

MASTERARBEIT

zur Erlangung des akademischen Grades
„Master of Science in Engineering“
im Studiengang Softwareentwicklung

Usability Evaluierung von komplexen Systemen zur Informationsanalyse im Bereich der Applikationsüberwachung

Ausgeführt von: Johanna Strutzenberger, BSc
Personenkennzeichen: 1510299006

1. BegutachterIn: Benedikt Salzbrunn, MSc
2. BegutachterIn: Dominik Mayr, BSc MA

Wien, 17.09.2017



Eidesstattliche Erklärung

„Ich, als Autor / als Autorin und Urheber / Urheberin der vorliegenden Arbeit, bestätige mit meiner Unterschrift die Kenntnisnahme der einschlägigen urheber- und hochschulrechtlichen Bestimmungen (vgl. etwa §§ 21, 46 und 57 UrhG idGF sowie § 11 Satzungsteil Studienrechtliche Bestimmungen / Prüfungsordnung der FH Technikum Wien).

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig angefertigt und Gedankengut jeglicher Art aus fremden sowie selbst verfassten Quellen zur Gänze zitiert habe. Ich bin mir bei Nachweis fehlender Eigen- und Selbstständigkeit sowie dem Nachweis eines Vorsatzes zur Erschleichung einer positiven Beurteilung dieser Arbeit der Konsequenzen bewusst, die von der Studiengangsleitung ausgesprochen werden können (vgl. § 11 Abs. 1 Satzungsteil Studienrechtliche Bestimmungen / Prüfungsordnung der FH Technikum Wien).

Weiters bestätige ich, dass ich die vorliegende Arbeit bis dato nicht veröffentlicht und weder in gleicher noch in ähnlicher Form einer anderen Prüfungsbehörde vorgelegt habe. Ich versichere, dass die abgegebene Version jener im Uploadtool entspricht.“

Ort, Datum

Unterschrift

Kurzfassung

Usability ist ein entscheidender Faktor für unterschiedlichste Produkte. Die üblichen Beispiele zur Beschreibung von Konzepten und Methoden der Usability Evaluierung sind jedoch oftmals einfache Anwendungen wie Online Shops oder Webseiten. Nur selten wird auf komplexe Beispiele, wie Systeme zur Informationsanalyse und Entscheidungsfindung, näher eingegangen. Dabei weisen diese Systeme auf Grund ihrer komplexen Aufgabenstellungen ganz andere Eigenschaften und Anforderungen an die Bedienung auf.

Diese Arbeit stellt sich der Frage, ob klassische Evaluierungsmethoden, wie Usability Tests und Metriken, auch zur Messung der Usability von komplexen Systemen zur Informationsanalyse effizient und effektiv genutzt werden können.

Zu diesem Zweck wurde die Usability einer realen firmeninternen Anwendung eines österreichischen Unternehmens evaluiert. Dabei handelt es sich um ein System zur Analyse von Daten, die während des Betriebs einer Softwareapplikation gesammelt werden, mit dem Ziel, den Zustand des Applikationsbetriebes zu bewerten. Zuerst wurde das bestehende System evaluiert, basierend auf diesen Ergebnissen wurde ein neues und verbessertes System entwickelt und dieses anschließend erneut evaluiert. Die Usability beider Systeme wurde durch moderierte Usability Tests und dem SUS Fragebogen überprüft. Anhand der gesammelten Daten wurden Usability Metriken wie Time on Task und Task Success ausgewertet und im Detail analysiert.

In der Gegenüberstellung beider Auswertungen zeigt sich eine deutliche Verbesserung im neuen System bei allen ausgewerteten Metriken zu den Usability Kriterien Effizienz, Effektivität und Zufriedenheit. Außerdem konnte damit dargelegt werden, dass die gemessenen Ergebnisse vergleichbar sind. Zusätzlich wurden im ersten Test wertvolle Usability Probleme erkannt und deren Verbesserungen in der zweiten Evaluierung des neuen Systems gemessen. Mit dieser Arbeit wird daher gezeigt, dass klassische Evaluierungsmethoden, wie Usability Tests und Metriken, auch bei komplexen Analysesystemen angewendet werden können und zu sinnvollen Ergebnissen führen.

Schlagwörter: Usability, Usability Evaluierung, Usability Tests, Usability Metriken, komplexe Systeme, Informationsanalyse, Applikationsüberwachung

Abstract

Usability is an important factor for different kinds of products. Common examples to describe concepts and methods for evaluating usability are often simple applications like online shops or websites. It is unusual that complex examples, such as a system for analysing information and supporting decision making, are used in this context. However, due to their complexity these systems exhibit different characteristics and requirements regarding their use.

This thesis explores the question if common methods for usability evaluation, such as usability tests and metrics, can also be used to effectively and efficiently measure the usability of complex system for information analysis.

For that purpose, the usability of a real system is evaluated. This system is used internally at an Austrian company to analyse application monitoring data and assess the state of the monitored application. First the current system was assessed, then a new and improved version was developed and finally the new system was measured. The methods used for both evaluations are moderated usability tests and the SUS questionnaire. Using the data collected with these methods usability metrics like time on task and task success have been calculated and analysed in detail.

For all evaluated metrics of the usability criterion efficiency, effectiveness and satisfaction a clear improvement could be seen in the new system. It also showed that the results of both evaluations are comparable. Additionally, valuable usability problems were discovered in the first evaluation and their improvements measured with the second evaluation of the new system. Therefore, this thesis shows that classic evaluation methods, such as usability tests and metrics, can be applied to complex systems for information analysis and still provide valid results.

Keywords: Usability, usability evaluation, usability tests, usability metrics, complex systems, information analysis, application monitoring

Danksagung

Ich möchte mich bei meinem FH Betreuer Benedikt Salzbrunn für die tolle Unterstützung und das wertvolle Feedback herzlich bedanken. Die vielen beigesteuerten Gedanken und Ideen zu dieser Arbeit, haben mir sehr weiter geholfen und mich auf den richtigen Weg gebracht.

Ein großes Dankeschön geht auch an meine lieben Arbeitskollegen bei Eidentity, die diesem Projekt nicht nur Interesse entgegen gebracht haben, sondern sich auch die Zeit genommen haben, an den Tests teilzunehmen. Ihre Begeisterung und Vorfreude auf das neue System hat meiner Motivation oft den notwendigen Schub verschafft. Insbesondere möchte ich mich bei Dominik Mayr als Firmenbetreuer bedanken und bei Martin Paulnsteiner, der mir für alle inhaltlichen Fragen und zur Diskussion von einigen Ideen immer tatkräftig zur Seite stand.

Von ganzem Herzen möchte ich mich besonders bei meinen wundervollen Eltern, Beatrix und Peter, sowie meinen großartigen Schwestern, Elisabeth und Antonia, für ihre endlose Unterstützung – nicht nur bei diesem Projekt sondern in allen Lebenslagen – bedanken. Ihre Zuversicht und Aufmunterung hat mich in vielen schwierigen Stunden beim Schreiben dieser Arbeit nicht verzweifeln lassen. Ihre Positivität und Vertrauen in mich waren oft ansteckend und haben mich nicht aufgeben lassen.

Außerdem möchte ich mich bei meinen besten Freundinnen, Melanie und Tina, für ihre Freundschaft und ihr Verständnis dafür, dass ich in den letzten Monaten oft nicht viel Zeit hatte, bedanken. Bei jedem Treffen, das sich doch ausgegangen ist, haben sie es immer geschafft, dass ich den ganzen Stress vergessen konnte.

Inhaltsverzeichnis

1	Einleitung.....	8
2	Grundlagen.....	10
2.1	Usability.....	10
2.2	Heuristische Evaluierung.....	12
2.3	Usability Tests.....	15
2.3.1	Charakteristiken.....	15
2.3.2	Arten von Usability Tests.....	17
2.3.3	Planung & Durchführung.....	18
2.4	Fragebögen.....	24
2.4.1	Zeitpunkte.....	24
2.4.2	Standardfragebögen.....	25
2.4.3	System Usability Scale.....	26
2.5	Usability Metriken.....	27
2.5.1	Effektivität.....	28
2.5.2	Effizienz.....	29
2.5.3	Zufriedenheit.....	31
2.5.4	Weitere Metriken.....	34
2.6	Komplexität von Analysesystemen.....	38
3	Methoden.....	41
3.1	Aktuelles System.....	41
3.1.1	Applikationsüberwachung.....	41
3.1.2	Systemlandschaft.....	43
3.2	Usability Tests.....	48
3.2.1	Teilnehmer und Teilnehmerinnen.....	49
3.2.2	Aufgabenstellungen.....	49
3.2.3	Testaufbau.....	52
4	Evaluierung des bisherigen Systems.....	54
4.1	Task Success.....	54
4.2	Time on Task.....	56
4.3	SUS Score.....	60

4.4	Usability Probleme	62
5	Neues System	66
5.1	Anforderungen	66
5.2	Architektur	67
5.3	Funktionalität & Oberflächen.....	69
6	Evaluierung des neuen Systems.....	78
6.1	Task Success	78
6.2	Time on Task.....	79
6.3	SUS Score.....	81
7	Diskussion	83
8	Zusammenfassung	88

1 Einleitung

Wenn man über Usability und Usability Evaluierung liest, werden die erläuterten Ansätze und Methoden oft mit praktischen Beispielen ergänzt. Viele dieser Beispiele befassen sich mit einfachen Systemen und Szenarien, wie zum Beispiel einem Online Shop und der Aufgabe ein Buch in diesem Shop zu bestellen. Auch im Rahmen der von Rolf Molich gegründeten und koordinierten „Comparative Usability Evaluation“ Studien wurden vorwiegend Usability Evaluierungen dieser Art von Systemen durchgeführt. Der Däne Rolf Molich beschäftigt sich bereits seit 1984 mit Usability, hat gemeinsam mit Jakob Nielsen die Heuristische Evaluierung erfunden und gilt als Usability Experte. Bei den neun bisherigen „Comparative Usability Evaluation“ Studien wurden Usability Fachleute aus der ganzen Welt gebeten, das selbe System zu evaluieren. Die Ergebnisse der Teilnehmer wurden anschließend von Rolf Molich verglichen, mit dem Ziel herauszufinden, ob die Ergebnisse vergleichbar sind und welche Praktiken am Effektivsten sind. Die vierte „Comparative Usability Evaluation“ Studie befasste sich zum Beispiel mit einer Hotelwebseite und die dritte, sechste und achte Studie jeweils mit der Usability von unterschiedlichen Autovermietungswebseiten. [1]

Wie sieht es jedoch bei komplexeren Systemen und Aufgabenstellungen aus? Hier sind die Beispiele weniger verbreitet und konkrete Anwendungsfälle für Usability Evaluierungen seltener vertreten. Eine Ausnahme stellt Barbara Mirel dar, die selber betont, dass sich komplexe Aufgabenstellungen und Systeme von den oben genannten Beispielen unterscheiden und andere Anforderungen an die Usability stellen. Barbara Mirel ist als Forscherin an der University von Michigan tätig und beschäftigt sich vor allem damit, wie Systeme für komplexe Aufgaben nützlich gestaltet und entworfen werden können. In ihrem Buch „Interaction Design for Complex Problem Solving“ [2] geht sie dieser Frage nach und zeigt Beispiele für komplexe Systeme auf, wie einem System für IT Spezialisten, das beim Auffinden und der Ursachenanalyse von Problemen im Betrieb von IT Infrastruktur und Applikationen unterstützen soll. Weitere Beispiele sind Systeme zur Ursachenanalyse für medizinische Fehler oder zur Informationsanalyse für strategische Markt- und Produktentscheidungen in Unternehmen.

In diesen komplexen Systemen dienen die Informationen nicht nur zur Beantwortung von einfachen Fragen, wie zum Beispiel, dass die Verfügbarkeit eines Hotelzimmers daran zu erkennen ist, dass die freien Tage im Kalender grün markiert sind. Stattdessen müssen Personen mit Hilfe dieser Informationen komplexe Probleme lösen und Entscheidungen auf Grund dessen treffen. In komplexen Systemen zur Informationsanalyse gibt es aber oft keinen eindeutigen idealen Lösungsweg und es zählt viel mehr, dass die Benutzer und Benutzerinnen in ihrer eigenen Vorgehensweise richtig unterstützt werden. Für diese Art von Systemen sind daher die klassischen Usability Kriterien wie Effektivität und Effizienz wesentlich wichtiger, als ein modernes und ansprechendes Aussehen.

Um die Usability und diese Kriterien bei komplexen Systemen zu überprüfen und verbessern, sind entsprechende Evaluierungen notwendig. Daher beschäftigt sich diese Masterthesis mit der Usability Evaluierung von komplexen Systemen zur Informationsanalyse im Bereich der Applikationsüberwachung. Dabei sollen folgende Forschungsfragen beantwortet werden:

- Kann mit klassischen Usability Methoden die Usability von komplexen Systemen zur Informationsanalyse evaluiert und Usability Metriken gemessen werden?
- Sind die Ergebnisse aus zwei Evaluierungen vergleichbar und können Maßnahmen daraus abgeleitet werden, die eine Verbesserung der Usability bewirken?

Dazu wurde die Usability eines bestehenden Systems zur Applikationsüberwachung bei einer österreichischen Softwarefirma evaluiert. Diese Firma entwickelt und betreibt eine Finanzapplikation für den Automobilbereich. Zur Überwachung dieser Applikation werden im Betrieb unterschiedlichste Informationen gesammelt. Die Aufgabe des internen Systems ist es, die Anwender und Anwenderinnen dabei zu unterstützen, die gesammelten Informationen zu überprüfen und eine Bewertung des Betriebszustands zu geben. Basierend auf den Ergebnissen der Evaluierung dieses Systems wurde ein neues System gebaut. Eine zweite Evaluierung des neuen Systems liefert Vergleichswerte und misst das Ausmaß der Verbesserungen.

Im folgenden Kapitel werden zuerst die Grundlagen der Usability und Usability Evaluierung erläutert. Im Anschluss beginnt der praktische Teil dieser Arbeit. Darin wird zuerst das aktuelle System und die eingesetzten Evaluierungsmethoden erläutert. Anschließend werden die Ergebnisse der ersten Evaluierungen präsentiert. Im nächsten Kapitel wird das neue System vorgestellt und daraufhin die Ergebnisse aus der Evaluierung des neuen Systems im Vergleich zur ersten Evaluierung dargestellt. Abschließend folgen die Diskussion und eine Zusammenfassung der Arbeit.

2 Grundlagen

In den folgenden Kapiteln werden die theoretischen Grundlagen zu Usability und den Möglichkeiten zur Evaluierung von Usability dargestellt.

2.1 Usability

Usability stellt ein Qualitätsmerkmal eines Produktes dar und steht für die Gebrauchstauglichkeit des Produktes. Es gibt eine eigene Norm, die sich dem Thema Usability widmet. Dabei handelt es sich um die Norm „ÖNORM EN ISO 9241: Ergonomie der Mensch-System-Interaktion“. In dieser Norm ist Usability definiert als:

„Ausmaß, in dem ein System, ein Produkt oder eine Dienstleistung durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden können, um festgelegte Ziele effektiv, effizient und zufriedenstellend zu erreichen.“ [3]

Die Benutzer und Benutzerinnen beziehen sich auf alle Personen, die mit dem System, dem Produkt oder der Dienstleistung in unterschiedlichen Formen interagieren. Neben der klassischen Nutzung zählen dazu zum Beispiel auch das Betreiben des Systems oder das Nutzen des Ergebnisses des Systems. Der Nutzungskontext stellt die bestimmte Kombination aus Benutzern und Benutzerinnen, Zielen, Arbeitsaufgaben, Ressourcen und der Umgebung dar, die für die Usability betrachtet wird. [4, p. 8 ff]

Joseph S. Dumas und Janice C. Redish nehmen in ihrer Definition in „A Practical Guide to Usability Testing“ [5] noch konkreter Bezug auf die wichtige Rolle der Benutzer und Benutzerinnen im Zusammenhang mit Usability. Ähnlich zur oben genannten Norm bedeutet Usability auch für sie, dass Benutzer und Benutzerinnen ein Produkt schnell und einfach nützen und so ihr Ziel erreichen können. Ihre Definition basiert zusätzlich auf folgenden vier Punkten:

- Usability bedeutet, dass die Benutzer und Benutzerinnen im Fokus stehen.
- Personen benutzen das Produkt, um produktiv zu sein.
- Die Anwender und Anwenderinnen sind beschäftigte Personen, die versuchen Aufgaben zu bewältigen.
- Die Benutzer und Benutzerinnen entscheiden, ob ein Produkt einfach zu benutzen ist. [5, p. 4]

Neben der Definition stellt sich aber auch die Frage, warum Usability wichtig ist. Für einfache Anwendungen wie Webseiten wird hier oft dargelegt, dass die Usability einer Anwendung entscheidend für ihren Erfolg sein kann. Jakob Nielsen beschreibt beispielsweise, dass Benutzer und Benutzerinnen Webseiten schnell wieder verlassen, wenn die Usability nicht passt. Die meisten Besucher und Besucherinnen sind nicht gewillt, sich lange mit einer

Webseite zu beschäftigen, um sie zu verstehen. Wenn die Benutzung schwierig ist, man sich nicht zurechtfindet oder die Informationen unverständlich sind, verlassen die meisten Personen die Seite wieder und sehen sich nach besseren Alternativen um. [6] Schlechte Usability kann bei Webseiten oder einfachen Applikation also schnell zum Verlust von Kunden und Kundinnen führen.

Aber wie sieht es bei komplexeren Applikationen aus? In diesem Bereich haben die Benutzer und Benutzerinnen meistens keine andere Wahl als die Applikation zu nutzen, weil diese beispielsweise den Mitarbeitern und Mitarbeiterinnen von der Firma vorgegeben wird oder ein Mangel an Alternativen herrscht. Im Gegensatz zu einfachen Anwendungen spielt Usability als Erfolgsfaktor zur Kundengewinnung in diesem Zusammenhang daher eine weniger wichtige Rolle. Dennoch ist es nicht so, dass Usability bei komplexen Systemen irrelevant ist. Anhand des Beispiels von Anwendungen im Intranet einer Firma legt Jakob Nielsen dar, dass die Usability bei diesen Applikationen ein entscheidender Faktor für die Mitarbeiterproduktivität ist. [6] Schlechte Usability führt in diesem Zusammenhang oft zu unnötigen Kosten, weil es die Produktivität der Angestellten senken und die Fehlerraten erhöhen kann. Zusätzlich kann es die Zufriedenheit der Mitarbeiter und Mitarbeiterinnen beeinflussen.

Gerade bei komplexen Systeme ist die Usability daher wichtig, um eine effiziente und effektive Nutzung zu ermöglichen. Effizienz, Effektivität und Zufriedenheit werden als Usability Kriterien ebenfalls in der ÖNORM EN ISO 9241 erfasst und ihre Definition wird im Folgenden erläutert.

Effektivität

Effektivität bezieht sich auf die Genauigkeit und Vollständigkeit, mit denen Benutzer und Benutzerinnen Aufgaben in einem System durchführen können. Die Genauigkeit stellt dabei dar, wie sehr das tatsächliche Ergebnis mit dem angestrebtem übereinstimmt. Mit Vollständigkeit ist gemeint, ob alle erwarteten Ergebnisse erreicht wurden. Für die Genauigkeit ist zusätzlich wichtig, dass möglichst wenige Fehler und negative Folgen bei der Nutzung entstehen. [4, p. 16 ff]

Effizienz

Effizienz betrachtet das Verhältnis der eingesetzten Ressourcen bei Nutzung eines Produktes zu den erreichten Ergebnissen. Die üblichste Ressource in diesem Zusammenhang ist Zeit. Das heißt, es wird gemessen, wie viel Zeit Personen für die Durchführung von Aufgaben benötigen. Eine andere Betrachtung ist, wie viel geistigen oder körperlichen Aufwand ein Benutzer oder eine Benutzerin leisten muss, um ein bestimmtes Ergebnis zu erreichen. Weitere mögliche Ressourcen sind Geld oder Materialien. [4, p. 18 f]

Zufriedenheit

Die Zufriedenheit bezieht sich auf die Wahrnehmung und Reaktionen von Personen, die aus der Nutzung eines Produktes entstehen. Je nachdem wie gut ein Produkt den Erwartungen und Bedürfnissen der Nutzer und Nutzerinnen entspricht, desto zufriedener sind sie damit. Zufriedenheit umfasst neben den Wahrnehmungen, Vorstellungen und Vorlieben der Benutzer und Benutzerinnen auch andere Elemente wie Emotionen oder physische und psychologische Reaktionen wie zum Beispiel Unbehagen oder Stress. [4, p. 19]

In „Handbook of Usability Testing“ nennen Jeff Rubin und Dana Chisnell neben diesen drei Kriterien noch die zwei folgenden Usability Kriterien.

Nützlichkeit

Nützlichkeit hängt davon ab, wie gut ein Produkt dem Benutzer oder der Benutzerin dabei hilft seine oder ihre Ziele zu erreichen. Es ist entscheidend für die Bereitschaft eines Benutzers oder einer Benutzerin das Produkt überhaupt zu verwenden. Selbst wenn es alle anderen Kriterien aber nicht den Zweck erfüllt, für den es eine Person benötigt, wird es nicht benutzt werden. [7, p. 4]

Erlernbarkeit

Die Erlernbarkeit eines Systems hängt mit der Effektivität zusammen und gibt an, wie schnell es Benutzern und Benutzerinnen möglich ist, ein Produkt effektiv zu nutzen. Während manche Produkte eine gewisse Einschulungs- und Einarbeitungsphase voraussetzen, ist es für andere Produkte wichtig, dass sie schon bei der ersten Nutzung verständlich sind oder unregelmäßige Anwender und Anwenderinnen auch nach längerer Inaktivität wieder schnell effektiv damit umgehen können. [7, p. 4]

2.2 Heuristische Evaluierung

Die Heuristische Evaluierung ist eine Methode der Usability Evaluierung, um Usability Probleme eines Produktes zu finden. Bei der Heuristischen Evaluierung inspizieren mehrere Personen die Benutzeroberflächen eines Produktes und überprüfen, ob die Gestaltung mit anerkannten Usability Prinzipien übereinstimmt. Diese Prinzipien werden auch Heuristiken genannt.

Wichtig beim Einsatz dieser Methode ist, dass mehrere Personen als Begutachter oder Begutachterinnen agieren. Eine evaluierende Person alleine kann nicht alle Usability Probleme finden. Mit mehreren Begutachtern oder Begutachterinnen erhöht sich die Trefferquote und es können mehr unterschiedliche Probleme gefunden werden. [8]

Die Evaluierung selbst funktioniert so, dass jede evaluierende Person zuerst alleine das Produkt begutachtet. Erst nachdem alle Personen die Evaluierung abgeschlossen haben,

wird kommuniziert und die Ergebnisse aggregiert. So ist sichergestellt, dass jeder unvoreingenommen und unabhängig vorgeht. [8]

Bei der Evaluierung wird das System nicht direkt durch die evaluierenden Personen benutzt, um eine richtige Aufgabe durchzuführen. Stattdessen gehen die Personen die unterschiedlichen Oberflächen – mitunter mehrmals – durch und inspizieren die verschiedenen Elemente und Dialoge. Dabei wird überprüft, ob die Gestaltung dieser Elemente und Dialoge bestimmten Usability Prinzipien entspricht. Bei diesen Prinzipien handelt es sich um Regeln, die allgemeine Eigenschaften von benutzerfreundlichen Oberflächen beschreiben. Als Ergebnis der Evaluierung entsteht eine Liste an Problemen und zu jedem Problem, welche Regel laut dem Bewerter verletzt wurde. Nachdem für die Evaluierung keine richtige Benutzung des Produktes notwendig ist, hat diese Methode den Vorteil, dass sie auch bereits mit sehr einfachen Prototypen durchgeführt werden kann. [8]

Zu den bei dieser Evaluierungsmethode sehr häufig eingesetzten und anerkannten Prinzipien zählen die 10 Heuristiken von Jakob Nielsen. [9] Diese zehn Regeln sollen dabei helfen, die wichtigsten allgemeinen Probleme in der Gestaltung von Benutzeroberflächen zu vermeiden. Im Folgenden werden diese zehn Heuristiken aufgelistet und näher beschrieben.

Sichtbarkeit des Systemstatus

Der Anwender oder die Anwenderin sollte zu jedem Zeitpunkt wissen, was das System gerade tut. Dazu sollte das System entsprechendes Feedback in angemessener Zeit liefern. [9]

Übereinstimmung zwischen dem System und der realen Welt

Die im System verwendete Bezeichnungen sollten der Sprache des Benutzers oder der Benutzerin entsprechen. Wenn diese Begriffe und Konzepte Beispielen aus der realen Welt entsprechen, erkennt der Benutzer oder die Benutzerin diese einfacher und findet sich so leichter zurecht. [9]

Benutzerkontrolle und -freiheit

Da es immer wieder passieren wird, dass Personen eine falsche Funktion auswählen, muss es möglich sein, ungewollte Funktionen einfach wieder zu verlassen oder falsche Aktionen rückgängig zu machen. [9]

Konsistenz und Standards

Durch das ganze System hinweg müssen die gleichen Begriffe und Konzepte einheitlich gestaltet sein. Zusätzlich sollten allgemeine Konventionen eingehalten werden, damit ein Anwender oder eine Anwenderin nicht dadurch verwirrt wird, dass sich etwas anders verhält, als er oder sie es gewohnt ist. [9]

Fehlervermeidung

Besser als jede Fehlermeldung ist es, wenn man verhindern kann, dass ein Fehler überhaupt eintritt. Deshalb soll das System durch zusätzliche Überprüfungen und Hinweise Fehlermöglichkeiten verhindern und die Benutzer und Benutzerinnen wichtige Aktionen extra bestätigen lassen. [9]

Wiedererkennen statt sich erinnern

Es sollte nicht notwendig sein, dass sich ein Anwender oder eine Anwenderin Informationen von einem Schritt im Prozess bis zu einem darauffolgenden merken muss. Instruktionen für den Prozess sowie wichtige Aktionen und Optionen sollten jederzeit sichtbar oder leicht abrufbar sein. [9]

Flexibilität und Effizienz der Benutzung

Das System sollte zwischen Anfängern oder Anfängerinnen und Experten oder Expertinnen unterscheiden. Obwohl manche Funktionen für neue Benutzer und Benutzerinnen zu Beginn verwirrend sein können, sind sie für erfahrene Personen wichtig, um das System effizienter zu benutzen. Deshalb sollte das System beide Gruppen berücksichtigen und dem Benutzer oder der Benutzerin die Möglichkeit geben, manche Funktionen zusätzlich zu aktivieren. [9]

Ästhetik und minimales Design

Um die Aufmerksamkeit der Benutzer und Benutzerinnen auf die wichtigen Informationen zu fokussieren, sollten die Dialoge im System auf ein Minimum an nötigen Informationen reduziert werden. Nicht relevante Informationen müssen vermieden werden und nur selten benötigte Informationen sollten nicht standardmäßig enthalten sein, sondern extra abrufbar sein. [9]

Hilfe beim Erkennen, Diagnostizieren und Beheben von Fehlern

Fehlermeldungen sollten für den Benutzer oder die Benutzerin verständlich und hilfreich gehalten werden. Das heißt, sie sollten in natürlicher Sprache formuliert sein und klar erkennbar machen, um was für ein Problem es sich handelt und welche Lösungen oder weitere Schritte möglich sind. [9]

Hilfe und Dokumentation

Auch wenn es das Ziel sein sollte, dass ein System ohne Dokumentation genutzt werden kann, sollte den Anwendern und Anwenderinnen eine Dokumentation zur Verfügung gestellt werden. Dabei ist wichtig, dass man einfach nach etwas suchen kann und sich die Informationen in der Dokumentation auf konkrete Aufgaben und Szenarien der Benutzer und Benutzerinnen beziehen sowie konkrete Schritte dazu aufgelistet werden. [9]

Neben dem Einsatz dieser Prinzipien im Rahmen einer Heuristischen Evaluierung können sie bereits beim Entwurf und der Entwicklung einer Anwendung als allgemeine Designregeln herangezogen werden.

2.3 Usability Tests

Im vorherigen Kapitel wurde die Heuristische Evaluierung als eine Methode zur Usability Evaluierung genannt. Eine weitere Methode sind Usability Tests. Der entscheidende Unterschied zur Heuristischen Evaluierung ist, dass im Rahmen dieser Methode das zu testende Produkt wirklich benutzt wird. Bei einem Usability Test beobachtet man Personen, wie sie unterschiedliche Aufgaben mit Hilfe des Produktes lösen.

Usability Tests sind außerdem besonders geeignet um die Usability Kriterien Effektivität und Effizienz zu evaluieren. Da die Teilnehmer und Teilnehmerinnen bei den Tests das System direkt nutzen, kann man sehr gut beobachten, wie effizient und effektiv die Personen Aufgaben im System bewältigen können.

In den folgenden Kapiteln werden zuerst die Charakteristiken von Usability Tests erläutert. Anschließend werden die unterschiedlichen Arten von Usability Tests beschrieben. Im letzten Kapitel wird dargestellt, was für einen Usability Test geplant und was bei der Durchführung beachtet werden soll.

2.3.1 Charakteristiken

In „A Practical Guide to Usability Testing“ [5] definieren Joseph Dumas und Janice Redish die folgenden fünf Charakteristiken von Usability Tests:

- Das Ziel ist, die Usability eines Produkts zu verbessern.
- Teilnehmer und Teilnehmerinnen repräsentieren reale Benutzer und Benutzerinnen.
- Teilnehmer und Teilnehmerinnen führen reale Aufgaben aus.
- Man beobachtet und protokolliert Handlungen und Aussagen der Teilnehmer und Teilnehmerinnen.
- Man analysiert die Daten, bestimmt die echten Probleme und empfiehlt Änderungen, um diese Probleme zu lösen [5, p. 22]

Ziel ist, die Usability eines Produkts zu verbessern

Das Hauptziel dieser Art von Tests ist es, die Usability des getesteten Produkts zu verbessern. Dadurch unterscheidet es sich bereits von anderen Arten von Tests, wie zum Beispiel funktionalen Tests, wo es vor allem darum geht, sicherzustellen, dass das Produkt der Spezifikation entspricht. Aus diesem allgemeinen Ziel können sich dann unterschiedliche spezifischere Ziele ergeben. Beispielsweise kann nur ein bestimmter Teil des Produkts, wie die Navigation auf einer Website, interessant sein und die Tests werden jeweils auf diesen

Bereich fokussiert. Des Weiteren kann darauf abgezielt werden, Unterschiede zwischen verschiedenen Zielgruppen oder Szenarien zu überprüfen. Oftmals haben die Tests auch das Ziel die Usability Kriterien Effizienz, Effektivität und Zufriedenheit zu überprüfen. Abhängig von diesen spezifischeren Zielen können dann entsprechende Aufgaben abgeleitet und passende Teilnehmer und Teilnehmerinnen ausgewählt werden. [5, p. 22 f]

Teilnehmer und Teilnehmerinnen repräsentieren reale Benutzer und Benutzerinnen

Eine weitere wichtige Charakteristik ist, dass die Teilnehmer und Teilnehmerinnen des Tests aus der tatsächlichen Zielgruppe für das Produkt stammen. Das heißt es soll jene Personen umfassen, die das Produkt bereits nutzen oder nutzen werden. Die Durchführung von Tests mit anderen Personengruppen kann auch Usability Probleme aufzeigen. Diese Probleme können für die Zielgruppe aber irrelevant sein, weil sie z.B. nur durch fehlende Kenntnisse im Fachgebiet entstehen. Man sieht also nicht, was wirklich passiert, wenn reale Benutzer und Benutzerinnen mit dem Produkt konfrontiert werden. [5, p. 23]

Teilnehmer und Teilnehmerinnen führen reale Aufgaben aus

Ebenso wichtig wie der Einsatz von realen Benutzern und Benutzerinnen ist es, dass diese Personen Aufgaben im Rahmen der Tests durchführen, die sie auch im echten Umgang mit dem Produkt durchführen müssten. Nur so können Probleme aufgedeckt werden, die dann wirklich bei der Arbeit hinderlich sind. Dazu ist aber notwendig, dass man das Aufgabengebiet und den Einsatzbereich des Produktes versteht, um anschließend passende Aufgaben ableiten zu können. Bei sehr umfangreichen oder komplexen Systemen, kann dies für Außenstehende schwierig sein und in manchen Fällen ist es nur möglich, einen Teil der realen Aufgaben zu testen. [5, p. 23 f]

Beobachtung und Protokollierung der Handlungen und Aussagen der Teilnehmer und Teilnehmerinnen

Der Grundgedanke eines Usability Tests ist es, die Teilnehmer und Teilnehmerinnen während der Durchführung zu beobachten. Abhängig von der Art des Tests kann dies auf unterschiedliche Weise passieren. Um weitere Analysen und Auswertungen zu ermöglichen, müssen zum Beispiel Auffälligkeiten im Verhalten, verbale Aussagen und Kommentare sowie andere Daten protokolliert werden. Das Ausmaß und die Art der Daten hängen davon ab, was man mit dem Test messen und herausfinden möchte. Zusätzlich zur Durchführung von Aufgaben können Teilnehmer und Teilnehmerinnen explizit nach ihrer Meinung gefragt werden, zum Beispiel in Form eines Fragebogens, um so weitere Daten zu sammeln. [5, p. 24]

Analyse der Daten, Bestimmung der echten Probleme und Empfehlung von Änderungen um diese Probleme zu lösen

Nach einem Test werden die gesammelten Daten gemeinsam mit den eigenen Beobachtungen ausgewertet, um so konkrete Usability Probleme ableiten und auch andere

Metriken berechnen zu können. Basierend darauf können anschließend Verbesserungsvorschläge erstellt werden, um diese Probleme zu beheben. [5, p. 24 f]

2.3.2 Arten von Usability Tests

Aus den unterschiedlichen Varianten, wie Usability Tests durchgeführt werden können, lassen sich sechs Kategorien ableiten:

- Formative und summative Tests
- Quantitative und qualitative Tests
- Moderierte und unmoderierte Tests

Formative Tests

Diese Tests werden durchgeführt während das Produkt noch entwickelt wird. Das Ziel ist es, bereits während der Entwicklung Probleme zu finden und diese damit auch möglichst schnell beheben zu können. [10, p. 14]

Summative Tests

Summative Tests werden im Vergleich zu formativen Tests an einem fertigen Produkt durchgeführt. Hier ist das Ziel zu überprüfen, ob das fertige Produkt bestimmte Voraussetzungen erfüllt oder um Basiswerte aus den Messungen zu erhalten, die später mit anderen Überprüfungen verglichen werden können. [10, p. 14]

Quantitative Tests

Bei dieser Art von Tests geht es vor allem um die Messung von Metriken, die sich in Zahlen ausdrücken lassen, wie zum Beispiel die Messung der benötigten Zeit für eine Aufgabe. Mit der Auswertung von konkreten Kennzahlen ist es einfacher bestimmte Annahmen oder die Erfüllung von Anforderungen zu bestätigen oder Vergleiche möglich zu machen. Meist erfordern diese Tests einen strikteren und wissenschaftlicheren Ablauf, damit die Ergebnisse valide und reproduzierbar sind. [11, p. 13 f]

Qualitative Tests

Im Vergleich zu quantitativen Tests liegt der Fokus hier stärker auf den qualitativen Aspekten, die sich schwerer in Zahlen ausdrücken lassen. Ein Beispiel hierfür sind die Usability Probleme, auf die Benutzer und Benutzerinnen gestoßen sind. Das Hauptziel ist es, Einblicke in das Benutzerverhalten zu erhalten und darauf basierende Verbesserungsmöglichkeiten abzuleiten. Bei diesen Tests geht es also vorrangig nicht darum, Vergleiche anzustellen oder etwas zu beweisen, sondern um Usability Probleme aufzudecken. [11, p. 14]

Moderierte Tests

Diese Tests finden oft in einem Usability Labor statt, sie können aber ebenfalls in einem informelleren Rahmen durchgeführt werden. Bei moderierten Usability Tests ist jedoch immer ein Testmoderator oder eine Moderatorin anwesend, der oder die die Teilnehmer und Teilnehmerinnen durch den Test begleitet, sie bei der Durchführung beobachtet und in manchen Fällen auch selber Notizen zu den Beobachtungen führt. [12, p. 53]

Unmoderierte Tests

Diese Art von Tests finden remote und meistens online statt. Dabei ist kein Moderator anwesend und die Teilnehmer und Teilnehmerinnen werden nicht direkt beobachtet. Stattdessen leitet sie eine Software durch den Test und sammelt dabei automatisch unterschiedliche Daten, wie zum Beispiel die Aufzeichnung des Bildschirminhalts oder die Messung der Zeit für jede Aufgabe. [12, p. 54]

2.3.3 Planung & Durchführung

Um sicherzustellen, dass das Richtige getestet wird und der Test möglichst reibungslos abläuft, ist es notwendig den Test und seine Bestandteile zu planen. Andernfalls kann es passieren, dass wichtige Aspekte des Produkts und dessen Nutzung nicht getestet werden, während dem Test Probleme auftreten oder die gesammelten Daten falsch bzw. unvollständig sind. Deshalb müssen im Vorhinein die Ziele, Teilnehmer, Aufgabenstellungen, Testumgebung und -material sowie die zu sammelnden Daten bedacht und definiert werden. Im Folgenden wird beschrieben, was bei der Planung und Durchführung dieser Bereiche zu beachten ist.

2.3.3.1 Ziele

Auf Grund von Zeit- und Ressourcenmangel ist es oft nicht möglich, alle Funktionen und Aspekte eines Produktes zu testen. Deshalb müssen Ziele definiert werden, die sich auf die wichtigsten Aspekte für den Test beziehen. Es muss festgelegt werden, was man über die Nutzung des Produktes durch reale Benutzer und Benutzerinnen wirklich lernen möchte. Je nach Produkt und Situation können diese Ziele daher ganz unterschiedlich ausfallen. Wenn ein Produkt zum ersten Mal getestet wird, muss man sich überlegen, was die wichtigsten Informationen sind, die man bei der Beobachtung der Benutzer und Benutzerinnen erfahren möchte. Bei einem wiederholten Test des gleichen Produkts wird der Fokus jedoch wahrscheinlich woanders liegen. Zum Beispiel möchte man Verbesserungen zu vorherigen Problemen überprüfen oder neue Funktionen testen. Oftmals beeinflussen auch Faktoren, wie das Feedback zum Produkt aus dem Kundensupport oder interne Diskussionen über unterschiedliche Lösungen, die Gestaltung der Ziele. Ebenso kann man sich bei der Festlegung der Ziele an den Usability Kriterien orientieren. Beispielsweise ist es wichtig, die Effizienz und Effektivität des Produkts zu überprüfen und der Frage nachzugehen, ob die Benutzer und Benutzerinnen eine bestimmte Aufgabe erledigen können und ob dies

innerhalb einer bestimmten Zeitspanne möglich ist. Daher sind die definierten Ziele und die damit verbundenen Fragen auch der Ausgangspunkt für die Definition der Aufgabenstellungen. Des Weiteren bestimmen sie welche Daten zur Beantwortung dieser Fragen gesammelt werden müssen. [10, p. 107 f]

Es ist auch wichtig, dass die Ziele so konkret, klar und messbar wie möglich definiert werden. Wenn die Ziele zu ungenau oder unklar formuliert sind, überprüft der Test vielleicht nicht die eigentlichen Bedenken zum Produkt. Wenn nicht klar hervorgeht, wie die Ziele gemessen werden können, ist es im Nachhinein schwer festzustellen, wo die konkreten Probleme liegen, und man tendiert dann dazu, die Ergebnisse besser darzustellen als sie sind. [7, p. 69 f]

2.3.3.2 Teilnehmer

Wie bereits in den Charakteristiken eines Usability Tests beschrieben wurde, ist es entscheidend, dass das Produkt mit echten Benutzern und Benutzerinnen getestet wird. In der Planung ist es daher notwendig die aktuellen und potenziellen Benutzer und Benutzerinnen zu identifizieren und entsprechende Teilnehmer und Teilnehmerinnen zu rekrutieren. Bei der Auswahl ist es wichtig, jene Charakteristiken der Personen zu beachten, die alle Mitglieder der Zielgruppe teilen, aber auch jene die Unterschiede darstellen und Untergruppen innerhalb der Zielgruppe definieren können. Beispielsweise könnte die Zielgruppe sowohl Anfänger und Anfängerinnen als auch erfahrene Nutzer und Nutzerinnen enthalten. Umso mehr Eigenschaften berücksichtigt werden, desto mehr Untergruppen lassen sich ausmachen. Für einen Test ist es meistens aber nicht möglich und oft nicht relevant, Personen aus allen möglichen Untergruppen zu inkludieren. Daher sollten jene Untergruppen ausgewählt werden, welche die höchste Relevanz in Bezug auf die Ziele haben und von denen man sich erwartet, dass sie die wichtigsten Usability Probleme aufzeigen werden. Bei einem ersten Usability Test würde man zum Beispiel die typischen Benutzer und Benutzerinnen einbeziehen, die den größten Anteil ausmachen, wohingegen man bei einem wiederholten Test eher an einer speziellen Untergruppe interessiert sein könnte. [5, pp. 120-125]

Die Anzahl der Teilnehmer und Teilnehmerinnen hängt von folgenden Aspekten ab [5, p. 127 f]:

- Wie viele Untergruppen sollen vertreten sein?
- Wie viel Zeit und Geld steht zur Verfügung?
- Sollen statistische Auswertungen gemacht werden?

Grundsätzlich kann man davon ausgehen, dass umso mehr Personen teilnehmen, desto mehr unterschiedliche Probleme zwar auftreten, aber die größten und wichtigsten Probleme bei jeder Person erneut beobachtet werden können. [5, p. 127 f] Laut Jakob

Nielsen und Thomas K. Landauer [13, pp. 206-213] können mit fünf Benutzern oder Benutzerinnen bereits ungefähr 80% der Usability Probleme aufgedeckt werden.

2.3.3.3 Aufgabenstellung

Wenn die Teilnehmer und Teilnehmerinnen ausgewählt sind, können die Aufgabenstellungen basierend auf den Zielen für die Nutzung des Produktes sowie der eigenen Ziele für den Usability Test definiert werden. Dabei ist es von großer Bedeutung, dass die Szenarien, die innerhalb der Aufgaben beschrieben werden, der realen Welt der Teilnehmer und Teilnehmerinnen entsprechen. Dazu sollten folgende Richtlinien beachtet werden:

- Die Aufgaben sollten in der Sprache der Benutzer und Benutzerinnen und nicht mit der Sprache des Produkts formuliert werden.
- Der Hintergrund einer Aufgabe sollte einem realen Szenario, welchem die Benutzer und Benutzerinnen bei ihrer Arbeit begegnen könnten, entsprechen.
- Die Aufgabenstellung sollte nicht aus den einzelnen Schritten zum Ziel bestehen, sondern das gewünschte Ergebnis beschreiben.
- Der Text sollte kurz gehalten werden und keine unnötigen Details enthalten. Die Informationen in der Aufgabenstellung sollten sich auf das Minimum beschränken, dass der Person auch im echten Leben zur Verfügung stehen würde.
- Falls die Eingabe von persönlichen Informationen notwendig ist, wie zum Beispiel ein Name oder die Emailadresse, sollten den Teilnehmern und Teilnehmerinnen Testdaten für die Eingabe zur Verfügung gestellt werden. Wenn die Eingabe von echten Daten notwendig ist, sollte das zu Beginn mit den Teilnehmern und Teilnehmerinnen geklärt werden und dennoch Testdaten zur Verfügung stehen.
- Für andere notwendige Eingaben sollten ebenfalls konkrete Werte in der Aufgabenstellung angegeben werden, wie zum Beispiel die Menge eines Produktes, die bestellt werden soll. [10, p. 128 ff]

Außerdem ist es wichtig, dass alle Aufgabenstellungen ein klares Ende haben. Den Teilnehmern und Teilnehmerinnen muss klar sein, wann sie eine Aufgabe abgeschlossen haben. Auch für die Auswertung der Aufgaben muss erkennbar sein, wann ein Teilnehmer oder eine Teilnehmerin denkt fertig zu sein. Schließlich kann es vorkommen, dass der Teilnehmer oder die Teilnehmerin nur annimmt fertig zu sein, obwohl die Aufgabe eigentlich noch nicht beendet ist. Genauso kann es passieren, dass eine Person weitermacht, obwohl sie schon fertig wäre, weil sie sich nicht sicher ist oder das Ende nicht erkannt hat. Beides sind Indikatoren für Usability Probleme, die einen positiven Abschluss einer Aufgabe beeinflussen und verhindern können. Wenn eine Aufgabe ein konkretes Ende definiert hat, können diese Fälle besser unterschieden werden. [10, p. 133 f]

Zusätzlich hat die eindeutige Definition des erwarteten Ergebnisses eine weitere Bedeutung bei der Auswertung der Tests. Nur so kann anschließend verlässlich bewertet werden, welche Aufgaben erfolgreich abgeschlossen wurden und welche nicht.

2.3.3.4 Messdaten

Bereits vor Durchführung eines Tests muss definiert werden, welche Daten während dem Test gesammelt bzw. aufgezeichnet werden sollen und welche Messungen dafür notwendig sind. Zum einen muss sichergestellt werden, dass die richtigen Informationen erfasst werden, die zur Auswertung der Testziele notwendig sind. Andererseits ist die Planung wichtig, um die Messung dieser Daten vorbereiten zu können. Beispielsweise soll in den meisten Fällen, die benötigte Zeit für die Durchführung der Aufgabe gemessen werden. Daher muss entsprechendes Equipment für die Zeitmessung vorbereitet werden. Wenn man anschließend Aussagen zu den einzelnen Usability Kriterien treffen möchte, muss man sich ebenfalls im Vorhin überlegen, mit welchen konkreten Daten die Effizienz, Effektivität und Zufriedenheit gemessen werden kann.

Welche Metriken für jedes Usability Kriterium herangezogen werden kann und welche Daten dazu notwendig sind, wird im Kapitel 2.5 Usability Metriken näher erläutert.

2.3.3.5 Testumgebung

Abschließend sollte man noch planen, wo und mit welchem Equipment der Test durchgeführt werden soll. In der klassischen Variante findet der Test in einem Usability Labor statt. Das bedeutet, es gibt einen fixen Raum, in dem Usability Tests durchgeführt werden und wo das Testequipment dauerhaft aufgebaut ist. Zusätzlich kann es auch einen zweiten Raum geben, in dem weitere Personen den Test beobachten können. Im Gegensatz dazu besteht die Möglichkeit mit einem portablen Usability Labor zu arbeiten. Dabei gibt es keinen dezidierten Raum für die Tests, sondern das notwendige Equipment wird zu unterschiedlichen Orten transportiert. So kann ein Usability Test auch direkt im Arbeitsumfeld der Benutzer und Benutzerinnen durchgeführt werden. [7, pp. 100-110]

In der einfachsten Form besteht das notwendige Equipment für einen Test aus einem Computer, an dem der Benutzer oder die Benutzerin arbeiten kann, sowie gegebenenfalls Geräten zur Audio- und Videoaufnahme. Außerdem ist es hilfreich eine Software zur Testaufzeichnung zu nutzen, um die Sammlung und Auswertung der Daten zu erleichtern. Je nach Produkt und Testszenario können zusätzlich andere Requisiten notwendig sein, um die Testumgebung möglichst ähnlich zum echten Arbeitsumfeld der Nutzer und Nutzerinnen zu gestalten. [5, p. 222]

Neben der Vorbereitung des technischen Equipments muss auch die zu testende Applikation vorbereitet werden. Für den Test müssen zum Beispiel bestimmte Ressourcen

wie Benutzerkonten oder Dokumente erstellt werden, die der Teilnehmer oder die Teilnehmerin bei der Nutzung des Produktes benötigt. Ebenso müssen Testdaten für die Durchführung der Aufgaben vorbereitet werden und eventuell bestimmte Situationen im System eingerichtet werden, auf denen manche Aufgaben basieren. Außerdem sollten Vorbereitungen für mögliche Probleme getroffen werden. Bei aufeinander aufbauenden Aufgaben kann es beispielsweise sinnvoll sein, Lösungen vorzubereiten, mit denen man bei einer bestimmten Aufgabe wieder einsteigen kann, falls die Software dazwischen abgestürzt ist. [5, p. 222 f]

2.3.3.6 Ablauf

Zu Beginn eines Usability Tests sollte der Teilnehmer oder die Teilnehmerin über die Testumgebung und den Ablauf informiert werden. Dazu zählen zum Beispiel:

- die Beschreibung des Testequipments inkl. Kamera und Mikrofon sowie des Raumes, falls spezielle Einrichtungen wie ein Einwegspiegel vorhanden sind
- die Information, ob Beobachter oder Beobachterinnen im selben oder einem getrennten Raum anwesend sind
- gegebenenfalls die Vorbereitung der Teilnehmer oder Teilnehmerinnen, dass sie während dem Test laut mitdenken sollen
- die Erklärung, wie der Teilnehmer oder die Teilnehmerin die Aufgaben erhält und wie er oder sie das Ende einer Aufgabe kundgeben soll
- die Vorbereitung der Teilnehmer und Teilnehmerinnen darauf, dass der Moderator oder die Moderatorin keine inhaltlichen Fragen zum Produkt beantworten oder den erfolgreichen Abschluss einer Aufgabe bestätigen wird
- die Information darüber, ob und wie Notizen aufgezeichnet werden, zum Beispiel handschriftlich oder mit einem Laptop
- die Betonung, dass jedes Feedback hilfreich ist und dass das Produkt getestet wird und nicht der Teilnehmer oder die Teilnehmerin
- die Information zum Status des zu testenden Produkt, also ob es zum Beispiel ein Prototyp oder ein fertiges System ist [10, p. 201 f]

Anschließend beginnt der Teilnehmer oder die Teilnehmerin mit der Durchführung der einzelnen Aufgabenstellungen. Dabei kann entweder der Moderator bzw. die Moderatorin der Testperson die Aufgabenstellungen geben oder sie werden automatisch durch eine Testsoftware präsentiert. Zwischen den Aufgaben kann der Moderator oder die Moderatorin bestimmte Aktionen der Benutzer oder Benutzerin näher hinterfragen. Grundsätzlich sollte der Moderator oder die Moderatorin aber nicht eingreifen, außer wenn etwas komplett schief läuft, wie zum Beispiel in folgenden Situationen:

- Das System hängt sich auf oder stürzt ab.
- Der Teilnehmer oder die Teilnehmerin hat sehr große Schwierigkeiten eine Aufgabe abzuschließen.

- Der Teilnehmer oder die Teilnehmerin schweift von der eigentlichen Aufgabenstellung ab. [10, p. 214]

Wie im Kapitel 2.4 Fragebögen näher beschrieben wird, ist es außerdem möglich nach einer Aufgabe oder nach Abschluss aller Aufgaben dem Teilnehmer oder der Teilnehmerin direkt Fragen zu stellen, um zusätzliches Feedback zu erhalten.

Nach Abschluss der Tests sollten die gesammelten Daten konsolidiert und ausgewertet werden. Das Kapitel 2.5 Usability Metriken erläutert näher, welche Metriken sich aus den gemessenen Daten berechnen lassen.

2.3.3.7 Richtlinien für Testmoderation

Das richtige Verhalten als Moderator oder Moderatorin während einer Testsession ist sehr wichtig, da man sonst das Verhalten der Teilnehmer und Teilnehmerinnen beeinflussen und so die Ergebnisse beeinträchtigen kann. Daher sollten folgende allgemeinen Richtlinien berücksichtigt werden:

- Die Testsession sollte objektiv und neutral moderiert werden.
- Man sollte sich bewusst sein, dass bestimmte Änderungen der Stimme oder der Körpersprache unbewusst Einfluss auf den Teilnehmer oder die Teilnehmerin haben können.
- Jeder Teilnehmer und jede Teilnehmerin sollten unabhängig von vorherigen Testpersonen betrachtet werden.
- Falls es angebracht ist, sollte die „Think Aloud“ Technik eingesetzt werden und die Testperson aufgefordert werden, ihre Gedanken während der Nutzung des Produktes laut auszusprechen.
- Die Interaktion mit einem Teilnehmer oder einer Teilnehmerin und das Hinterfragen von Aktionen sollte zu angemessener Zeit und in angemessener Form passieren.
- Man sollte objektiv bleiben aber einen lockeren Ton beibehalten, um eine entspannte Atmosphäre für die Teilnehmer und Teilnehmerinnen zu schaffen.
- Wenn eine Testperson Probleme bei der Nutzung hat, sollte nicht zu schnell eingegriffen und Hilfestellung gegeben werden.
- Nur als letzter Ausweg sollte ein Teilnehmer oder eine Teilnehmerin bei der Lösung einer Aufgabe unterstützt werden.
- Fehler, die einem bei der Moderation passieren, sollten einfach ignoriert werden.
- Wenn die Aufgabenstellungen mündlich gegeben werden, sollte man nach Beobachtung der Fertigstellung ein paar Momente warten, bevor man zur nächsten Aufgabe weitergeht. Dadurch erhält der Teilnehmer oder die Teilnehmerin nicht frühzeitig die Information, dass die Aufgabe erfolgreich gelöst wurde. [7, pp. 202-211]

2.4 Fragebögen

Zusätzlich zu den Erkenntnissen aus der Beobachtung von Teilnehmern und Teilnehmerinnen während dem Test, kann es auch weitere hilfreiche Einblicke liefern, wenn man die Teilnehmer und Teilnehmerinnen explizit zu bestimmten Punkten befragen kann. Oft ist es zum Beispiel interessant den Hintergrund und die Erfahrungen der Teilnehmer und Teilnehmerinnen sowie deren Meinung zu einer bestimmten Aufgabe oder ganz allgemein zur Usability des Produktes zu erfahren. Mit Hilfe von Fragebögen können genau diese Informationen eingeholt werden. Der Einsatz von Fragebögen ist auch die einfachste Methode um das Kriterium Zufriedenheit für ein Produkt zu messen. So können die Benutzer und Benutzerinnen direkt zu den unterschiedlichen Aspekten der Zufriedenheit befragt werden.

Obwohl eine mündliche Befragung möglich ist, haben schriftlichen Fragebögen den Vorteil, dass man jeden Teilnehmer und jede Teilnehmerin die exakt selbe Frage stellt und keine Fragen vergessen werden können. Andernfalls wird die Vergleichbarkeit der Antworten erschwert. Eine weitere wichtige Rolle zur Vereinfachung der Vergleichbarkeit spielt die Form, in der die Fragen gestellt werden. Antworten auf offene Fragen können sehr unterschiedlich ausfallen, was die Interpretation und einheitliche Auswertung erschwert. Dahingegen lassen sich Antworten in Form von Bewertungsskalen einfacher auswerten und vergleichen. [5, p. 208]

2.4.1 Zeitpunkte

Die Befragung der Teilnehmer und Teilnehmerinnen kann zu unterschiedlichen Zeitpunkten während einem Usability Test stattfinden. Je nachdem, welche Information man erhalten möchte, gibt es drei übliche Zeitpunkte, an denen den Personen Fragen gestellt werden:

- Vor dem Test (Pre-Test)
- Nach Abschluss einer Aufgabe (Post-Task)
- Nach dem Test (Post-Test)

Pre-Test Fragebögen

Eine Befragung vor dem Test hat meistens das Ziel, allgemeine Informationen zu den Nutzern und Nutzerinnen zu sammeln und deren Qualifikationen zu verifizieren. So können Unterschiede zwischen den Personen ausgemacht werden, die für die Ergebnisse relevant und später bei der Interpretation bestimmter Aktionen und des Benutzerverhaltens hilfreich sein können. [10, p. 173 f]

Post-Task Fragebögen

Wenn ein Usability Tests aus mehreren Aufgabenstellungen besteht, können die Teilnehmer und Teilnehmerinnen nach jeder Aufgabe befragt werden. Im Vergleich zur Befragung der

Personen zu einer bestimmten Aufgabe nachdem der ganze Test bereits abgeschlossen ist, hat es dazwischen den Vorteil, dass die Eindrücke noch frisch sind und die Personen sich besser erinnern können. Oft wird der Benutzer oder die Benutzerin anschließend an eine Aufgabe, um die Bewertung einer der folgenden Aspekte gebeten:

- Wie einfach die Durchführung der Aufgabe war.
- Ob weniger oder mehr Zeit als erwartet benötigt wurde.
- Wie wahrscheinlich man diese Funktion wiederverwenden möchte. [10, p. 176]

Dabei kann man nach jeder Aufgabe die gleichen Fragen stellen oder diese unterschiedlich gestalten und an die einzelnen Aufgaben spezifisch anpassen.

Post-Test Fragebögen

Die letzte Möglichkeit, die Teilnehmer oder Teilnehmerinnen zu befragen, bietet sich nach dem der Test abgeschlossen wurde an. Bei dieser Art von Befragung geht es vor allem darum, die allgemeine Wahrnehmung des Produktes sowie die Meinung der Personen zur Usability des Produktes und dem Nutzungserlebnis einzufangen. [10, p. 176 f]

Im folgende Kapitel werden einige Standardfragebögen genannt, die als Post-Task oder als Post-Test Fragebögen eingesetzt werden können.

2.4.2 Standardfragebögen

Es gibt einige Standardfragebögen, die während oder unmittelbar nach einem Usability Test eingesetzt werden können, um die Zufriedenheit von Nutzern und Nutzerinnen sowie deren Wahrnehmung der Usability eines Produktes zu ermitteln. Diese Fragebögen sind in einem bestimmten Format gestaltet und definieren meistens bestimmte Richtlinien für die Auswertung von Metriken aus den Ergebnissen. Dadurch ist ein wiederholter Einsatz auch bei unterschiedlichen Produkten möglich und die Ergebnisse vergleichbar. [14, p. 185]

Auch wenn diese Fragebögen nicht im Detail auf das zu testende Produkt zugeschnitten sind, hat die Nutzung von Standardfragebögen laut Jeff Sauro und James R. Lewis [14] folgende Vorteile:

- Objektivität: Ergebnisse können von anderen unabhängig verifiziert werden.
- Reproduzierbarkeit: Studien sind einfacher durch andere reproduzierbar und verlässlicher.
- Quantifizierung: Es ermöglicht eine detailliertere statistische Auswertung.
- Wirtschaftlichkeit: Der Einsatz ist weniger aufwändig und kostengünstiger als die Entwicklung eines eigenen Fragebogens.
- Kommunikation: Die Ergebnisse können einfacher und standardisiert kommuniziert werden.

- Wissenschaftliche Generalisierung: Standardisierung unterstützt dabei, allgemeine Schlüsse aus den Ergebnissen ziehen zu können. [14, p. 165 f]

Zusätzlich zeigt sich, dass mit Hilfe von Standardfragebögen auch Vergleiche von technologisch und inhaltlich sehr unterschiedlichen Systemen möglich ist. Auf Grund der Technologieunabhängigkeit können sie sowohl für ältere Anwendungen als auch für moderne und neuartige Produkte eingesetzt werden.

Jeff Sauro und James R. Lewis nennen in „Quantifying the User Experience“ [14] folgende Standard-Fragebögen für Post-Task und Post-Test Fragebögen.

Post-Task Fragebögen

- After-Scenario Questionnaire (ASQ)
- Expectation ratings (ER)
- Usability Magnitude Estimation (UME)
- Single Ease Question (SEQ)
- Subjective Mental Effort Question (SMEQ) [14, p. 186]

Post-Test Fragebögen

- Questionnaire for User Interaction Satisfaction (QUIS)
- Software Usability Measurement Inventory (SUMI)
- Post-Study System Usability Questionnaire (PSSUQ)
- Software Usability Scale (SUS) [14, p. 186]

Im Rahmen dieser Arbeit wurde der Standardfragebogen „Software Usability Scale“ eingesetzt und dieser wird daher im folgenden Kapitel näher beschrieben.

2.4.3 System Usability Scale

Der Fragebogen „Software Usability Scale“, kurz SUS, wurde im Jahr 1986 von John Brooke entwickelt. Das Ziel war es, einen kurzen und verlässlichen Fragebogen zu erstellen, der unmittelbar nach einem Usability Test eingesetzt werden kann. Er hat den Zweck die subjektive Wahrnehmung der Benutzer und Benutzerinnen zur Usability eines Systems zu messen. [15]

Der SUS Fragebogen besteht aus zehn Sätzen, wovon jeweils die Hälfte positiv und die andere Hälfte negativ formuliert sind. Zu jedem Satz kann der Benutzer oder die Benutzerin seinen oder ihren Zustimmungsgrad in einer Skala bestehend aus fünf Punkten ausdrücken. [15]

Zusätzlich kann der SUS auch in einer abgeänderten Form mit nur positiv formulierten Sätzen eingesetzt werden. Jeff Sauro und James R. Lewis konnten mit einem rein positiven SUS Fragebogen keinen signifikanten Unterschied in den Ergebnissen zum originalen Fragebogen finden. [16]

Beim SUS handelt es sich um einen der meist genutzten und weit verbreitetsten Usability Fragebögen. Auch John Brooke selbst konnte dies nach über 25 Jahren bestärken, da der Fragebogen bereits in über 1.200 Publikationen zitiert wurde. Darüber hinaus ist er bereits fixer Bestandteil in manchen kommerziellen Anwendungen zur Usability Evaluierung wie der Software Morae von TechSmith. [17] Jeff Sauro sammelte über 3.000 Antworten auf den SUS Fragebogen aus öffentlich verfügbaren Datenquellen und Publikationen, um einen globalen Benchmark für SUS Ergebnisse zu generieren. Diese Studien decken die unterschiedlichsten Produkte ab und reichen von großen kommerziellen Applikation über Webseiten bis hin zu mobilen Applikationen. [18] Diese Vielseitigkeit zeigt wie universell der SUS eingesetzt werden kann. Auf Grund dieser Eigenschaften wurde dieser Standardfragebogen auch zur Messung der Zufriedenheit im Rahmen dieser Arbeit ausgewählt.

Die originale englische Version des SUS Fragebogens, sowie die in dieser Arbeit eingesetzte deutsche Version [19] ist im Anhang A einzusehen. Im Kapitel 2.5.3 „Zufriedenheit“ wird die Auswertung und Berechnung eines Gesamtwerts für den SUS näher erläutert.

2.5 Usability Metriken

Usability Metriken bieten eine Möglichkeit, um die Usability eines Produkts in Zahlen zu messen und so ein quantitatives Ergebnis zu erhalten. Ähnlich wie bei einem Längenmaß kann damit in Zahlen ausgedrückt werden, wie einfach ein System zu benutzen ist. Für die Evaluierung von Metriken wird ein verlässliches System zur Messung vorgegeben, das sicherstellt, dass die Ergebnisse reproduzierbar und vergleichbar sind. [12, p. 6 f]

Durch Auswertung von Metriken bieten sich einige Möglichkeiten und Vorteile. Zum einen ermöglichen sie es, die Ergebnisse in einer strukturierten Form zu dokumentieren, zum anderen machen sie Vergleiche einfacher. So kann beispielsweise in konkreten Zahlen festgestellt werden, welche Version eines Produkts die bessere Usability hat, indem beide Versionen getestet, anschließend die gleichen Metriken ausgewertet und einander gegenübergestellt werden. Wenn man einen erneuten Usability Test macht, nachdem Verbesserungsmaßnahmen zu vorher erhobenen Problemen umgesetzt wurden, lässt sich mit Hilfe von Metriken auch klarer erkennen, ob diese Maßnahmen den gewünschten positiven Effekt erzielt haben. Außerdem lässt sich nicht nur erkennen, ob es eine positive Veränderung gab, sondern das Ausmaß der Veränderung wird in konkreten Zahlen

ausgedrückt. Zum Beispiel können so die daraus resultierenden Kostenersparnisse, Zeitersparnisse oder der Return On Investment (kurz ROI) abgeleitet werden. Diese Möglichkeiten führen dazu, dass man weitere Erkenntnisse aus den Ergebnissen gewinnen kann, die eine verlässlichere Basis für Entscheidungen bilden und Entscheidungen auf Grund von falschen Annahmen oder Bauchgefühlen verhindern. Ein weiterer Vorteil ist, dass sich manche Muster oft besser in Zahlen und Grafiken erkennen lassen, und mit Hilfe von Metriken können diese Darstellungsmöglichkeiten genutzt werden. [12, p. 8 f]

Es gibt unterschiedliche Usability Metriken, die bei einem Usability Test ausgewertet werden können, um die drei Usability Kriterien Effektivität, Effizienz und Zufriedenheit für ein Produkt zu messen und so in Zahlen auszudrücken. Im Folgenden werden für jedes dieser Kriterien einige mögliche Metriken beschrieben.

2.5.1 Effektivität

Die Effektivität bezieht sich auf die Genauigkeit und Vollständigkeit mit denen Benutzer und Benutzerinnen, Aufgaben in einem System durchführen können. Um die Effektivität sinnvoll auswerten zu können, müssen sich die gemessenen Metriken immer auf ein konkretes Szenario oder eine Aufgabe beziehen. Ohne einer genauen Definition der Aufgabenstellung ist es beispielsweise nicht möglich festzustellen, ob ein Nutzer oder eine Nutzerin bereits fertig ist und ob die Aufgabe erfolgreich abgeschlossen wurde.

Im Folgenden werden zwei Metriken näher beschrieben, mit denen die Effektivität eines Produktes gemessen werden kann. Diese Metriken zeigen auf, wo etwas in einem System nicht passt, aber nicht warum es nicht passt. Dazu müssen andere Daten aus den Beobachtungen und Befragungen der Teilnehmer und Teilnehmerinnen herangezogen werden.

2.5.1.1 Task Success

Mit dieser Metrik wird ausgewertet, ob ein Benutzer oder eine Benutzerin eine Aufgabe erfolgreich abschließen konnte. Somit kann diese Metrik sehr universell und bei den meisten Usability Tests angewandt werden. Wichtig ist nur, dass jede Aufgabe ein klares Endergebnis definiert hat, damit man eindeutig erkennen kann, ob dieses Ergebnis erreicht und damit die Aufgabe erfolgreich abgeschlossen wurde. [12, p. 65]

In der einfachsten Variante wird das Ergebnis für Task Success in binärer Form ermittelt. Das heißt eine Person kann eine Aufgabe entweder erfolgreich oder nicht erfolgreich abgeschlossen haben. Für jede Person und jede Aufgabe wird daher eine 1 oder eine 0 als Ergebnis notiert. Diese Auswertungsform sollte besonders dann gewählt werden, wenn es bei einem System eine sehr große Rolle spielt, dass Benutzer eine Aufgabe fehlerfrei durchführen können, weil jeder Fehler problematisch ist. Ein Beispiel dafür ist ein

Defibrillator, hier kann jeder Fehler gravierende Auswirkungen haben und muss daher im Test als ein Fehlschlagen der Aufgabe gewertet werden. [12, p. 66 f]

Bei anderen Anwendungen hingegen muss nicht jeder Fehler gleich ein kompletter Misserfolg sein. In einigen Fällen möchte man einen Task nicht wegen so einem Fehler als komplett fehlgeschlagen werten, sondern eine feinere Abstufung wählen. Anstatt einer binären Auswertung können daher auch unterschiedliche Levels für den Erfolg einer Aufgabe definiert werden. Ein Beispiel für diese Levels wäre eine Abstufung in erfolgreich, teilweise erfolgreich und nicht erfolgreich. Das liefert oft ein besseres Bild zu einer Aufgabe und diese zusätzlichen Informationen gehen nicht verloren. [12, p. 70 f]

2.5.1.2 Fehler

Fehler sind falsche Aktionen oder Schritte, die ein Benutzer oder eine Benutzerin durchführt und die dazu führen können, dass eine Aufgabe nicht erfolgreich abgeschlossen wird. Usability Probleme und Fehler sind nicht das Gleiche. Ein Usability Problem ist die Ursache für ein Problem, während ein Fehler die Folge eines Usability Problems ist. [12, p. 82]

Die unterschiedlichsten Aktionen und Ereignisse können einen Fehler darstellen. Im Allgemeinen sind es Aktionen, die den Benutzer oder die Benutzerin vom optimalen Pfad zur erfolgreichen Lösung einer Aufgabe abbringen. Das inkludiert unterschiedliche Arten von Handlungen, wie zum Beispiel:

- Eine falsche Eingabe von Werten
- Eine falsche Auswahl eines Buttons, Menüeintrags, Links, ...
- Eine falsche Reihenfolge an Schritten
- Ein Auslassen einer wichtigen Aktion [12, p. 83]

Die Auswertung von Fehlern kann aufzeigen, wie viele Fehler und wo diese passiert sind. Vor allem bei einem Vergleich von verschiedenen Designs oder Produktversionen kann man so gut die unterschiedliche Häufigkeit von Fehlern oder die unterschiedlichen Arten von Fehlern erkennen. Das Ermitteln dieser Metrik kann in folgenden Situationen besonders nützlich sein:

- Wenn ein Fehler die Effizienz stark beeinflusst, weil dadurch beispielsweise Aktionen doppelt durchgeführt werden müssen und es die Ausführung einer Aufgabe wesentlich verlangsamt.
- Wenn ein Fehler zusätzliche Kosten verursacht.
- Wenn ein Fehler zum Misserfolg einer Aufgabe führt. [12, p. 82 f]

2.5.2 Effizienz

Die Effizienz bezieht sich auf das Ausmaß der eingesetzten Ressourcen bei der Nutzung eines Produktes. Am Häufigsten wird dabei betrachtet, wie viel Zeit für eine bestimmte

Aufgabe benötigt wird. Neben der Zeit werden im Folgenden aber auch weitere Beispiele an Metriken vorgestellt, die andere Ressourcen betrachten.

2.5.2.1 Time on Task

Diese Metrik misst, wie lange ein Benutzer oder eine Benutzerin braucht, um eine bestimmte Aufgabe durchzuführen. Damit bietet Time on Task eine gute Möglichkeit, um die Effizienz eines Systems zu messen. In den meisten Fällen ist es umso besser für ein System, je schneller eine Aufgabe erledigt werden kann. Das hat eine noch größere Bedeutung, umso öfter diese Aufgabe durchgeführt werden muss. Bei regelmäßigen Aktionen ist eine effiziente Handhabung sehr wichtig und diese Zeitersparnisse kann man oft direkt in Kosteneinsparungen umrechnen. [12, p. 74 f]

Bei der Zeitmessung selbst ist zu beachten, dass der Start und das Ende richtig gewählt werden. Die Messung sollte erst beginnen, sobald der Teilnehmer oder die Teilnehmerin die Aufgabenstellung gelesen hat und wirklich mit der Durchführung startet. Für das Ende einer Aufgabe muss der Person klar sein, wie sie angibt, dass sie fertig ist. In einer Testsoftware kann das meistens automatisch passieren, indem man auf einen Antwort- oder Ende-Button klickt. Wenn man die Zeit manuell misst, sollte man dies vorher genau festlegen und dem Teilnehmer oder der Teilnehmerin erklären, in welcher Form die Antwort gegeben werden soll, zum Beispiel schriftlich oder mündlich. Die Zeitmessung sollte dann beendet werden, wenn die Person ihre Antwort gibt. [12, p. 75 f]

2.5.2.2 Anzahl an Aktionen

Time on Task wurde bereits als Metrik genannt, um die Effizienz in Form von Zeit zu messen. Eine andere Möglichkeit ist es, den Aufwand in der Anzahl an Aktionen oder Schritten zu zählen, die ein Nutzer oder eine Nutzerin bei der Durchführung einer Aufgabe gemacht hat. Umso weniger Schritte eine Person ausführen muss, desto produktiver und effizienter kann sie die Aufgabe lösen. [12, p. 87]

2.5.2.3 Kognitiver und physischer Aufwand

Der Aufwand einer Aufgabe kann auch in kognitiven und physischen Aufwand unterteilt werden. Der kognitive Teil betrifft das richtige Auffinden, Erkennen und Interpretieren, welche Aktion als nächstes notwendig ist. Die Durchführung dieser Aktion, wie zum Beispiel der Klick mit der Maus oder die Eingabe mit der Tastatur, sind der physische Aufwand. Manchmal kann es angebracht sein, die Effizienz in diesem Detailgrad auszuwerten. Bei einer Anwendung, wie einem Navigationssystem, ist es zum Beispiel sehr wichtig, dass der kognitive und physische Aufwand möglichst gering ist, weil der Hauptfokus des Benutzers oder der Benutzerin woanders bleiben und nicht auf die Nutzung des Systems fallen sollte. [12, p. 87]

2.5.2.4 Verlorenheit

Bei Webseiten bzw. webbasierten Applikation gibt es einen weiteren möglichen Indikator für die Effizienz, der als Verlorenheit bzw. im Englischen als „Lostness“ bezeichnet wird. Dieser Faktor beruht auf der Annahme, dass umso mehr Seiten eine Person während einer Aufgabe aufruft, desto wahrscheinlicher kennt sich diese Person nicht aus und fühlt sich verloren. Durch Messung der Anzahl verschiedener besuchter Webseiten (N), der Gesamtanzahl an allen besuchten Seiten inklusiver der doppelten Aufrufe (S) und der Minimumanzahl an notwendigen Seitenaufrufen (R) kann anschließend der Verlorenheitsgrad mit folgender Formel (1) berechnet werden:

$$(1) \quad L = \sqrt{\left(\frac{N}{S} - 1\right)^2 + \left(\frac{R}{N} - 1\right)^2}$$

Das perfekte Ergebnis wäre 0. Dies würde bedeuten, dass sich der Nutzer oder die Nutzerin optimal zu recht gefunden hat. [12, p. 89 f]

2.5.3 Zufriedenheit

In den vorherigen Kapiteln wurden Metriken beschrieben, die aus Daten ausgewertet werden, welche aus der Beobachtung der Personen bei der Nutzung des Produktes entstehen. Eine weitere übliche Methode, um Informationen über die Usability eines Produktes zu erhalten, ist die direkte Befragung der Benutzer und Benutzerinnen. Dabei werden Personen gebeten zu unterschiedlichen Aspekten, wie der Zufriedenheit, der Einfachheit der Benutzung, dem Erscheinungsbild, der Wahrnehmung bestimmter Bereiche oder dem Verständnis von Begriffen, ihre Meinung zum Beispiel in Form von offenen Fragen oder Bewertungsskalen zu geben. Deshalb handelt es sich bei der Auswertung dieser Antworten um Metriken, die Aufschluss über die Zufriedenheit der Benutzer und Benutzerinnen geben.

Diese Metriken liefern wichtige Informationen darüber, wie die Benutzer und Benutzerinnen ein System wahrnehmen und was sie über das System denken. Es ist immer das Ziel, dass ein Produkt in positiver Erinnerung bleibt und Personen zuerst an dieses Produkt denken, wenn sie eine Lösung in dessen Einsatzgebiet benötigen. Dafür ist es oft weniger entscheidend, wie lange eine Person zur Durchführung von Aufgaben gebraucht hat. Wenn sie trotzdem zufrieden mit der Nutzung war, ist eine erneute Benutzung wahrscheinlicher, als wenn es schneller ging, aber kein positiver Eindruck hinterlassen wurde. [12, p. 122 f]

Die häufigste Form, in der Fragen im Rahmen einer Usability Evaluierung zu beantworten sind, ist eine Bewertungsskala. Als klassische Varianten lassen sich hier die Likert Skala und die semantische Differenzialskala einsetzen. Bei der Likert Skala erhalten die Teilnehmer und Teilnehmerinnen eine Aussage und sollen ihre Zustimmung dazu bewerten.

Üblicherweise wird eine 5-Punkte-Skala eingesetzt, die sich aus folgenden Zustimmungsggraden zusammensetzt:

- trifft zu
- trifft eher zu
- weder noch
- trifft eher nicht zu
- trifft nicht zu [12, p. 123 f]

Beim semantischen Differenzial hingegen werden den Teilnehmern und Teilnehmerinnen Paare aus gegensätzlichen Begriffen oder Eigenschaften präsentiert. Zwischen den Begriffen wird eine Skala bestehend aus fünf oder sieben Punkten dargestellt, auf der die Personen ihre Tendenz zu einem Begriff abgeben können. Beispiele für solche Paare sind:

- Klein und groß
- Schwach und stark
- Kalt und warm [12, p. 123 f]

Es gibt unterschiedliche Möglichkeiten, wie die Antworten der Teilnehmer und Teilnehmerinnen gesammelt werden können. Die Personen können zum Beispiel mündlich befragt werden, sie können ihre Antworten in einem Fragebogen auf Papier niederschreiben oder den Fragebogen mit Hilfe einer Software elektronisch ausfüllen. Obwohl mündlich für den Teilnehmer oder die Teilnehmerin am einfachsten ist, hat es den Nachteil, dass die Personen meistens weniger ehrlich antworten. Viele Menschen antworten mündlich positiver oder sagen eher, was sie glauben, dass man hören möchte. Auf Papier oder in elektronischer Form ist es anonym, weil es ohne Beobachtung ausgefüllt werden kann und daher geben die meisten Personen direkteres und ehrlicheres Feedback. Außerdem lassen sich elektronische Fragebögen einfacher auswerten, da in der Software meistens bereits Funktionen dafür vorgesehen sind. [12, p. 125]

Im Rahmen dieser Arbeit wurde am Ende einer Testsession den Teilnehmern und Teilnehmerinnen der SUS Fragebogen vorgelegt. Bei diesem Fragebogen lässt sich aus den Antworten ein Gesamtwert – der sogenannte SUS-Score – als Metrik berechnen. Jeff Sauro beschreibt in „A Practical Guide to the System Usability Scale“ [18] folgende Schritte zur Berechnung dieser Metrik:

1. Die Antworten der Teilnehmer und Teilnehmerinnen erhalten die Werte 1 bis 5 auf der Skala von „Ich stimme überhaupt nicht zu“ (1) bis „Ich stimme vollständig zu“ (5).
2. Für Fragen mit ungerader Nummerierung wird 1 vom Antwortwert abgezogen. Für alle Fragen mit gerader Nummerierung wird der Antwortwert von 5 abgezogen. Nach diesem Schritt befindet sich das Ergebnis für jede Frage im Bereich von 0 bis 4.
3. Diese konvertierten Ergebnisse aller Fragen werden anschließend für jeden Teilnehmer und jede Teilnehmerin aufsummiert und mit dem Wert 2,5 multipliziert.

So erhält man einen Gesamtwert pro Teilnehmer oder Teilnehmerin im Bereich von 0 bis 100, der den SUS Score dieses Teilnehmers oder dieser Teilnehmerin darstellt. [18, p. 19]

Die Reaktionen von Benutzern und Benutzerinnen auf ein Produkt und ihre Zufriedenheit damit, zeigen sich aber nicht nur in den Antworten auf Fragen. Während der Nutzung eines Produktes zeigen Personen ganz unbewusst unterschiedliche Reaktionen, wie zum Beispiel lachen, unruhig im Sessel sitzen oder unaufgeforderte verbale Ausdrücke. Dieses Verhalten wird ausgelöst, weil Benutzer und Benutzerinnen die unterschiedlichsten Emotionen, wie Stress, Aufregung, Frustration, Ärger oder Überraschung, fühlen können. Diese Aspekte können ebenfalls einen wichtigen Einblick in das Nutzungserlebnis eines Produktes liefern und für die Zufriedenheit mit dem Produkt entscheidend sein. [12, p. 163] Im Folgenden werden drei Beispiele für Metriken genannt, um auch diese Aspekte der Zufriedenheit zu messen.

Unaufgeforderte verbale Ausdrücke

Wenn ein Teilnehmer oder eine Teilnehmerin während einem Test etwas unaufgefordert sagt, kann das hilfreiche Informationen zu seinem oder ihrem emotionalen und mentalen Zustand liefern. Meistens lassen sich diese Ausdrücke in positive, negative und neutrale Kommentare unterteilen. So lässt sich die Anzahl der Kommentare für jede dieser drei Kategorien ermitteln und anschließend das Verhältnis zueinander berechnen. Umso mehr negative Kommentare gegeben werden, desto eher deutet dies auf eine Unzufriedenheit beim Teilnehmer oder der Teilnehmerin hin. [12, p. 163 f]

Emotionen

Die Messung von Emotionen ist schwierig. Ein explizites Fragen der Benutzer und Benutzerinnen zu Ihren Gefühlen ist oft nicht sehr effektiv, weil Personen in dieser Situation dazu neigen, dass zu sagen, was sie denken, dass man hören möchte. Außerdem können sie Schwierigkeiten beim Ausdrücken ihrer Gefühle haben oder ihre Gefühle grundsätzlich ungern einem Fremden mitteilen. Um sich nicht nur auf die Aussagen der Personen verlassen zu müssen, gibt es drei andere Möglichkeiten um Emotionen zu messen. Es können Gesichtsausdrücke, die Hautleitfähigkeit oder ein EEG herangezogen werden, um aus den damit gemessenen Reaktionen Emotionen abzuleiten. [12, p. 176]

Stress

Wenn während der Nutzung eines Produktes schwierige oder unklare Situationen für einen Benutzer oder eine Benutzerin entstehen, kann das Stress verursachen. Dadurch kann ein negatives Gefühl gegenüber dem Produkt entwickelt werden und daher ist Stress ein wichtiger Faktor für das Benutzererlebnis. Am Häufigsten wird Stress über die Herzfrequenz oder genauer die Herzfrequenzvariabilität – kurz HRV – gemessen. Dabei werden die Zeitintervalle zwischen den Herzschlägen bestimmt und die Veränderung dieser Frequenz gemessen. Bestimmte Muster in der Veränderung können auf Stress hindeuten. Trotzdem

lässt sich Stress schwer auswerten und konkrete Ursachen für Stress ermitteln. Bei einem Usability Test muss diese Messung nicht aussagekräftig sein, weil auch andere Faktoren die Ursache dafür sein können. Mögliche andere Ursachen sind zum Beispiel Nervosität, weil man sich in einer Art Testsituation befindet, oder die Sorge, dass man etwas falsch machen könnte. [12, p. 182]

2.5.4 Weitere Metriken

Die bisher beschriebenen Metriken konnten einem der Usability Kriterien klar zugeordnet werden. Es gibt aber auch Auswertungsmöglichkeiten, die für mehr als nur ein Kriterium eingesetzt werden können. Mit Metriken zur Erlernbarkeit und aus dem Eye Tracking können beispielsweise Aspekte der Effektivität und Effizienz gemessen werden. Die Auswertung der Usability Probleme hingegen ist keine direkte Messung dieser Kriterien, sondern liefert vielmehr die Ursachen und Gründe für eine schlechte Effektivität, Ineffizienzen oder Unzufriedenheit. Daher werden diese drei Beispiele in den folgenden Kapiteln beschrieben.

2.5.4.1 Erlernbarkeit

Die Metrik zur Erlernbarkeit wird gemessen, in dem man sich ansieht, wie viel Zeit und Aufwand notwendig sind, bis eine Person die Nutzung des Systems gut beherrscht. Die Zeitspanne für das Erlernen eines Systems kann kurz oder lange sein.

Kurze Zeitspannen belaufen sich auf Minuten, Stunden oder wenige Tage. Sie sind meistens dann der Fall, wenn etwas sehr häufig benutzt wird, wie zum Beispiel die tägliche Nutzung einer Zeiterfassung. Durch den häufigen Einsatz entwickeln die Benutzer und Benutzerinnen schnell ein mentales Model, wie die Anwendung funktioniert, und Strategien, um es effizient zu nutzen. Das Gedächtnis spielt dabei meist eine geringe Rolle. Das Gegenteil ist der Fall bei einer langen Zeitspanne. Bei manchen Anwendungen erstreckt sich das Erlernen über mehrere Wochen, Monate oder sogar Jahre, weil zwischen der Nutzung sehr große Pausen existieren. Umso größer die Pausen zwischen den Einsätzen sind, desto wichtiger wird das Gedächtnis. Wenn einem die Anwendung nicht gut im Gedächtnis bleibt, kann es passieren, dass man sie bei jeder Nutzung neu erlernen muss. [12, p. 92 f]

Um diese Metrik zu ermitteln, werden im Grunde die zuvor beschriebenen Metriken ausgewertet. Allerdings passiert dies nicht nur einmal sondern im Rahmen von mehreren Usability Tests. Der Abstand zwischen den Tests kann Stunden, Tage oder Wochen betragen und sollte davon abhängen, welche Häufigkeit in der Realität zu erwarten ist. Bei jedem Test werden die gleichen Metriken gemessen und durch den Lernprozess sollten sich die Werte verbessern. [12, p. 94]

2.5.4.2 Eye Tracking

Beim Eye Tracking werden die Augenbewegungen der Benutzer und Benutzerinnen aufgezeichnet und so gemessen, worauf die Personen am Bildschirm genau hinsehen.

Dazu sind eine Infrarotkamera und eine Infrarotlichtquelle notwendig. Die Lichtquelle erzeugt eine Reflektion am Auge des Teilnehmers oder der Teilnehmerin, welche die Kamera einfängt und dann wird die Distanz zwischen dem Reflektionspunkt und der Pupille berechnet. Wenn die Person ihre Augen bewegt, ändert sich die Distanz zwischen diesen zwei Punkten. Um daraus anschließend die genaue Stelle am Bildschirm zu ermitteln, auf die der Teilnehmer oder die Teilnehmerin geblickt hat, muss das System zuerst kalibriert werden. Dazu muss die Person auf bestimmte vordefinierte Punkte am Bildschirm schauen. Dabei wird die oben beschriebene Distanz zwischen Reflektionspunkt und Pupille gemessen und daraus kann später interpoliert werden, welche Stelle am Bildschirm der Benutzer oder die Benutzerin gerade ansieht. Eine passende Kalibrierung spielt dabei eine bedeutende Rolle. Nur so kann man sicherstellen, dass die Stellen in der anschließenden Auswertung auch wirklich den Punkten entsprechen, auf die der Benutzer oder die Benutzerin geblickt hat. Während der Messung ist es wichtig, dass die Person sich nicht viel im Sitz bewegt, da die Kalibrierung sonst nicht mehr passen könnte. [12, p. 165 ff]

Eye Tracking kann bei einem Usability Test hilfreiche Information liefern, weil man zum Beispiel erkennen kann, ob ein Element in der Applikation gar nicht gesehen wird oder zwar wahrgenommen aber nicht verstanden wird. Durch den Einsatz von Eye Tracking lassen sich folgende Metriken auswerten:

- Verweildauer: gesamte Zeit, die eine Person auf einen bestimmten Bereich geschaut hat
- Anzahl der Fixierungen: Anzahl, wie oft sich der Blick einer Person auf einen bestimmten Bereich fixiert hat
- Fixierdauer: durchschnittliche Zeit pro Fixierung auf einen bestimmten Bereich
- Abfolge: Reihenfolge, in der die unterschiedlichen Bereiche angesehen wurden
- Zeit zur ersten Fixierung: Zeit bis ein bestimmter Bereich zum ersten Mal angesehen wurde
- Erneute Besuche: Anzahl, wie oft eine Person zu einem Bereich zurückgekehrt ist, nachdem sie dazwischen einen anderen Bereich angesehen hat
- Trefferquote: Anzahl der Personen, die zumindest eine Fixierung auf einen bestimmten Bereich hatten [12, p. 172 ff]

2.5.4.3 Usability Probleme

Die Hauptaufgabe einer Usability Evaluierung ist es Usability Probleme und ihre Ursachen zu ermitteln, um darauf basierend Verbesserungsvorschläge zu ermitteln. Tom Tullis und

Bill Albert nennen in „Measuring the User Experience“ [12] folgende übliche Arten von Usability Problemen:

- Das Verhalten, das den Abschluss einer Aufgabe verhindert
- Das Verhalten, das jemanden vom normalen Kurs abbringt
- Der Ausdruck von Frust beim Teilnehmer oder der Teilnehmerin
- Das Nicht-Erkennen von etwas, das bemerkt werden sollte
- Die Aufgabe wird als abgeschlossen angesehen, obwohl sie es noch nicht ist.
- Die Ausführung einer Aktion, die nicht zur richtigen Lösung einer Aufgabe führt
- Die Fehlinterpretation von Informationen
- Die Auswahl eines falschen Links zur Navigation zwischen Webseiten [12, p. 100]

Von besonderer Bedeutung bei der Identifikation von Usability Problemen ist es, dass echte von falschen Problemen unterschieden werden. Dabei sind mit echten Problemen jene gemeint, die nachvollziehbare Schwierigkeiten darstellen und von denen auszugehen ist, dass sie mehrere Benutzer und Benutzerinnen betreffen können. Andere Beobachtungen von falschen Aktionen eines Teilnehmers oder einer Teilnehmerin, können hingegen auch nur eine Anomalie oder Eigentümlichkeit bei dieser Person darstellen. Das ist vor allem der Fall, wenn es sich nicht logisch nachvollziehen lässt und keine sinnvolle Begründung für diese Aktion vorhanden ist. [12, p. 101]

Zusätzlich sollte man auch den Schweregrad eines Usability Problems berücksichtigen. Nicht jeder Fall ist gleich gravierend wie ein anderer, da die Probleme unterschiedlich starke Auswirkungen haben können. Die Bewertung des Schweregrads ist daher wichtig, damit die gefundenen Punkte priorisiert und der Fokus vorrangig auf die schwerwiegendsten Probleme gelegt werden kann. [12, p. 103 f]

Zur Einteilung des Schweregrads gibt es zwei Varianten. Die üblichste Möglichkeit ist es, die Einteilung auf das Maß der Auswirkungen zu basieren. Dabei könnte die Einschätzung zum Beispiel mit den folgenden drei Levels passieren:

- Niedrig: umfasst Probleme, die Teilnehmer stören aber den Erfolg einer Aufgabe nicht behindern und Effizienz und Zufriedenheit nur leicht beeinflussen
- Mittel: umfasst Probleme, die eine bedeutende Erschwernis für die Teilnehmer und Teilnehmerinnen darstellen aber keinen Misserfolg verursachen
- Hoch: umfasst jedes Problem, dass zum Scheitern einer Aufgabe führt [12, p. 104]

Die zweite Variante zur Einstufung des Schweregrads berücksichtigt noch andere Faktoren neben der Auswirkung auf die Usability. Diese anderen Faktoren können sein:

- Häufigkeit der Nutzung
- Auswirkung auf Unternehmensziele oder Geschäftsinteressen
- Implementierungskosten für die Problembhebung

- Eintrittshäufigkeit [12, p. 105 f]

Obwohl Usability Probleme oft als eine rein qualitative Sache angesehen werden, gibt es dennoch einige Metriken, die in diesem Bereich gemessen werden können, und im Folgenden beschrieben werden.

Häufigkeit von eindeutigen Problemen

Diese Metrik drückt einfach nur die Anzahl der eindeutigen Probleme aus. Zu beachten ist dabei, dass nicht die Gesamtanzahl der angetroffenen Probleme gezählt wird, sondern wenn mehrere Personen das gleiche Problem haben, dies nur einmal gezählt wird. Wenn die Probleme auch anhand ihres Schweregrads eingestuft wurden, kann die Anzahl auch pro Schweregrad ausgewertet werden. So kann man erkennen, wie viele schwerwiegende Probleme es im Vergleich zu unkritischen gibt. Daher ist diese Metrik hilfreich beim Vergleich von unterschiedlichen Versionen eines Produkts oder mehreren Designs, um aufzuzeigen, dass die Anzahl sinkt bzw. wo sie am geringsten ist. [12, p. 108 f]

Häufigkeit von Problemen pro Teilnehmer oder Teilnehmerin

Eine weitere Auswertungsmöglichkeit ist es, sich die Gesamtanzahl der Probleme pro Teilnehmer oder Teilnehmerin im Vergleich von Produktversionen oder Designs anzusehen. Auch hier kann überprüft werden, ob sich die Anzahl pro Person verringert. [12, p. 109]

Häufigkeit der Teilnehmer oder Teilnehmerinnen pro Problem

Es kann auch eine Rolle spielen, wie viele Teilnehmer oder Teilnehmerinnen auf ein bestimmtes Problem gestoßen sind. Wenn man ein Element speziell beachten möchte, kann man so zum Beispiel auswerten, dass in Design A die Hälfte der Teilnehmer ein Problem bei der Benutzung dieses Elements hatten, während es bei Design B nur ein Viertel war. Diese Metrik hilft daher gut, Verbesserungen an bestimmten Elementen oder Bereichen einer Applikation zu messen. Eine andere Einsatzmöglichkeit entsteht aus der Kombination der Häufigkeit mit dem Schweregrad. In diesem Fall zählt man die Anzahl der Personen, die ein schwerwiegendes Problem hatten. Damit lässt sich zwar keine Aussage mehr zu einem konkreten Problem treffen, aber man kann die Usability so pauschal zwischen unterschiedlichen Versionen vergleichen. [12, p. 109 f]

Probleme pro Kategorie

Da man manchmal den Fokus auf bestimmte Bereiche eines Produktes legen möchte, kann es hilfreich sein, Kategorien für diese Bereiche zu definieren. Beispiele für solche Kategorien wären Navigation, Inhalt oder Begrifflichkeiten. Die aufgetretenen Probleme können ebenfalls in diese Kategorien eingeteilt werden. Dadurch kann anschließend die Anzahl der Probleme pro Kategorie ausgewertet werden. [12, p. 110]

Probleme pro Aufgabe

Neben der Auswertung pro Kategorie ist ebenso eine Auswertung pro Aufgabe möglich. Dafür gibt es zwei Varianten. Es können die eindeutigen Probleme der Teilnehmer und Teilnehmerinnen pro Aufgabe gezählt werden. Das zeigt auf, welche Aufgaben am meisten Probleme verursachen, sodass anschließend bei den Verbesserungen der Fokus auf diese gelegt werden kann. Als zweite Möglichkeit kann die Anzahl der Personen, die ein Problem bei einer Aufgabe hatten, herangezogen werden. Dadurch kann man erkennen, wie beständig sich manche Punkte bei allen Teilnehmern und Teilnehmerinnen als Problem herausstellen. Bei der ersten Variante könnten auch nur wenige Personen besonders viele Probleme bei einer Aufgabe gehabt haben, während die Anderen keine hatten. [12, p. 111]

2.6 Komplexität von Analysesystemen

Analysesysteme zeichnen sich meist dadurch aus, dass sie die Nutzer und Nutzerinnen bei der Lösung von komplexen Problemen unterstützen sollen. In "Complex Problem Solving and Content Analysis" [20] beschreibt Michael J. Albers die unterschiedliche Komplexität von Aufgaben als strukturierte und unstrukturierte bzw. schlecht strukturierte Probleme.

In einem stark strukturierten Umfeld sind die Probleme der Benutzer und Benutzerinnen meist klar definiert. Es besteht eine limitierte Anzahl an Lösungsansätzen, die von den Anwendern und Anwenderin gewählt werden und in einer Software abgebildet werden sollten. Für die Nutzer und Nutzerinnen geht es vor allem darum, die Aufgaben so effizient wie möglich durchzuführen. In diesen Fällen gibt es einen eindeutigen optimalen Weg zur Lösung, der sich aus einer klaren Abfolge von Schritten zusammensetzt. Daher geht es in der Softwareimplementierung besonders darum, die Personen auf diesem Weg zum Ziel zu führen. [20, p. 267]

Wenn man das Beispiel eines Geldautomaten betrachtet, so möchte der Benutzer oder die Benutzerin Geld abheben oder den Kontostand abfragen. Bei einem Online Shop steht der Kunde oder die Kundin meist vor dem Problem ein bestimmtes Produkt zu suchen. In beiden Fällen handelt es sich um einfache und klare Probleme, die ohne hohe Komplexität in einer Applikation abgebildet und von den Benutzern und Benutzerinnen gelöst werden können.

Wenn das Ziel der Benutzer und Benutzerinnen aber die Analyse von Daten bzw. das Lösen eines komplexen Problems ist, existiert diese eindeutige optimale Schrittabfolge zur Lösung nicht. In diesem Fall sind das Umfeld und das Problem selbst nicht gut strukturiert und es existiert eine große Anzahl an möglichen Lösungswegen und Vorgehensweisen. Abhängig von jeder neuen Information, die eine Person betrachtet, wählt sie die unterschiedlichsten nächsten Schritte und Aktionen. So gelangen verschiedene Personen auf unterschiedlichen Wegen zu einem Ergebnis oder erhalten vielleicht sogar unterschiedliche valide Ergebnisse. Man kann daher nicht davon ausgehen, dass alle Anwender und Anwenderinnen die

gleichen Informationen benötigen oder in welcher Reihenfolge sie diese Informationen berücksichtigen. [20, p. 267]

Ein Beispiel für diesen Bereich ist ein Manager oder eine Managerin, die die Verkaufszahlen des letzten Quartals analysieren und herausfinden möchte, warum die Umsätze zurückgegangen sind. Auch im medizinischen Umfeld gibt es Einsatzgebiete für diese Art von Analysesystemen. Beispielsweise sollen sie beim Stellen einer Diagnose oder bei der Ermittlung der Ursache für medizinische Fehler wie einer Überdosierung helfen. Für diese Probleme und Anwendungsfälle kann die passende Vorgehensweise nicht so klar und eindeutig definiert werden und bringt daher eine Komplexität in die Systeme zur Unterstützung dieser Aufgaben.

Ein weiteres Merkmal von komplexen Systeme ist, dass es sich oft nicht nur um eine Anwendung handelt, mit der die Benutzer und Benutzerinnen arbeiten müssen. Stattdessen ist eine ganze Systemlandschaft bestehend aus mehreren Komponenten und Tools involviert. [21] Häufig entstehen diese Systemlandschaften auch einfach mit der Zeit. Wenn eine Applikation nicht alle Funktionen enthält oder in manchen Bereichen, die Usability nicht ausreichend ist, so weichen die Benutzer und Benutzerinnen möglicherweise für Teilaufgaben auf andere Anwendungen aus. Mit wachsender Anzahl an Komponenten wird das ganze System aber immer komplexer, da viele Tools und Übergänge zwischen den Tools von den Benutzern und Benutzerinnen beachtet werden müssen.

Die durch die Komplexität verursachten Umstände bedingen auch, dass unterschiedliche Aspekte bei Systemen zur Lösung von einfachen und komplexen Problemen berücksichtigt werden müssen. Wie bereits erwähnt, enthalten einfache Systeme typischerweise vordefinierte Lösungswege bzw. Abläufe und liefern den Nutzern und Nutzerinnen eine einfache und gute Bedienung, um diese Prozesse durchzuführen. Bei komplexen Systemen ist diese Vorgehensweise aber nicht möglich. Eine fixe Definition der Abläufe würde die Betroffenen einschränken anstatt zu helfen. Daher müssen den Personen andere Werkzeuge und Möglichkeiten für ihre Arbeit zur Verfügung gestellt werden. Der Gedanke hinter einem Analysesystem sollte daher der eines Partners sein, der die Benutzer und Benutzerinnen bei ihrer Arbeit unterstützt und zur Seite steht, anstatt ihre Arbeit möglichst stark zu automatisieren. Dazu ist es wichtig, dass entsprechende Anwendungen flexibel gestaltet sind und sich auf das Fachwissen der Anwender und Anwenderinnen bzw. deren Prozesse zur Entscheidungsfindung stützen. Sie sollten den Benutzern und Benutzerinnen möglichst viele verschiedene Möglichkeiten und Funktionen bieten und ihnen die Freiheit geben, selbst zu entscheiden wann und wie sie welche Funktionen einsetzen möchten. Der Nutzer oder die Nutzerin erhält also die Kontrolle darüber, wie er oder sie mit den Daten umgehen möchte und welcher Lösungsansatz bzw. Ablauf genutzt wird. [22, p. 233 f]

Der Hauptfokus bei der Gestaltung solcher Analysesysteme sollte daher auf der Nützlichkeit liegen und diese sollten höher priorisiert werden als die Einfachheit der Benutzung. Es muss vor allem sichergestellt werden, dass ein Analysesystem die richtigen Konzepte und Funktionen enthält, um die Nutzer und Nutzerinnen bei ihrer Arbeit und Erreichung ihrer Ziele optimal zu unterstützen. Ansonsten könnte ein System zwar einfach und verständlich zu bedienen sein, aber keinen Mehrwert für die Anwender und Anwenderinnen in ihrer Arbeit bringen. Bei Systemen für Aufgaben mit hoher Komplexität wird dieser Punkt oft vernachlässigt, weil man sich dazu besonders stark mit den Zielen und Aufgaben der Zielgruppe auseinandersetzen muss. Man muss die Informationen und diversen Varianten, wie Personen mit diesen Informationen umgehen, verstehen und berücksichtigen, um ein nützliches System zu entwerfen. [22, p. 235 f]

3 Methoden

Nachdem im vorherigen Kapitel die Grundlagen zur Themenstellung dieser Arbeit und der eingesetzten Methoden erläutert wurden, befasst sich das folgende Kapitel mit der praktischen Anwendung der Methoden. Dazu wird im ersten Teil der Aufgabenbereich und Prozess des konkreten Anwendungsfalls vorgestellt und das aktuelle System beschrieben. Anschließend wird näher auf die Usability Tests eingegangen, die für dieses System durchgeführt wurden.

3.1 Aktuelles System

Das für die Usability Evaluierung in dieser Arbeit herangezogene System, stammt von der Firma Edentity Software Solutions GmbH (im folgenden nur Edentity genannt). Wie der Name bereits ausdrückt, entwickelt diese Firma aus Wien Softwarelösungen. Eines der wichtigsten Produkte der Firma Edentity ist das Dealer Benchmarking System kurz DBS. Dabei handelt es sich um eine Webapplikation, die es Automobilkonzernen ermöglicht, die Finanzkennzahlen ihrer Autohändler und Filialen zu vergleichen und daraus fundierte Managemententscheidungen abzuleiten. [23]

Die Firma Edentity entwickelt und vertreibt DBS nicht nur, sondern hostet und wartet die Webapplikation selbst für mehrere Kunden weltweit. Unter diese Aufgaben fällt auch die Verantwortung, einen stabilen Betrieb der Software zu gewährleisten. Zu diesem Zweck hat die Firma eigene interne Anwendungen entwickelt, mit denen der Betrieb der Applikation überwacht und die gesammelten Daten analysiert werden können. So soll ermöglicht werden, dass man Probleme rechtzeitig erkennt und möglichst frühzeitig darauf reagieren kann.

3.1.1 Applikationsüberwachung

Bei DBS handelt es sich um eine umfangreiche und verteilte Anwendung, die auf mehreren Servern betrieben wird. Bei mehreren Kunden kommt es so zu einer Vielzahl an unterschiedlichen Komponenten und Servern, die man überwachen muss. Außerdem wird DBS ständig weiterentwickelt, weshalb sich nicht verhindern lässt, dass immer wieder neue Fehler in der Software auftreten. Manche Fehler machen sich allerdings nicht über die Oberfläche bemerkbar und können nur auf Grund einer Fehlermeldung der Software erkannt werden, die im Hintergrund protokolliert wird. Mit der Applikationsüberwachung wird bei Edentity daher die Überwachung all dieser Aspekte bezeichnet. Dieses Aufgabengebiet beinhaltet, dass im ersten Schritt diverse Daten, die sowohl von den Servern als auch aus der Software stammen können, gesammelt werden. Anschließend müssen diese Daten analysiert und damit überwacht werden, ob der Betrieb der Applikation reibungslos verläuft.

Beispiele für die gesammelten Daten sind:

- RAM & CPU Verbrauch
- Freier Speicherplatz
- Fehlermeldungen der Applikation
- Ergebnis von automatisierten Tests
- User Aktivität
- Status von Services und automatisierten Prozessen

Manche dieser Daten werden bereits automatisch überprüft. Das ist dort der Fall, wo die Überprüfung recht einfach ist und diese auch in Echtzeit passieren sollte. Beispielsweise können pro Server bestimmte RAM und CPU Limits definiert werden und sobald diese Grenzen überschritten werden, schickt das System Benachrichtigungen an die zuständigen Personen. Eine Benachrichtigung wird ebenfalls automatisch ausgeschickt, wenn zum Beispiel ein wichtiger Service nicht mehr aktiv ist.

Andere Daten werden allerdings nicht automatisch überprüft, sondern müssen durch Mitarbeiter und Mitarbeiterinnen analysiert werden. Dabei handelt es sich zum Beispiel um die Fehlermeldungen der Applikation. Hier ist eine automatische Einschätzung, wie kritisch ein Fehler ist, schwierig, weshalb diese Entscheidung durch einen Mitarbeiter oder eine Mitarbeiterin getroffen werden sollte.

Die Analyse dieser Daten passiert jedoch nicht in Echtzeit, sondern wird einmal am Tag durchgeführt. Pro Kundensystem werden dazu die Daten vom jeweiligen Vortag gebündelt und dann täglich in der Früh von einer Person überprüft. Teilweise werden die Informationen für diese Arbeit bereits vorgefiltert und aufbereitet, aber der Großteil der Arbeit ist manuell. Die Hauptaufgabe ist Informationen zu kritischen Problemen herauszufiltern und alle Auffälligkeiten zu überprüfen und zu kommentieren. Zusätzlich muss basierend auf den Erkenntnissen der Analyse eine allgemeine Aussage zum Betriebszustand getroffen werden. Diese Einschätzung ist in Form von drei Levels – ok, mittel und kritisch – zu treffen.

Diese Tätigkeit wird durch Mitarbeiter und Mitarbeiterinnen des Support Teams durchgeführt. Das Ergebnis der Analyse, mit allen Auffälligkeiten und der allgemeinen Einschätzung, erhält anschließend die Abteilung Operations, die für den Betrieb der Kundensysteme zuständig ist. Die Mitarbeiter und Mitarbeiterinnen dieses Teams überprüfen die gefundenen Probleme und erstellen entsprechende Maßnahmen.

Das grundsätzliche Ziel dieses Prozesses ist, Probleme frühzeitig zu erkennen und proaktiv handeln zu können, bevor der Kunde oder die Kundin ein Problem bemerkt bzw. ein größerer Schaden entsteht. Für das System, welches diese Arbeit unterstützen soll, sind daher auch die drei Usability Kriterien Effizienz, Effektivität und Zufriedenheit wichtig. Die Fehler von

Mitarbeitern und Mitarbeiterinnen sollten minimiert werden, da kritische Probleme sonst übersehen oder falsch eingeschätzt werden könnten. Da die Arbeit täglich durchgeführt wird, spielt die Effizienz eine wichtige Rolle und eine Unzufriedenheit mit dem System kann sich auf die grundsätzliche Mitarbeiterzufriedenheit auswirken. Mit einer Usability Evaluierung des bisherigen Systems sollen deshalb diese Aspekte überprüft werden.

Im folgenden Kapitel wird das System vorgestellt, das bisher für die zuvor beschriebenen Aufgaben eingesetzt wurde.

3.1.2 Systemlandschaft

Die Systeme, die für das im vorherigen Kapitel genannte Aufgabengebiet eingesetzt werden, lassen sich grob in zwei Bereiche unterteilen. Zum einen gibt es eine Anwendung, die alle relevanten Daten von den Servern und der Applikation sammelt und in einer Datenbank erfasst. Der zweite Teil umfasst jenes System, welches die Mitarbeiter und Mitarbeiterinnen bei der manuellen Analyse der Daten nutzen. Da es sich beim ersten Punkt um einen Hintergrundprozess handelt, der keine direkte Nutzerinteraktion beinhaltet, wurde für die Usability Evaluierung nur der zweite Bereich berücksichtigt. Im Folgenden wird daher dieser Teil näher beschrieben.

Das bisherige System besteht eigentlich aus bis zu vier verschiedenen Anwendungen, die im Rahmen der manuellen Analyse eingesetzt werden. Dabei handelt es sich um die folgenden Programme:

- Desktop Applikation zur Generierung der Email mit den Daten
- Microsoft Outlook zur Bearbeitung der Email
- Webseite zur Anzeige der Fehlermeldungen
- Microsoft Access Anwendung zum Abruf von weiteren Daten

Der gesamte Prozess dreht sich um eine E-Mail, die unterschiedliche Daten zur Analyse enthält. Eine einfache Desktopapplikation generiert diese Email. In der in Abbildung 3.1 zu sehenden Applikation kann man ein bestimmtes Kundensystem und die eigene Emailadresse auswählen. Anschließend ruft dieses Programm die relevanten Daten des Vortages für das ausgewählte System ab, bereitet sie in einer HTML Struktur auf und sendet sie anschließend als E-Mail an die gewählte Person.

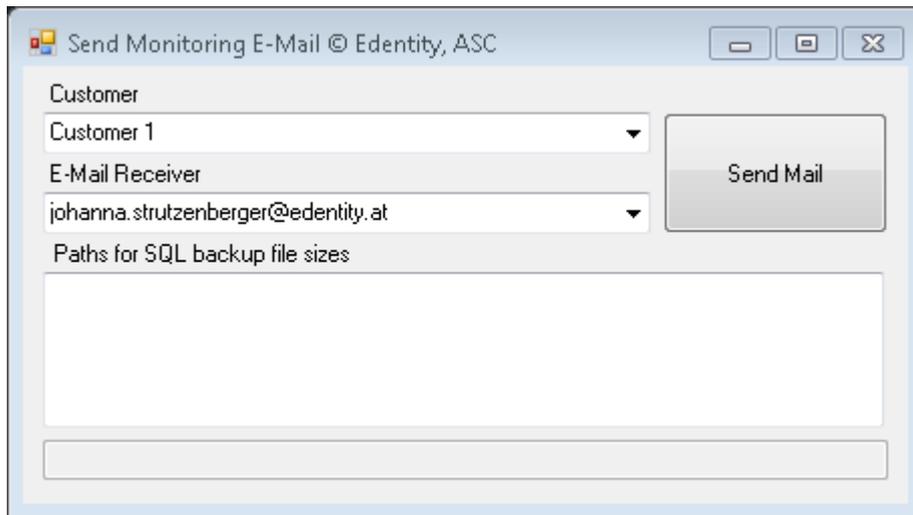


Abbildung 3.1: Applikation zum Versand der Datenemail

Diese generierte E-Mail enthält dann unterschiedliche Bereiche mit den jeweils zum Bereiche gehörenden Daten in einer Tabellenstruktur. Die Abbildung 3.2 zeigt einen beispielhaften Ausschnitt dieser Email mit allen Fehlermeldungen der Software in einer Tabelle.

Von: Johanna Strutzenberger
 An: Johanna Strutzenberger
 CC:
 Betreff: TEST - DBS Quickcheck Monitoring

Gesendet: Do 03.08.2017 21

Applikation-Monitoring Status

Eventlog: Errors & Warnings:

Datum	System	Anzahl	Message	Comment
22.11.2016 08:17:52	DBS-P-ABC-WEB01.edyhosting2.intra - APP	2	#Error @32 Error CopyNoData [no_data] Source : Help link : Data : System.Collections.ListDictionaryInternal TargetSite : Stack Trace : The stack trace is unus	
22.11.2016 09:38:12	DBS-P-ABC-WEB02.edyhosting2.intra - APP	8	Could not find file to export with designated filename 'Data_1_1_PerFmYear_' and GUID " Source : Identity.DBS.Web Help link : Data : System.Collections.ListDiction	gilt nicht mehr für Standard Report (Ticket #191278) - Problem in ABCCN mit Findata Export, Ticket #12909: Findata Export schlägt fehl, aufgrund von Verbindungsproblemen in CN! Sollte das Problem nicht in CN stattfinden, muss es separat betrachtet werden.
22.11.2016 09:58:12	DBS-P-ABC-APP15.edyhosting2.intra - APP	1	Validation process finished with errors. Source : Identity.DBS.Service.Calculation Help link : Data : System.Collections.ListDictionaryInternal TargetSite : System.G	bitte prüfen, ob die Validation.exe abstürzt -> Ursache könnte ein fehlerhafter Validation-FailedMaster sein (muss dann in ERMF gelistet werden) & ABCCHK
22.11.2016 10:17:45	DBS-P-ABC-WEB01.edyhosting2.intra - APP	8	Internal Error(@66 Error GeneralDataAccess)The communication object, System.ServiceModel.Channels.ServiceChannel, cannot be used for communication because it is in the F	
22.11.2016 10:17:45	DBS-P-ABC-APP19.edyhosting2.intra - APP	6	Error while reading KeyValues: Error while Reading file: Cannot recognize current file type. Source : SqFileImporter Help link : Data : System.Collections.ListDiction	
22.11.2016 10:17:45	DBS-P-ABC-APP19.edyhosting2.intra - APP	6	Error while Reading file: Cannot recognize current file type. Source : SqFileImporter Help link : Data : System.Collections.ListDictionaryInternal TargetSite : Syst	
22.11.2016 10:17:46	DBS-P-ABC-WEB01.edyhosting2.intra - APP	8	#Error @66 Error GeneralDataAccess [The communication object, System.ServiceModel.Channels.ServiceChannel, cannot be used for communication because it is in the Failed s	
23.11.2016 07:28:02	DBS-P-ABC-APP15.edyhosting2.intra - APP	3	No Entry for Method: GetCurrencies found Source : Help link : Data : System.Collections.ListDictionaryInternal TargetSite : HResult : -2146233088 Stack Trace : T	Job fehlt, muss eingetragen werden
23.11.2016 07:28:02	DBS-P-ABC-APP15.edyhosting2.intra - APP	3	GetCurrencies Job not found in Job-Table Source : Help link : Data : System.Collections.ListDictionaryInternal TargetSite : HResult : -2146233088 Stack Trace : T	Job fehlt, muss eingetragen werden.

Jobs:

Task	Begin	End	Dauer	Area/Run	All_Errors	Interpreted_Errors	All_Warnings	Interpreted_Warnings	NotInterpreted
Identity.DBS-Export Cockpit	23.11.2016 02:31:00	23.11.2016 02:31:42	0,7		1	1	0	0	84
Identity.DBS-Export Crowe	15.11.2016 15:36:18	15.11.2016 20:04:33	268,25		0	0	0	0	2258
Identity.DBS-Export Erabase	23.11.2016 06:00:23	23.11.2016 06:00:29	0,1		0	0	0	0	9
Identity.DBS-Import Currency	23.11.2016 00:02:01	23.11.2016 00:02:01	0		0	0	1	1	6
Identity.DBS-Import Hat	23.11.2016 07:15:00	23.11.2016 07:15:39	0,65		0	0	0	0	36
Identity.DBS-Service-DealerImportHook	23.11.2016 09:45:00	23.11.2016 10:00:01	15,02		0	0	0	0	0
Identity.DBS-Service-MarketImportHook	15.11.2016 12:29:56	16.11.2016 09:44:21	11274,42		994	994	0	0	510

Automatischer Web Click Test:

[Bitte hier den Selenium Screenshot aufrufen](#)

Freier Speicherplatz der Server:

Task	Begin	End	Dauer	Area/Run	All_Errors	Interpreted_Errors	All_Warnings	Interpreted_Warnings	NotInterpreted
DBS-P-ABC-APP01 - Disk	25.11.2016 00:10:18	25.11.2016 08:10:11		DiskSpaceInfo	0	0	0	0	5
DBS-P-ABC-APP02 - Disk	25.11.2016 00:10:18	25.11.2016 08:10:11		DiskSpaceInfo	0	0	0	0	5
DBS-P-ABC-APP03 - Disk	25.11.2016 00:10:18	25.11.2016 08:10:13		DiskSpaceInfo	0	0	0	0	5
DBS-P-ABC-APP04 - Disk	25.11.2016 00:19:54	25.11.2016 08:10:13		DiskSpaceInfo	0	0	0	0	4
DBS-P-ABC-APP05 - Disk	25.11.2016 00:19:54	25.11.2016 08:10:13		DiskSpaceInfo	0	0	0	0	4
DBS-P-ABC-APP06 - Disk	25.11.2016 00:19:54	25.11.2016 08:10:11		DiskSpaceInfo	0	0	0	0	5
DBS-P-ABC-APP07 - Disk	25.11.2016 00:19:54	25.11.2016 08:10:11		DiskSpaceInfo	0	0	0	0	4
DBS-P-ABC-APP08 - Disk	25.11.2016 00:19:54	25.11.2016 08:10:11		DiskSpaceInfo	0	0	0	0	4
DBS-P-ABC-APP09 - Disk	25.11.2016 00:19:54	25.11.2016 08:10:12		DiskSpaceInfo	0	0	0	0	4
DBS-P-ABC-APP10 - Disk	25.11.2016 00:19:54	25.11.2016 08:10:12		DiskSpaceInfo	0	0	0	0	4

Abbildung 3.2: Beispiel einer generierten Email mit Daten

Von diesem Zeitpunkt an arbeitet der Mitarbeiter oder die Mitarbeiterin hauptsächlich mit dieser E-Mail und analysiert die enthaltenen Daten. Für manche Aufgaben sind die Informationen in der E-Mail allerdings nicht vollständig, weshalb noch weitere Anwendungen notwendig sein können.

Der wichtigste und umfangreichste Teil der Tätigkeit ist die Analyse der Fehlermeldungen. Jedoch enthält die generierte E-Mail nur die ersten 128 Zeichen jeder Fehlermeldung. Für eine sinnvolle Einschätzung reicht diese Information oft nicht aus, weshalb die Nutzer und Nutzerinnen die vollständige Fehlermeldung gesondert raussuchen müssen. Dazu stehen zwei Möglichkeiten zur Verfügung. Zum einen gibt es eine weitere E-Mail, die täglich automatisch versendet wird, in welcher alle Fehlermeldungen des letzten Tages komplett enthalten sind. Außerdem gibt es eine interne Webseite, die auch alle kompletten Fehlermeldungen auflistet. Der Benutzer oder die Benutzerin kann also mit einer dieser zwei Möglichkeiten die entsprechende Fehlermeldung raussuchen. Die Abbildung 3.3 stellt einen Ausschnitt der Email dar und Abbildung 3.4 zeigt die Webseite.

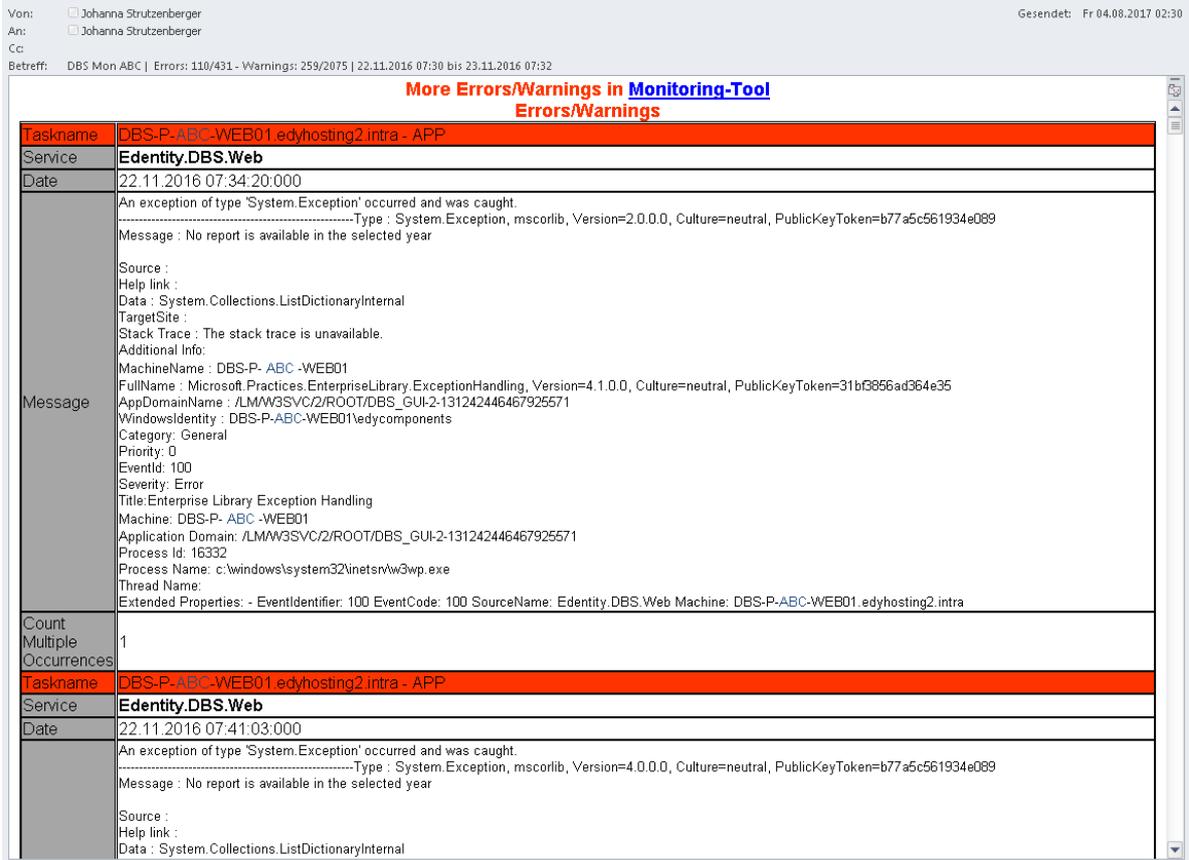


Abbildung 3.3: Email mit kompletten Fehlermeldungen des vorangehenden Tages

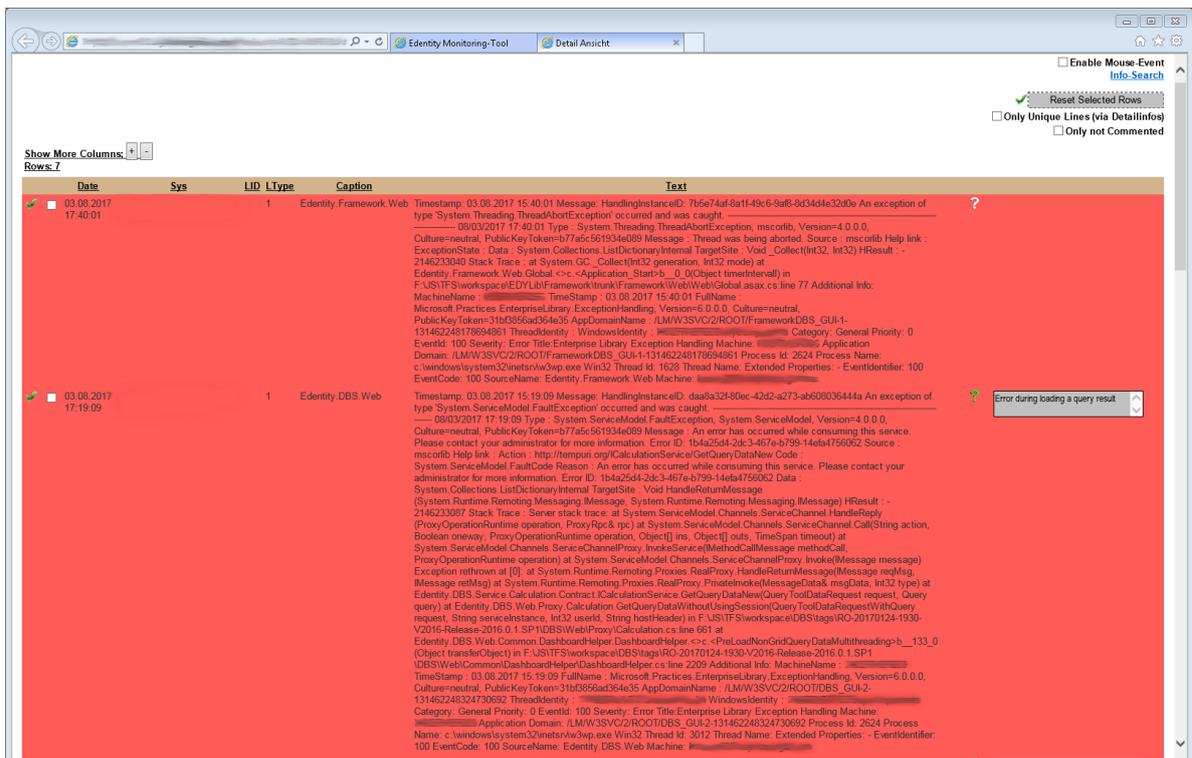


Abbildung 3.4: Webseite mit kompletten Fehlermeldungen

Zusätzlich sind seit der Entwicklung der Desktopanwendung weitere neue Datengruppen dazugekommen, die ebenfalls überprüft werden sollen. Da diese allerdings noch nicht in der Anwendung für die E-Mail berücksichtigt werden, müssen die Nutzer und Nutzerinnen diese Informationen selber abfragen und in der E-Mail hinzufügen. Dazu steht ihnen eine Microsoft Access Anwendung zur Verfügung, mit der auf diese Daten aus der Datenbank zugegriffen werden kann. In Abbildung 3.5 sieht man das Ergebnis der Abfrage zur Benutzeraktivität, welches auch in der E-Mail erfasst werden soll.

Timestamp	Machine	UniqueUserC	CallsCount	AvgRuntime	AvgOpentime	StdevRuntime	Time
03.08.2017 23:00:00		2	765	539	-71	2339	04.08.2017 00:02:13
03.08.2017 22:00:00		1	368	1008	-102	5493	03.08.2017 23:01:06
03.08.2017 21:00:00		3	800	616	-84	3148	03.08.2017 22:01:11
03.08.2017 20:00:00		2	319	1133	-78	5945	03.08.2017 21:01:02
03.08.2017 19:00:00		4	1151	528	-44	2567	03.08.2017 20:01:17
03.08.2017 18:00:00		9	1360	803	-28	3910	03.08.2017 19:01:05
03.08.2017 17:00:00		15	4128	491	-49	2263	03.08.2017 18:01:04
03.08.2017 16:00:00		33	19288	771	-36	7164	03.08.2017 17:01:03
03.08.2017 15:00:00		31	11400	841	-79	6617	03.08.2017 16:01:13
03.08.2017 14:00:00		24	9254	690	-102	4193	03.08.2017 15:01:10
03.08.2017 13:00:00		26	9843	852	-81	5123	03.08.2017 14:01:12
03.08.2017 12:00:00		31	8821	692	-79	4798	03.08.2017 13:01:00
03.08.2017 11:00:00		42	14977	628	-61	5848	03.08.2017 12:02:49
03.08.2017 10:00:00		30	12602	603	-68	4002	03.08.2017 11:01:05
03.08.2017 09:00:00		26	5737	658	-43	4514	03.08.2017 10:01:14
03.08.2017 08:00:00		19	6997	554	-26	3863	03.08.2017 09:01:01
03.08.2017 07:00:00		19	4400	511	16	4304	03.08.2017 08:01:02
03.08.2017 06:00:00		9	1594	514	-19	2511	03.08.2017 07:01:03
03.08.2017 05:00:00		9	2357	276	-96	999	03.08.2017 06:01:07
03.08.2017 04:00:00		10	3759	364	-78	1607	03.08.2017 05:01:00
03.08.2017 03:00:00		15	3661	404	-69	1613	03.08.2017 04:01:35
03.08.2017 02:00:00		2	240	465	-47	1918	03.08.2017 03:01:07
03.08.2017 01:00:00		1	1	6	60		03.08.2017 02:01:01
03.08.2017		3	1099	629	17	2773	03.08.2017 01:01:03
02.08.2017 23:00:00		3	2864	582	-10	4290	03.08.2017 00:01:59
02.08.2017 22:00:00		3	1891	521	0	4165	02.08.2017 23:01:06
02.08.2017 21:00:00		1	232	706	-15	3006	02.08.2017 22:01:11
02.08.2017 20:00:00		2	794	415	-101	1538	02.08.2017 21:01:06
02.08.2017 19:00:00		5	1041	714	-97	3471	02.08.2017 20:01:19
02.08.2017 18:00:00		13	2171	655	-38	7067	02.08.2017 19:01:04
02.08.2017 17:00:00		19	6465	649	-60	5462	02.08.2017 18:01:11
02.08.2017 16:00:00		23	5971	539	-52	5226	02.08.2017 17:01:06
02.08.2017 15:00:00		36	8268	564	19	4703	02.08.2017 16:01:01
02.08.2017 14:00:00		34	11950	767	6	7655	02.08.2017 15:01:00
02.08.2017 13:00:00		26	4749	495	4	2201	02.08.2017 14:01:05
02.08.2017 12:00:00		31	7039	400	-37	1764	02.08.2017 13:01:04
02.08.2017 11:00:00		50	6812	403	171	1409	02.08.2017 12:02:31
02.08.2017 10:00:00		44	13029	740	86	7044	02.08.2017 11:01:02
02.08.2017 09:00:00		41	8296	521	-36	3581	02.08.2017 10:01:00
02.08.2017 08:00:00		26	10077	720	-82	6214	02.08.2017 09:01:01
02.08.2017 07:00:00		14	6385	554	-65	4016	02.08.2017 08:01:05
02.08.2017 06:00:00		10	11352	457	-45	2348	02.08.2017 07:01:17
02.08.2017 05:00:00		10	12323	328	-28	1391	02.08.2017 06:01:02
02.08.2017 04:00:00		12	2399	471	-7	1826	02.08.2017 05:01:05
02.08.2017 03:00:00		10	1749	403	-42	1546	02.08.2017 04:01:57
02.08.2017 02:00:00		7	1206	689	-106	2546	02.08.2017 03:01:06
02.08.2017 01:00:00		2	912	515	-92	2034	02.08.2017 02:01:01
02.08.2017		1	4	14	-42	9	02.08.2017 01:01:08
01.08.2017 23:00:00		1	87	135	44	355	03.08.2017 00:01:35

Abbildung 3.5: Beispielhafte Abfrage der Useraktivität in der MS Access Anwendung

Bereits seit längerem besteht die Vermutung, dass diese Systemlandschaft den aktuellen Prozess nicht optimal unterstützt. Die Aufteilung auf mehrere Anwendungen erschwert die Arbeit und macht auch eine Einschulung von neuen Mitarbeitern und Mitarbeiterinnen

schwieriger. Durch neue Kundensysteme ist die benötigte Zeit für diese Tätigkeiten zusätzlich stark gestiegen. Außerdem wurde von unterschiedlichen Mitarbeitern und Mitarbeiterinnen eine Unzufriedenheit mit dem aktuellen System ausgedrückt und von anderen die Qualität der Ergebnisse bemängelt. Aus dieser Situation heraus entstand der Wunsch nach einem neuen System. Um die richtigen Maßnahmen zur Steigerung der Effektivität und Effizienz zu setzen und die Zufriedenheit zu verbessern, ist es daher notwendig, eine Evaluierung des bisherigen Systems durchzuführen.

Das nächste Kapitel beschreibt die Usability Tests, die zu diesem Zweck durchgeführt wurden und konkrete Probleme in der aktuellen Systemlandschaft aufzeigen sollen.

3.2 Usability Tests

Im Rahmen dieser Arbeit wurde die Usability des im vorherigen Kapitel beschriebenen Systems evaluiert. Dazu wurden moderierte Usability Tests und der SUS-Fragebogen eingesetzt.

Im Kapitel 2.4.3 „System Usability Scale“ wurde bereits dargelegt, dass man sich für diesen Fragebogen auf Grund seiner weiten Verbreitung und universellen Einsatzmöglichkeit entschieden hat. Weiters wurde keine Heuristische Evaluierung durchgeführt, sondern Usability Tests zur Evaluierung des Systems gewählt, da so die Effizienz und Effektivität besser gemessen werden kann. Gerade bei komplexen Systemen sind diese Kriterien wichtig und das informelle Feedback der Mitarbeiter und Mitarbeiterinnen deutet auf Probleme in diesen Bereichen hin. Um einen Vergleich zwischen altem und neuem System zu ermöglichen und das Ergebnis in Zahlen rechtfertigen zu können, spielen Metriken eine wichtige Rolle. Im Rahmen einer Heuristischen Evaluierung können aber keine Metriken in Bezug auf Effizienz und Effektivität gemessen werden. Die Entscheidung moderierte und nicht unmoderierte Usability Tests durchzuführen, fiel aus praktischen Gründen. So konnten die Tests sehr einfach direkt in der Firma gemeinsam mit den betroffenen Mitarbeitern und Mitarbeiterinnen durchgeführt werden. Zusätzlich ermöglicht einem die direkte Beobachtung, dass Ursachen für Probleme schneller und einfacher erkannt werden können.

Die folgenden drei Unterkapitel beschreiben den konkreten Aufbau und die Vorgehensweise dieser Usability Tests. Das erste Kapitel befasst sich mit den Teilnehmern und Teilnehmerinnen, welche den Test durchgeführt haben. Anschließend werden die Aufgabenstellungen erläutert, die diese Personen im Rahmen der Tests durchführen mussten. Zuletzt werden allgemeine Informationen zur Organisation, Testumgebung und Ausstattung beschrieben.

3.2.1 Teilnehmer und Teilnehmerinnen

Die Zielgruppe des Systems besteht aus zwei Benutzergruppen. Zum einen gibt es die Supportmitarbeiter und -mitarbeiterinnen, die täglich die beschriebene Analysetätigkeit durchführen und dafür das System verwenden. Zusätzlich gibt es das Team „Operations“, das vorrangig mit dem Ergebnis der Analyse arbeitet. Bei der weiteren Recherche zu den Ergebnissen kann es jedoch dazu kommen, dass sie ebenfalls das System nutzen. Des Weiteren müssen diese Mitarbeiter, im Falle eines Ausfalls im Supportteam, als Vertreter agieren und die Analysetätigkeit selbst durchführen können. Diese zwei Gruppen lassen sich also vor allem dadurch unterscheiden, dass die eine Gruppe täglich mit dem System arbeitet, während die zweite Gruppe es nur hin und wieder nützt.

Der Hauptfokus dieser Usability Evaluierung liegt auf den Aufgaben der Benutzer und Benutzerinnen des Supportteams, da diese die wichtigste Zielgruppe des Systems darstellen. Dennoch wurden auch Mitglieder vom Team „Operations“ berücksichtigt, um festzustellen, wie einfach ihnen die Nutzung fällt, obwohl sie das System nicht täglich nützen.

Das Supportteam der Firma Eidentity besteht aus vier Mitarbeitern und Mitarbeiterinnen. Zusätzlich hat das Operationsteam drei Mitglieder. Die Usability Tests konnten mit all diesen Mitarbeitern und Mitarbeiterinnen durchgeführt werden. Somit haben insgesamt sieben Personen an der Evaluierung teilgenommen und es konnte die gesamte Zielgruppe dieses Systems abgedeckt werden.

3.2.2 Aufgabenstellungen

Neben der Qualität der Ergebnisse ist bei täglicher Nutzung des Systems auch die benötigte Zeit für die Tätigkeiten von großer Bedeutung. Daher liegt ein besonderer Fokus bei dieser Evaluierung auf der Effizienz und Effektivität bei der Arbeit mit dem System. Somit ist das Hauptziel der im folgenden beschriebenen Aufgaben, diesen Prozess zu überprüfen und Problemquellen zu finden.

Dazu wurden die täglichen Aufgaben und der Ablauf gemeinsam mit betroffenen Personen näher beleuchtet und aus diesen Informationen konkrete Aufgabenstellungen für die Usability Tests abgeleitet. Die Aufgabenstellungen wurden so gewählt, dass sie den Prozess möglichst realitätsnah widerspiegeln. Um eine strukturierte Vorgehensweise und vergleichbare Ergebnisse zu ermöglichen, wurden manche Teile angepasst und auf das Wesentliche reduziert.

So führen die Teilnehmer und Teilnehmerinnen im Grunde die gleichen Aufgaben wie bei ihrer normalen Arbeit durch, aber die Datenmenge zur Analyse wurde beispielsweise reduziert. Zusätzlich wurden manche Aufgaben so definiert, dass sie nur mit der Identifizierung eines möglichen Problems enden. Normalerweise findet eine genauere

Analyse statt, um anschließend eine detaillierte Aussage zu Bedeutung und Ursache des Problems liefern zu können. Dieser Schritt wurde ausgelassen, da er sehr stark von der Erfahrung des Mitarbeiters oder der Mitarbeiterin abhängig und weniger stark durch das System beeinflusst wird. Ein Vergleich der Ergebnisse wäre deshalb schwer möglich. Außerdem wurde die Analyse mancher Daten auf mehrere Aufgaben aufgeteilt, um so besser auf die unterschiedlichen Aspekte in der gesamten Analyse der Daten eingehen zu können. Dadurch kann besser festgestellt werden, wo genau die Probleme liegen. Neben den Anpassungen der Aufgabenstellungen, um eine strukturiertere Durchführung der Tests zu ermöglichen, wurden manche Aufgaben auch mit neuen Punkten ergänzt. Dabei handelt es sich um Schritte und Fragen, die aktuell zwar noch nicht in der täglichen Arbeit behandelt werden, aber es besteht die Anforderung, dass es in der Zukunft passieren soll.

Insgesamt wurden so neun Aufgaben ermittelt. Der genaue Text jeder Aufgabe, der den Teilnehmern und Teilnehmerinnen vorgelegt wurde, befindet sich im Anhang B. Bei jeder Aufgabe muss der Benutzer oder die Benutzerin unterschiedliche Daten analysieren und die Erkenntnisse jeder Aufgabe sind in einer Zusammenfassung zu erfassen. So kann bei der Auswertung der Tests überprüft werden, ob die Personen zum richtigen Ergebnis gekommen sind.

Bei der Formulierung der Aufgabenstellungen wurde versucht die Aufgabe so systemneutral wie möglich zu formulieren. Im Grunde sollte nur definiert werden, welche Fragen der Benutzer oder die Benutzerin auf Grund der Datenanalyse beantworten soll und welche Informationen dazu im Ergebnis erwartet werden. Dennoch war es nicht möglich, die exakt selbe Formulierung für die Evaluierung des alten und des neuen Systems zu verwenden, da die Bedienung bei manchen Aufgaben im neuen System unterschiedlich ist.

Im Folgenden werden die einzelnen Aufgaben kurz beschrieben und dabei falls vorhanden auf die Unterschiede zwischen beiden Evaluierungen eingegangen.

Aufgabe 1: Einstieg

Beim Einstieg in das System ist der Unterschied zwischen beiden Systemen am Größten. Daher unterscheidet sich diese Aufgabenstellung in beiden Evaluierungen. Im ersten Test wird der Teilnehmer oder die Teilnehmerin beim Einstieg gebeten, das Programm zum Senden der E-Mail zu starten und die E-Mail für ein bestimmtes Kundensystem schicken zu lassen. Die Aufgabe ist beendet, sobald die E-Mail im Postfach eingelangt ist. Beim zweiten Test wird der Teilnehmer oder die Teilnehmerin gebeten, die neue Applikation aufzurufen und ein bestimmtes Kundensystem auszuwählen. Sobald alle Daten geladen sind, ist die Aufgabe in diesem Fall beendet.

Aufgabe 2: Jobs

In dieser Aufgabe geht es darum, dass der Benutzer oder die Benutzerin überprüft, ob bestimmte automatisierte Prozesse in der Software – intern als Jobs bezeichnet – durchgeführt wurden. Ziel ist es, dass in der überarbeiteten E-Mail fehlgeschlagene Jobs zur besseren Erkennung hervorgehoben sind und der Benutzer oder die Benutzerin in den Fehlermeldungen jene Fehler ermittelt, welche die Ursache für den Fehlschlag sind.

Aufgabe 3-5: Errors

Wie bereits erwähnt wurde, ist die Hauptaufgabe der täglichen Tätigkeit die Analyse der Fehlermeldungen. Dabei sollen aber unterschiedliche Aspekte überprüft werden, um die wichtigsten Problemfälle zu ermitteln. Deshalb wurde dieser Teil in drei Aufgaben aufgeteilt:

1. In der ersten Aufgabe sollen alle aufgetretenen Datenbank-Blockings ermittelt und zusammengefasst werden.
2. In der zweiten Aufgabe geht es darum, dass die Häufigkeit der Fehler überprüft wird. Eine hohe Anzahl kann auf kritische Probleme hindeuten und würde daher weitere Recherche benötigen.
3. In der dritten Aufgabe sollen die Fehler noch anhand der weiteren üblichen Kriterien überprüft werden. Dazu zählt zum Beispiel, ob sie in einem kritischen Teil der Applikation aufgetreten sind. Zusätzlich wurde hier ein neuer Schritt ergänzt, der aktuell auf Grund des hohen manuellen Aufwands nicht durchgeführt wird. Es soll auch überprüft werden, welche Fehler noch nie davor aufgetreten sind. Allerdings wurde für den ersten Test eingeschränkt, dass es ausreicht die letzten drei Tage zu überprüfen, um den Aufwand in einem zumutbaren Rahmen zu halten.

Aufgabe 6: Selenium Tests

Im Rahmen dieser Aufgabe ist das Ergebnis von automatisierten Tests zu überprüfen. Da diese Ergebnisse im bisherigen System noch nicht automatisch erfasst werden können, müssen die Resultate von den Benutzern und Benutzerinnen zuerst aus den generierten Logdateien kopiert werden. Im neuen System können die Daten direkt in der Anwendung aufgerufen werden. Zusätzlich wurde auch diese Aufgabe um einen neuen Schritt erweitert. Die Durchlaufzeiten der Tests sollen mit dem Vormonat verglichen werden, um festzustellen ob es eine starke Veränderung gab.

Aufgabe 7: Disk Space

Die im bisherigen System generierte E-Mail enthält pro Server die Information, ob der Speicherplatz in Ordnung ist oder es Probleme gibt. Der Teilnehmer soll erkennen, wo es Probleme gab und die betroffenen Server hervorheben.

Aufgabe 8: Schnittstellen

Die Information zum Status der Schnittstellen ist in der bisher generierten E-Mail nicht enthalten. Im Rahmen des ersten Tests muss der Benutzer oder die Benutzerin diese

Informationen daher in der Microsoft Access Anwendung abrufen und in der E-Mail ergänzen, bevor sie überprüft werden können. Beim Test des neuen Systems mussten die Personen diese Informationen nur direkt im selben System finden und überprüfen.

Aufgabe 9: DBS User Activity

Abschließend soll die Benutzeraktivität für das gewählte Kundensystem überprüft werden. Da auch diese Information im bisherigen System noch nicht automatisch erfasst wird, muss sie manuell vom Benutzer oder der Benutzerin eingetragen werden. Die Information kann entweder mit der Microsoft Access Anwendung ermittelt oder direkt auf der Datenbank des Kundensystems abgefragt werden. Im neuen System muss man die Daten für diese Aufgabe nicht mehr kopieren, sondern nur den entsprechenden Bereich für diese Daten aufrufen. Zusätzlich wurde diese Aufgabe mit einem weiteren neuen Schritt ergänzt. Es soll eine Aussage zur Stärke der Benutzeraktivität im Vergleich zu einem anderen Tag mit den Kategorien hoch, mittel und niedrig erfasst werden.

Allgemein ist noch anzumerken, dass in der ersten Evaluierung nur für die Aufgaben 1 bis 7 die Daten zur Analyse simuliert wurden. Für die Aufgaben 8 und 9 war eine Simulation leider nicht möglich, da die Daten nicht in der E-Mail vorkommen und die Microsoft Access Anwendung nicht mit einer Testversion ersetzt werden konnte. Für diese Aufgaben musste daher individuell ausgewertet werden, ob das Ergebnis richtig ist.

3.2.3 Testaufbau

Die Tests wurden nicht in einem Usability Labor, sondern im Firmenbüro an einem normalen Arbeitsplatz durchgeführt. Die Ausstattung bestand daher aus einem PC mit zwei Bildschirmen plus einer Maus und einer Tastatur. Auf dem PC waren alle für die Aufgaben nötigen und im Kapitel 3.1.2 „Systemlandschaft“ beschriebenen Programme eingerichtet. Zusätzlich wurde die Software Morae Recorder [24] von TechSmith installiert. Mit Hilfe dieser Software wurde den Benutzern und Benutzerinnen die einzelnen Aufgabenstellungen nach einander präsentiert, die Zeit für die Durchführung jeder Aufgabe gemessen und der gesamte Bildschirminhalt während der Tests aufgezeichnet. Während den Tests war die Autorin dieser Arbeit als Moderatorin anwesend.

Neben dem Einrichten der Testumgebung waren auch einige Anpassungen und Vorbereitungen des zu testenden Systems notwendig. Wie bereits im vorherigen Kapitel kurz angemerkt wurde, mussten einige Daten für den Test simuliert werden. Damit trotzdem eine sinnvolle Auswahl an Daten analysiert wird, wurden dennoch echte Daten, jedoch aus der Vergangenheit, herangezogen. Diese Daten wurden auf eine geringere Anzahl gekürzt, die für den Test ausreichend ist.

Für die erste Evaluierung des bisherigen Systems wurde mit diesen Daten anschließend eine fixe E-Mail für den Test vorbereitet. Damit die Teilnehmer und Teilnehmerinnen dennoch die Applikation zur Generierung der E-Mail wie im normalen Prozess benutzen können, wurde das Programm für den Test manipuliert, sodass es immer die vorbereitete E-Mail ausschickt.

Da die Arbeit im neuen System nicht mehr auf einer E-Mail basiert, sondern direkt in der neuen Applikation passiert, war die Vorbereitung der E-Mail für die zweite Evaluierung nicht notwendig. Stattdessen wurde direkt das neue System für den Test manipuliert, so dass es für alle Personen immer die gleichen Daten abrufen und den Teilnehmern und Teilnehmerinnen anzeigen. So ist in beiden Evaluierungen sichergestellt, dass jeder Teilnehmer und jede Teilnehmerin mit den exakt selben Daten wie die Anderen arbeitet und das Ergebnis am Ende für alle gleich ausgewertet werden kann.

4 Evaluierung des bisherigen Systems

Repräsentativ für jedes Usability Kriterium wurden im Rahmen der Usability Tests für das bisherige System drei Metriken gemessen:

- Effektivität → Task Success
- Effizienz → Time on Task
- Zufriedenheit → SUS Score

Die folgenden Kapitel präsentieren die Ergebnisse für diese drei Metriken. Zusätzlich dazu wurde aus den Beobachtungen während der Tests Usability Probleme abgeleitet und die wichtigsten Probleme werden im letzten Kapitel näher beschrieben.

4.1 Task Success

Zur Auswertung, ob eine Aufgabe erfolgreich abgeschlossen wurde, ist keine binäre Einteilung gewählt worden. Die Aufgaben wurden daher nicht nur als erfolgreich bei einem optimalen Ergebnis oder fehlgeschlagen in den restlichen Fällen bewertet. Bei diesem geringen Detailgrad würden zu viele Informationen verloren gehen. Auf Grund der Komplexität der Aufgaben bestehen die Ergebnisse aus verschiedenen Punkten, welche die Lösungen der Benutzer und Benutzerinnen richtig enthalten sollen. Beispielsweise sollen in einer Aufgabe die häufigsten drei Fehler gefunden werden. Wenn eine Person nur zwei von diesen drei Fehlern findet, wäre sie in einer binären Einteilung wohl als fehlgeschlagen bewertet worden. Allerdings ist es ein interessanter Unterschied, ob die Person keinen Punkt erreicht hat oder doch zumindest zwei von drei Fehlern gefunden hat. Um diese Information nicht zu verlieren und einen besseren Grad der Aufgabenerfüllung darzustellen, wurden mehr als zwei Erfolgslevel für die Bewertung gewählt.

Es gibt unterschiedliche Möglichkeiten, wie diese Erfolgslevels definiert werden können. Tom Tullis und Bill Albert nennen in „Measuring the User Experience“ [12] die folgenden zwei Ansätze. Die Levels basieren auf dem Benutzererlebnis während der Durchführung der Aufgaben. Die Antwort wird in diesem Fall immer noch nur als richtig oder falsch gewertet, aber es können unterschiedliche Aspekte während der Durchführung dazu führen, dass man Unterscheidungen in den Levels treffen möchte. Beispielsweise können manche Personen die Aufgabe nur mit Hilfe lösen, haben Probleme oder machen größere Umwege bei der Lösung, während andere Personen die Aufgabe ohne Anstrengungen lösen können. Beim zweiten Ansatz hängen die Levels stärker vom Ergebnis ab und drücken aus, dass Benutzer und Benutzerinnen die Aufgabe in unterschiedlichen Ausprägungen gelöst haben. Manche Personen lösen eine Aufgabe zum Beispiel optimal und komplett richtig, während andere es weniger optimal lösen und die Antwort nicht zu 100 Prozent korrekt ist. [12, p. 71 f]

Bei der Auswertung der Aufgabenergebnisse in dieser Arbeit wurden die Erfolgslevels an den zweiten Ansatz angelehnt definiert. Das richtige Ergebnis zu jeder Aufgabe besteht aus einer bestimmten Anzahl an Punkten, die der Benutzer oder die Benutzerin richtig erfasst haben sollte. Dadurch kann für jede Aufgabe ausgewertet werden, wie viele dieser Punkte die Person richtig hat und so der Erfüllungsgrad bestimmt werden. Wenn eine Person beispielsweise die drei häufigsten Fehler erfassen sollte und alle erfasst wurden, wird die Aufgabe mit der vollen Punktezahl – also in diesem Fall drei – bewertet. Während eine Person, die nur einen Fehler richtig erkannt hat, nur einen Punkt erhält. So lässt sich genau bestimmen, zu welchem Grad der Benutzer oder die Benutzerin die Aufgabe erfolgreich abgeschlossen hat. In der Tabelle 4.1 sind die erreichten Punkte jedes Teilnehmers und jeder Teilnehmerin für jede Aufgabe dargestellt. Die Zahl in den Klammern in der Spaltenüberschrift stellt die mögliche Gesamtzahl dar. Der Erfüllungsgrad der Aufgabe wird zusätzlich über die grüne Befüllung der Zelle veranschaulicht.

	Start Aufgabe 1 (1)	Jobs Aufgabe 2 (2)	DB Blockings Aufgabe 3 (8)	Häufige Fehler Aufgabe 4 (3)	Neue/kritische Fehler Aufgabe 5 (5)	Selenium Aufgabe 6 (3)	Disk space Aufgabe 7 (2)	Schnittstellen Aufgabe 8 (1)	User Activity Aufgabe 9 (2)
T1	1	2	8	3	2	3	2	1	2
T2	1	2	7	3	1	3	2	1	1
T3	1	2	6	3	4	2	2	1	2
T4	1	1	4	3	1	2	2	0	2
T5	1	2	3	2	3	3	2	0	2
T6	1	2	8	2	5	2	2	0	2
T7	1	2	8	3	2	2	2	0	1

Tabelle 4.1: Erfüllungsgrad jeder Aufgabe pro TeilnehmerIn

Tom Tullis und Bill Albert nennen in „Measuring the User Experience“ [12] zusätzlich eine übliche Einteilung in folgende drei Levels:

- Kompletter Erfolg
- Teilweiser Erfolg
- Misserfolg [12, p. 71]

Das Ergebnis der Aufgaben wurde auf diese drei Levels zusammengefasst, indem nur eine volle Punkteanzahl als kompletter Erfolg gewertet wurde. Alles, wo zumindest die Hälfte der Punkte erreicht wurde, ist ein teilweiser Erfolg und der Rest wurde als Misserfolg gezählt. Das Diagramm in Abbildung 4.1 zeigt, wie sich die Teilnehmer und Teilnehmerinnen pro Aufgabe auf diese drei Kategorien aufteilen.

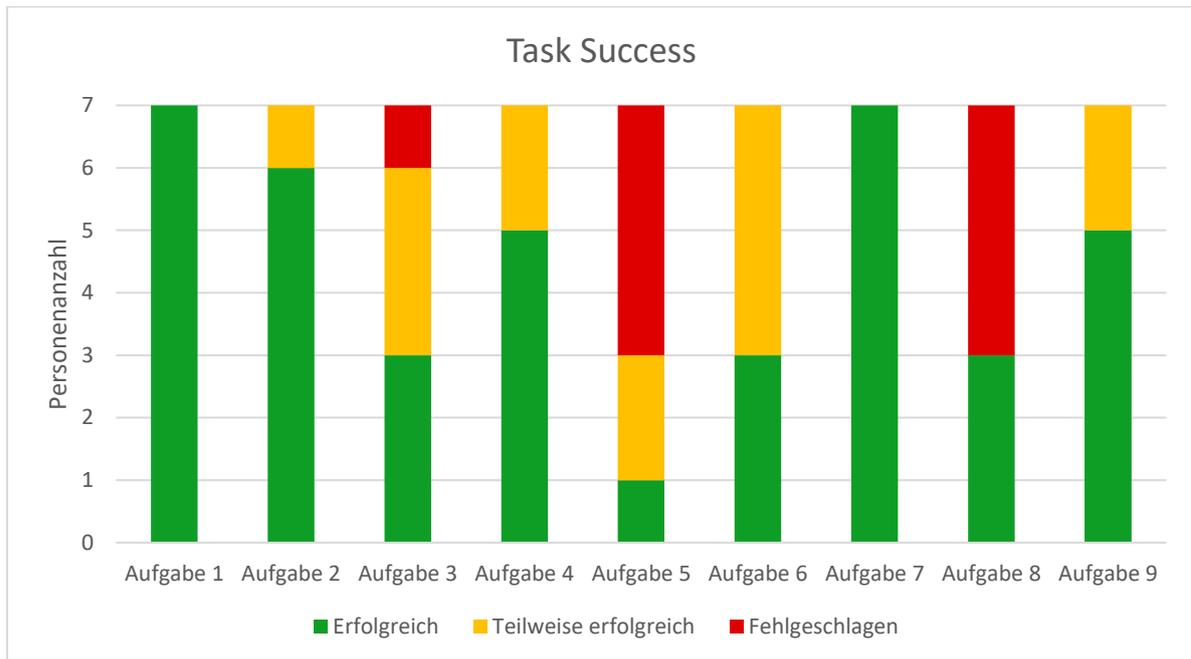


Abbildung 4.1: Task Success pro Aufgabe

Lediglich zwei Aufgaben erreichen in der ersten Evaluierung eine Erfolgsrate von 100. In der Abbildung 4.1 lässt sich gut erkennen, dass besonders die Aufgaben 3, 5, 6 und 8 schlecht abschneiden, weil weniger als die Hälfte der Teilnehmer und Teilnehmerinnen die Aufgabe komplett erfolgreich abschließen konnten. Die Tabelle 4.1 zeigt zusätzlich auf, dass die Probleme bei Aufgabe 8 gravierende sein müssen. In dieser Aufgabe gab es nur einen zu erreichenden Punkt und dennoch ist die Hälfte der Teilnehmer und Teilnehmerinnen nicht erfolgreich.

4.2 Time on Task

Wenn man die gemessene Zeit auswertet, die Teilnehmer und Teilnehmerinnen für die Aufgaben benötigt haben, ist die üblichste Variante die Berechnung des Mittelwerts über alle Zeiten pro Aufgabe. In dieser Variante können Ausreißer das Ergebnis jedoch verzerren. Daher gibt es auch die Möglichkeit den Median zu berechnen. Der grundsätzliche Trend in den Ergebnissen und die interessanten Muster darin zeigen sich aber meistens in beiden Varianten. Die gleichen Aufgaben zeigen also in beiden Varianten die längste oder kürzeste Zeit auf. [12, p. 78 f]

Des Weiteren stellt sich bei der Auswertung die Frage, ob nur die Zeit von erfolgreichen oder allen Aufgaben inkludiert werden sollen. Nicht erfolgreiche Aufgaben können starke Schwankungen in der benötigten Zeit darstellen, weil Benutzer und Benutzerinnen zum Beispiel sehr lange versucht haben, die Aufgabe zu lösen oder viel schneller waren, weil sie zu einem falschen Ergebnis kamen. Wenn nur erfolgreiche Aufgaben einberechnet werden, birgt es daher weniger Risiko, dass falsche Ausreißer das Ergebnis verzerren. Das

Auslassen der fehlgeschlagenen Daten kann aber ebenfalls zu einem falschen Bild führen. Die Analyse von allen Daten spiegelt das gesamte Benutzererlebnis oft besser wieder. Beispielsweise könnte eine Aufgabe einen sehr guten Durchschnitt aufweisen, wenn ein geringer Teil der Teilnehmer oder Teilnehmerinnen in kurzer Zeit erfolgreich war, obwohl der Rest nach längeren Zeiten gescheitert ist. Dieser Fall könnte zu einer falschen Interpretation führen, dass die Time on Task für diese Aufgabe gut sei. In diesem Fall würde die Betrachtung der Zeiten also nur in Kombination mit den Ergebnissen der Metrik Task Success ein richtiges Bild vermitteln. Wenn hingegen alle Aufgaben ausgewertet werden, können die Ergebnisse unabhängig von anderen Metriken betrachtet werden. [12, p. 81]

In dieser Arbeit wurden für die Auswertung der Metrik Time on Task die fehlgeschlagenen Teilnehmer und Teilnehmerinnen zwar nicht mit eingerechnet, aber es wurden die teilweise erfolgreichen Teilnehmer und Teilnehmerinnen inkludiert. Manche Aufgaben wurden abgebrochen, weil die Personen nicht wussten, wo die notwendigen Informationen zu finden seien und in manchen Fällen war das Ergebnis nur weniger als die Hälfte richtig. Bei diesen Fällen ist davon auszugehen, dass die Zeiten keine repräsentativen Werte aufweisen. Bei den Personen mit teilweisem Erfolg ist mit höherer Wahrscheinlichkeit damit zu rechnen, dass die benötigten Zeiten realistischen Werten entsprechen.

Die Abbildung 4.2 zeigt pro Aufgabe den Vergleich zwischen dem Mittelwert und dem Median der gemessenen Zeiten der sieben Teilnehmer und Teilnehmerinnen. Die Zeiten sind in Sekunden angegeben. Vor allem bei Aufgabe 5 und 6 lassen sich größere Unterschiede zwischen beiden Werten erkennen. Diese Unterschiede sind auf Schwankungen und Ausreißer in den Zeiten der einzelnen Benutzer und Benutzerinnen zurückzuführen, wie in Abbildung 4.3 zu erkennen ist.

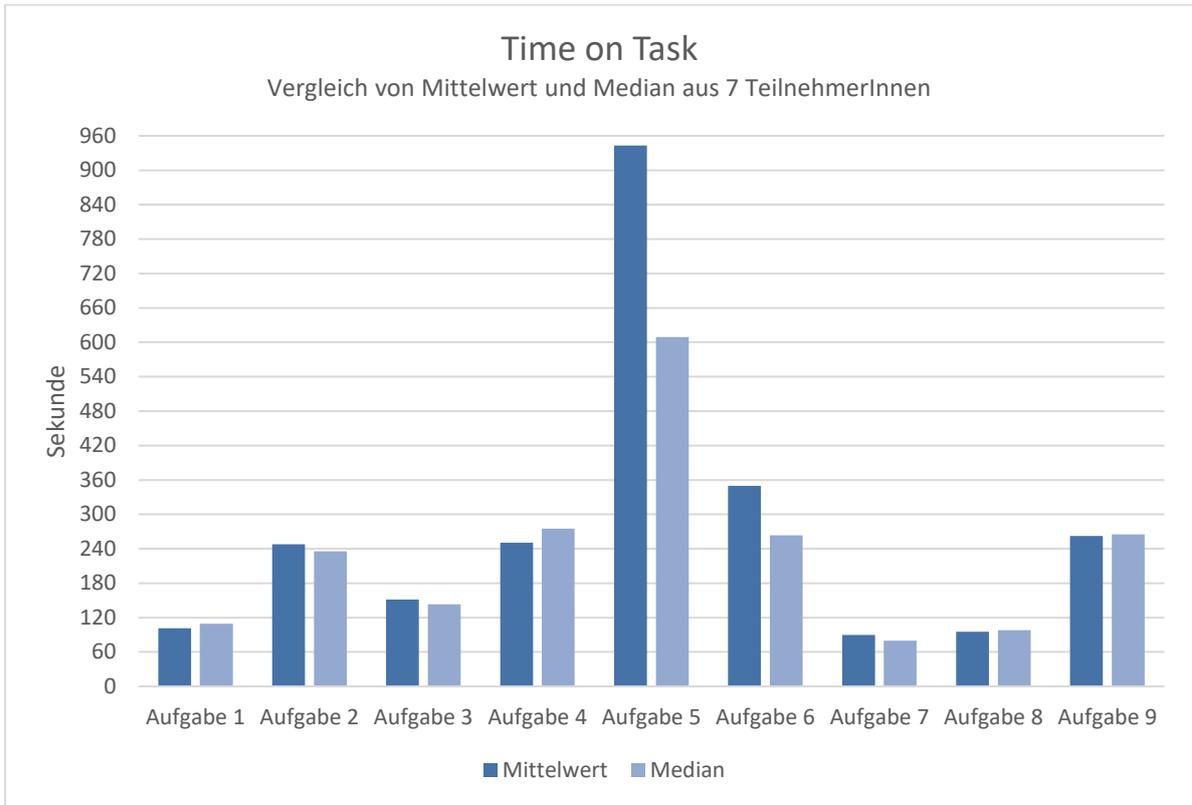


Abbildung 4.2: Time on Task als Mittelwert und Median pro Aufgabe

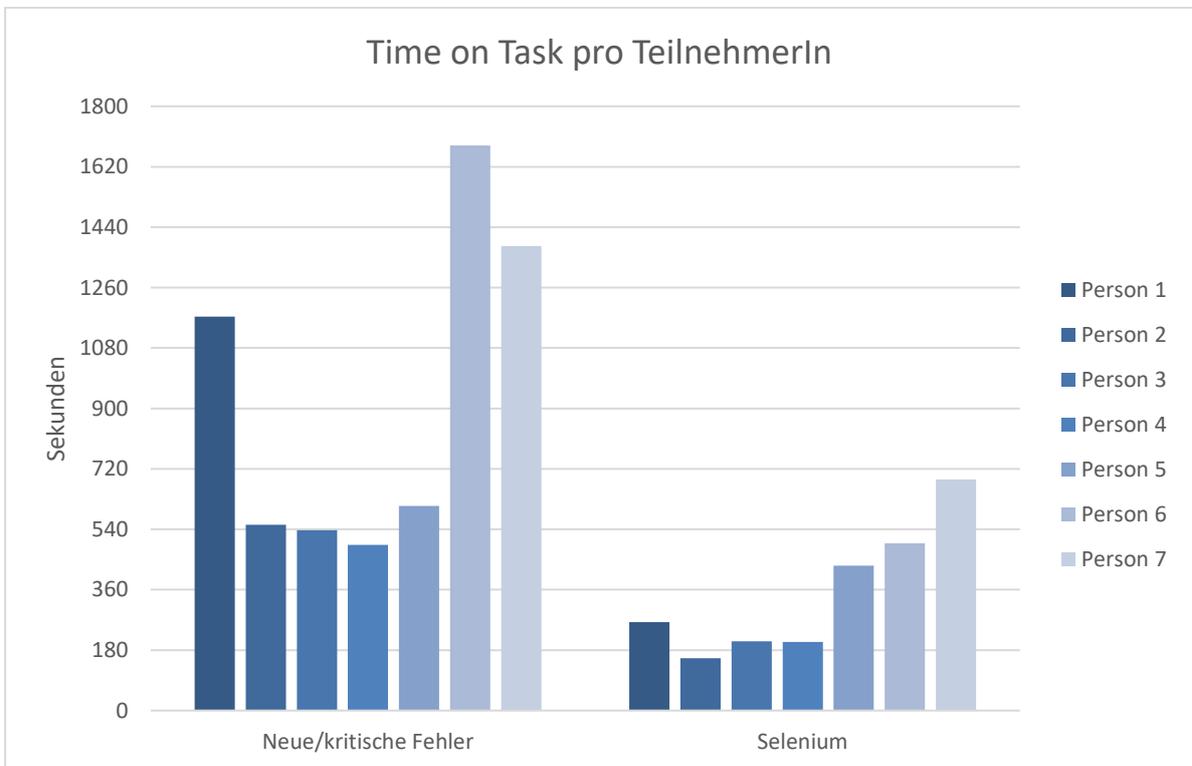


Abbildung 4.3: Time on Task pro TeilnehmerIn für Aufgabe 5 & 6

Abgesehen von der durchschnittlich benötigten Zeit für eine Aufgabe, ist ebenso interessant, wie viele Benutzer und Benutzerinnen die Aufgabe innerhalb einer bestimmten Zeit erfolgreich beenden konnten. Schließlich ist es ein wichtiges Ziel, die Anzahl der Personen mit hohen Zeiten zu minimieren und es zu ermöglichen, dass man eine Aufgabe innerhalb einer akzeptablen Zeit durchführen kann. Zur Auswertung muss für jede Aufgabe ein Zeitlimit bestimmt werden. Anschließend kann die Anzahl der Teilnehmer und Teilnehmerinnen gezählt werden, die eine Aufgabe innerhalb dieses Limits beendet haben. [12, p. 80]

Die große Schwierigkeit bei dieser Auswertung – vor allem in dem Anwendungsbereich der Informationsanalyse – ist die Festlegung von sinnvollen Zeitlimits. Für jede Aufgabe muss überlegt werden, was die gewünschte oder akzeptable Durchführungszeit ist. Das kann zum Beispiel mit Produktexperten definiert werden oder indem man die Aufgaben selber ausführt und diese Zeit dann verdoppelt. Jedoch kommt bei Aufgaben der Informationsanalyse die Abhängigkeit von der Menge an Daten dazu. Eine pauschale Aussage zu gewünschten Zeiten für eine Aufgabe ist daher oft nicht möglich, weil die Menge der zu analysierenden Daten unterschiedlich sein kann und allein dadurch die Aufgabe schon länger oder kürzer dauern wird. Für die Evaluierung können die Zeitlimits also nur als sinnvolle Zeiten für das im Test eingesetzte Set an Daten definiert werden. Allerdings lassen sich aus der Auswertung dieser Limits dennoch allgemeine Erkenntnisse ableiten. Wenn der Benutzer oder die Benutzerin für diese Datenmenge schon länger als erwartet braucht, wird es auch bei einer anderen Datenmenge zu lange dauern.

Darüber hinaus wurden für diese Auswertung nur die Zeiten herangezogen, bei welchen die Aufgabe komplett erfolgreich durchgeführt wurde. Bei teilweisem Erfolg ist anzunehmen, dass die Benutzer oder Benutzerinnen länger gebraucht hätten, wenn sie es komplett richtig gemacht hätten. Die Abbildung 4.4 zeigt, wie viele Teilnehmer und Teilnehmerinnen die einzelne Aufgabe innerhalb der definierten Zeitlimits erfolgreich abschließen konnten. Die Aufgaben 1 und 5 wurden von keiner Person innerhalb einer akzeptablen Zeit beendet und auch bei den restlichen Aufgaben konnte ein Großteil der Personen das Limit nicht einhalten.

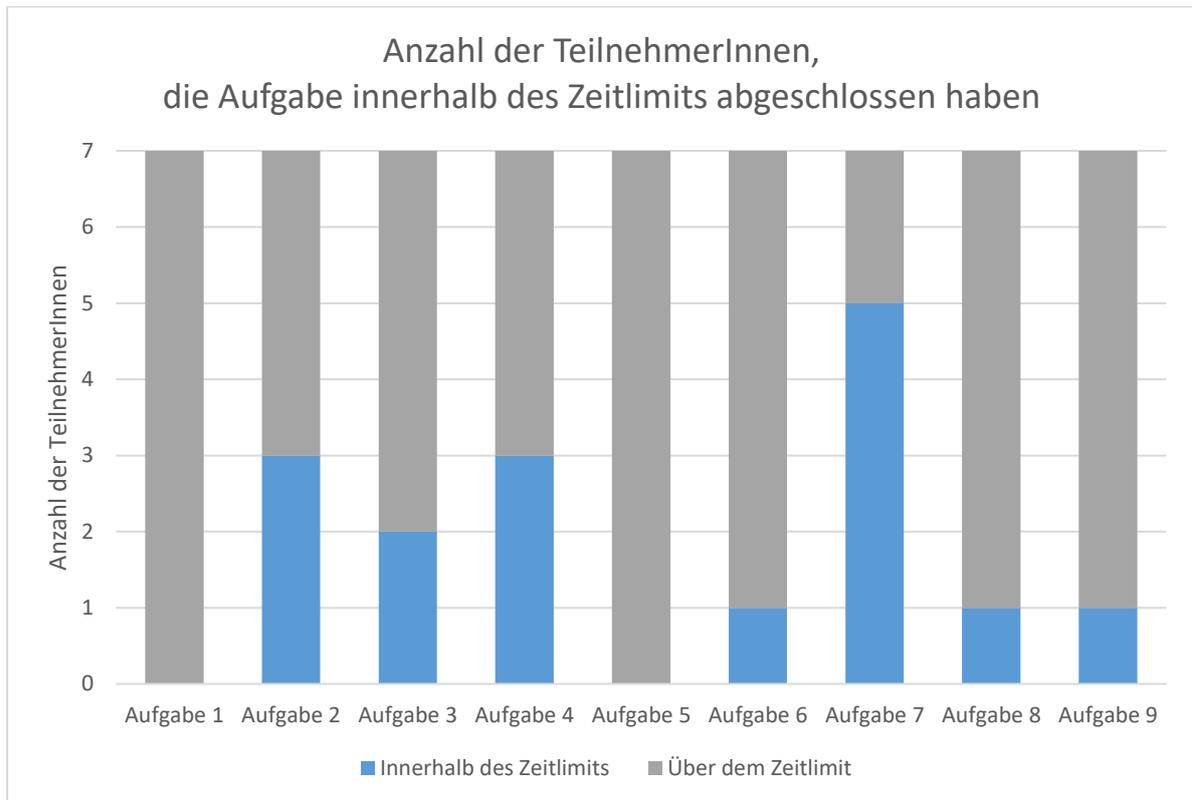


Abbildung 4.4: Anzahl der TeilnehmerInnen pro Aufgabe mit Abschluss innerhalb des Zeitlimits

4.3 SUS Score

Auf Grund der Antworten der Teilnehmer und Teilnehmerinnen für den SUS Fragebogen wurde für jeden Teilnehmer und jede Teilnehmerin, wie im Kapitel 2.4.3 „System Usability Scale“ beschrieben, der SUS Score berechnet.

Der berechnete SUS Score fällt in einen Bereich zwischen 0 und 100. Allerdings kann er nur 41 unterschiedliche Werte annehmen und das Ergebnis wird in der Berechnung auf 100 skaliert wird. [18, p. 31]

Obwohl der SUS Score auf Werte zwischen 0 und 100 fällt, ist er kein Prozentsatz. Eine Interpretation als Prozentsatz könnte zu einer falschen Aussage führen. Da der durchschnittliche SUS Score auf 68 fällt, würde eine Interpretation eines SUS Scores von 70 als 70% möglicherweise zur falschen Annahme führen, dass die Usability überdurchschnittlich bewertet wurde, obwohl es eigentlich fast dem Durchschnitt entspricht. [18, p. 36]

Der durchschnittliche SUS Score stammt aus einer globalen Benchmark Auswertung von Jeff Sauro in „A Practical Guide to the System Usability Scale“ [18]. Dazu wurden Studien

aus drei verschiedenen Quellen kombiniert und so ein Gesamtset an Antworten aus 446 Umfragen erhalten. Die Auswertung dieser Daten hat einen durchschnittlichen SUS Score von 68 mit einer Standardabweichung von 12,5 ergeben. [18, p. 53]

Der SUS Score kann jedoch in einen Prozentsatz umgerechnet werden. Wenn man das Ergebnis mit Daten von Benchmarks vergleichen möchte, ist es besser den Wert in einen Prozentsatz umzuwandeln. Wenn man den SUS Score mit bereits für das Produkt gemessenen Ergebnissen vergleichen möchte, reicht der berechnete SUS Score aus. [18, p. 36]

Wenn man, wie Jeff Sauro in „A Practical Guide to the System Usability Scale“ [18] beschreibt, die möglichen SUS Scores mit ihren Prozenträngen betrachtet, ergibt sich die in Abbildung 4.5 dargestellte Verteilung. In dieser Abbildung ist zum Beispiel zu sehen, dass ein SUS Score von 60 dem Prozentrang von 29% entspricht. Damit drückt diese Darstellung aus, dass die Usability eines Produkts mit einem Ergebnis von 60 besser als 29 Prozent aller Produkte ist, die mit dem SUS Fragebogen evaluiert wurden. [18, p. 57 f] In dieser Abbildung lässt sich so auch nochmal zeigen, dass ein SUS Score von 68 den Durchschnitt darstellt, weil er bei 50 Prozent platziert ist.

Der aus den einzelnen SUS Scores der Teilnehmer und Teilnehmerinnen berechnete Mittelwert ist 20. In Abbildung 4.5 ist dieses durchschnittliche Ergebnis für das bisherige System im Vergleich zur gesamten Verteilung in Gelb hervorgehoben. Die Platzierung lässt erkennen, dass es sich dabei um eine sehr niedrige Bewertung handelt. In diesem Vergleich ist das bisherige System gerade einmal besser als ein Prozent der getesteten Produkte.

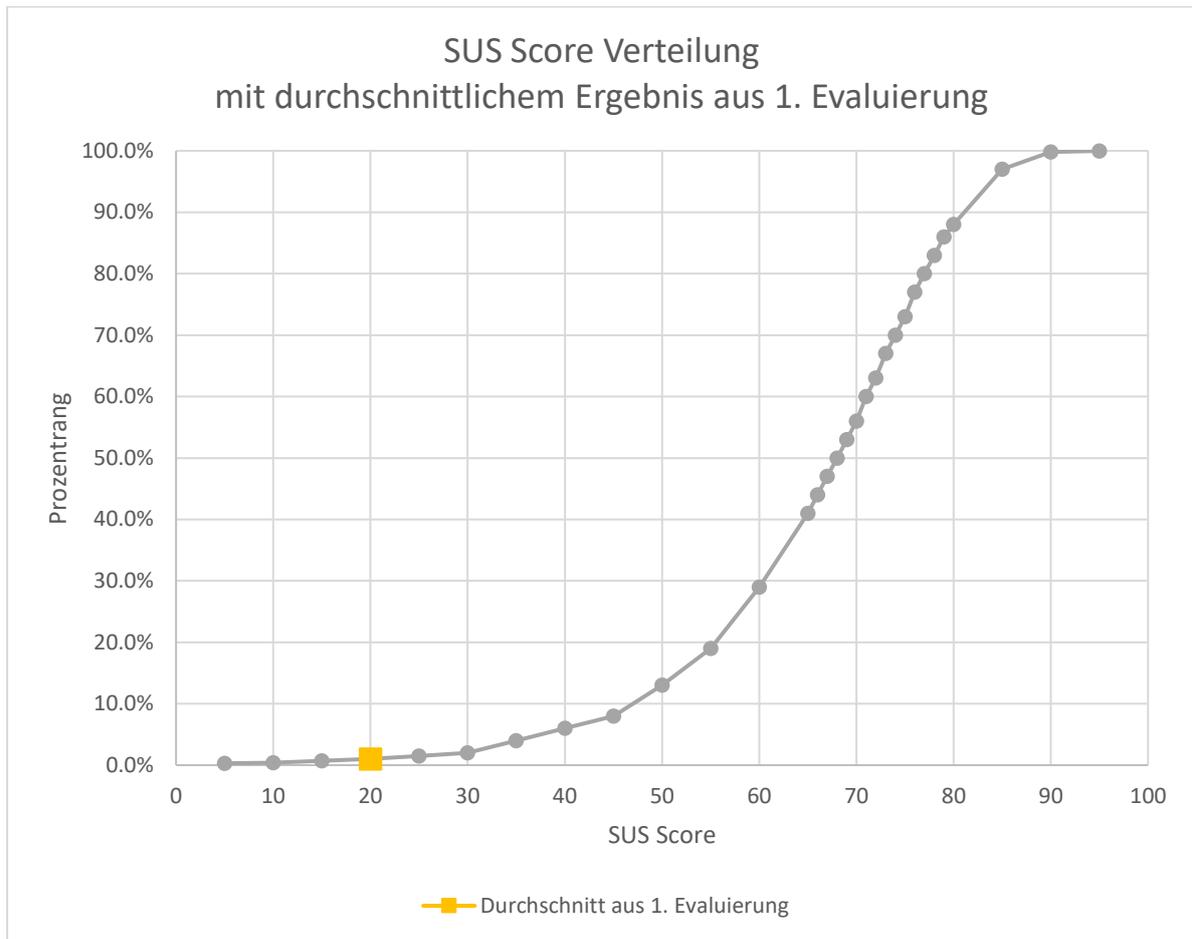


Abbildung 4.5: SUS Score Verteilung mit dem durchschnittlichen Ergebnis aus der ersten Evaluierung hervorgehoben

Neben dem Mittelwert könnte auch der Median über die SUS Scores aller Teilnehmer und Teilnehmerinnen berechnet werden. Bei der Frage, ob besser der Mittelwert oder der Median herangezogen wird, empfiehlt Jeff Sauro in „A Practical Guide to the System Usability Scale“ [18] den Mittelwert. Er hat die Antworten von 3000 Personen aus 104 Datensets ausgewertet und den Mittelwert und Median verglichen. Dabei konnte kein signifikanter Unterschied zwischen beiden ermittelt werden. [18, p. 22 f]

4.4 Usability Probleme

Die vorherigen Metriken drücken zwar aus, in welchen Bereichen oder Aufgaben es Probleme im bisherigen System gibt, jedoch nicht was die Probleme sind, die zum Beispiel zu längeren Zeiten oder fehlgeschlagenen Aufgaben führen. Dazu müssen aus den Beobachtungen des Verhaltens der Teilnehmer und Teilnehmerinnen während des Tests die konkreten Probleme ermittelt werden. Diese Erkenntnisse sind besonders für die Gestaltung des neuen Systems wichtig, damit dort die richtigen Maßnahmen zur Verbesserung der Usability durchgeführt werden können. Im Folgenden werden die

häufigsten und gravierendsten Probleme aus der Evaluierung des alten Systems näher erläutert.

Häufiger Wechsel zwischen Bereichen

Für den Hauptteil der Aufgaben arbeiten die Benutzer und Benutzerinnen mit der E-Mail, die unterschiedlich zu analysierende Daten enthält. Allerdings handelt es sich dabei um ein statisches Dokument. Somit gibt es keine Navigationselemente und die Benutzer und Benutzerinnen können nur durch Scrollen zu unterschiedlichen Bereichen gelangen. Im Rahmen einiger Aufgaben ist es notwendig, dass der Benutzer oder die Benutzerin mehrere Abschnitte in der E-Mail gemeinsam nutzt. Das führt dazu, dass der Benutzer oder die Benutzerin oft in der E-Mail hin und her scrollen muss. Beispielsweise muss man bei jeder Aufgabe Informationen in die Zusammenfassung schreiben. Das bedeutet, die Person sieht sich die Daten an, ihr fällt etwas auf, sie kopiert es und scrollt zum Anfang der E-Mail, um die Information in die Zusammenfassung einzufügen. Anschließend muss diese Person wieder zurück zu den Daten scrollen, um diese weiter zu analysieren. Für jede Auffälligkeit wird dieser Vorgang wiederholt. Bei mehreren Benutzern und Benutzerinnen konnte man beobachten, dass umso länger die Zusammenfassung geworden ist oder umso weiter unten in der E-Mail die Daten zu finden waren, desto eher wurde die gesuchte Stelle nicht gleich gefunden und zu weit gescrollt. Der Benutzer oder die Benutzerin musste dann suchen bis er oder sie die richtige Stelle wiedergefunden hat.

Bei anderen Aufgaben mussten Daten aus anderen E-Mails oder Programmen berücksichtigt werden. Hier kommt also zusätzlich das Wechseln zwischen verschiedenen Bereichen in unterschiedlichen Applikationen hinzu.

Fehlender Überblick und Zusammenhang

Das im vorherigen Punkt genannte Problem des häufigen Wechselns zwischen Bereichen führt auch dazu, dass der Benutzer oder die Benutzerin schnell den Überblick verliert. Zum Beispiel durch das Scrollen zur Zusammenfassung und wieder zurück zu den Daten ist einem schnell unklar, wo genau man in den Daten zuvor stehen geblieben ist und welche schon überprüft wurden. Bei manchen Benutzern und Benutzerinnen konnte man beobachten, dass sie mehrmals zwischen Daten und Zusammenfassung gewechselt haben, um zu überprüfen, was schon erfasst wurde, weil kein Überblick mehr vorhanden war.

Zusätzlich kann es innerhalb der Daten bei manchen Aufgaben schwierig sein, den Überblick zu bewahren. Bei Aufgabe 3 mussten beispielsweise die Datenbank-Blockings überprüft werden. Ein Filtern der Daten ist in der E-Mail nicht möglich, daher muss man die Liste einzeln durchgehen und die Blockings raussuchen. Besonders beim Erfassen der zusätzlich notwendigen Informationen für die Blockings, wie die Dauer oder Anzahl, ist es leicht möglich, etwas zu übersehen. Beim Ermitteln der drei häufigsten Fehler in Aufgabe 4, tritt dieses Problem ebenfalls auf. Hier müssen die Benutzer und Benutzerinnen die Liste, die in der E-Mail nach Datum sortiert ist, durchgehen und im Kopf behalten, was die bisher

höchsten Zahlen waren. Manche Benutzer und Benutzerinnen nutzen andere Programme, wie Microsoft Excel oder Notepad++, für diese Aufgaben, indem sie die Daten dorthin kopieren und anschließend die Funktionen dieser Programme nutzen. Dies erhöht aber wiederum den Wechsel zwischen Bereichen und Applikationen.

Zusätzlich war es bei manchen Aufgaben notwendig, unterschiedliche Daten zu verknüpfen. Beispielsweise musste für eine genauere Analyse die komplette Fehlermeldung in einer anderen E-Mail gesucht werden oder für einen fehlgeschlagenen Job, der dazugehörige Fehler ermittelt werden. Das System hat in diesem Fall keine Unterstützung geboten, um den Zusammenhang zwischen diesen Informationen herzustellen. Stattdessen musste es vom Benutzer oder der Benutzerin manuell ermittelt werden.

Manueller Aufwand

Im bisherigen System müssen viele Arbeiten durch die Benutzer und Benutzerinnen manuell durchgeführt werden, die einfach zu automatisieren wären. Beispielsweise sind nicht alle Daten in der E-Mail vorhanden und müssen von den Benutzern und Benutzerinnen gesondert aufgerufen und eingefügt werden. Die farbliche Hervorhebung, die bei manchen Daten notwendig ist, basiert auf sehr einfachen Regeln, die daher einfach zu programmieren wären.

Die Einschätzung, wie stark sich die Benutzeraktivität oder Durchlaufzeiten von automatisierten Tests geändert haben, wurde von den meisten Benutzern und Benutzerinnen nur geschätzt. Genaueres Kopfrechnen wäre kognitiver Mehraufwand und zusätzlich einen Rechner zu öffnen ein weiterer manueller Aufwand, zu dem die Benutzer und Benutzerinnen nicht bereit waren. Gerade Kalkulationen sind aber etwas, das der Computer einfach und wesentlich schneller erledigen kann.

Die fünfte Aufgabe war mit Absicht so ausgelegt, dass sie einen besonders starken manuellen Aufwand aufzeigt. Dies ist auch der Grund, weshalb diese Aufgabe im aktuellen Betrieb nicht so durchgeführt wird. Im Rahmen dieser Aufgabe müssen die Benutzer und Benutzerinnen für jeden Fehler herausfinden, ob dieser zum ersten Mal aufgetreten ist. Die einzige Möglichkeit, um diese Aufgabenstellung im alten System zu lösen, ist die Fehler händisch in den vorherigen E-Mails zu suchen. Trotz der Einschränkung beim Test, nur die E-Mails der letzten drei Tage zu überprüfen, zeigte sich, dass es ein enormer manueller Aufwand und sehr fehleranfällig ist.

Fehlende Infos und Unklarheiten

In den vorherigen Punkten wurde bereits erwähnt, dass nicht alle Daten in der E-Mail vorhanden sind. Dies führt zu einem höheren manuellen Aufwand, weil diese erst in die E-Mail kopiert werden müssen. Viel problematischer ist jedoch, dass die Benutzer und Benutzerinnen selber wissen müssen, wo sie diese Informationen finden können, und sie

keine Hinweise dazu erhalten. Bei der achten Aufgabe hat dieser Aspekt dazu geführt, dass vier von sieben Benutzer und Benutzerinnen nicht wussten, wo sie die notwendige Information finden können.

Darüber hinaus werden zusätzliche Informationen, die zwar nicht kopiert werden müssen, aber für die Analyse der Daten notwendig sind, viel leichter übersehen, wenn sie nicht direkt in der E-Mail vorhanden sondern an anderen Stellen abgelegt sind. Beispielsweise können manche Fehler ignoriert werden, wenn sie zu einer Zeit stattgefunden haben, wo ein Update der überwachten Applikation durchgeführt wurde. Diese Information befindet sich allerdings in anderen E-Mails, wodurch sie von den meisten Benutzern und Benutzerinnen übersehen wird.

Neben diesen vier Punkten ist noch ein weiterer negativer Aspekt des aktuellen Systems während den Tests aufgefallen. Die im Rahmen der Aufgaben erstellte Zusammenfassung sieht bei allen Benutzern und Benutzerinnen sehr unterschiedlich aus, obwohl im Grunde die selben Informationen erfasst wurden. Für die Aufgaben im Test hatte dies keine Auswirkung, solange die richtigen Punkte irgendwie erfasst wurden. Unter der Berücksichtigung, dass andere Personen jedoch anschließend mit dem Ergebnis dieser E-Mail arbeiten müssen, erschwert dieser Umstand die Arbeit. Es ist anzunehmen, dass sich eine einheitlichere Zusammenfassung positiv auf die Usability der Folgeprozesse auswirken würde.

5 Neues System

Anschließend zur ersten Evaluierung wurde unter Berücksichtigung von allgemeinen Anforderungen und den Erkenntnissen aus den ersten Usability Tests eine neue Applikation entwickelt, die das bisherige System ersetzen soll. In den folgenden Unterkapiteln werden zuerst die Anforderungen an das neue System beschrieben. Im zweiten Unterkapitel wird die technische Implementierung und die Architektur des neuen Systems präsentiert. Abschließend werden die wichtigsten Funktionen und Oberflächen des neuen Systems dargestellt.

5.1 Anforderungen

An das neue System wurden drei Hauptanforderungen gestellt. Im Vergleich zum bisherigen System sollten alle Tätigkeiten in einem einzigen System durchführbar sein, um die mehreren bisherigen Anwendungen für die unterschiedlichen Daten zu ersetzen.

Eine weitere Anforderung ist, dass die Applikation die Daten aus der bestehenden Datenbank abfragen kann. Das neue System soll die Daten nicht direkt von den Servern oder aus Dateien auslesen. Alle Daten der Server und Software werden in der Datenbank gesammelt. Die Sammlung der Daten ist keine Aufgabe des neuen Systems, weil dafür bereits eine ausreichende Implementierung existiert. Die neue Applikation soll lediglich diese Datenbank als Datenquelle nutzen und alle zu analysierenden Information dort abfragen können.

Die dritte, entscheidende Anforderung an das neue System ist eine konfigurierbare und flexible Lösung. Da sich die zu überprüfenden Daten ändern bzw. neue hinzukommen können, soll es möglich sein, die Applikation entsprechend anzupassen. Ebenso soll die Applikation pro Kundensystem unterschiedliche Konfigurationen hinterlegen können, um somit unterschiedliche Daten überprüfen zu können.

Neben diesen Kriterien wurden auch noch folgende Ansprüche an die eingesetzten Technologien gestellt. Sowohl die Softwareentwicklung, als auch alle internen Tools laufen auf Windows und sind auf das Microsoft .NET Framework aufgebaut. Das neue System soll ebenfalls diesen Standards entsprechen. Zusätzlich soll es als interne Webapplikation implementiert werden, damit keine lokalen Installationen nötig sind. Um aber den Zugriff auf diese Webapplikation nur für berechtigte User sicherzustellen, ist eine Authentifizierung notwendig. Damit die Benutzerverwaltung im Unternehmen weiterhin zentral verwaltet werden kann, muss die Applikation entweder seine eigene Benutzerverwaltung mit dem Active Directory der Firma synchronisieren oder die Windows Authentifikation implementieren.

5.2 Architektur

Im ersten Abschnitt dieses Kapitels werden die eingesetzten Technologien genannt, auf welchen die Implementierung basiert. Der anschließende Teil geht auf den Aufbau der Applikation und die wichtigsten Konzepte der Implementierung ein.

Wie bereits im vorherigen Kapitel beschrieben wurde, basiert das neue System auf dem Microsoft .NET Framework [25] – in diesem konkreten Fall auf der Version 4.6 – und ist mit der objekt-orientierten Programmiersprache C# [25] implementiert worden. Bei der Implementierung handelt es sich um eine ASP.NET MVC 5 [26] Webapplikation die zusätzlich das ASP.NET WebAPI Framework [27] nutzt. ASP.NET MVC 5 stellt das Grundgerüst der Webapplikation dar und liefert die notwendigen Seiten an den Benutzer oder die Benutzerin. Die auf ASP.NET WebAPI basierenden HTTP Services liefern weitere Daten, die vom Client angefragt werden können. Clientseitig wird zusätzlich das Javascript Framework AngularJS [28] eingesetzt, mit dem es möglich ist, Inhalte einer Seite dynamisch nachzuladen ohne dabei ein komplettes Neuladen der Seite zu benötigen.

Die Architektur der Applikation ist in zwei wesentliche Bereiche unterteilt. Ein Teil stellt das Webinterface dar, bestehend aus Client und serverseitigem Code. Der zweite Teil ist eine Service-DLL, welche die eigentliche Logik und Datenzugriffe abbildet. Die Webapplikation nutzt diese DLL also zum Laden bzw. Manipulieren der Daten und Ausführen jeder Logik, welche nicht unmittelbar mit der visuellen Darstellung zu tun hat.

Eine weitere Anforderung ist die Konfigurierbarkeit und Flexibilität der Anwendung. Um diese Anforderung zu erfüllen, wurden kaum direkte Informationen zu den Daten in die Software einprogrammiert. Stattdessen wurden allgemeine Konstrukte geschaffen, um Daten zu laden, zu verarbeiten und anzuzeigen. Um welchen Daten es sich dann schlussendlich wirklich handelt, kann mit XML Dateien konfiguriert werden. Jede XML Datei stellt ein Template dar und enthält eine konkrete Konfiguration für ein Set an verschiedenen Daten. Zum Beispiel gibt es pro Kundensystem ein eigenes Template, das unterschiedlichste Daten und Einstellungen haben kann. Die Abbildung 5.1 zeigt einen Auszug aus einer möglichen XML Konfiguration.

```

<?xml version="1.0" encoding="utf-8"?>
<Template xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <Name>customer1</Name>
  <Title>Customer 1</Title>
  <Description>Daily monitoring of DBS system for customer 1</Description>
  <DefaultLookbackHours>26</DefaultLookbackHours>
  <EmailSettings>
    <MessageTemplate>DBSMailTemplate</MessageTemplate>
    <SubjectTemplate><![CDATA[{0} DBS Quickcheck Monitoring --> {1}]]></SubjectTemplate>
    <ToAddress />
    <CcAddress />
    <BccAddress />
  </EmailSettings>
  <Sections>
    <Section>
      <Name>errors</Name>
      <Title><![CDATA[Errors & Warnings]]></Title>
      <Description><![CDATA[Errors & Warnings from Server Eventlogs]]></Description>
      <Blocks>
        <Block>
          <Name>errors</Name>
          <Title><![CDATA[EventLogs]]></Title>
          <SqlSetting>
          <DataDefinition>
          <Rules>
          <Filters>
        </Block>
      </Blocks>
    </Section>
    <Section>
      <Name>jobs</Name>
      <Title>Jobs</Title>
      <Description></Description>
      <Blocks>
    </Section>
    <Section>
      <Name>selenium</Name>
      <Title>Automatische Test</Title>
      <Description></Description>
      <Blocks>
    </Section>
    <Section>
      <Name>disk_space</Name>

```

Abbildung 5.1: Auszug aus einer XML Konfigurationsdatei

In der XML Datei können beliebig viele Bereiche festgelegt werden, die der Benutzer oder die Benutzerin überprüfen soll. Innerhalb dieser Bereiche wird definiert, welche Daten für diesen Bereich geladen und wie sie dargestellt werden sollen. So bleibt der Inhalt flexibel und kann einfach angepasst bzw. ergänzt werden, wenn es notwendig ist. Für den Anwender oder die Anwenderin werden die Daten dann in Form einer Tabelle präsentiert, deren Darstellung und Funktionen sich dynamisch an die jeweiligen Daten und Einstellungen anpassen. Zu den Funktionen zählen beispielsweise:

- Sortierbarkeit
- Vordefinierte Filterauswahl
- Manuelle Filtermöglichkeiten
- Farbliche Markierungen
- Verknüpfung mit anderen Daten

Sowohl die Templates als auch die bereits geladenen Daten werden für eine bestimmte Zeit in einem Cache gespeichert, um unnötige mehrfache Datenbank- und Dateizugriffe zu vermeiden. So kann der Benutzer oder die Benutzerin beliebig zwischen Bereichen wechseln und muss nicht jedes Mal erneut warten, bis die Daten geladen sind.

Der Zugriff auf die Applikation ist nur für authentifizierte Benutzer und Benutzerinnen möglich. Allerdings wurde keine eigene Benutzerverwaltung entwickelt, sondern Windows Authentication implementiert. So können die Anwender und Anwenderinnen ohne gesonderte Anmeldung auf die Applikation zugreifen, solange sie in der entsprechenden Windows Domain angemeldet sind. [29] Außerdem kann so auf bereits bestehende Informationen zur zugreifenden Person, wie Name und E-Mail-Adresse, zugegriffen werden.

5.3 Funktionalität & Oberflächen

Dieses Kapitel präsentiert die wichtigsten Funktionen und Oberflächen des neu entwickelten Systems.

Beim Einstieg in das System erhält der Benutzer oder die Benutzerin zuerst die in Abbildung 5.2 dargestellte Seite, die alle verfügbaren Templates auflistet. Ein Template entspricht in diesem Fall der Konfiguration eines bestimmten Kundensystems, für das die Analyse der konfigurierten Daten durchgeführt werden soll. Der Zugriff auf die gesamte Applikation ist grundsätzlich nur für autorisierte Benutzer und Benutzerinnen möglich. Die Anmeldung erfolgt automatisch über Windows Authentication und der aktuelle Benutzer oder die aktuelle Benutzerin wird in der rechten oberen Ecke angezeigt.

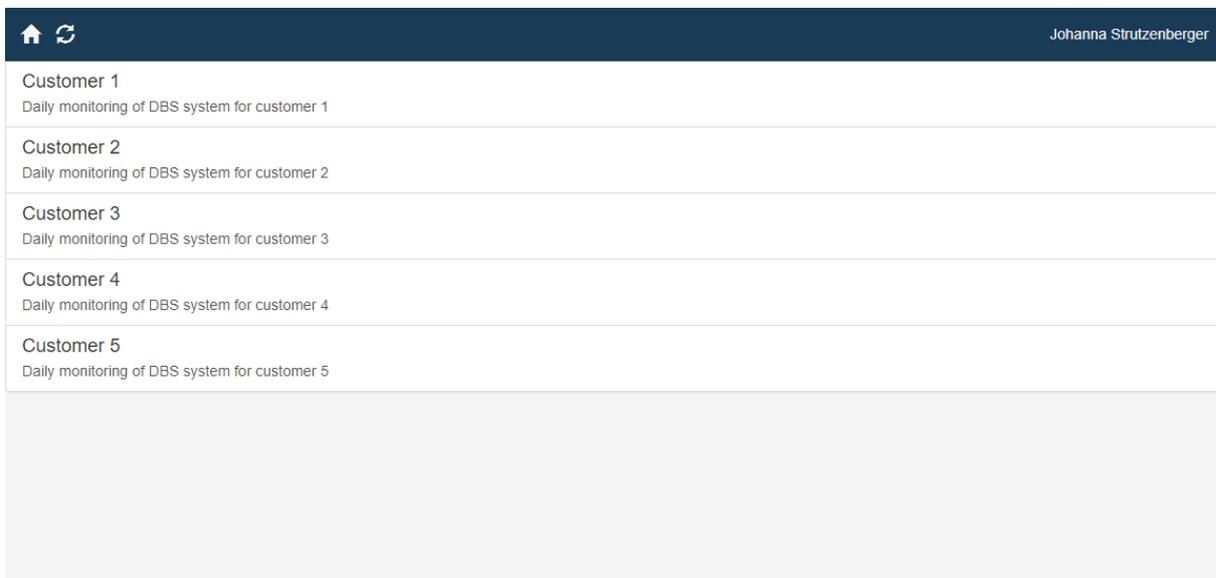


Abbildung 5.2: Startseite des neuen Systems

Bei Auswahl eines Template wird der Benutzer oder die Benutzerin nach dem gewünschten Zeitfenster für die Daten gefragt. Nach der Bestätigung werden die Daten aus diesem Zeitraum geladen. Da das Laden der Daten einige Sekunden dauern kann, sieht der Benutzer oder die Benutzerin in der Zwischenzeit die in Abbildung 5.3 angezeigte

Ladeanimation. Dem Benutzer oder der Benutzerin wird sowohl in der seitlichen Navigation für alle Bereiche als auch im Hauptfenster für den aktuell ausgewählten Bereich angezeigt, wenn die Daten noch laden.

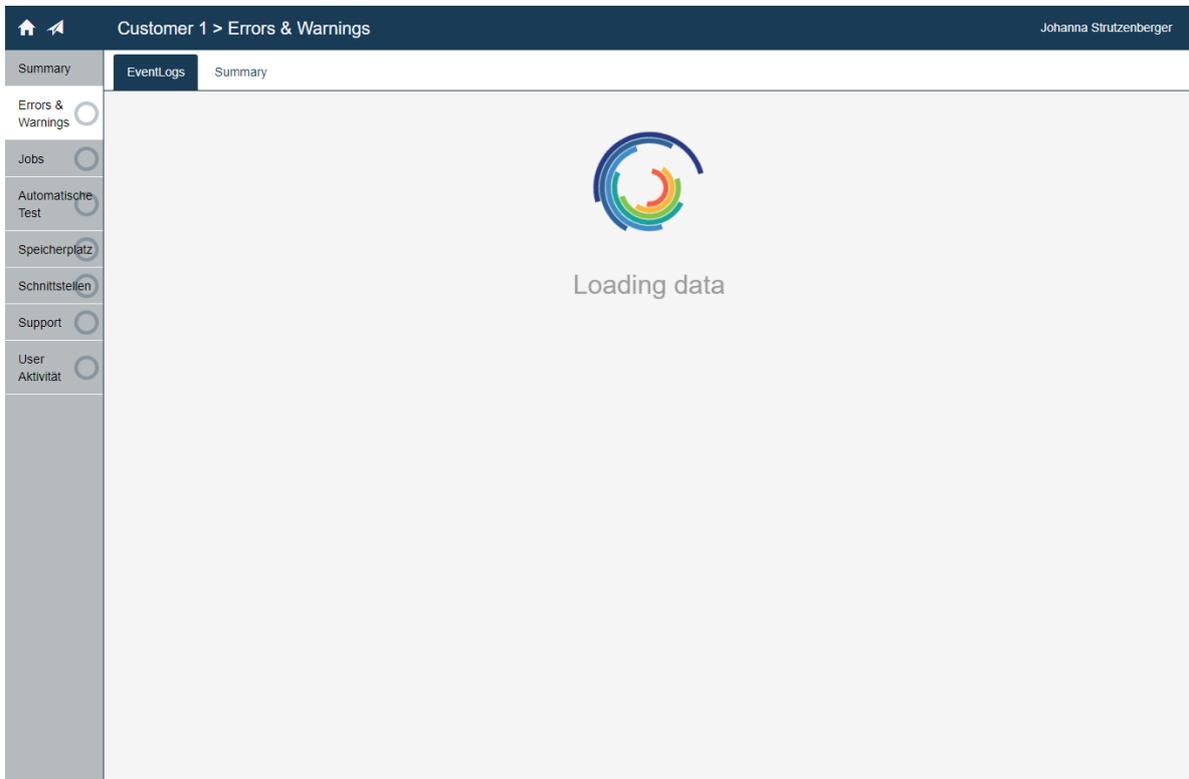


Abbildung 5.3: Laden der Daten im neuen System

Grundsätzlich ist die Seite so aufgebaut, dass in der seitlichen Navigation alle konfigurierten Bereiche aufgelistet sind, für welche die Daten analysiert werden sollen. Im Hauptfenster werden die jeweiligen Daten aus dem ausgewählten Bereich angezeigt. Wenn ein Bereich mehrere unterschiedliche Datengruppen enthalten sollte, können diese zusätzlich innerhalb des Bereichs auf mehrere Tabs aufgeteilt werden.

Sobald die Daten geladen sind, werden diese in einer Tabellenstruktur dargestellt. Für alle Daten bietet diese Tabelle folgende Funktionen:

- Das Filtern der Daten durch Eingabe von Suchbegriffen in den jeweiligen Spalten
- Das Sortierung nach einer bestimmten Spalte
- Die Auswahl von Zeilen für weitere Aktionen
- Das Hinzufügen einer oder mehrere Zeilen zur Zusammenfassung

In Abbildung 5.4 ist beispielhaft die Tabelle mit den Fehlermeldungen dargestellt. Die Tabelle selbst enthält nur den Anfang einer Fehlermeldung. In der Abbildung ist jedoch zu sehen,

dass der gesamte Fehlertext als Tooltip angezeigt werden kann. Abbildung 5.5 zeigt den Dialog, in dem ebenfalls die gesamte Fehlermeldung auf Wunsch angezeigt werden kann.

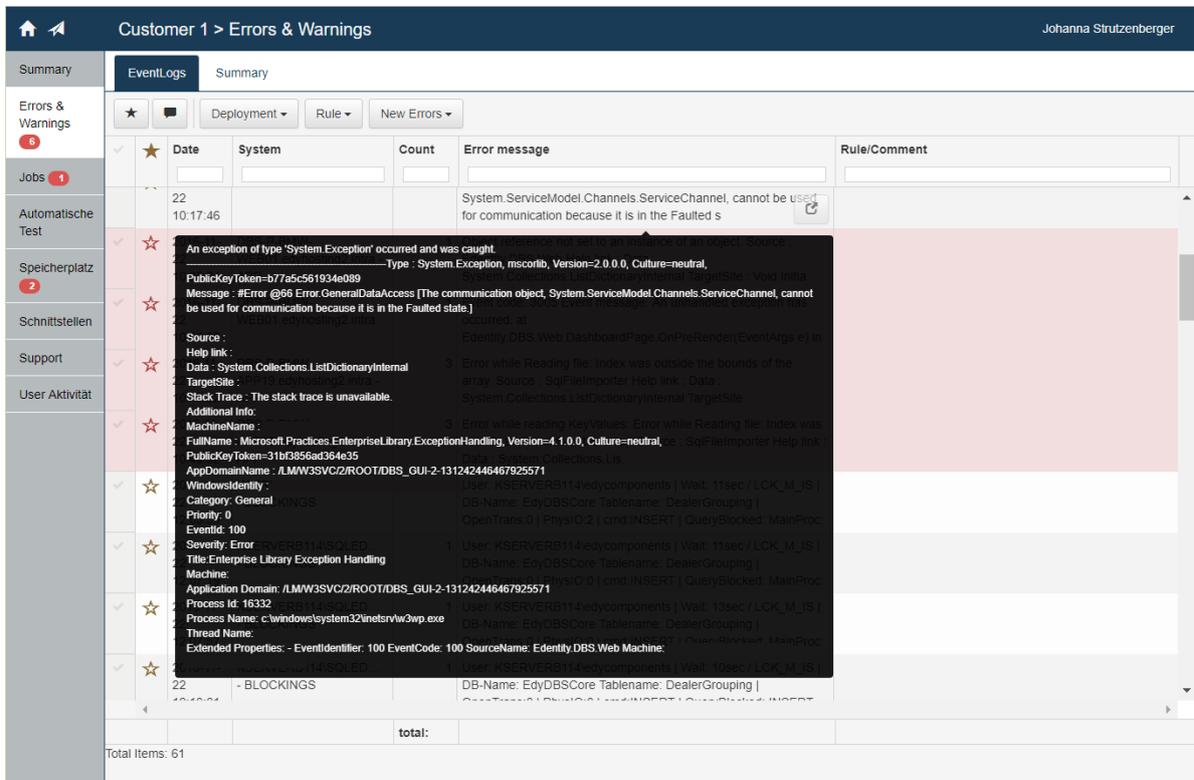


Abbildung 5.4: Fehleransicht im neuen System

Customer 1 > Jobs Johanna Strutzenberger

Summary

Jobs Filter entries by: Status

Errors & Warnings Status

Jobs Error

Automatische Test

Task	Begin	End	Duration	Area/Run	All errors	Interpreted errors	All warnings	Interpreted warnings	Not interpreted	Actions
★ Eidentity.DBS.Export.Cockpit	2016-11-23	2016-11-23	0.70		1	1	0	0	84	Errors
★ Eidentity.DBS.Export.Crowe	2016-11-15	2016-11-15	268.25		0	0	0	0	2258	Errors
★ Eidentity.DBS.Export.Essbase	2016-11-23	2016-11-23	0.10		0	0	0	0	9	Errors
★ Eidentity.DBS.Import.Currency	2016-11-23	2016-11-23	0.00		0	0	1	1	6	Errors
★ Eidentity.DBS.Import.Hst	2016-11-23	2016-11-23	0.65		0	0	0	0	36	Errors
★ Eidentity.DBS.Service.DealerImp...	2016-11-23	2016-11-23	15.02		0	0	0	0	0	Errors
★ Eidentity.DBS.Service.MarketImp...	2016-11-15	2016-11-16	1274.42		994	994	0	0	510	Errors

Total Items: 7

total: 995 total: 995 total: 1 total: 1 total: 2,903

Abbildung 5.6: Filter- und Darstellungsmöglichkeiten im neuen System

Zusätzlich ist es oft notwendig sich Daten aus anderen Bereich anzusehen, um einen Zusammenhang zu den aktuellen Daten zu überprüfen und die gefundenen Daten gegebenenfalls zu verknüpfen. In Abbildung 5.7 ist zum Beispiel dargestellt, wie man für einen fehlerhaft durchgeführten Job nach der Ursache innerhalb der Applikationsfehler aus der dazu passenden Zeitspanne suchen und passende Einträge verknüpfen kann.

The screenshot displays a software interface with an 'Errors' window. The window includes a search bar with 'Start' and 'End' date and time pickers. Below the search bar is a table with the following columns: Date, System, Cou..., Error message, and Rule. The table contains several rows of error logs, including messages about 'Error while reading KeyValues' and 'Wait: 11sec / LCK_M_IS | DB-Name: EdyDBSCore Table-Name:'. At the bottom of the table, there is a 'total:' row. The interface also shows a sidebar on the left with navigation options like 'Summary', 'Jobs', 'Errors & Warnings', and 'User Aktivität'. A 'Johanna Strutzenberger' name is visible in the top right corner.

Date	System	Cou...	Error message	Rule
	APP		outside the bounds of the array. Source : SqlFileImporter Help link : Data :	
22.11.2016 10:48:02	- APP	3	Error while reading KeyValues: Error while Reading file. Index was outside the bounds of the array. Source :	
22.11.2016 12:05:59	- BLOCKINGS	1	User: Wait: 11sec / LCK_M_IS DB-Name: EdyDBSCore Table-Name:	
22.11.2016 12:07:19	- BLOCKINGS	1	User: Wait: 11sec / LCK_M_IS DB-Name: EdyDBSCore Table-Name:	
22.11.2016 12:07:57	- BLOCKINGS	1	User: Wait: 13sec / LCK_M_IS DB-Name: EdyDBSCore Table-Name:	
22.11.2016 12:10:01	- BLOCKINGS	1	User: Wait: 10sec / LCK_M_IS DB-Name: EdyDBSCore Table-Name:	
22.11.2016 12:10:01	- BLOCKINGS	2	User: Wait: 10sec / LCK_M_IS DB-Name: EdyDBSCore Table-Name:	
		total:		

Total Items: 61 (Selected Items: 2)

Abbildung 5.7: Suche in verknüpften Daten im neuen System

Wenn Datenspalten numerische Werte enthalten, kann für diese Daten definiert werden, dass sie auf bestimmte Art und Weise aggregiert und das Ergebnis in der Fußzeile der Tabelle ausgegeben werden soll. Zum Beispiel könnte die Summe oder der Durchschnitt der Werte berechnet werden, wie in Abbildung 5.8 dargestellt. Diese Abbildung zeigt zusätzlich, dass eine Spalte definiert werden kann, die den Trend der Differenz aus zwei anderen Spalten ausgibt.

Customer 1 > User Aktivität Johanna Strutzenberger

Summary **User Aktivität** Summary

Errors & Warnings 6

Jobs 1

Timestamp	Machine	UniqueUserCount	LastMonth	Trend	CallsCount	AvgRuntime	AvgOpentime	StdevRuntime
2016-12-09		60	212	-72% ↓	23633	860	245	5864
2016-12-09		113	220	-49% ↓	35140	966	-7	5259
2016-12-09		141	227	-38% ↓	53269	1563	54	9732
2016-12-09		140	236	-41% ↓	59068	1390	-3	8767
2016-12-09		133	216	-38% ↓	47331	1009	-67	6498
2016-12-09		105	203	-48% ↓	37568	632	-112	3775
2016-12-09		81	276	-71% ↓	21633	596	-137	3435
2016-12-09		132	302	-56% ↓	54211	792	-158	5279
2016-12-09		124	331	-63% ↓	52835	698	-96	4544
2016-12-09		115	212	-46% ↓	40077	523	-11	3350
2016-12-09		49	9	444% ↑	15196	405	-27	2301
2016-12-09		5	6	-17% ↓	1208	1297	2303	10883
2016-12-08		4	5	-20% ↓	786	336	70	912
2016-12-08		3	9	-67% ↓	1277	189	1610378	620
2016-12-08		3	5	-40% ↓	1473	130	4586067	516
2016-12-08		2	9	-78% ↓	1225	634	4176469	10182
2016-12-08		4	21	-81% ↓	2557	949	1220333	8757
2016-12-08		5	50	-90% ↓	2080	3168	555182	23934
2016-12-08		8	73	-89% ↓	4173	1767	10545	14516
		avg: 60	avg: 132	avg: -41	avg: 21,848	avg: 942	avg: 419,340	avg: 6,694

Total Items: 29

Abbildung 5.8: Trendspalte und Aggregation von numerischen Werten im neuen System

Ein essentieller Teil der Tätigkeit ist es, alle Auffälligkeiten in einer Zusammenfassung zu kommentieren. Dazu kann in jeder Datentabelle eine oder mehrere Zeilen ausgewählt werden, die anschließend mit einem Kommentar in die Zusammenfassung eingetragen werden können. In Abbildung 5.9 wird der Dialog dargestellt, in dem die ausgewählten Daten kommentiert und hinzugefügt werden können. Die Abbildung 5.10 zeigt die fertige Zusammenfassung mit allen hinzugefügten Einträgen, die über einen eigenen Punkt in der Navigation aufgerufen werden können. Zusätzlich gibt es in jedem Bereich einen eigenen Tab mit dem alle Einträge für die Zusammenfassung aus diesem Bereich angezeigt werden.

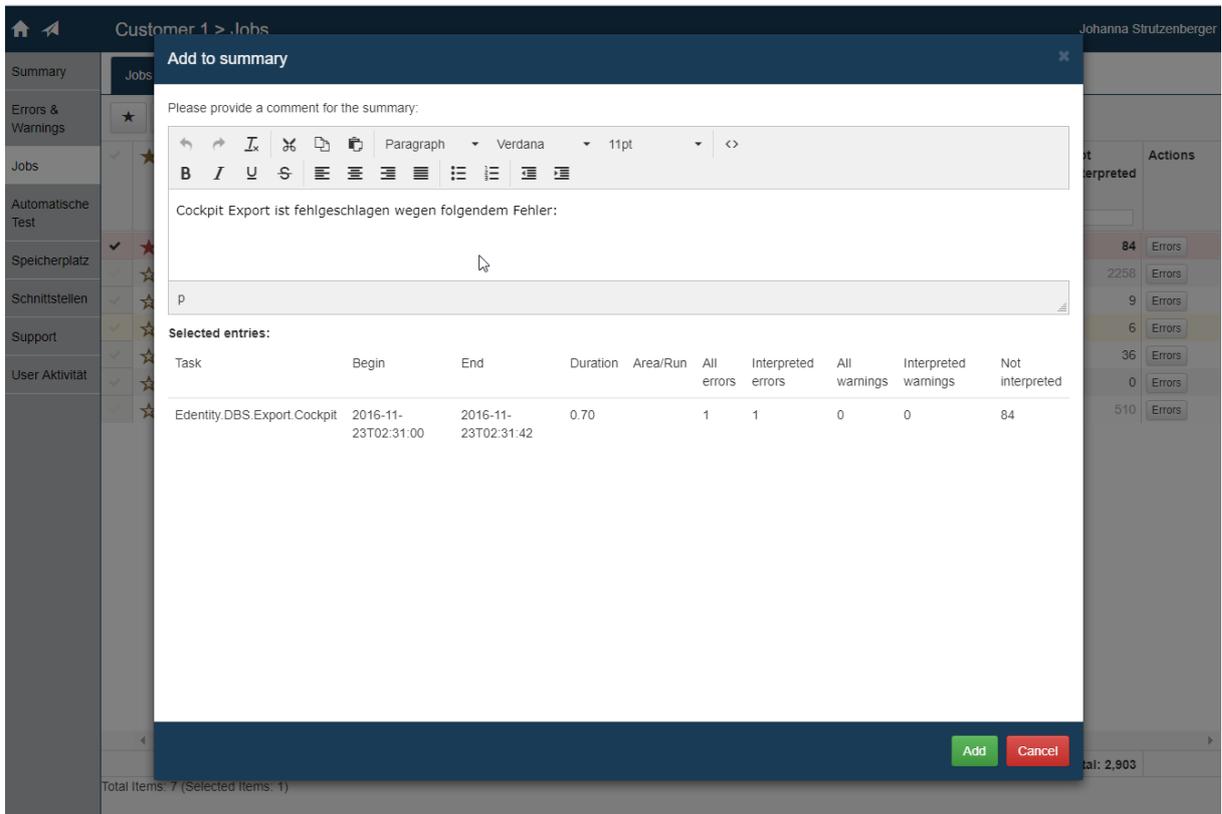


Abbildung 5.9: Kommentieren von Problemen für die Zusammenfassung im neuen System

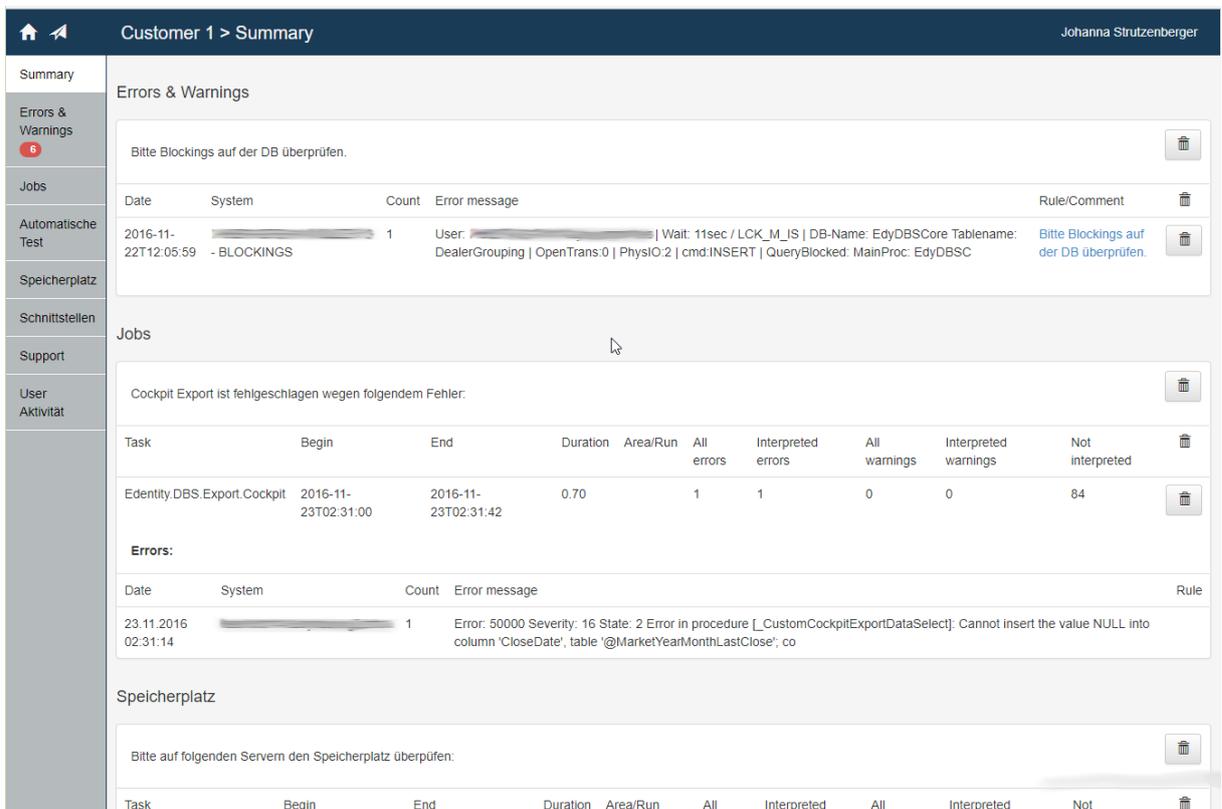


Abbildung 5.10: Zusammenfassung im neuen System

Sobald der Benutzer oder die Benutzerin mit der Analyse der Daten und dem Erfassen aller Auffälligkeiten fertig ist, kann er oder sie das Ergebnis der Arbeit als E-Mail versenden. In dem in Abbildung 5.11 zu sehenden Dialog können die Benutzer und Benutzerinnen noch ihre Gesamtbewertung zum Status des Applikationsbetriebes auswählen und die Empfänger und Empfängerinnen der E-Mail bestimmen. Anschließend werden alle Informationen in der Zusammenfassung, gemeinsam mit allen ursprünglichen Daten aus jedem Bereich, in einer E-Mail an diese Empfänger und Empfängerinnen versendet.

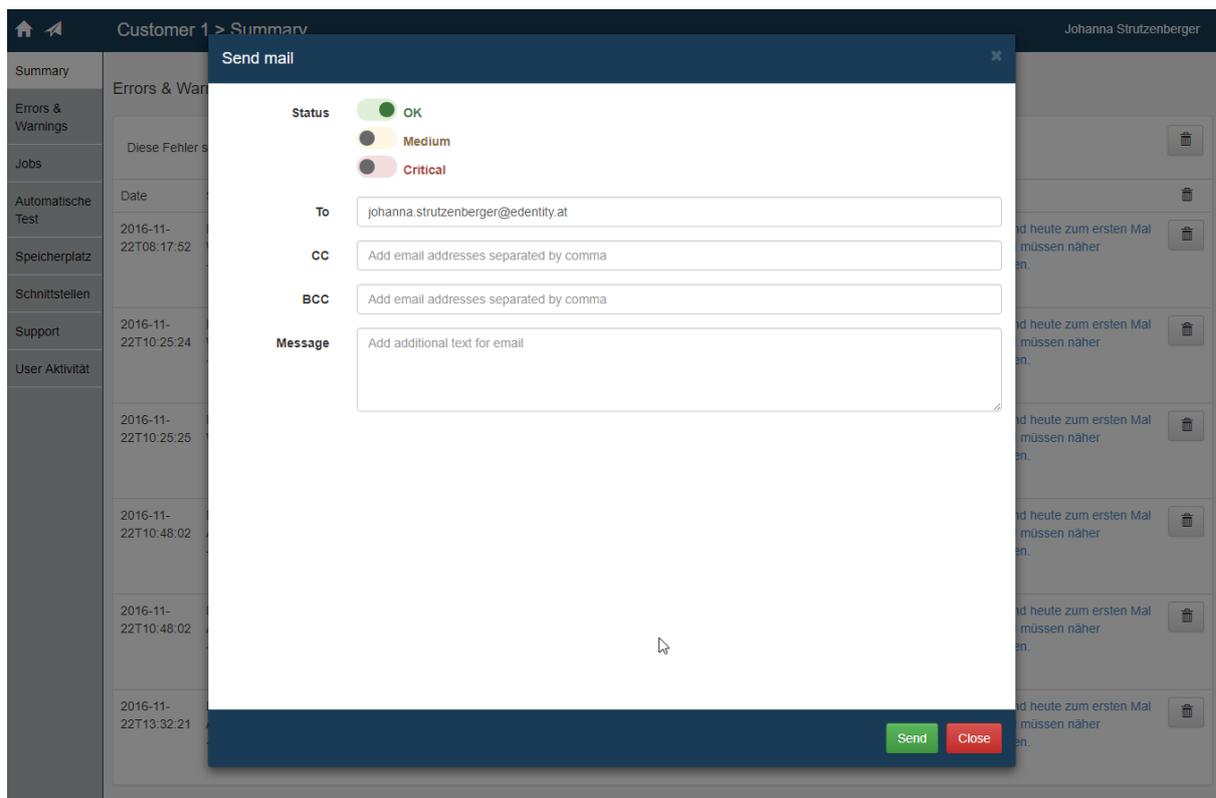


Abbildung 5.11: Senden der Email im neuen System

6 Evaluierung des neuen Systems

Wie schon bei der ersten Evaluierung wurden aus den Usability Tests des neuen Systems die Metriken Task Success, Time on Task und der SUS Score ausgewertet. In den folgenden drei Unterkapiteln werden die Ergebnisse dieser Metriken im Vergleich zu den Ergebnissen der ersten Evaluierung präsentiert.

6.1 Task Success

Die Abbildung 6.1 zeigt die Aufteilung der Teilnehmer und Teilnehmerinnen auf die Kategorien „erfolgreich abgeschlossen“, „teilweise erfolgreich abgeschlossen“ und „fehlgeschlagen“ pro Aufgabe. Die erste Spalte stellt dabei jeweils die Ergebnisse aus der Evaluierung des alten Systems dar, während die zweite Spalte das Resultat für das neue System repräsentiert.

Der Vergleich zeigt, dass mehr Aufgaben im neuen System von allen Teilnehmern und Teilnehmerinnen erfolgreicher abgeschlossen werden konnten als im alten System. Nur bei den Aufgaben 3 und 4 ist die Erfolgsrate nicht 100%, aber dennoch konnte auch hier mehr als die Hälfte der Personen die Aufgabe erfolgreich abschließen. Insgesamt scheiterte lediglich eine Person bei Aufgabe 4, während die anderen drei Personen bei Aufgabe 3 und 4 zumindest teilweise erfolgreich waren.

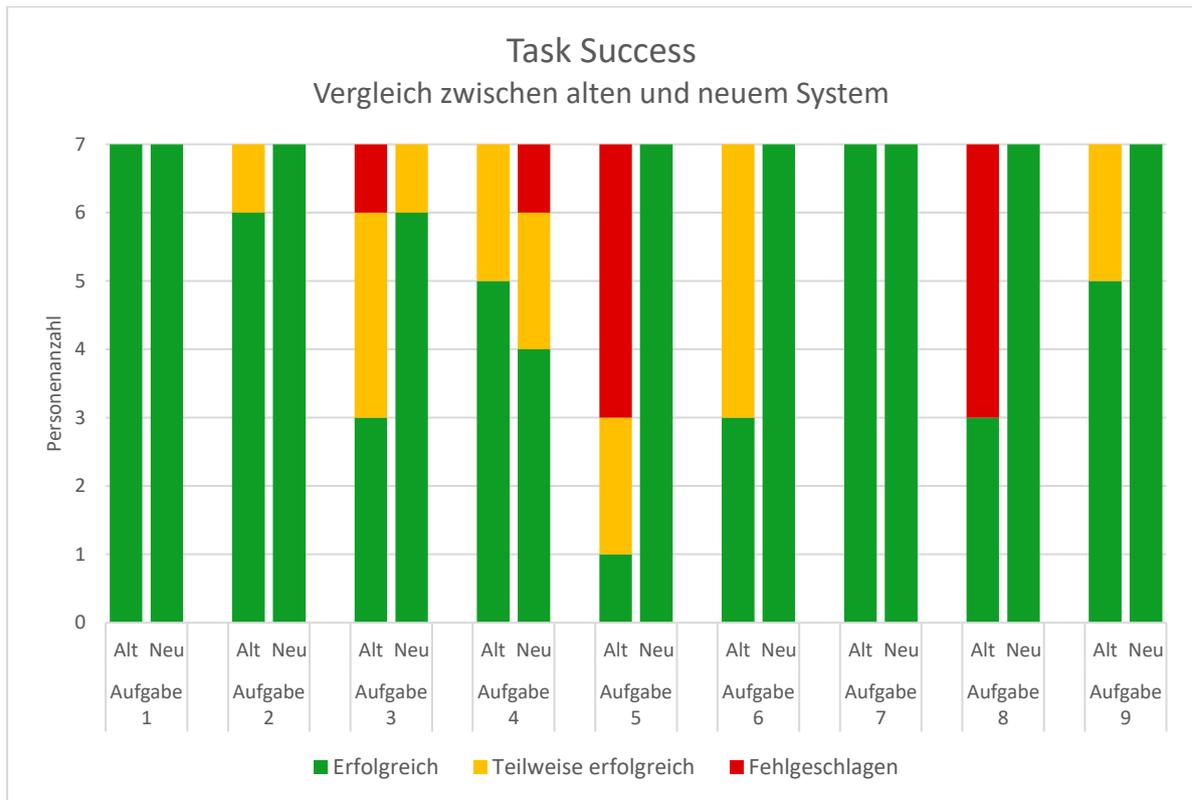


Abbildung 6.1: Vergleich von Task Success zwischen altem und neuem System

6.2 Time on Task

Bei der ersten Evaluierung benötigten die Teilnehmer und Teilnehmerinnen im Durchschnitt für alle Aufgaben insgesamt 41 Minuten (2478 Sekunden). Für die Evaluierung des neuen Systems beträgt die durchschnittliche Gesamtzeit für alle Aufgaben 16 Minuten (957 Sekunden). Die Teilnehmer und Teilnehmerinnen führten demnach die selben Aufgaben im neuen System schneller durch als im alten System.

Die Abbildung 6.2 zeigt den Mittelwert der gemessenen Zeiten pro Aufgabe, jeweils für das alte System in der ersten Spalte und das neue System in der zweiten Spalte. Im Vergleich der durchschnittlichen Zeiten pro Aufgabe sticht nur eine Aufgabe hervor, deren Ergebnis sich im neuen System um 15 Sekunden von 152 auf 167 Sekunden verschlechtert hat. Dafür ist bei anderen Aufgaben, wie zum Beispiel Aufgabe 5 oder 6, zu sehen, dass die Zeit um mehr als die Hälfte reduziert werden konnte.

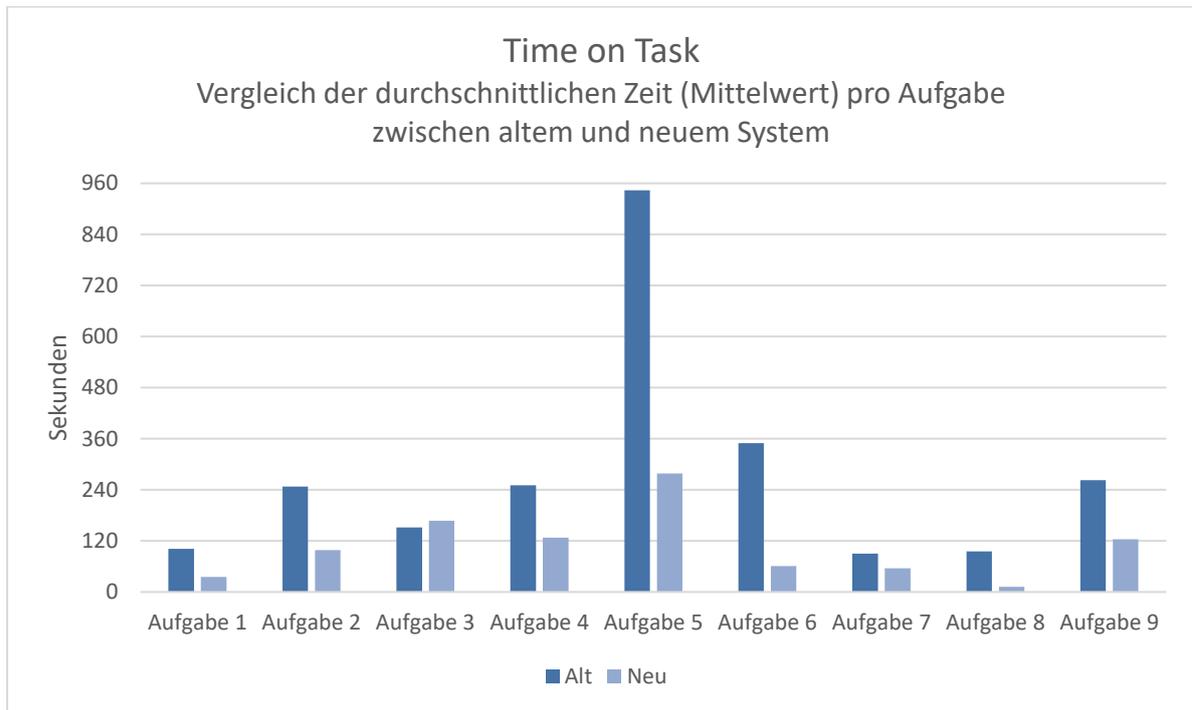


Abbildung 6.2: Vergleich der durchschnittlichen Time on Task (Mittelwert) zwischen altem und neuem System

Neben der durchschnittlichen Zeit pro Aufgabe ist auch der Vergleich der Aufgabenschlüsse innerhalb der definierten Zeitlimits interessant. Die Abbildung 6.3 zeigt die Anzahl der Teilnehmer und Teilnehmerinnen, die die jeweilige Aufgabe innerhalb einer bestimmten Zeit durchführen konnten. Im Vergleich zum alten System (links) ist zu erkennen, dass im neuen System (rechts) zumindest bei jeder Aufgabe die Hälfte der Personen das Zeitlimit einhalten konnten.

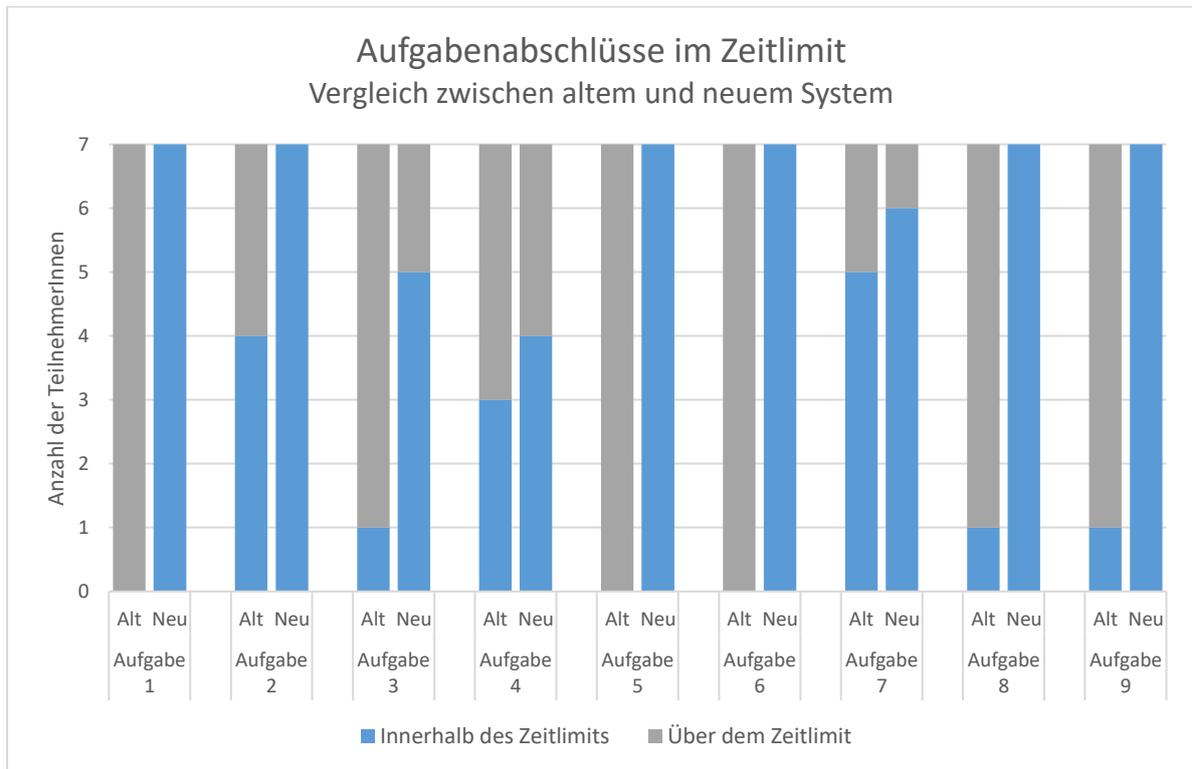


Abbildung 6.3: Vergleich der Anzahl an TeilnehmerInnen mit Abschluss im Zeitlimit zwischen altem und neuem System

6.3 SUS Score

In der Evaluierung des neuen Systems wurden die Teilnehmer und Teilnehmerinnen erneut gebeten den SUS Fragebogen, diesmal in Bezug auf das neue System, auszufüllen. Aus den Antworten jedes Teilnehmers oder jeder Teilnehmerin wurde erneut der SUS Score berechnet und auch bei dieser Metrik konnte eine Verbesserung festgestellt werden. Im Durchschnitt erhielt das alte System einen SUS Score von 20, wohingegen es sich beim neuen System auf 91 beläuft. Die Abbildung 6.4 zeigt erneut die SUS Score Verteilung auf die Prozentränge. Das durchschnittliche Ergebnis für das alte System ist in Gelb und das Ergebnis für das neue System in Grün hervorgehoben. Im Vergleich zum alten System ist die Usability des neuen Systems damit besser bewertet worden als über 90 Prozent aller Produkte, die Jeff Sauro in „A Practical Guide to the System Usability Scale“ aus 446 Studien ausgewertet hat. [18, p. 53]

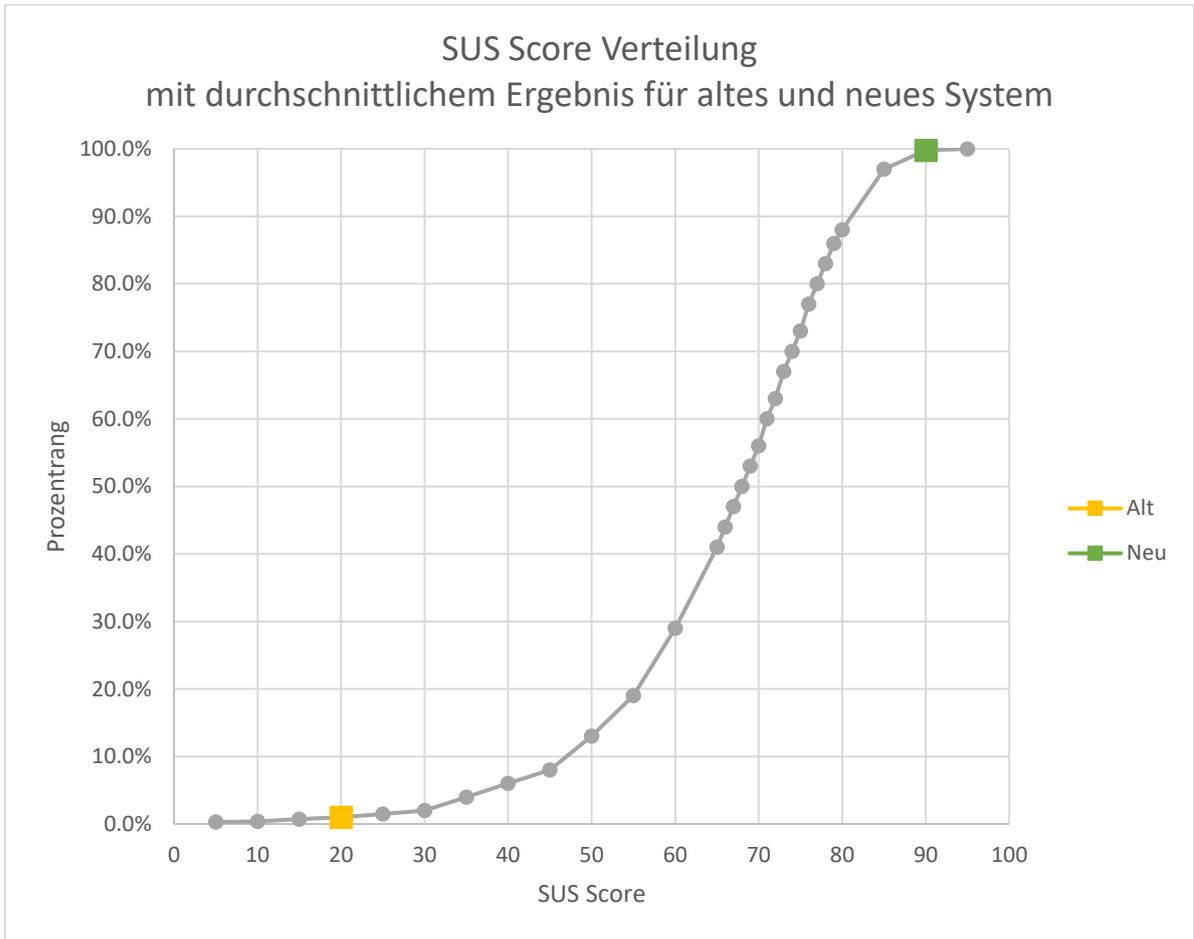


Abbildung 6.4: SUS Score Verteilung mit den durchschnittlichen Ergebnissen für das alte und neue System hervorgehoben

7 Diskussion

Diese Arbeit beschäftigt sich mit der Usability Evaluierung von komplexen Systemen zur Informationsanalyse und geht speziell auf die praktische Anwendung im Bereich der Applikationsüberwachung ein. Zu Beginn dieser Arbeit wurden folgende Forschungsfragen aufgestellt:

- Kann mit klassischen Usability Methoden die Usability von komplexen Systemen zur Informationsanalyse evaluiert und Usability Metriken gemessen werden?
- Sind die Ergebnisse aus zwei Evaluierungen vergleichbar und können Maßnahmen daraus abgeleitet werden, die eine Verbesserung der Usability bewirken?

Zur Beantwortung dieser Fragen wurde die Usability eines bestehenden Analysesystems evaluiert und dafür Usability Tests und der SUS Fragebogen als Methoden eingesetzt.

Wenn man zuerst nur die in Kapitel 4 „Evaluierung des bisherigen Systems“ präsentierten Ergebnisse zur ersten Evaluierung betrachtet, so lässt sich in allen ausgewerteten Metriken der selbe Trend erkennen. Time on Task in Kombination mit Task Success zeigt auf, dass Usability Probleme zu Ineffizienzen und nicht erfolgreich gelösten Aufgaben führen. Obwohl alle Teilnehmer und Teilnehmerinnen das System bereits kannten, dauerten die Aufgaben länger als erwartet und führten teilweise zu einer hohen Fehlerrate.

Wie in „Measuring the User Experience“ [12, p. 71] beschrieben, wurde das Ergebnis der Teilnehmer und Teilnehmerinnen pro Aufgabe anhand von drei unterschiedlichen Erfolgslevels ausgewertet:

- Erfolgreich
- Teilweise erfolgreich
- Fehlgeschlagen

Diese Auswertung zeigte, dass nur bei zwei von neun Aufgaben, die im Rahmen der Tests durchgeführt wurden, alle Teilnehmer und Teilnehmerinnen diese Aufgaben erfolgreich abschließen konnten. Hinzu kommt, dass bei vier Aufgaben mehr als die Hälfte der Personen die Aufgabe nur teilweise oder gar nicht erfolgreich abschließen konnten.

Die Auswertung der durchschnittlichen Zeit pro Aufgabe ohne Vergleichswerte liefert noch keine konkrete Aussage darüber, ob diese Zeiten in Ordnung sind oder die Personen auf Grund von Usability Problemen länger gebraucht haben. Deshalb wurden akzeptable Zeitlimits für jede Aufgabe definiert und zusätzlich ausgewertet, wie viele Personen eine Aufgabe innerhalb dieses Zeitlimits abschließen konnten. [12, p. 80] Mit dieser Vorgehensweise deutet auch diese Metrik auf Usability Probleme hin, da nur bei einer Aufgabe mehr als die Hälfte der Teilnehmer und Teilnehmerinnen unter dem Zeitlimit

geblieben sind. Bei zwei anderen Aufgaben war es keiner Person möglich, die Aufgabe innerhalb einer akzeptablen Zeit abzuschließen.

Ebenso deutet die Auswertung der Antworten des SUS Fragebogen auf eine schlechte Usability des bisherigen Systems hin. Im Durchschnitt erhielt das bisherige System von den Benutzern und Benutzerinnen nur einen SUS Score von 20. Wenn man diesen Wert mit dem von Jeff Sauro ausgewerteten durchschnittlichen SUS Score von 68 vergleicht, so wurde dieses System stark unter dem Durchschnitt bewertet. [18, p. 53] Selbst der von zwei Personen vergebene maximale SUS Score von 33 kommt nicht an den Durchschnitt heran.

Die erste wichtige Frage dieser Arbeit befasst sich damit, ob klassische Methoden der Usability Evaluierung wie Usability Tests auch bei komplexen Systemen zur Informationsanalyse eingesetzt werden können. Wenn man nur nach Betrachtung der ersten Evaluierung auf diese Frage eingeht, so deutet der einheitliche Eindruck zur Usability, der von allen Auswertungen vermittelt wird, bereits daraufhin. Es ist ein positives Anzeichen für die Gültigkeit der eingesetzten Methoden, dass diese Ergebnisse in sich schlüssig sind und keine Widersprüche enthalten.

Zur weiteren Validierung der Methoden spielt es dennoch eine wichtige Rolle, dass eine zweite Evaluierung durchgeführt und ausgewertet wird. Da die erste Evaluierung Usability Probleme im bisherigen System aufgezeigt hat, wurde ein neues System entwickelt. Das Ziel des neuen Systems ist, die im Kapitel 4.4 beschriebenen Usability Probleme zu beseitigen und die Metriken aus der ersten Evaluierung zu verbessern. Mit einer zweiten Evaluierung des neuen Systems wurde anschließend überprüft, ob dieses Ziel erreicht werden konnte. Im Rahmen dieser Evaluierung führten die selben Teilnehmer und Teilnehmerinnen die gleichen neun Aufgaben wie bereits bei den ersten Tests durch.

Der Vergleich der ausgewerteten Metrik Task Success aus beiden Tests zeigt eine deutliche Verbesserung im neuen System. Statt nur zwei Aufgaben mit einer Erfolgsrate von 100 Prozent, konnten im neuen System sieben von neun Aufgaben von allen Teilnehmern und Teilnehmerinnen erfolgreich abgeschlossen werden. Die verbleibenden zwei Aufgaben weisen ein bzw. drei Personen auf, die diese Aufgaben entweder nur teilweise oder gar nicht abschließen konnten. Insgesamt verzeichnet das neue System nur einen Misserfolg, während es mit dem alten System neun gab.

Bei der Auswertung der durchschnittlich benötigten Zeit für alle Aufgaben zeigt sich im Vergleich zum alten System eine starke Verbesserung. Bis auf eine Aufgabe mit minimal höheren Zeiten im neuen System konnten die Zeiten für alle anderen Aufgaben verbessert werden. Insgesamt konnte die Dauer für die Durchführung aller Aufgaben mit dem neuen System um 60 Prozent reduziert werden.

Auch bei der Betrachtung der Aufgabenabschlüsse innerhalb der definierten Zeitlimits zeigte sich eine Verbesserung. Während mit dem bisherigen System bei keiner Aufgabe alle Personen das Zeitlimit einhalten konnten, so konnten bei sechs von neun Aufgaben alle Teilnehmer und Teilnehmerinnen im neuen System die Aufgabe innerhalb der Zeit abschließen. Bei den restlichen drei Aufgaben war es mindestens die Hälfte der Personen.

Sowohl der Vergleich der durchschnittlichen Zeiten als auch der Abschlüsse innerhalb der Zeitlimits legen dar, dass im neuen System eine Effizienzsteigerung erreicht werden konnte. Zusätzlich ist zu berücksichtigen, dass die Teilnehmer und Teilnehmerinnen bei der Durchführung der Tests das alte System bereits länger kannten als das Neue. Die Erlernbarkeit des neuen Systems wurde im Rahmen dieser Evaluierungen nicht gemessen. Bei einer kurzen Zeitspanne, wie bei der täglichen Nutzung, ist es aber das Ziel der meisten Benutzer und Benutzerinnen, so schnell wie möglich eine höhere Effizienz zu erreichen. [12, p. 93] Deshalb ist damit zu rechnen, dass sich die Zeiten noch weiter steigern können.

Ein weiterer interessanter Aspekt ist die finanzielle Seite dieser Usability Verbesserungen. Neben den bereits beschriebenen Usability Metriken nennen Tom Tullis und Bill Albert in „Measuring the User Experience“ auch den Return on Investment oder kurz ROI. Die Idee eines Usability ROI ist es, die finanziellen Vorteile von Usability Optimierungen auszudrücken. [12, p. 233] Mit der gemessenen Effizienzsteigerung durch das neue System ist eine Zeitersparnis verbunden, die zu Kosteneinsparungen führen kann. Bei täglicher Durchführung der Arbeit für sechs Kundensysteme kommt man auf eine jährliche Zeitersparnis von über 600 Stunden und umgerechnet eine jährliche Kostenersparnis von rund € 8.800,-.¹ Für die Implementierung und Durchführung der Evaluierung wurden ungefähr 350 Stunden benötigt, was umgerechnet circa € 6.200,-² entspricht. Wenn man also die Kostenersparnis den entstandenen Kosten gegenüberstellt so zeigt, sich, dass sich der Aufwand bereits innerhalb eines Jahres gerechnet hat. In diesem Fall wurde jedoch nur die Effizienzsteigerung für die ROI Berechnung berücksichtigt. Man kann aber davon ausgehen, dass die Verbesserung der Effektivität den ROI noch weiter verbessern würde, jedoch ist es schwieriger diese Optimierungen in direkten Kostenersparnissen auszudrücken.

Wie schon in der ersten Evaluierung wurden die Teilnehmer und Teilnehmerinnen erneut gebeten den SUS Fragebogen auszufüllen. Im zweiten Durchlauf sollten sie die Fragen in Bezug auf das neue System beantworten. Die Auswertung dieser Ergebnisse gibt einen durchschnittlichen SUS Score von 91. Im Vergleich zur durchschnittlichen Bewertung von 20 für das bisherige System befindet sich dieses Ergebnis also am anderen Ende des Spektrums. Jeff Sauro schlägt außerdem ein System vor, um die SUS Scores in Noten

¹ Ausgehend vom Grundgehalt laut IT Kollektivvertrag 2017 für Einstufung AT & Regelstufe

² Ausgehend vom Grundgehalt laut IT Kollektivvertrag 2017 für Einstufung ST1 & Regelstufe

auszudrücken, wobei die Bestnote A+ ist und F die schlechteste Note. [18, p. 33 f] Wenn man die für beide Systeme gemessenen SUS Scores auf diese Noteneinteilung überträgt, so würde das alte System die Note F erhalten, während das neue System mit A benotet wird. Auch diese Interpretation zeigt somit, dass eine Verbesserung der Zufriedenheit der Benutzer und Benutzerinnen erzielt werden konnte.

Die beiden Auswertungen der Daten aus den durchgeführten Usability Tests schließen darauf, dass beide zu Beginn gestellten Forschungsfragen positiv beantwortet werden können. Die erste Frage handelt davon, ob klassische Usability Methoden wie Usability Tests zur Evaluierung von komplexen Systemen zur Informationsanalyse angewendet werden können und ob Usability Metriken gemessen werden können. Zusätzlich wurde die Frage gestellt, ob die Ergebnisse der Evaluierungen vergleichbar sind und darauf basierte Maßnahmen zu Usability Verbesserungen führen können. Mit der ersten Evaluierung konnten erste Daten ermittelt und Usability Probleme festgestellt werden. Anschließend wurde ein neues System mit Usability Verbesserungen, aufbauend auf die Ergebnisse des Usability Tests des alten Systems, implementiert. Um den Fragen entsprechend sicherzustellen, dass mit den eingesetzten Methoden sinnvolle und vergleichbare Ergebnisse erzielt werden können, muss in diesem Fall daher eine positive Verbesserung der Usability gemessen werden. Diese Verbesserung konnte mit der Auswertung der zweiten Evaluierung bestätigt werden. Es ist daher möglich, bei komplexen Analysesystemen Usability Tests so zu gestalten und durchzuführen, dass Daten auf verlässliche und reproduzierbare Weise erzeugt und miteinander verglichen werden können. So konnten die richtigen Usability Probleme entdeckt werden und mit den daraus entwickelten Verbesserungsmaßnahmen die gewünschten Effekte erzielt werden.

Zusätzlich zeigte sich im Rahmen dieser Arbeit, dass besonders bei komplexen Systemen die drei klassischen Usability Kriterien Effektivität, Effizienz und Zufriedenheit keineswegs an Relevanz verloren haben. Heutzutage spielen bei mobilen Applikationen oder anderen modernen Anwendungen für Endkunden und -kundinnen andere Usability Aspekte wie die Einfachheit der Benutzung oder Gamification eine immer größere Rolle. Bei komplexen Analysesystemen ist dies jedoch kaum der Fall. Hier zählt vor der einfachen Benutzung, dass die Anwendung nützlich ist und ihren Zweck erfüllt. Für die Usability dieser Art von Systemen ist es daher entscheidend, dass vor allem die Effektivität und Effizienz der Interaktion mit dem System evaluiert wird.

Dennoch gibt es auch besondere Herausforderungen in der Anwendung von Usability Tests bei komplexen Systeme zur Informationsanalyse. Wie in Kapitel 2.6 beschrieben, dienen diese Systeme der Lösung von komplexen Problemen. Michael J. Albers definiert diese komplexen oder unstrukturierten Probleme damit, dass es keinen idealen Lösungsweg gibt, der als fixe Schrittfolge in einem System abgebildet werden kann. Jede Person wählt eine andere Vorgehensweise und wählt diese abhängig von den vorhandenen Informationen. [20,

p. 267] Da diese Arbeit also stark mit den Informationen verbunden ist, können die Problemstellungen und Vorgehensweisen oft nur mit einem entsprechenden Domainwissen nachvollzogen werden.

Wie Joseph Dumas und Janice Redish als Charakteristik eines Usability Test definieren, sollen die Teilnehmer und Teilnehmerinnen reale Aufgaben durchführen. [5, p. 22] Im Bereich von komplexen Analysesystemen ist dabei aber ein gewisses Domainwissen notwendig, um diese Aufgaben definieren zu können. Auch Parmit K. Chilana, Jacob O Wobbrock und Andrew J. Ko sind in „Understanding Usability Practices in Complex Domains“ bereits zu der Erkenntnis gekommen, dass dieser Aspekt eine Hürde bei der Usability Evaluierung sein kann, da es oft schwierig und aufwändig ist sich mit der Domain vertraut zu machen. Die Unterstützung durch Domainexperten kann nötig und auf jeden Fall sehr hilfreich sein. [30]

Das Domainwissen kann ebenso bei der Auswertung der Ergebnisse dieser Aufgaben fehlen. Grundsätzlich sollten bereits im Vorhinein für alle Aufgaben klar definiert sein, was das zu erwartende korrekte Ergebnis ist. [10, p. 133 f] Auf Grund der Komplexität können aber dennoch bei der Auswertung Fragen und Unklarheiten auftauchen. Um keine Fehler bei der Auswertung zu machen bzw. die richtigen Schlüsse zu ziehen, muss man unter Umständen erneut einen Domainexperten zu Rate ziehen.

Auf Grund des speziellen Einsatzgebietes des evaluierten Systems konnten nur sieben Personen für die Usability Tests eingebunden werden. Für dieses Aufgabengebiet sind bestimmte Vorkenntnisse notwendig, die nur bei bestehenden Benutzern und Benutzerinnen vorhanden sind. Allerdings konnte in diesem Fall mit den sieben Teilnehmern und Teilnehmerinnen die gesamte Zielgruppe abgedeckt werden. Weitere Usability Evaluierungen für komplexe Systeme mit einer größeren Zielgruppe und daher einer größeren Teilnehmeranzahl könnten aber zusätzliche interessante Erkenntnisse bringen.

Komplexe Systeme zur Informationsanalyse befinden sich nicht nur im Einsatzgebiet der Applikationsüberwachung. Zur weiteren Behandlung dieses Themas wäre auch die Anwendung der beschriebenen Usability Evaluierungsmethoden auf andere Arten von Analysesysteme interessant. Ein anderes Einsatzgebiet zeigt beispielsweise Jean Scholtz in „Progress and Challenges in Evaluating Tools for Sensemaking“. [31] Mit „Sensemaking“ sind Aktivitäten gemeint, bei denen unterschiedliche Informationen analysiert und eine Bedeutung bzw. ein Sinn daraus gewonnen werden soll.

Im Rahmen dieser Arbeit wurde behandelt, ob klassische Usability Tests auf komplexe Systeme angewendet werden können. Wie Janice Redish in „Expanding Usability Testing to Evaluate Complex Systems“ beschreibt können jedoch auch neue Testverfahren nützlich sein, um die Usability von komplexen Systemen zu evaluieren. [21]

8 Zusammenfassung

Diese Arbeit beschäftigt sich mit der Frage, ob klassische Usability Methoden genutzt werden können, um die Usability von komplexen Systemen zur Informationsanalyse zu evaluieren. Eine gute Usability bedeutet, dass Benutzer und Benutzerinnen eines Produktes, dieses schnell und einfach nutzen können und mit der Nutzung zufrieden sind. Besonders wichtig bei der Usability von komplexen Systemen sind dabei die Usability Kriterien Effektivität und Effizienz.

Mit Hilfe von Usability Tests wird gemessen, wie gut die Usability eines Produktes ist. Dabei werden die Benutzer und Benutzerinnen direkt beim Umgang mit dem Produkt beobachtet. Besonders wichtig für Usability Tests ist, dass die Teilnehmer und Teilnehmerinnen reale Benutzer und Benutzerinnen repräsentieren und im Rahmen der Tests reale Aufgaben durchführen. Ebenso können während oder nach einem Test Fragebögen zum Einsatz kommen, um zusätzlich die Meinung und Wahrnehmung der Teilnehmer und Teilnehmerinnen zu ermitteln.

Aus den gesammelten Daten können anschließend unterschiedliche Usability Metriken ausgewertet werden. Zu den wichtigsten Metriken für die Evaluierung von Effektivität und Effizienz gehören Task Success und Time on Task. Die Metrik Task Success drückt aus, wie viele Personen die Aufgaben erfolgreich abschließen konnten, während Time on Task angibt, in welcher Zeit der Abschluss einer Aufgabe möglich war.

Systemen zur Informationsanalyse unterscheiden sich aber in ihrer Komplexität von anderen Systemen, die üblicherweise als Beispiel im Zusammenhang mit Usability Tests und Metriken gebracht werden. Diese Analysesysteme müssen Benutzer und Benutzerinnen bei der Lösung von komplexen Problemen unterstützen. Diese Probleme zeichnen sich dadurch aus, dass es keine eindeutige optimale Lösung gibt. Daher kann auch kein schrittweiser Prozess in einem Analysesystem definiert werden, der die Benutzer und Benutzerinnen zur Lösung führen soll. Stattdessen müssen unterschiedliche Werkzeuge und Aktionsmöglichkeiten geschaffen und dem Benutzer oder der Benutzerinnen die Kontrolle überlassen werden, wie er mit den Informationen umgehen möchte.

Um festzustellen, ob die klassischen Methoden der Usability Evaluierung auch für diese Art von System geeignet sind, wurde die Usability eines Analysesystems für die Applikationsüberwachung evaluiert. Dazu wurden moderierte Usability Test durchgeführt und der SUS Fragebogen eingesetzt. Zuerst wurde das bestehende System mit sieben Personen evaluiert. Basierend auf den Ergebnissen dieser Evaluierung wurde ein neues System entwickelt und dessen Usability erneut durch Usability Tests mit den selben Benutzern und Benutzerinnen evaluiert. Aus den Daten beider Evaluierungen wurden die Metriken Task Success, Time on Task und der SUS Score berechnet und verglichen.

Im Vergleich zeigte sich, dass die Usability im neuen System verbessert wurde. Während im ersten Test nur bei zwei von neun Aufgaben alle Personen die Aufgabe erfolgreich abschließen konnten, waren es im Test des neuen Systems sieben Aufgaben. Die durchschnittliche Zeit für alle neuen Aufgaben eines Tests konnte von 41 Minuten auf 16 Minuten reduziert werden. Dieses Bild der Verbesserung zeigt sich auch beim SUS Score, wo der Durchschnitt von 20 auf 91 verbessert wurde.

Als Herausforderung beim Einsatz von Usability Tests zur Evaluierung von Analysesystemen zeigte sich, dass ein starkes Domainwissen bei der Planung und Auswertung von den Tests notwendig sein kann.

Dennoch konnte mit der Durchführung dieser Usability Tests sowie der anschließenden Auswertung und dem Vergleich der Ergebnisse die Forschungsfragen dieser Arbeit bestätigt werden. Es konnte dargelegt werden, dass die Usability von komplexen Systemen zur Informationsanalyse mit klassischen Methoden evaluiert werden kann, die Ergebnisse vergleichbar sind und Maßnahmen für Verbesserungen aus den Ergebnissen abgeleitet werden können.

Literaturverzeichnis

- [1] R. Molich, „CUE (Comparative usability evaluation),“ DialogDesign, [Online]. Available: <http://www.dialogdesign.dk/CUE.html>. [Zugriff am 8 Jänner 2017].
- [2] B. Mirel, Interaction Design for Complex Problem Solving: Developing Useful and Usable Software, San Francisco: Morgan Kaufmann Publishers Inc., 2003.
- [3] Österreichisches Normungsinstitut, *ÖNORM EN ISO 9241-210: Ergonomie der Mensch-System-Interaktion - Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme*, 2010.
- [4] Österreichisches Normungsinstitute, *ÖNORM EN ISO 9241-11: Ergonomie der Mensch-System-Interaktion - Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte*, 2016.
- [5] J. S. Dumas und J. C. Redish, A Practical Guide to Usability Testing, Portland, Oregon: Intellect Books, 1999.
- [6] J. Nielsen, „Usability 101: Introduction to Usability,“ Nielsen Norman Group, 4 Jänner 2012. [Online]. Available: <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>. [Zugriff am 6 Juni 2017].
- [7] J. Rubin und D. Chisnell, Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests, 2nd Hrsg., Indianapolis: Wiley Publishing, Inc., 2008.
- [8] J. Nielsen, „How to Conduct a Heuristic Evaluation,“ Nielsen Norman Group, 1 Jänner 1995. [Online]. Available: <https://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/>. [Zugriff am 3 September 2017].
- [9] J. Nielsen, „10 Usability Heuristics for User Interface Design,“ Nielsen Norman Group, 1 Jänner 1995. [Online]. Available: <https://www.nngroup.com/articles/ten-usability-heuristics/>. [Zugriff am 7 Juni 2017].
- [10] C. M. Barnum, Usability Testing Essentials, Burlington: Morgan Kaufmann, 2011.
- [11] S. Krug, Rocket Surgery Made Easy: The Do-It-Yourself Guide to Finding and Fixing Usability Problems, Berkeley: New Riders, 2010.
- [12] T. Tullis und B. Albert, Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics, San Francisco, CA: Morgan Kaufmann, 2013.

- [13] J. Nielsen und T. K. Landauer, „A mathematical model of the finding of usability problems,“ in *Proceedings of ACM INTERCHI'93 Conference*, Amsterdam, 1993.
- [14] J. Sauro und J. R. Lewis, *Quantifying the User Experience: Practical Statistics for User Research*, Waltham: Morgan Kaufmann, 2012.
- [15] J. Brooke, „SUS - A quick and dirty usability scale,“ *Usability evaluation in industry*, Bd. 189, Nr. 194, pp. 4-7, 1996.
- [16] J. Sauro und J. R. Lewis, „When Designing Usability Questionnaires, Does It Hurt to Be Positive?,“ in *Proceedings of the Conference in Human Factors in Computing Systems*, Vancouver, 2011.
- [17] J. Brooke , „SUS: A Retrospective,“ *Journal of Usability Studies*, Bd. 8, Nr. 2, pp. 29-40, 2013.
- [18] J. Sauro, *A Practical Guide to the System Usability Scale*, Denver: Measuring Usability LLC, 2011.
- [19] K. Lohmann, „System Usability Scale (SUS) – An Improved German Translation of the Questionnaire,“ CoreMedia, 18 September 2013. [Online]. Available: <https://minds.coremedia.com/2013/09/18/sus-scale-an-improved-german-translation-questionnaire/>. [Zugriff am 5 September 2017].
- [20] M. J. Albers, „Complex Problem Solving and Content Analysis,“ in *Content and Complexity*, Mahwah, New Jersey, Lawrence Erlbaum Associates, 2003, pp. 263-284.
- [21] J. G. Redish, „Expanding usability testing to evaluate complex systems,“ *Journal of Usability Studies*, Bd. 2, Nr. 3, pp. 102-111, 2007.
- [22] B. Mirel, „Dynamic Usability: Designing Usefulness Into Systems for Complex Tasks,“ in *Content and Complexity*, Mahwah, NJ, Lawrence Erlbaum Associates, 2003, pp. 233-262.
- [23] Edentity Software Solutions GmbH, „Dealer Benchmarking System,“ Edentity Software Solutions GmbH, 2017. [Online]. Available: <http://edentity.at/en/solution/dealer-benchmarking-system/>. [Zugriff am 17 September 2017].
- [24] TechSmith, „Morae Funktionen,“ TechSmith, 2017. [Online]. Available: <https://www.techsmith.de/morae-funktionen.html#Recorder>. [Zugriff am 11 Juli 2017].
- [25] B. Wagner, T. Pratt, L. Latham und M. Wenzel, „Introduction to the C# Language and the .NET Framework,“ Microsoft, 27 Mai 2017. [Online]. Available:

<https://docs.microsoft.com/en-us/dotnet/csharp/getting-started/introduction-to-the-csharp-language-and-the-net-framework>. [Zugriff am 12 Juli 2017].

- [26] Microsoft, „ASP.NET MVC,“ Microsoft, 2017. [Online]. Available: <https://www.asp.net/mvc>. [Zugriff am 12 Juli 2017].
- [27] Microsoft, „ASP.NET Web API,“ Microsoft, 2017. [Online]. Available: <https://www.asp.net/web-api>. [Zugriff am 12 Juli 2017].
- [28] Google, „AngularJS,“ Google, 2017. [Online]. Available: <https://angularjs.org/>. [Zugriff am 12 Juli 2017].
- [29] M. Wasson, „Integrated Windows Authentication,“ Microsoft, 18 Dezember 2012. [Online]. Available: <https://docs.microsoft.com/en-us/aspnet/web-api/overview/security/integrated-windows-authentication>. [Zugriff am 7 August 2017].
- [30] P. K. Chilana, J. O. Wobbrock und A. J. Ko, „Understanding Usability Practices in Complex Domains,“ in *Conference on Human Factors in Computing Systems - Proceedings*, Atlanta, 2010.
- [31] J. Scholtz, „Progress and Challenges in Evaluating Tools for Sensemaking,“ in *ACM Computer Human Information (CHI) conference Workshop on Sensemaking*, Florence, 2008.

Abbildungsverzeichnis

Abbildung 3.1: Applikation zum Versand der Datenemail	44
Abbildung 3.2: Beispiel einer generierten Email mit Daten	45
Abbildung 3.3: Email mit kompletten Fehlermeldungen des vorangehenden Tages	46
Abbildung 3.4: Webseite mit kompletten Fehlermeldungen	46
Abbildung 3.5: Beispielhafte Abfrage der Useraktivität in der MS Access Anwendung	47
Abbildung 4.1: Task Success pro Aufgabe	56
Abbildung 4.2: Time on Task als Mittelwert und Median pro Aufgabe.....	58
Abbildung 4.3: Time on Task pro TeilnehmerIn für Aufgabe 5 & 6.....	58
Abbildung 4.4: Anzahl der TeilnehmerInnen pro Aufgabe mit Abschluss innerhalb des Zeitlimits.....	60
Abbildung 4.5: SUS Score Verteilung mit dem durchschnittlichen Ergebnis aus der ersten Evaluierung hervorgehoben	62
Abbildung 5.1: Auszug aus einer XML Konfigurationsdatei	68
Abbildung 5.2: Startseite des neuen Systems	69
Abbildung 5.3: Laden der Daten im neuen System.....	70
Abbildung 5.4: Fehleransicht im neuen System.....	71
Abbildung 5.5: Detailansicht eines Fehlers im neuen System	72
Abbildung 5.6: Filter- und Darstellungsmöglichkeiten im neuen System.....	73
Abbildung 5.7: Suche in verknüpften Daten im neuen System	74
Abbildung 5.8: Trendspalte und Aggregation von numerischen Werten im neuen System	75
Abbildung 5.9: Kommentieren von Problemen für die Zusammenfassung im neuen System	76
Abbildung 5.10: Zusammenfassung im neuen System.....	76
Abbildung 5.11: Senden der Email im neuen System.....	77
Abbildung 6.1: Vergleich von Task Success zwischen altem und neuem System	79
Abbildung 6.2: Vergleich der durchschnittlichen Time on Task (Mittelwert) zwischen altem und neuem System	80
Abbildung 6.3: Vergleich der Anzahl an TeilnehmerInnen mit Abschluss im Zeitlimit zwischen altem und neuem System.....	81
Abbildung 6.4: SUS Score Verteilung mit den durchschnittlichen Ergebnissen für das alte und neue System hervorgehoben	82

Tabellenverzeichnis

Tabelle 4.1: Erfüllungsgrad jeder Aufgabe pro TeilnehmerIn55

Anhang A: SUS Fragebögen

Englische Fragen

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

Deutsche Fragen

1. Ich denke, dass ich dieses System gerne regelmäßig nutzen würde.
2. Ich fand das System unnötig komplex.
3. Ich denke, das System war leicht zu benutzen.
4. Ich denke, ich würde die Unterstützung einer fachkundigen Person benötigen, um das System benutzen zu können.
5. Ich fand, die verschiedenen Funktionen des Systems waren gut integriert.
6. Ich halte das System für zu inkonsistent.
7. Ich glaube, dass die meisten Menschen sehr schnell lernen würden, mit dem System umzugehen.
8. Ich fand das System sehr umständlich zu benutzen.
9. Ich fühlte mich bei der Nutzung des Systems sehr sicher.
10. Ich musste viele Dinge lernen, bevor ich mit dem System arbeiten konnte. [19]

Anhang B: Aufgabenstellungen

1. Evaluierung

Monitoring Mail/Template

Bitte rufen Sie zur Prüfung die Monitoring Daten für das Kundensystem auf.

Senden Sie dazu die Monitoring Email an die Adresse johanna.strutzenberger@edentity.at.

Sie finden die E-Mail anschließend im Posteingangsortner „00_MonitoringTool_Tests“.

Jobs

Bitte überprüfen Sie den Status der durchgeführten Jobs. Heben Sie fehlerfrei gelaufene und fehlgeschlagene Jobs unterschiedlich hervor und kommentieren Sie mögliche fehlgeschlagene Jobs mit ihrer Ursache in der Zusammenfassung.

Errors 1

Überprüfen Sie, ob es Datenbank-Blockings im System gab und erfassen Sie in der Zusammenfassung welche Tabellen mit welcher Blocking-Dauer betroffen waren.

Errors 2

Überprüfen Sie die Häufigkeit der aufgetretenen Fehler. Die 3 häufigsten Fehler, bei denen die hohe Anzahl nicht bereits als unkritisch eingestuft wurde, sollten in der Zusammenfassung erfasst werden.

Error 3

Überprüfen Sie bitte, ob neue bzw. unbekannte Fehler aufgetreten sind oder andere Fehler als kritisch einzustufen sind. Erfassen Sie diese Fehler in der Zusammenfassung.

(Für die Bestimmungen der neuen/unbekannten Fehler ist es ausreichend, wenn diese in den letzten 3 Tagen nicht aufgetreten sind.)

Selenium Tests

Bitte überprüfen Sie den Status der Selenium Tests für das Kundensystem. Erfassen Sie das Ergebnis der Testläufe (inkl. Hervorhebung des Status) und die Durchlaufzeiten der Tests. Falls es fehlgeschlagene Tests gab oder sich die Durchlaufzeiten im Vergleich zum Vormonat stark verändert haben (>20%), erfassen Sie diese Info in der Zusammenfassung.

Disk Space

Bitte überprüfen Sie den freien Speicherplatz aller Server und heben Sie Server, deren Speicherplatz in Ordnung ist, und Server mit Speicherplatz-Problemen unterschiedlich hervor. Ergänzen Sie gegebenenfalls Probleme in der Zusammenfassung.

Schnittstellen

Überprüfen Sie bitte den Status der Schnittstellen und erfassen Sie das Ergebnis entsprechend.

DBS User Activity

Überprüfen Sie bitte die User-Aktivität im Kundensystem der letzten 24 Stunden. Erfassen Sie das Ergebnis und liefern Sie in der Zusammenfassung eine Bewertung, ob die Aktivität im Vergleich zum 16.11.2016 hoch, mittel oder niedrig ist.

Stufen:

Hoch → mehr als 70%

Mittel → 30 bis 70%

Niedrig → weniger als 30%

2. Evaluierung

Monitoring Mail/Template

Bitte rufen Sie zur Prüfung die Monitoring Daten für das Kundensystem auf und stellen Sie sicher, dass alle Daten geladen sind.

Jobs

Bitte überprüfen Sie den Status der durchgeführten Jobs. Stellen Sie sicher, dass fehlerfrei gelaufene und fehlgeschlagene Jobs unterschiedlich hervorgehoben sind, und kommentieren Sie mögliche fehlgeschlagene Jobs mit ihrer Ursache in der Zusammenfassung.

Errors 1

Überprüfen Sie, ob es Datenbank-Blockings im System gab und erfassen Sie in der Zusammenfassung, welche Tabellen mit welcher Blocking-Dauer betroffen waren.

Errors 2

Überprüfen Sie die Häufigkeit der aufgetretenen Fehler. Die 3 häufigsten Fehler, bei denen die hohe Anzahl nicht bereits als unkritisch eingestuft wurde, sollten in der Zusammenfassung erfasst werden.

Error 3

Überprüfen Sie bitte, ob neue bzw. unbekannte Fehler aufgetreten sind oder andere Fehler als kritisch einzustufen sind. Erfassen Sie diese Fehler in der Zusammenfassung.

Selenium Tests

Bitte überprüfen Sie den Status der Selenium Tests für das Kundensystem. Falls es fehlgeschlagene Tests gab oder sich die Durchlaufzeiten im Vergleich zum Vormonat stark verändert haben (>20%), erfassen Sie diese Info in der Zusammenfassung.

Disk Space

Bitte überprüfen Sie den freien Speicherplatz aller Server und stellen Sie sicher das Server, deren Speicherplatz in Ordnung ist, und Server mit Speicherplatz-Problemen unterschiedlich hervorgehoben sind. Ergänzen Sie gegebenenfalls Probleme in der Zusammenfassung.

Schnittstellen

Überprüfen Sie bitte den Status der Schnittstellen und erfassen Sie das Ergebnis entsprechend.

DBS User Activity

Überprüfen Sie bitte die User-Aktivität im Kundensystem der letzten 24 Stunden und liefern Sie in der Zusammenfassung eine Bewertung, ob die Aktivität im Vergleich zum 16.11.2016 hoch, mittel oder niedrig ist.

Stufen:

Hoch → mehr als 70%

Mittel → 30 bis 70%

Niedrig → weniger als 30%