7 Prüfen von Zusammenhängen mit externen Daten

Im letzten Schritt werden nun zusätzlich zur Systemnutzung externe Daten herangezogen, um mögliche Zusammenhänge zu identifizieren.

Hierzu wurde das Klausurergebnis verwendet, bei dem ein Kolmogorov-Smirnov-Test zeigt, dass die Normalverteilungshypothese nicht zurückgewiesen werden sollte ($p=0,50$). Im Mittel wurden von den 173 Klausurteilnehmern 23,7 von 60 Punkten erreicht. Von diesen Teilnehmern haben 76 das webbasierte System genutzt, was einem Anteil von 44 % an den potentiellen Benutzern entspricht. Man kann nun anhand dieses Kriteriums die Verteilung in zwei Gruppen aufsplitten (Abbildung 3). Beide Untergruppen (links die Nutzer, rechts die Nicht-Nutzer) zeigen wieder Normalverteilungseigenschaften (Kolmogorov-Smirnov-Test, $p=0,59$ bzw. $p=0,45$).

Bezogen auf die Mittelwerte der beiden Verteilungen zeigt sich anhand des T-Tests ein hochsignifikanter Unterschied, da die Mittelwertgleichheitshypothese der Grundgesamtheit mit Irrtumswahrscheinlichkeit $p=0,004$ zurückgewiesen werden kann. Die Nutzer des Systems erreichten im Mittel 26,1 Punkte in der Klausur, die Nicht-Nutzer nur 21,8. Im Mittel wurden also von den Nutzern 20 % mehr Klausurpunkte erreicht (95%-Konfidenzintervall für die Mittelwertdifferenz: [7 % ; 33 %]).

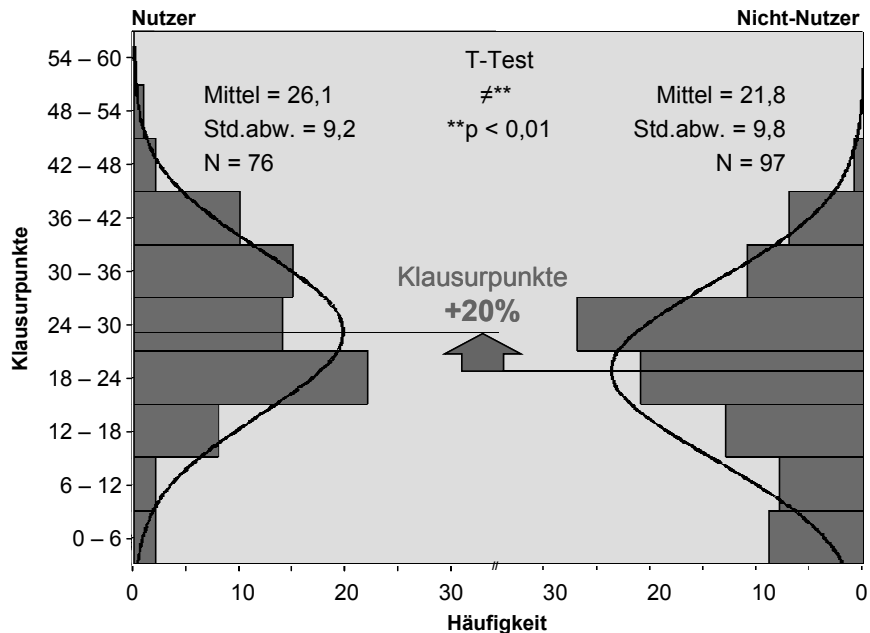


Abbildung 3: Ergebnis der Klausur nach Nutzung des Systems

In einer weiteren Studie, die ebenfalls mit dem INTEGRAL-II-System durchgeführt wurde [CK05], wurden vergleichbare Ergebnisse festgestellt. Hier ergab sich bei einer Klausur mit 105 Teilnehmern ein hochsignifikanter Unterschied ($p < 0,01$) zwischen den in etwa gleich großen Gruppen der Nutzern und der Nicht-Nutzer des Lernmoduls „Zeitermittlungsverfahren“, wobei allerdings die Nutzung durch Befragung und nicht per Logfile-Analyse erhoben wurde. Die Nutzer des Lernmoduls erreichten insgesamt durchschnittlich 15 % mehr Punkte als die Nicht-Nutzer. Bei den nicht im INTEGRAL-II-Lernmodul aufgegriffenen Themenbereichen der Klausur waren hingegen keine signifikanten Unterschiede zwischen beiden Gruppen zu verzeichnen. Dies stützt die Hypothese, dass sich die unterschiedlichen Ergebnisse der beiden Gruppen nicht nur auf eine insgesamt bessere Vorbereitung oder höhere Motivation der E-Learning-Nutzergruppe zurückführen lassen, sondern auch ursächlich mit der INTEGRAL-II-Nutzung zusammenhängen.

Zu einem anderen Ergebnis kommt hingegen eine Studie, die den Einsatz eines E-Learning-Moduls an der FH Berlin und der TU Berlin untersuchte [Sch03]. Die hierbei eingesetzte Software „E-Kreide“ gab das Tafelbild und den Vortrag des Lehrenden zeitgleich zur Präsenzveranstaltung auch im Internet wieder und konnte dort auch asynchron und wiederholt abgerufen werden. In dieser Studie ergab sich im Klausurergebnis kein signifikanter Zusammenhang zwischen der Nutzung des Systems und der Klausurleistung, was jedoch auch durch die unveränderte Form der Inhaltsaufbereitung begründet sein kann.

Bei der INTEGRAL-II-Plattform sind die Inhalte der Lernmodule speziell auf das E-Learning-Medium zugeschnitten, während die Lerninhalte beim E-Kreide-Projekt zum Großteil aus der nicht modifizierten Wiedergabe der aufgezeichneten Präsenzveranstaltung bestanden. Dies führte außerdem bei einem Teil der Nutzer zu technischen Schwierigkeiten beim Zugriff auf das bereitgestellte Lernmaterial.

3 Zusammenfassung

Zum Schluss der Betrachtungen sollen die hier vorgestellten exemplarischen Ergebnisse zusammenfassend den eingangs erwähnten Evaluationsfragestellungen zugeordnet werden:

Woher und wie gelangen die Benutzer zum betrachteten System? Unmittelbar aus dem Logfile ist über die Referrer-Angabe feststellbar, ob eine direkte Adresseingabe oder ein Bookmark, ein externer Hyperlink oder eine Suchmaschine einschließlich der verwendeten Begriffe auf das webbasierte System geführt hat. Nach dem DNS-Look-Up ergeben sich Anhaltspunkte für die geographische oder institutionelle Herkunft der Benutzer und damit auch Informationen über das Erreichen der Zielgruppe.

Wie wird das Angebot genutzt? Nach der Vorfilterung und dem Sessionizing lässt sich die Systemnutzung anhand von Häufigkeiten (Anzahl Visits) und Intensitäten (Zeit pro Visit, Zeit pro PageView) beurteilen. Die semantische Kodierung der Interaktionen schafft die Grundlage für Aussagen zum Navigationsverhalten und der Frequentierung verschiedener Inhalte und Funktionsbereiche hinsichtlich Spektrum und Zeitanteil. Nach der Zielfilterung lassen sich einzelne Benutzergruppen, die durch sie selektierten Kategorien und die damit verbundenen Bearbeitungszeiten, oder der Anteil der Nutzer an der Zahl der potentiellen Nutzer näher betrachten. Insgesamt lassen sich diese Nutzungsdaten verwenden, um Rückschlüsse auf die Akzeptanz des Systems zu ziehen, Gestaltungspotentiale genauer zu lokalisieren und zu begründen, relative Vergleiche zwischen Benutzergruppen zu ermöglichen und die Kennzahlen mit anderen Systemen zu vergleichen.

Wann und wie verlassen die Benutzer das System? Der Anteil einmaliger und wiederkehrender Besucher lässt sich feststellen. Anhand der Ausstiegsseite lässt sich möglicherweise der Abbruch oder die erfolgreiche Nutzung beurteilen. Wenn man externe Daten hinzuzieht, lassen sich weitere Zusammenhänge mit dem Ausstiegszustand der Benutzer feststellen, wie die signifikanten Zusammenhänge zwischen Systemnutzung und Klausurerfolg exemplarisch deutlich machten.

Beim hier vorgestellten Datenaufnahmeverfahren handelt sich um ein nicht-reaktives Beobachtungsinstrument für Interaktionen von Benutzern mit Prototypen und Software-Produkten. Insbesondere server-seitige Logfiles zeichnen sich durch den geringen Erhebungsaufwand für objektive quantitative Daten aus, die mit anderen Verfahren nicht erhoben werden können.

Die flexible semantische Kodierung des implizit erfassten Benutzerverhaltens lässt vielfältige Untersuchungen zu und bestimmt im Wesentlichen auch den Auswertungsaufwand. Anhand der realen Nutzung webbasierter Systeme lassen sich Gestaltungsalternativen im Feld und Labor vergleichen und deren Erfolg in der Praxis bewerten.

Literaturverzeichnis

- [CK05] Corth, C.; Kirchner, J.H.: Zeitermittlung für Arbeitsabläufe. In (Luczak, H.; Schmidt, L. Hrsg.): E-Learning Kooperation in der Arbeitswissenschaft. Ergonomia Verlag, Stuttgart, 2005; S. 107-113.
- [FG05] Fuller, R.; de Graaf, J.J.: Measuring User Motivation from Server Log Files. <http://www.microsoft.com/usability/webconf/fuller/fuller.htm> (21.02.2005).
- [HB05] Hallam-Baker, P.M.; Behlendorf, B.: Extended Log File Format : W3C Working Draft WD-logfile-960323. <http://www.w3.org/TR/WD-logfile.html> (21.02.2005).
- [HR00] Hilbert, D.M.; Redmiles, D.F.: Extracting Usability Information from User Interface Events. In: ACM Computing Surveys 32 (2000), Nr. 4; S. 384-421.
- [LS05] Luczak, H.; Schmidt, L. (Hrsg.): E-Learning-Kooperation in der Arbeitswissenschaft. Ergonomia Verlag, Stuttgart, 2005.
- [SR03] Schmidt, L.; Rötting, M.: Vernetzung arbeitswissenschaftlicher Hochschullehre - Das E-Learning-Kooperationsprojekt Integral II. In (Strasser, H.; Kluth, K.; Rausch, H.; Bubb, H. Hrsg.): Qualität von Arbeit und Produkt in Unternehmen der Zukunft. Ergonomia Verlag, Stuttgart, 2003; S. 1085-1088.
- [Sc03] Schmidt, L.: Webbasiertes Hochschul-E-Learning - Deutschlandweite Kooperation in der Arbeitswissenschaft. In (Beck, U.; Sommer, W. Hrsg.): Tagungsband der Learntec 2003 - 11: Europäischer Kongress und Fachmesse für Bildungs- und Informationstechnologie, Karlsruhe 2003. Holler Verlag, Karlsruhe, 2003; S. 201-206.
- [Sch03] Schulte, J.: Evaluation des Einsatzes der Software „E-Kreide“ in der universitären Lehre. Magisterarbeit, TU Berlin, Fakultät I - Geisteswissenschaften - Institut für Sprache und Kommunikation, 2003; <http://kazan.inf.fu-berlin.de/echalk/docs/jschulte2003.pdf>. (25.02.2005).