

# Swordfish: Eine Software zur asynchronen Remote-Evaluation der Usability von Webseiten

**Prof. Dr. Norbert Gerth**  
Fakultät Informatik  
Hochschule Augsburg  
Postfach 11 06 05  
86031 Augsburg  
gerth@informatik.fh-augsburg.de  
www.hs-augsburg.de

**Thomas Krach**  
allevia GmbH  
Rehmstraße 4  
86161 Augsburg  
t.krach@allevia.de  
www.allevia.de

**Florian Pilottek**  
Fakultät Informatik  
Hochschule Augsburg  
Postfach 11 06 05  
86031 Augsburg  
florian@pilottek.de  
www.hs-augsburg.de

## Abstract

Anspruch dieses eigenen Ansatzes ist es, das Grundprinzip des klassischen Usability Tests im Labor in eine Online-Anwendung zu „portieren“. Auf diese Weise soll ein Remote-Testing ermöglicht werden, welches die Charakteristik und die Vorteile des klassischen Labor-

tests weitgehend erhält und die Vorteile aus der unmittelbaren Online-Umgebung (z.B. geringe Rekrutierungs- und Durchführungskosten, hohe Nutzeranzahl, Nutzer unmittelbar aus der Zielgruppe) nutzt.

## Keywords

Asynchron, Remote-Evaluation, Usability-Test, Website-Evaluation

## 1.0 Einleitung

Der Hintergrund des Konzeptes der Software war eine Übertragung des bewährten Usability-Tests im Labor auf das Internet. Der Ablauf sollte dabei automatisiert werden, woraus sich ein asynchroner Remote-Test ergibt. Aus technischer Sicht waren die Voraussetzungen an die Software die einer Webanwendung, also kein Medienbruch zum Testobjekt. Somit mussten einerseits Wege gefunden werden, die Vorteile des Labortests auf das Internet zu übertragen und dabei entstehende Nachteile auszugleichen. Dabei standen geringe Kosten des Tests und eine hohe Effizienz im Vordergrund.

Das Einsatzgebiet der Software sind funktionstüchtige Prototypen oder fertige Websites. Diese Software ermöglicht eine permanente Evaluation der Website und unterstützt ein iteratives Design.

Die Software stellt eine Kombination bekannter Internet-Techniken dar, worauf das eigene Konzept beruht. Durch die gezielte Online-Rekrutierung geeigneter Testpersonen direkt von der zu testenden Webseite, lassen sich alle realen Zielgruppen erreichen. Dabei gibt

es keine Einschränkung bzgl. des Aufenthaltsortes. Der Nutzungskontext entspricht genau dem, in dem die Webseite normalerweise auch verwendet wird.

Da die Software selbst eine Webanwendung ist, werden außer einem Webserver und einer Datenbank keine weiteren Anforderungen gestellt. Somit kann die Software sehr kostengünstig eingesetzt werden und es bedarf keines weiteren technischen Aufwands.

Ziel ist es ein Tool zu erstellen, das, ähnlich wie z.B. Google Analytics, einfach in die Zielseite einzubauen ist und ein Webinterface für die Auswertung bietet.

Der komplette Testablauf findet dabei On-Site statt, so dass die Testpersonen die eigentliche Webseite nie verlassen müssen und auch weiterhin die volle Kontrolle über den Browser und den Besuch der Seite haben.

## 2.0 Testablauf

Die fünf Ablaufphasen orientieren sich an denen eines Labortests:

1. Entwicklung Persona und Aufgabenszenario

2. Rekrutierung und Motivation
3. Beobachtung und Feedback
4. Fragebogen (ISO-Norm oder Szenario bezogen)
5. Auswertung und Ergebnisse

Zu Beginn des Tests werden die allgemeinen Rahmenbedingungen wie Ein- und Ausstiegsseite, Zeitraum und Fallzahl festgelegt. Für die genaue Filterung der gewünschten Testpersonen durch das Online-Panel stehen technische Informationen wie der Referrer, Browser, Sprache, bereits besuchte Seiten der Website oder wiederkehrende Benutzer zur Verfügung. Im Rahmen des Pre-Test-Fragebogens, der bewusst auf ein Minimum reduziert wurde, findet dann die endgültige Auswahl anhand der Nutzerangaben bzgl. des Geschlechts, Alters, Erfahrungsstandes mit der Website, sowie des eigenen Zieles statt. Durch die sehr gering gehaltene Einstiegshürde wird versucht, auch neue Nutzergruppen zu erreichen die, i.d.R. nicht an Umfragen oder Tests teilnehmen, da ihnen der Aufwand zu groß ist oder die Motivation dazu fehlt. Genau an diesem Punkt setzt der virtuelle Moderator an, der die eigentliche Aufgabe des Testleiters im Labor übernimmt. Er leitet dabei die Testperson durch den kompletten

Test, angefangen von der Begrüßung (es wird z.B. eine Belohnung in Form eines 5 oder 10 € Gutscheins für Amazon in Aussicht gestellt) bis hin zum abschließenden Fragebogen. Dieser Moderator ist dabei eine reale Person in Form von Bildern oder einzelnen Videosequenzen, die über Flash abgespielt werden.



Abb 1: Virtueller Moderator

Somit ermöglicht die Software eine gezielte Filterung bestehender Nutzer der Website und Rekrutierung von geeigneten Probanden, die auf der Website eine reale Aufgabe erfüllen.

Erklärt sich der Nutzer dazu bereit, an dem Test teilzunehmen, und erfüllt die entsprechenden Kriterien der gewünschten Zielgruppe, erfolgt eine kleine Einweisung in den Ablauf des Tests.

Während des Tests kann der Proband sich frei auf der Seite bewegen und seiner Aufgabe nachgehen. Der Moderator ist dabei permanent an einer passenden Stelle mit Hilfe eines Layers über der Seite eingeblendet. Möchte der Proband ein Feedback abgeben oder den Test beenden, kann er dieses durch einen Klick erreichen. Das Feedback ist in drei Schritte aufgeteilt. Zuerst ordnet die Testperson das Feedback in negativ oder positiv ein. Die Skalierung ist dabei ohne feste Stufen angelegt, da sich bereits in anderen Studien gezeigt hat, dass die Einordnung der Testpersonen nicht aussagekräftig ist (vgl. auch And

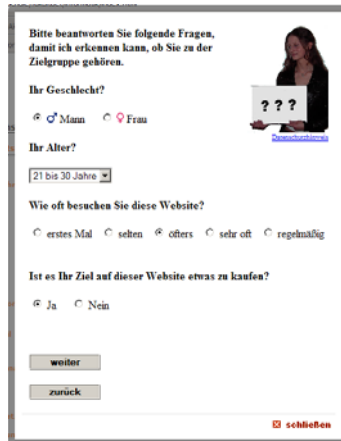


Abb 2: Pre-Test

reasen et al. 2007) und dient primär der Erfassung der gefühlten Bewertung durch den Probanden.

Im zweiten Schritt besteht die Möglichkeit, das Feedback zu kategorisieren. Die verschiedenen Kategorien basieren dabei auf den 10 Heuristiken von Jakob Nielsen und wurden zu sechs Aussagen umformuliert, die sich auf Funktionalität, Text, Sprache und Inhalt beziehen. Eine Mehrfachauswahl ist dabei möglich. Der dritte Schritt ist ein Freitextfeld, in dem der Proband seine Gedanken in schriftlicher Form abgeben kann. Das Feld ist am Anfang drei Zeilen hoch, um einen gewissen Umfang an Text zu fordern, aber auch nicht größer, um den Nutzer nicht abzuschrecken. Das Feld passt die Höhe automatisch während des Schreibens an, falls der Proband mehr eingeben möchte, und verhindert so ein sonst notwendiges Scrollen innerhalb des Textfeldes. Die Kategorien sind je nach dem, ob positives oder negatives Feedback abgeben wird, verschieden.

Die negativen Kategorien sind wie folgt:

- Direkter Link zu Seite ... fehlt
- Der Inhalt ist unübersichtlich
- Bedienung ist unverständlich und kompliziert

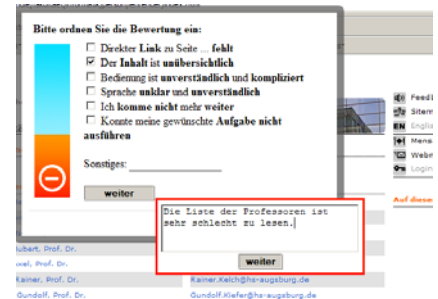


Abb 3: Feedback

- Sprache unklar und unverständlich
- Ich komme nicht mehr weiter
- Konnte meine gewünschte Aufgabe nicht ausführen
- Sonstiges (freies Textfeld)

Jeder Nutzer, der die Website das erste Mal besucht, wird als neuer Nutzer erkannt und mit einer eindeutigen ID versehen. Alle Seitenaufrufe werden dann gespeichert, was einerseits für die Filterung passender Testpersonen zu Beginn des Tests notwendig ist und während des Tests zur Aufzeichnung des Nutzerverhaltens. Bei den Probanden werden zusätzlich der Inhalt der Webseite und alle Interaktionen (Maus und Tastatur) chronologisch gespeichert. Mit Hilfe dieser Daten kann dann später eine Wiedergabe ähnlich einer Videoaufzeichnung rekonstruiert werden.

Die Software sieht zusätzlich zu dem freiwilligen Feedback noch die Möglichkeit vor, von dem Nutzer ein Feedback zu fordern, sofern bestimmte Kriterien wie Verweildauer, Maus- und Scrollwege einen definierten Sollwert überschreiten. Damit dies nicht ständig zu einer Beeinträchtigung des Probanden führt, wirken bereits abgegebene Feedbacks oder Formulareingaben diesem Mechanismus entgegen.

Hat die Testperson den Ausstiegspunkt des Tests erreicht oder diesen selber beendet, leitet der Moderator den Probanden durch den abschließenden Fragebogen.

Die Standardfragen basieren auf dem ISO-Norm Fragebogen (Prümper & Anft 1993), der entsprechend für Websites umformuliert wurde. Die Fragen wurden in drei Blöcke zu je zwei Abschnitten und je drei Fragen aufgeteilt. Ähnlich wie der ISO-Norm Fragebogen werden dabei die jeweiligen Abschnitte immer mit einem erklärenden Satz eingeleitet, um dem Probanden den Hintergrund der Fragen näher zu bringen. Dadurch soll eine Art Dialog zwischen der Testperson und dem Moderator vermittelt werden. Die Beantwortung der Fragen ist in fünf Stufen von absolut „nicht zutreffend“ bis „trifft zu“ aufgeteilt. Beantwortete Fragen werden grau markiert, was der Testperson eine Rückmeldung über den Stand der Beantwortung gibt. Der Proband kann auch nach Belieben zwischen den drei Seiten des Fragebogens vor- und zurückgehen. Die wichtigen Begriffe in den Fragen sind hervorgehoben, was ein schnelleres Erfassen der Aussage unterstützt.

Am Ende des Tests bedankt sich der Moderator für die Teilnahme und überreicht den Gutschein. Wenn die Testperson eine Kontaktaufnahme wünscht, kann sie an dieser Stelle auch ihre E-Mail Adresse hinterlassen.

Informationen wie wiederkehrende Besuche und Seitenaufrufe dieser Testperson werden auch weiterhin aufgezeichnet und ermöglichen so eine Beobachtung des konkreten Verhaltens dieses Nutzers über den eigentlichen Test hinaus.

### 3.0 Auswertung und Ergebnisse

Die gesammelten Daten werden in drei verschiedenen Auswertungen auf

bereitet, die sich an den drei Dimensionen einer Website orientieren.

- Gesamteindruck



Abb 4: Post-Test Fragebogen

- Content-Design
- Informationsarchitektur

Bei der ersten Auswertung handelt sich um die Ergebnisse des Fragebogens, die Rückschlüsse auf die ganze Website zulassen.

Dabei werden der Mittelwert, die Standardabweichung, absolute und relative Werte pro Frage ausgegeben. Die Software bietet die Möglichkeit den Fragebogen entsprechend anzupassen, um so bei den einzelnen Tests genauere Ergebnisse bzgl. des Test-



Abb 5: Auswertung Fragebogen

ziels zu ermitteln. Die zweite Auswertung ist der Replay-Modus, in dem der Test ähnlich einem Video abgespielt wird. Dabei sind die Mausbewegungen und Tastatureingaben der Testperson zu sehen. Mausklicks werden in Form von immer (chronologisch) größer werdenden Kreisen dargestellt. Zudem können aktivierte Elemente der Webseite farblich markiert werden, um zu erkennen, welche Bereiche genutzt und welche nicht genutzt wurden. Unter dem Wiedergabefenster werden die

vom Probanden abgegebenen Kommentare und Bewertungen angezeigt.

Die gesammelten Daten bzgl. Navigationswege und Bewertungen werden in einer Baumstruktur grafisch aufbereitet. Dabei ist die Wurzel die erste Seite des Tests (es kann auch jede andere Seite gewählt werden). Alle weiteren Seiten, die von dieser aus besucht wurden, werden je nach Zeitabstand und Häufigkeit weiter entfernt und mit einer dickeren Linie dargestellt. Zur besseren visuellen Erfassung sind die einzelnen Webseiten in Form von Thumbnails dargestellt und überwiegend positiv bewertete Seite mit einem grünen Hacken oder entsprechend negative mit einem roten „X“ versehen.

Die Visualisierung findet dabei ebenfalls im Browser mit Hilfe eines Flash-PlugIns statt. Somit sind auch bei der Auswertung keine weiteren technischen Anforderungen notwendig und diese kann von jedem Rechner mit Internetzugang aus erfolgen. Wurde im Rahmen der Entwicklung ein Navigationsmodell erstellt, kann dieses direkt mit dem tatsächlichen verglichen werden.

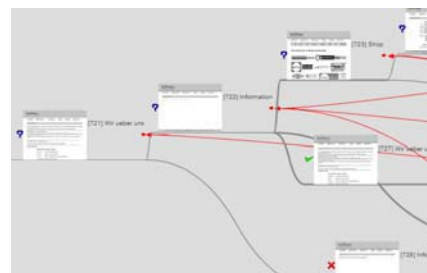


Abb 6: Auswertung Navigationswege

### 4.0 Technik

Um keine weiteren Anforderungen an den Rechner der Probanden zu stellen wurden als Technik nur JavaScript eingesetzt und aktivierte Cookies bei der Testperson vorausgesetzt. Ein installierter Flash-Player ist optional vorgesehen.

## 5.0 Fazit

Die Software hat in ersten Tests gezeigt, dass Sie funktionieren kann. Interessant werden weitere Tests und eine genaue Evaluation der Software, auch im direkten Vergleich mit einem Labortest.

Die gesammelten Daten können in weiteren Auswertungen noch detaillierter visualisiert werden, um z.B. die Erstellung von Klick-Maps oder ähnliche Auswertungen darzustellen.

Der Ablauf des Tests spiegelt die Aspekte Beobachtung, Moderator und Feedback der Testperson wider. Das laute Denken wird durch Kategorien und ein Textfeld ersetzt. Fehlende Kontrolle durch den Testleiter und die Möglichkeit, Probleme in einem anschließenden Interview zu klären, werden durch die gezielte Auswahl an Probanden aus den realen Nutzern und eine höhere Fallzahl ausgeglichen.

Abschließend lassen sich die Nachteile dieser Testmethode wie folgt aufzählen:

- Wenig Kontrolle über den Probanden
- Interpretationsschwierigkeiten
- Repräsentanzproblem

Die Vorteile dagegen sind:

- Kosten- und Zeitersparnis
- Leichte Rekrutierung
- Natürliches Nutzerverhalten
- Automatisierung, kein Medienbruch
- Unabhängigkeit von Ort und Zeit
- Konstante Testbedingungen
- Verknüpfung von quantitativen und qualitativen Daten

## 6.0 Literaturverzeichnis

Andreasen, M. S.; Nielsen, H. V.; Schrøder, S. O.; Stage, J. (2007): What Happened to Remote Usability Testing? An Empirical Study of Three Methods. CHI 2007, San Jose, California, USA.

Prümper, J.; Anft, M. (1993): Beurteilung von Software auf Grundlage der Internationalen

Ergonomie-Norm ISO 9241/10. FHTW-Berlin, Fachgebiet Wirtschaftspsychologie.

Nielsen J. (1994): Ten Usability Heuristics. [www.useit.com/papers/heuristic](http://www.useit.com/papers/heuristic).