Gottfried Wilhelm Leibniz Universität Hannover
Fakultät für Elektrotechnik und Informatik
Institut für Praktische Informatik
Fachgebiet Software Engineering

Auswirkung von Sicherheitserklärungen auf die Verständlichkeit und die Nutzbarkeit von Registrierungsdialogen

Effects of Security Explanations on the Understandability and Usability of Registration Dialogues

Masterarbeit

im Studiengang Informatik

von

Chris Noah Burmeister

Prüfer: Prof. Dr. Kurt Schneider
Zweitprüfer: Prof. Dr. Markus Dürmuth
Betreuer: M. Sc. Jakob Droste

Hannover, 25. April 2024

Erklärung der Selbstständigkeit

Hiermit versichere ich, dass ich die vorliegende Masterarbeit selbstständig und ohne fremde Hilfe verfasst und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel verwendet habe. Die Arbeit hat in gleicher oder ähnlicher Form noch keinem anderen Prüfungsamt vorgelegen.

Hannover, den 25. April 2024

Chris Noah Burmeister

Zusammenfassung

Passwörter sind heutzutage die am häufigsten verwendete Authentifizierungsmethode und werden dies voraussichtlich auch noch eine Weile bleiben. Wie durch verschiedene Arbeiten gezeigt wurde, haben diese jedoch einige Schwächen. Unter anderem aus diesem Grund gibt es viele Forschungsarbeiten, die sich mit der Verbesserung von Passwörtern auseinandersetzen. Eine bereits etablierte Herangehensweise ist die Verwendung von Passwortstärkeanzeigen. Diese zeigen Nutzern die Stärke ihres momentan eingegebenen Passworts bei der Erstellung dessen an und versuchen so, dafür zu sorgen, dass die Nutzer stärkere Passwörter wählen. Ein in diesem Zusammenhang noch wenig untersuchtes Thema ist die Erklärbarkeit. Diese beschäftigt sich damit, einem Nutzer ein System verständlich zu machen. Passwortstärkeanzeigen um Erklärbarkeit zu erweitern könnte positive Auswirkungen auf das System bzw. die Nutzer haben.

Diese Arbeit untersucht die Auswirkungen, welche Sicherheitserklärungen auf die Verständlichkeit und Nutzbarkeit von Passwortstärkeanzeigen in Registrierungsdialogen haben. Zu diesem Zweck wurde zuerst eine Literaturrecherche zu den dafür relevanten Themen durchgeführt. Mit dem daraus erhaltenen Wissen wurde ein Konzept für erklärbare Registrierungsdialoge erarbeitet und dieses in Form eines Software-Prototyps umgesetzt. Um die Auswirkungen der in diesem enthaltenen Erklärungen zu untersuchen, wurde daraufhin eine Studie mit 28 Teilnehmern durchgeführt. Dabei wurden besonders die Auswirkungen auf Vertrauenswürdigkeit, Verständlichkeit und Nutzbarkeit des Systems sowie die Auswirkungen auf die Passwortstärke der von den Nutzern gewählten Passwörter untersucht. Die Ergebnisse zeigen, dass Erklärungen in Registrierungsdialogen sich positiv auf die Verständlichkeit des Systems und die Passwortstärke der gewählten Passwörter bei sicherheitskritischen Systemen auswirken, während die Nutzbarkeit nur geringfügig verschlechtert wird. Daraus folgt, dass Erklärungen eine sinnvolle Erweiterung für Registrierungsdialoge sein können.

Abstract

Effects of Security Explanations on the Understandability and Usability of Registration Dialogues

Passwords are the most commonly used authentication method today and are likely to remain so for some time. However, as various studies have shown, they do have some weaknesses. This is one of the reasons why there is a lot of research into improving passwords. One already established approach is the use of password strength meters. These meters show users the strength of their currently entered password when it is created and thus try to ensure that users choose stronger passwords. A topic that has not yet been investigated much in this context is explainability. This deals with making a system understandable to a user. Adding explainability to password strength meters could have a positive impact on the system and users.

This thesis examines the effects that security declarations have on the understandability and usability of password strength meters in registration dialogs. To this end, a literature review was first conducted on the relevant topics. With the knowledge gained from this, a concept for explainable registration dialogs was developed and implemented in the form of a software prototype. A study with 28 participants was then carried out to investigate the effects of the explanations contained in the prototype. In particular, the effects on trustworthiness, understandability and usability of the system as well as the effects on the password strength of the passwords chosen by the users were examined. The results show that explanations in registration dialogs have a positive effect on the comprehensibility of the system and the password strength of the chosen passwords for high-security systems, while usability is only slightly impaired. It follows that explanations can be a useful extension for registration dialogs.

Inhaltsverzeichnis

Kapitel 1

Einleitung

1.1 Motivation

In der heutigen Zeit haben die meisten Menschen viele verschiedene Konten, die es zu sichern gilt. Dies können eher unwichtige Konten, wie die für den Gratis-Service einer Website sein, oder wichtige, wie für das Onlinebanking. Für die meisten Arten von Konten werden heutzutage Passwörter verwendet ([33], [13], [10]). Trotz der weiten Verbreitung sind Passwörter als Authentifizierungsmethode, dennoch fehlerbehaftet und führen somit zu einigen Problemen. ([33], [13]) Aus diesem Grund gibt es viel Forschung zu verschiedenen Aspekten von Passwörtern, um diese zu verbessern und sicherer zu machen. Zwar gibt es auch Forschung zu Mechanismen, welche Passwörter ersetzen könnten, jedoch ist es wahrscheinlich, dass Passwörter auch in absehbarer Zukunft noch relevant sein werden ([33], [13], [10]). So ist also die Forschung zur Verbesserung von Passwörtern für viele Menschen von Bedeutung, ob es nun darum geht, den Nutzer dazu zu bringen, ein sichereres Passwort zu wählen, oder ihm eine angenehme Nutzererfahrung zu bereiten.

1.2 Problemstellung

Passwörter wurden aufgrund der weiten Verbreitung bereits viel erforscht, jedoch gibt es noch einige offene Fragen in diesem Forschungsbereich. Ein solcher, noch eher unerforschter Aspekt ist die Erklärbarkeit von Passwörtern, denn obwohl Passwörter so häufig genutzt werden, wissen viele Nutzer eher wenig über sie und verstehen ihre Funktionsweise nicht vollständig. Dieses fehlende Verständnis könnte unter anderem ein Grund sein, weshalb von Nutzern gewählte Passwörter häufig eher unsicher sind. Ein möglicher praktischer Lösungsansatz, welcher auch ausgiebig erforscht wurde, sind Passwortstärkeanzeigen. Diese sorgen bereits durch das Anzeigen der sogenannten Passwortstärke für eine Verhaltensänderung bei der Passwortwahl von Nutzern ([33], [13]). Ziel dieser Arbeit ist es, die

bereits erforschten Registrierungsdialoge mitsamt Passwortstärkeanzeigen, um Erklärungen zu erweitern und die Auswirkungen der Erklärungen auf die Registrierungsdialoge zu untersuchen. Im Fokus sind dabei die Aspekte Vertrauen, Verständlichkeit, Passwortstärke und Nutzbarkeit.

1.3 Lösungsansatz

Um Passwortstärkeanzeigen im Hinblick auf Erklärbarkeit zu untersuchen, werden in dieser Arbeit zuerst thematisch verwandte und relevante Arbeiten in einer systematischen Literaturrecherche (SLR) untersucht. Dabei wird insbesondere auf Arbeiten zu Passwortstärkeanzeigen und Arbeiten zu Erklärbarkeit geachtet. Danach wird auf Grundlage der in der SLR erlangten Erkenntnisse ein Konzept für erklärbare Registrierungsdialoge entwickelt und dieses mit dem Ziel einer empirischen Studie prototypisch umgesetzt. Diese empirische Studie umfasst einen Nutzbarkeit-Test sowie eine Befragung in zwei Gruppen mit je 14 Teilnehmenden. Die Ergebnisse der Studie werden mithilfe der gängigen statistischen Tests ausgewertet und eingeordnet.

1.4 Wissenschaftlicher Beitrag

Während es sehr viel Forschung zu Passwortstärkeanzeigen gibt, gibt es noch zu wenig Forschung zu diesen in Kombination mit Erklärungen. Generell gibt es noch relativ wenig Forschung zur Erklärbarkeit in Verbindung mit Sicherheit. Dieses Feld, welches als erklärbare Sicherheit (engl. Explainable Security) bezeichnet werden kann, befindet sich noch in einer frühen Phase und hat noch viele Aspekte, die erforscht werden müssen. So möchte diese Arbeit in diesem Bereich weiterhelfen und die Auswirkungen von Erklärungen in Registrierungsdialogen mit Passwortstärkeanzeigen genauer untersuchen, besonders im Hinblick auf Vertrauen, Verständlichkeit, Passwortstärke und Nutzbarkeit.

1.5 Struktur der Arbeit

Nach diesem einführenden Kapitel sind in Kapitel 2 alle für das Verständnis dieser Arbeit notwendigen Grundlagen zu finden. So wird auf andere Arbeiten mit wichtigen Erkenntnissen eingegangen und es werden für die Thematik dieser Arbeit relevante Begriffe geklärt. Dabei wird auf die Themen Nutzbarkeit (engl. Usability), Erklärbarkeit (engl. Explainability) und Passwörter eingegangen. Im darauf folgenden Kapitel 3 sind Informationen zu der für diese Arbeit durchgeführten Literaturrecherche zu finden. Danach kann in Kapitel 4 das entwickelte Konzept zu erklärbaren Registrierungen nachvollzogen werden, während in Kapitel 5 die zur Überprüfung dieses Konzepts durchgeführte Studie einschließlich des dazu entwickelten Prototyps

betrachtet werden kann. In Kapitel 6 werden die in der Studie gesammelten Daten präsentiert und in ihren Kontext eingeordnet, sodass ihre Bedeutung dann in Kapitel 7 diskutiert werden kann. Letztlich wird in Kapitel 8 noch auf verwandte Arbeiten eingegangen, bevor in Kapitel 9 diese Arbeit abschließend zusammengefasst wird.

Kapitel 2

Grundlagen

In diesem Kapitel werden die Grundlagen für diese Arbeit beschrieben. Dabei wird besonders auf Nutzbarkeit, Erklärbarkeit und Passwörter eingegangen.

2.1 Nutzbarkeit

Die Nutzbarkeit (engl. Usability) ist eine Eigenschaft bzw. ein Maß eines Systems, welches sich mit den Möglichkeiten und der Erfahrung des Nutzers bei der Nutzung dieses Systems beschäftigt. Formal definiere ich die Nutzbarkeit, analog zu der Definition in der ISO 9241-11:2018 ([12]), wie folgt:

Definition 1 *Nutzbarkeit ist das Ausmaß, zu dem ein System, Produkt oder Service von einem bestimmten Nutzer genutzt werden kann, um bestimmte Ziele effektiv, effizient und befriedigend in einem bestimmten Kontext zu erreichen.*

Die Nutzbarkeit ist ein sehr großes Forschungsfeld, welches viele verschiedene Teilaspekte hat. Deswegen kann es schwierig sein, die Nutzbarkeit eines Systems genau zu erheben, ohne eine sehr umfangreiche Befragung durchzuführen. Eine Möglichkeit, die Nutzbarkeit nur mit wenigen Fragen durchzuführen, schafft dabei K. Finstad in seiner Arbeit ([9]), mit der „Usability Metric for User Experience“ (UMUX). Dies ist eine Metrik, welche nur mit 4 Fragen die Nutzbarkeit eines Systems erhebt. Berkman und Karahoca zeigen dabei in ihrer Arbeit zu dieser Metrik mithilfe einer Studie, dass sie trotz der geringen Anzahl der Fragen ein verlässliches und valides Tool zur Erhebung der Nutzbarkeit ist ([2]).

2.1.1 Nutzbare Sicherheit

Die nutzbare Sicherheit (engl. Usable Security) ist ein Bereich der Sicherheit, welcher sich dem Namen entsprechend mit der Nutzbarkeit beschäftigt. Dabei steht der Mensch bzw. der Nutzer im Fokus, da in den meisten

Systemen ein Mensch in irgendeiner Form involviert ist. Zwar werden Nutzbarkeit und Sicherheit häufig als einander zuwiderlaufende Aspekte eines Systems angenommen, also dass eine höhere Sicherheit für eine schlechtere Nutzbarkeit sorgen würde und andersherum ([14]), jedoch zeigt die nutzbare Sicherheit, dass dies nicht zwingend der Fall ist. So wird auch davon ausgegangen, dass das falsche Benutzen eines Systems für seine Sicherheit sehr gefährlich sei. Dies kann durch eine schlechte Nutzbarkeit auf mehrere Weisen bewirkt werden. Zum einen kann es sein, dass ein Nutzer das System wegen schlechter Nutzbarkeit aus Versehen falsch benutzt und somit Sicherheitsprobleme auftreten. Zum anderen kann es sein, dass ein Nutzer Sicherheitsmechanismen, welche ihm bei der Erfüllung seiner Aufgabe im Wege stehen, absichtlich umgeht. Demnach kann auch eine geringere Nutzbarkeit für geringere Sicherheit sorgen und eine höhere Nutzbarkeit verschlechtert nicht zwingend die Sicherheit, sondern kann diese sogar erhöhen ([16]).

Ein wichtiges Feld der Usable Security ist die Authentifizierung von Nutzern, besonders jene mit Passwörtern. Dies liegt daran, dass eine hohe Sicherheit in einem Passwort oft mit einer guten Nutzbarkeit bei der Passworterschaffung zusammenhängt. Aus diesem Grund wird in diesem Forschungsfeld stetig daran gearbeitet, Passwortsysteme mit einer besseren Nutzbarkeit zu entwickeln.

In „Usable Security“ ([26]) geben Wash und Zurko eine Einführung in die nutzbare Sicherheit und geben dabei eine Motivation für Forschung in diesem Bereich.

Payne und Edwards geben in ihrer Arbeit ([16]) einen Überblick über die in den vergangenen Jahren bereits durchgeführte Forschung im Bereich der nutzbaren Sicherheit. Dabei gehen sie besonders auf die E-Mail-Verschlüsselung und die Authentifizierung von Nutzern ein. Außerdem listen sie wichtige Erkenntnisse und Richtlinien im Bereich der nutzbaren Sicherheit auf.

In der Arbeit von Naqvi und Seffah ([14]) wird genauer auf die Wechselwirkungen zwischen Nutzbarkeit und Sicherheit eingegangen. Dabei bemängeln sie den momentanen Umgang dieser zwei Bereiche miteinander und schlagen die Nutzung von „Design Patterns“ als eine Verbesserung vor. In ihrer Arbeit kommen sie auch zu dem Schluss, dass wirkliche Sicherheit nicht erreicht werden könne, ohne dass sie nutzbar für den Benutzer sei.

In „Perceptions of Beauty in Security Ceremonies“ ([1]) stellen Bella et al. einen anderen Ansatz zur Betrachtung von Nutzbarkeit und Sicherheit vor. So betrachten sie die Schönheit von sogenannten Sicherheitszeremonien. Eine Sicherheitszeremonie ist dabei eine Interaktion zwischen einem Nutzer und einem Sicherheitssystem, wobei es auch alle damit in Kontakt stehenden Dinge und Tätigkeiten umfasst, die sonst meist als außerhalb angesehen würden. So würden zum Beispiel bei einem Parkautomaten nicht nur die Aktionen des Nutzers zum Ziehen eines Tickets betrachtet, sondern auch

unter anderem das Ticket selbst, seine Aufbewahrung und seine weitere Verwendung. Ziel ist es nun, um die Nutzbarkeit von solchen Sicherheitszeremonien zu verbessern, diese schön zu machen. Was genau dies im Kontext von Sicherheitszeremonien bedeutet, wird als Teil der Arbeit erkundet. So werden anhand des Beispiels eines Fahrkartenautomaten drei Studien durchgeführt. Es wird zu dem Ergebnis gekommen, dass in diesem Kontext Schönheit durch Einfachheit, Bequemlichkeit und Modernität erreicht werden kann. Während Einfachheit und Bequemlichkeit lediglich gängige Designprinzipien seien, wie sie bereits vorher bekannt waren, sei Modernität etwas Neues. Hier wird Modernität noch mit Nachhaltigkeit in Verbindung gebracht und soll den Unterschied zwischen Schönheit und einfachem Design darstellen. Zwar werden die hier vorgestellten Konzepte keine direkte Anwendung in der vorliegenden Arbeit finden, jedoch zeigt diese Arbeit, dass Konzepte aus anderen Disziplinen und auf den ersten Blick eher unpassende Konzepte sich als sinnvolle Erweiterungen der digitalen Sicherheit herausstellen können.

2.2 Erklärbarkeit

Erklärbarkeit (engl. Explainability) beschreibt die Fähigkeit des Bereitstellens von Informationen, um so für ein besseres Verständnis zu sorgen. Ein großer Teil der Erklärbarkeit sind somit die eigentlichen Erklärungen, welche die Informationen bereitstellen. Eine formale Definition kann (nach L. Chazette, [3]) wie folgt gegeben werden:

Definition 2 *Explainability is the ability or act of disclosing information that is necessary for an addressee to understand a particular aspect of a system in a given context, which can be accomplished by providing explanations.*

L. Chazette setzt sich in ihrer Arbeit ([3]) mit der Erklärbarkeit als eine Qualitätsanforderung an ein System auseinander. Dabei wird unter anderem der Begriff genau definiert, sodass ein gemeinsames Verständnis geschaffen wird. Außerdem wird darauf eingegangen, wie Erklärbarkeit die weitere Qualität eines Systems beeinflusst und wie ein erklärbares System sinnvoll entwickelt werden kann.

Rosenfeld und Richardson ([19]) gehen in ihrer Arbeit auf mehrere wichtige Fragen in Bezug auf die Erklärbarkeit ein. So untersuchen sie die Fragen, warum Erklärbarkeit wichtig ist, für wen sie von Bedeutung ist und welche Erklärungen für diese Personen generiert werden sollten. Zudem wird darauf eingegangen, wann einem Nutzer eine Erklärung gezeigt werden sollte und wie die Erklärbarkeit eines Systems gemessen werden könnte. Während hier der Fokus auf der Erklärbarkeit im Zusammenhang mit künstlicher Intelligenz liegt, haben die Erkenntnisse dennoch durchaus auch Bedeutung außerhalb dieses Bereiches.

In „Explanation and trust: what to tell the user in security and AI?“ ([17]) geht Wolter Pieters unter anderem auf die Unterschiede der Erklärbarkeit im Sicherheitskontext und im Kontext der künstlichen Intelligenz ein. Dabei unterscheidet er zwischen Erklärungen für „Confidence“, welche im Folgenden als Zuversicht übersetzt wird und Erklärungen für „Trust“, welches nun mit Vertrauen übersetzt wird. Wenn im Kontext von künstlicher Intelligenz von Erklärbarkeit die Rede ist, werden meist Erklärungen für Zuversicht gemeint. Dies sind Erklärungen, welche die Ergebnisses eines Systems Erklären wollen, um so dem Nutzer Zuversicht in die vom System getroffene Entscheidung zu vermitteln. Im Kontext der Sicherheit hingegen sind Erklärungen für Vertrauen relevant. Dabei geht es darum, dem Nutzer durch die Erklärungen etwas transparent zu machen, um ihm so z.B. Vertrauen in die Sicherheitsmaßnahmen des Systems zu geben. In der vorliegenden Arbeit sind die Erklärungen für Vertrauen deutlich relevanter. Die Unterscheidung zwischen diesen beiden Arten der Erklärbarkeit ist wichtig, da es ansonsten zu Missverständnissen kommen könnte. So kann einem System, welches im Kontext von künstlicher Intelligenz als erklärbar angesehen wird, im Kontext der Sicherheit diese Eigenschaft abgesprochen werden. Wenn in dieser Arbeit von Erklärbarkeit gesprochen wird, wird der Kontext der Sicherheit angenommen, falls nicht zuvor explizit ein anderer Kontext genannt wurde.

2.2.1 Erklärbare Sicherheit

Erklärbare Sicherheit (engl. Explainable Security) ist ein Unterbereich der Sicherheit, bei welchem angestrebt wird, die Sicherheit von Systemen bzw. Software mit dem Forschungsbereich der Erklärbarkeit zu vereinen. Demnach geht es in diesem Bereich darum, die sehr wichtigen und häufig auch sehr komplexen Themen der Sicherheit zu erklären, um sie so verständlich zu machen.

In dieser Art ausformuliert wurde der Begriff von Vigano und Magazzeni in „Explainable Security“ ([25]), auch wenn es bereits vorher Forschung gab, besonders im Bereich der künstlichen Intelligenz, welche heutzutage der erklärbaren Sicherheit zugeordnet werden würde. In ihrer Arbeit definieren sie den Bereich der erklärbaren Sicherheit und seinen Zweck. Außerdem liefern sie eine „Roadmap“ möglicher Forschungsrichtungen in diesem Bereich.

Gefahren von Erklärungen

Erklärungen, besonders im Kontext der Sicherheit, bergen auch Gefahren. So kann ein eventueller Angreifer jegliche Informationen nutzen, um seine Angriffe zu stärken. Ein Beispiel hierfür ist die Mitteilung einiger Anmeldedialoge bei einer versuchten Anmeldung, dass kein Konto mit dem eingegebenen Benutzernamen existiert. Während dies für einen normalen

Nutzer sinnvoll ist, da ihm so Tippfehler oder Ähnliches auffallen können, kann eine solche Aussage bereits als Sicherheitsverstoß angesehen werden ([25]). So ist eine solche Information für einen eventuellen Angreifer nützlich, da er sie sehr einfach erlangen kann (durch einen einfachen Anmeldeversuch) und sie ihm verrät, ob unter einem gewissen Nutzernamen ein Konto existiert, für welches er versuchen könnte, das Passwort herauszufinden. Ein Beispiel für einen solchen gefährlichen Anmeldedialog ist in Abbildung 2.1 zu sehen.

Viele Erklärungen können gewisse Einsichten über das System bieten oder etwas über die inneren Vorgänge verraten, was ein Angreifer dann gegen das System verwenden könnte. Zwar ist es demnach sinnvoll Erklärungen, welche offensichtlich etwas verraten, was besser nicht für alle Nutzer sichtbar sein sollte, zu ändern oder gar zu entfernen, jedoch können nicht immer alle möglichen Angriffe vorhergesehen werden und es kann somit auch nicht immer jedes Sicherheitsrisiko erkannt werden. Es muss also bei Erklärungen im Sicherheitskontext immer abgewogen werden, wie viel diese verraten müssen, um ihren Sinn zu erfüllen, und ob die Vorteile der Erklärung die eventuellen Risiken wert sind ([25]).

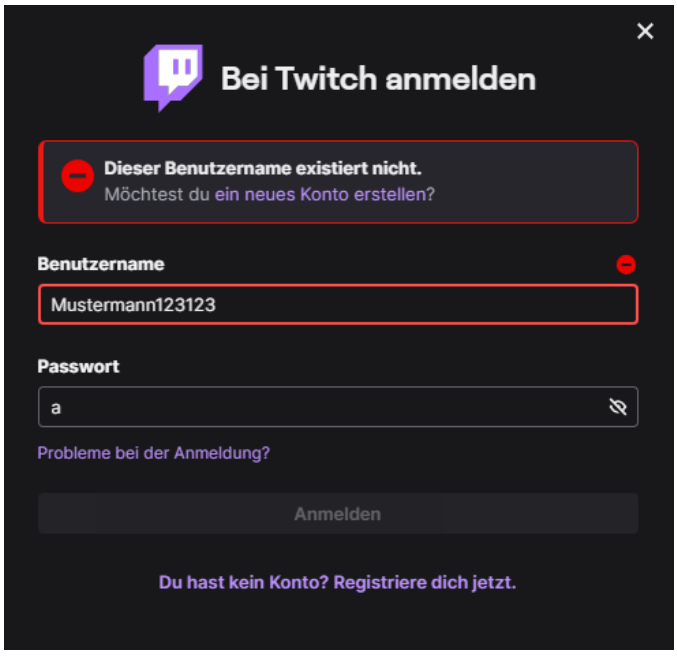

Abbildung 2.1: Beispiel für eine Erklärung, welche einem Angreifer Informationen mitteilt („twitch.tv“)

2.3 Passwörter

Passwörter werden heutzutage sehr häufig zur Authentifizierung von Nutzern verwendet und sind somit fast jedem bekannt. Auch wenn Passwörter bereits viel erforscht wurden, werden sie, wegen der großen Verbreitung und auch weil sie keine perfekte Art der Authentifizierung bieten, weiterhin stark erforscht.

2.3.1 Passwortstärke

Die Stärke eines Passworts ist gleichbedeutend mit der Sicherheit, die dieses Passwort bietet. So ist ein Passwort sehr stark, wenn es für einen beliebigen Angreifer sehr schwer ist, dieses zu erraten bzw. er sehr lange dafür brauchen würde. Ein schwaches Passwort ist somit eines, welches von einem Angreifer schnell erraten wird und somit nicht genügend Sicherheit bietet. Die Passwortstärke mit einer Skala und somit mit konkreten Zahlen zu versehen, ist eher schwierig und wird nicht einheitlich umgesetzt. Ein Passwort, welches in einer Anwendung als stark gilt, kann in einer anderen als schwach gelten. Dies liegt unter anderem daran, dass jede Anwendung die Passwortstärke nach anderen Regeln bestimmt und somit keine vollkommen sicheren Erkenntnisse über die Berechnung der Passwortstärke existieren. Auch ist nicht in allen Kontexten die gleiche Passwortstärke zwingend benötigt oder gewünscht. So ist der Schutz des Zugangs eines kaum genutzten Social Media Kontos wahrscheinlich weniger wichtig als der eines Bankkontos und stellt somit eventuell auch weniger strenge Anforderungen an die Stärke eines Passworts. Zwar wäre es eventuell für die Sicherheit optimal für jedes Konto, ein möglichst starkes Passwort zu nutzen, jedoch ist dies für einen normalen Nutzer nicht machbar. So ist ein stärkeres Passwort z. B. auch meist schwerer zu merken und hat somit eine schlechtere Erinnerungsfähigkeit (engl. Memorability).

Eine Möglichkeit, die Stärke eines Passworts objektiv zu beurteilen, ist über die Modellierung eines Angreifers. So kann ein starker Angreifer mit einer bestimmten Taktik und einer bestimmten Rechenleistung bzw. Rategeschwindigkeit abhängig von dem momentanen Stand der Technik angenommen werden. Nun wird betrachtet, wie lange der modellierte Angreifer voraussichtlich benötigen würde, um das betrachtete Passwort zu erraten. Über diese Dauer kann eine Einschätzung der Stärke gebildet werden.

In „What is in Your Password? Analyzing Memorable and Secure Passwords using a Tensor Decomposition“ ([21]) untersuchen Shin und Woo den Aufbau von Passwörtern. Sie versuchen, durch die Untersuchung einer großen Menge starker und erinnerungsfähiger Passwörter, gemeinsame Faktoren zwischen diesen zu finden. Dazu nutzen sie die mathematische Methode der sogenannten „tensor decomposition“. Das Ziel ist es dabei, gängige

Passworterstellungsrichtlinien zu bestätigen. Shin und Woo kommen zu dem Ergebnis, dass es wichtig für die Erstellung starker und erinnerungsfähiger Passwörter sei, dass diese lang und ohne Wörter aus Wörterbüchern seien.

2.3.2 Erinnerungsfähigkeit

Eine wichtige Eigenschaft von Passwörtern ist die Erinnerungsfähigkeit (engl. Memorability). Die Erinnerungsfähigkeit beschreibt die Eigenschaft eines Passworts, von einem Nutzer gemerkt zu werden. So ist ein Passwort mit einer hohen Erinnerungsfähigkeit leicht für einen Nutzer zu merken. Diese Eigenschaft wird heutzutage viel untersucht, da sie als wichtiger Bestandteil von Passwörtern und Registrierungssystemen anzusehen ist. So kann ein Passwort noch so sicher sein, wenn der Nutzer es sich nicht merken kann, ist es für ihn nutzlos und muss wieder von ihm geändert werden.

2.3.3 Passwortstärkeanzeigen

Passwortstärkeanzeigen (engl. Password Strength Meters) sind bereits seit einiger Zeit eine gängige Erweiterung von Registrierungsdialogen. Dabei wird während der Registrierung die Stärke des momentan eingegebenen Passworts angezeigt, um so den Nutzer bei der Passwortwahl zu unterstützen und ihn dazu zu animieren, ein stärkeres Passwort zu wählen. Dabei gibt es verschiedene Arten, die Darstellung der Passwortstärke zu visualisieren, wobei die gängigste wohl jene als Balken ist, welcher sich bei höherer Stärke verlängert bzw. mehr befüllt und eventuell noch die Farbe wechselt. Da Passwortstärkeanzeigen bereits weit verbreitet sind, Nutzer jedoch noch immer teilweise schwache Passwörter wählen, werden Passwortstärkeanzeigen weiterhin erforscht und es wird danach gestrebt, sie zu verbessern. Abbildung 2.2 zeigt ein Beispiel für eine Passwortstärkeanzeige.

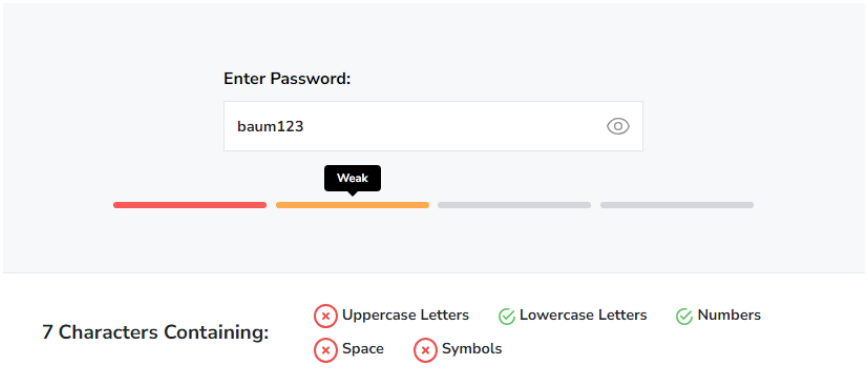

Abbildung 2.2: Beispiel für eine Passwortstärkeanzeige von der Webseite „smallseotools.com“

Egelman et al. überprüfen in ihrer Arbeit ([8]) die allgemeine Wirksamkeit von Passwortstärkeanzeigen. So wurde das Verhalten von Nutzern bei der Neuwahl eines Passworts untersucht, wobei die Anwesenheit einer Passwortstärkeanzeige zwischen den Gruppen variiert wurde. Die Nutzer, welche nicht darüber informiert waren, dass ihre Passwörter untersucht werden, wählten signifikant stärkere Passwörter durch die Anwesenheit einer Passwortstärkeanzeige. Dies war jedoch nur bei sicherheitskritischen Konten der Fall, während bei unwichtigen Konten kein solcher Effekt sichtbar wurde.

In „From Very Weak to Very Strong: Analyzing Password-Strength Meters“ ([6]) untersuchen de Carné de Carnavalet und Mannan verschiedene in der Praxis von Webseiten genutzte Passwortstärkeanzeigen. Dabei werden diese miteinander verglichen und ihre Konsistenz wird untersucht. Sie kommen zu dem Ergebnis, dass viele der genutzten Passwortstärkeanzeigen einige schwache Passwörter dennoch als stark einstufen und somit nicht sehr zuverlässig sind. Außerdem wird somit den Nutzern kein schlüssiges Feedback geliefert, wodurch diese keine sinnvolle Entscheidung zur Wahl ihres Passworts treffen können.

In „On the Accuracy of Password Strength Meters“ ([10]) vergleichen Golla und Dürmuth die Genauigkeit verschiedener Algorithmen zur Bestimmung der Passwortstärke. Dazu definieren sie zuerst einmal die Eigenschaften der Genauigkeit eines solchen Algorithmus und entwickeln anschließend eine Art, diese zu messen. Schließlich vergleichen sie 45 verschiedene Passwortstärkeanzeigen in der Genauigkeit der von ihnen genutzten Algorithmen.

Shay et al. untersuchen in ihrer Arbeit ([20]) den allgemeinen Einfluss von Feedback im Passworterstellungsprozess. Ihre Ergebnisse zeigen, dass eine Art von Feedback bei einem solchen Prozess einen positiven Effekt auf Teilkonzepte der Nutzbarkeit hat. Nach ihrer Aussage gibt es noch zu wenig Forschung, welche sich mit der Art der Präsentation von solchem Feedback über Passwortstärkeanzeigen hinweg beschäftigt.

Golla et al. stellen in ihrer Arbeit [11] einige verschiedene Arten vor, wie Passwortstärkeanzeigen visualisiert werden können und untersuchen in einer Studie, welchen Effekt die verschiedenen Arten auf Benutzer haben. Genauer werden dabei drei Visualisierungen betrachtet, die von dem üblichen Balken abweichen. Diese drei sind ein „High Score Meter“, ein „Badges Meter“ und ein „Peer-Pressure Meter“. Dabei wird die Idee der sogenannten „Gamification“ verwendet, wobei Elemente aus Spielen übernommen werden bzw. Systemteile wie ein Spiel entwickelt werden, um so einen entsprechenden Einfluss auf den Nutzer zu haben. Es wurde zu dem Ergebnis gelangt, dass die Art der Visualisierung wenig bis keinen Effekt auf die Stärke der von Nutzern gewählten Passwörter habe.

In der Arbeit von Vance et al. ([24]) werden die Auswirkungen von sogenannten „Angst-Appellen“, welche eine Art der Schubser (engl. Nudges, siehe 2.3.5 im Folgenden) sind, untersucht. Hierbei werden Elemente

verwendet, welche einem Nutzer Angst machen sollen, um ihn so in seinem Verhalten zu beeinflussen. In diesem Fall sind die „Angst-Appelle“ in der Form eines Textes, welcher dem Nutzer mitteilt, wie lange ein Angreifer brauchen würde, um sein Passwort herauszufinden. Dies soll dem Nutzer Angst davor machen, ein zu schwaches Passwort zu wählen und ihn so dazu bringen, die Passwortstärke des von ihm gewählten Passworts zu verbessern. Während „Angst-Appelle“ schon länger erforscht werden, werden in der Arbeit von Vance et al. interaktive „Angst-Appelle“ verwendet. Diese geben dem Nutzer konkrete Zahlen, wie lange ein Angreifer brauchen würde, abhängig von dem momentan eingegebenen Passwort, anstatt wie in der bisherigen Forschung nur einen allgemeinen Zeitrahmen anzugeben. Um die Effektivität der interaktiven „Angst-Appelle“ zu überprüfen, wurde eine Nutzerstudie durchgeführt, in welcher die Passwortstärke der von Nutzern gewählten Passwörter betrachtet wurde, während die Anwesenheit von „Angst-Appellen“ und die Interaktivität des gesamten Registrierungsdialoges variiert wurden. Die Ergebnisse zeigen, dass interaktive „Angst-Appelle“ normalen „Angst-Appellen“ gegenüber deutlich wirkungsvoller sind und somit eine sinnvolle Erweiterung darstellen. Generell könnte Interaktivität bei der Passworterstellung auch ein wichtiger Faktor sein, unabhängig von der Anwesenheit von „Angst-Appellen“.

Dupuis und Khan versuchen in ihrer Arbeit ([7]) Passwortstärkeanzeigen um eine soziale Komponente zu erweitern. Zu diesem Zweck untersuchen sie den Effekt, den sogenanntes „Peer-Feedback“ in Passwortstärkeanzeigen hat. „Peer-Feedback“ bedeutet dabei in diesem Zusammenhang eine Rückmeldung an den Nutzer darüber, wie stark die gewählten Passwörter anderer Nutzer sind. Dieser Vergleich mit anderen soll den Nutzer dazu bringen, tendenziell stärkere Passwörter zu wählen. Um diese These zu überprüfen, wurde eine Studie durchgeführt, in welcher eine traditionelle Passwortstärkeanzeige mit einer Passwortstärkeanzeige mit „Peer-Feedback“ verglichen wurde. Eine solche Passwortstärkeanzeige mit „Peer-Feedback“ kann in Abbildung 2.3 gesehen werden. Es zeigt sich, dass Passwortstärkeanzeigen mit „Peer-Feedback“ zu stärkeren Passwörtern führen, wenn eine explizite Anweisung zum Erstellen eines neuen, einzigartigen Passworts gegeben wurde.

Abbildung 2.3: Passwortstärkeanzeigen mit „Peer-Feedback“, entnommen aus der Arbeit von Dupuis und Khan ([7])

In der Arbeit von Khern-am-nuai et al. ([13]) werden drei erweiterte Passwortstärkeanzeigen in mehreren Studien getestet. Dabei wurden diese einmal um „Angst-Appelle“, einmal um Vergleiche mit anderen Nutzern und einmal um Nachrichten, welche ein gemeinsames Band zu anderen herstellen sollen, erweitert. Nach ihren Ergebnissen seien diese Erweiterungen sinnvoll und sorgen für signifikant stärkere Passwörter.

In der Arbeit von Zimmermann et al. ([33]) wurde zuerst eine Literaturrecherche zum Thema der Passwortstärkeanzeigen durchgeführt. Die bei dieser Literaturrecherche gefundenen Passwortstärkeanzeigen werden im Folgenden betrachtet und analysiert. So wird ein Überblick über einen Teil der bisherigen Forschung in diesem Bereich geschaffen. Aus der gesammelten Literatur wird zum einen gefolgert, dass Passwortstärkeanzeigen ein vielversprechender Ansatz sind. Außerdem wurde festgestellt, dass effektive Passwortstärkeanzeigen meist sowohl irgendeine Art von sogenannten Schubsern (engl. Nudges, siehe 2.3.5 im Folgenden) beinhalten, wie z. B. einen farbigen Balken, welcher Nutzer indirekt zu stärkeren Passwörtern beeinflusst, als auch irgendeine Art von zusätzlichen Informationen, wie ein starkes Passwort zu erstellen ist. Passwortstärkeanzeigen, welche beide dieser Elemente beinhalten, werden von Zimmermann et al. übersetzt hybride Passwortstärkeanzeigen (engl. hybrid password meters) genannt. Es wurde auch eine Nutzerstudie durchgeführt, in welcher mehrere verschiedene hybride Passwortstärkeanzeigen miteinander verglichen wurden. Diese bestätige erneut die Ergebnisse der Literaturrecherche. Auch wurde zu dem Schluss gekommen, dass die verschiedenen getesteten Variationen der hybriden Passwortstärkeanzeigen meist wenig bis keinen Einfluss hatten.

2.3.4 Spielbarmachung

Spielbarmachung (engl. Gamification) ist ein Begriff, welcher nicht direkt aus der Informatik stammt. Spielbarmachung beschreibt dabei den Prozess, etwas mit Elementen aus der Spieleentwicklung zu versehen, was jedoch aus einem anderen Kontext als dem der Spieleentwicklung stammt ([18]). Ziel ist es dabei die Motivation eines Nutzers, durch die spielerische Gestaltung eines Prozesses, zu erhöhen. Im Kontext der Informatik und von Software könnte dies bedeuten, einen Nutzer zu mehr Benutzung der Software oder zu einem anderen Verhalten zu bewegen. Ein Beispiel für Spielbarmachung im Bereich der Softwareentwicklung ist die Sprachlern-App „Duolingo“. Diese verwendet mehrere gängige Elemente der Spieleentwicklung, wie ein Punktsystem, Abzeichen und Ranglisten. Weiterhin orientiert sich auch das generelle Aussehen der App sehr an dem eines Spiels, wie in Abbildung 2.4 zu sehen ist. Auch im Kontext der Sicherheit und somit auch im Kontext von Passwörtern werden die Auswirkungen von Spielbarmachung untersucht (z. B. in [11]).

Abbildung 2.4: Ausschnitt aus der App „Duolingo“

2.3.5 Schubser

Die sogenannten Schubser (engl. Nudges) werden (im Kontext der Informatik) durch bestimmte Entwurfsentscheidungen hervorgerufen und dienen dazu, den Nutzer indirekt in eine Richtung zu lenken. So könnte mithilfe von Schubsern ein Nutzer zu einem stärkeren Passwort gelenkt werden. Dieser Begriff wurde von Thaler und Sunstein ([22]) geprägt und muss laut ihnen einige Kriterien erfüllen. Erneut aufgegriffen und zusammengefasst werden diese Kriterien in der Arbeit von Zimmermann et al. ([33]). Die Kriterien sind die Folgenden:

1. Vorhersagbarkeit
2. Bewahrung der Wahl
3. Gleichheit der Kosten
4. automatischer kognitiver Prozess
5. ethische Anwendung

Nach 1. muss ein Schubser einen vorhersehbaren Effekt haben, da dieser ansonsten nicht sinnvoll genutzt werden könnte. Es muss also immer vorher genau eingeschätzt und im besten Fall auch getestet werden, welchen Einfluss ein bestimmter Schubser hat. Nach 2. darf ein Schubser nicht die Auswahl eines Nutzers direkt durch die Limitierung der Möglichkeiten beeinflussen. So wäre demnach die strikte Richtlinie, dass ein Passwort mindestens

acht Zeichen haben muss, kein Schubser, da die Auswahlmöglichkeiten des Nutzers eingeschränkt wurden. Nach 3. dürfen Schubser nicht die Kosten oder Gewichtung der einzelnen Optionen direkt beeinflussen. Ein negatives Beispiel dafür wäre es, wenn einem Nutzer größerer Speicherplatz von einem entsprechenden Anbieter versprochen wird, sollte er ein Passwort mit einer gewissen Stärke wählen. Dies würde zwar wahrscheinlich zu stärkeren Passwörtern führen, wäre jedoch kein Schubser. 4. besagt, dass sich Schubser auf einen automatischen kognitiven Prozess beziehen sollen, wie er z. B. durch eine bestimmte Farbwahl hervorgerufen wird. So wird die Farbe Rot meist mit eher schlechten oder gefährlichen Dingen in Verbindung gebracht, während die Farbe Grün eher für positive Dinge steht. Der 5. und letzte Punkt besagt, dass Schubser ethisch korrekt eingesetzt werden sollten. So bürgen diese eine ethische Gefahr, da Nutzer eventuell beeinflusst werden, ohne dies zu bemerken. Es sollte sichergestellt werden, dass bei der Nutzung von Schubsern diese nur für ethische korrekte Zwecke verwendet werden. Außerdem sollten, um eine unbemerkte Beeinflussung zu vermeiden, z. B. textuelle Hinweise verwendet werden, um den Schubser für Nutzer transparent zu gestalten. Eine gängige, ethisch eher fragwürdige, Verwendung wäre so z.B. das Ausstatten von bestimmten Vorschaubildern auf ein Video-Website (z.B. YouTube) mit farblich besonders auffälligen Pfeilen, um so den Nutzer zum Anklicken des eigenen Videos zu bringen (bekannt als sogenannter „Clickbait“).

Ein Beispiel für einen Schubser ist z. B. der Balken einer Passwortstärkeanzeige. Dieser beeinflusst durch Farbe und Füllgrad (Kriterium 4) den Nutzer in Richtung eines stärkeren Passworts, wodurch der Effekt vorhersehbar (Kriterium 1) ist. Die Anwesenheit des Balkens schränkt jedoch in keiner Art und Weise die Wahlmöglichkeiten des Nutzers ein und bietet ihm auch keinen direkten Vorteil durch das Wählen eines stärkeren Passworts (Kriterien 2 und 3). Durch Hinzufügen von Text, welcher den Balken erklärt, z. B. indem die Passwortstärke als Zahl angegeben wird, wird er für Nutzer transparent und da er auch ein gutes Ziel vertritt (Schutz des Nutzern), ist er ethisch korrekt (Kriterium 5).

Das in dieser Arbeit genutzte Konzept, welches später in Kapitel 4 vorgestellt wird, nutzt mehrere Schubser. Diese werden dort jedoch nicht noch einmal explizit erwähnt, da sie nicht direkt Gegenstand der hier durchgeführten Forschung sind.

2.3.6 Algorithmen zur Beurteilung von Passwortstärke

Bei jedem System gibt es andere Richtlinien, was bei der Erstellung eines Passwortes beachtet werden muss. Lange Zeit wurde dabei „LUDS“ als Methode genutzt. Dies steht für: „counts of lower- and uppercase letters, digits and symbols“. Dabei wurde also die Stärke eines Passworts anhand der Anzahl von Kleinbuchstaben, Großbuchstaben, Zahlen und Sonderzeichen

beurteilt, was, wie heutzutage in der Forschung bekannt ist, eine eher ineffektive Methode ist ([27], [15]), da es zu nicht sehr zuverlässigen Einschätzungen der Passwortstärke führt.

Eine Alternative zu „LUDS“ stellt D. L. Wheeler in seiner Arbeit mit seiner Version des „zxcvbn“-Algorithmus vor ([27]). Dieser Algorithmus sieht Passwörter als eine Aneinanderkettung mehrerer Muster an und analysiert sie dementsprechend, um ihre Stärke zu beurteilen. Während auch die Länge des Passworts ein wichtiger Faktor ist, werden außerdem mehrere Datenkörper an häufigen Passwörtern durchsucht. Weiterhin wird auch betrachtet, ob gängige Vor- und Nachnamen im Passwort oder offensichtliche Abfolgen wie „zxcvbn“ (eine Abfolge auf einer „qwerty“-Tastatur, woher wohl auch der Name des Algorithmus stammt) enthalten sind. Dem gegenüber stellt der Algorithmus vier theoretische Angreifer mit verschiedener Stärke und beurteilt, abhängig von ihrer vermuteten Fähigkeit, das Passwort zu knacken, wie stark dieses ist. Weiteres zu der Verwendung dieses Algorithmus und den von ihm überprüften Aspekten kann in Kapitel 4 gefunden werden. Außerdem ist in Abbildung 2.5 eine schematische Darstellung des Algorithmus zu finden.

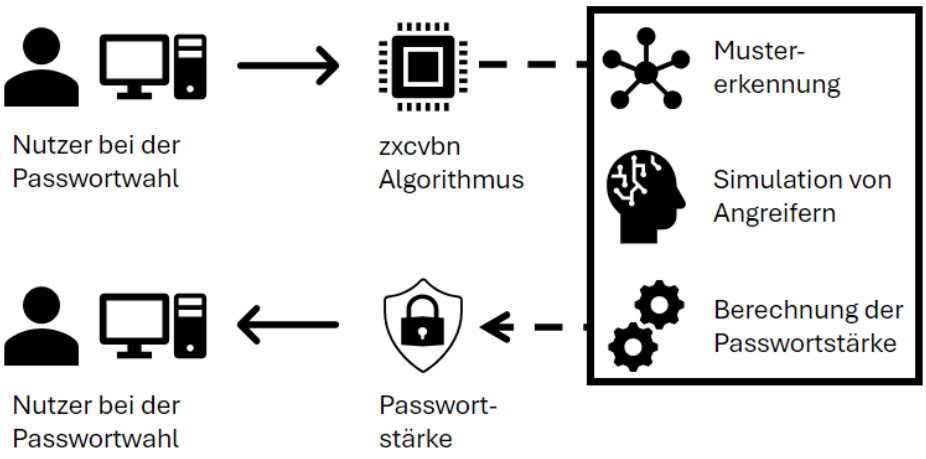

Abbildung 2.5: Schematische Darstellung der Funktionsweise des „zxcvbn“-Algorithmus

Kapitel 3

Literaturrecherche

In diesem Kapitel wird beschrieben, wie für diese Arbeit Literatur gesammelt und begutachtet wurde. Es wurde Literatur auf mehreren Wegen gesammelt, wobei eine systematische Literatur Recherche (SLR) die Hauptquelle darstellt. Aus diesem Grund wird ein großer Teil dieses Kapitels darauf verwendet, die Planung, den Ablauf und die Ergebnisse der durchgeführten SLR zu besprechen. Darüber hinaus wird auf weitere Quellen eingegangen, aus welchen Literatur bezogen wurde.

3.1 Planung der SLR

Als Grundlage für die systematische Literatur Recherche wurden die von Wohlin ([28]) aufgestellten Richtlinien und der von ihm definierte „Snowballing"-Prozess genutzt. Ergänzend wurden auch Teile aus der Arbeit von Wolfswinkel et al. ([30]) genutzt. Zuerst wurden Inklusionskriterien definiert, die eine Arbeit erfüllen muss, um für die Aufnahme in Erwägung gezogen zu werden:

1. Die Arbeit muss sich mit mindestens einem der im Folgenden als relevant definierten Themen beschäftigen.
2. Die Arbeit muss auf Englisch oder Deutsch zugänglich sein.
3. Die Arbeit muss kostenlos zugänglich sein.
4. Die Arbeit muss „Peer-reviewed" sein.

Als relevant wurden dabei folgende Themen definiert:

- Erklärbarkeit
- Nutzbarkeit
- Sicherheit

- Passwortstärke

Ein Fokus wurde dabei auf Arbeiten aus dem Feld der Informatik und besonders auf Arbeiten, die sich mit der Authentifizierung durch Passwörter befassen, gelegt. Für die während des „Snowballings“ identifizierten Arbeiten sowie auch für jene aus anderen Quellen wurde „Google Scholar“ verwendet, um auf sie zu zugreifen. Diese Entscheidung wurde aufgrund der einfachen Benutzung und der erfahrungsgemäß guten Ergebnisse getroffen.

3.2 Startset

Zu Beginn des von Wohlin ([28]) beschriebenen Prozesses muss ein Startset gewählt werden. Für diese Arbeit wurden zwei Arbeiten als Startset gewählt. Die Arbeit von Golla et al. „Bars, Badges, and High Scores: On the Impact of Password Strength Visualizations“ ([11]) behandelt verschiedene Arten, Passwortstärke darzustellen, und wie sich diese auf die von Nutzern gewählten Passwörter auswirken. Diese Arbeit war bereits im Vorfeld bekannt und diente als Grundlage für die Motivation dieser Arbeit. Aus diesem Grund und wegen ihrer Abdeckung für diese Arbeit relevanter Themen wurde sie als Teil des Startsets gewählt.

Die zweite Arbeit wurde durch eine Datenbanksuche zum Thema „explainable security“ entdeckt. Der Begriff war zu diesem Zeitpunkt noch nicht als formal definierter Begriff bekannt und entstammte lediglich eigenen Überlegungen. Die Datenbanksuche führte zu der Arbeit „Explainable Security“ von Vigano und Magazzeni, welche den Begriff der Erklärbare Sicherheit (engl. Explainable Security) formal definiert ([25]). Diese Arbeit bildet den zweiten Teil des Startsets, da sie eine Verbindung zu der Erklärbarkeit darstellt. Die Erklärbarkeit ist für diese Arbeit sehr wichtig, jedoch war über die erste Arbeit des Startsets zu dieser noch keine Verbindung gegeben.

3.3 Snowballing

Der nächste Schritt der SLR war das „Snowballing“. Dabei wird nacheinander, beginnend beim Startset, immer eine Arbeit betrachtet und ihre Referenzen und Referenzierungen werden auf Relevanz untersucht. Das „Snowballing“ wird dabei in „Backward Snowballing“ und „Forward Snowballing“ aufgeteilt. Während beim „Backward Snowballing“ die von der betrachteten Arbeit referenzierten Arbeiten angeschaut werden, werden beim „Forward Snowballing“ die Arbeiten angeschaut, welche die momentan betrachtete Arbeit referenzieren. Die durch das „Snowballing“ gefundenen Arbeiten werden dann darauf überprüft, ob sie alle Inklusionskriterien erfüllen, und jene Arbeiten, die diese nicht erfüllen, werden verworfen.

Danach werden die verbleibenden Arbeiten in zwei Phasen noch einmal überprüft. Zunächst werden sie nur oberflächlich überprüft und es werden thematisch irrelevante Arbeiten verworfen. Danach werden die übrigen Arbeiten noch einmal gründlich darauf geprüft, ob sie für das Ziel der jeweiligen SLR sinnvoll genutzt werden können und werden ansonsten ebenfalls verworfen. Dies wird zunächst mit allen Arbeiten des Startsets im Fokus durchgeführt und danach mit den neu gefundenen relevanten Arbeiten fortgesetzt ([28], [30]). Dieser Prozess wird solange fortgesetzt, bis durch ausreichende Abdeckung der relevanten Themen eine Saturation nach Wolfswinkel et al. ([30]) erreicht wurde. Dieser Ablauf wird in Abbildung 3.1 graphisch dargestellt.

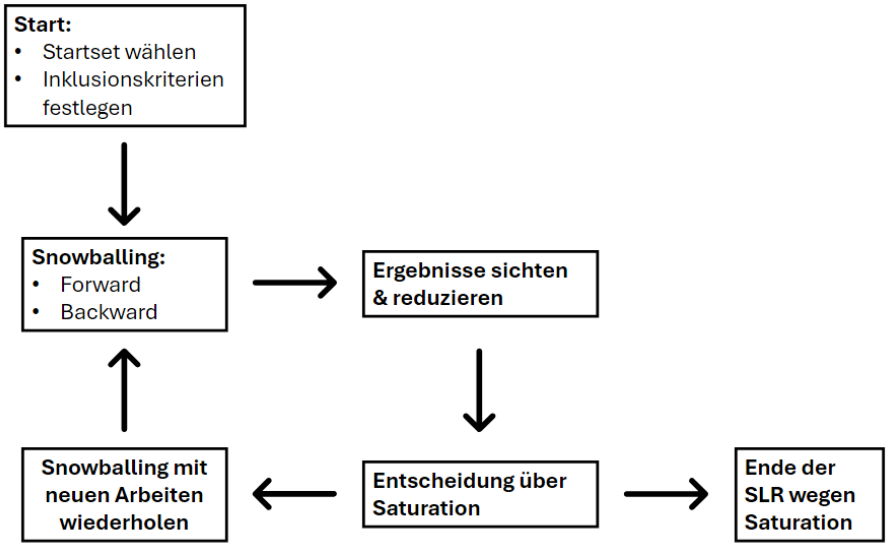

Abbildung 3.1: Allgemeiner Ablauf einer SLR mit Snowballing nach Wohlin ([28]), erweitert nach Wolfswinkel et al. ([30])

3.4 Ergebnisse

Die im Zuge dieser Masterarbeit durchgeführte SLR begann bei 2 Arbeiten als Startset, welche zusammen 70 andere Arbeiten referenzieren und von 88 weiteren Arbeiten referenziert wurden. Nachdem diese 158 Arbeiten in zwei Phasen gesichtet wurden, sind 16 relevante Arbeiten identifiziert worden. SLRs sind im Regelfall niemals vollständig, da theoretisch immer mehr Arbeiten gesichtet werden könnten. Da allerdings eine SLR irgendwann beendet werden muss, wurde die hier durchgeführte SLR nach der Auswertung der Arbeiten des Startsets als nach Wolfswinkel et al. ([30])

saturiert angenommen. Dies hängt mit der Abdeckung der wichtigen Themen zusammen. Da also alle relevanten Themen ausreichend abgedeckt wurden und keine neuen interessanten Themen absehbar waren, wurde die Recherche beendet. Die für diese Arbeit durchgeführte SLR kann auch grafisch in Abbildung 3.2 nachvollzogen werden.

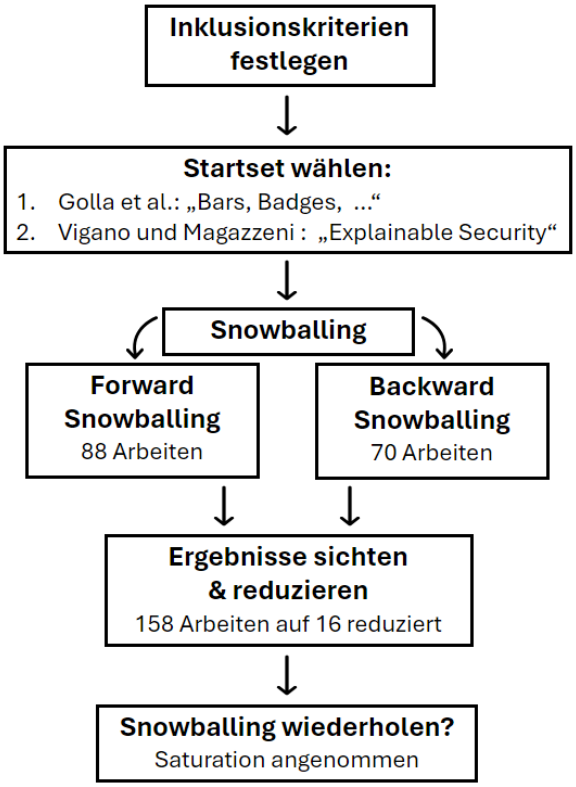

Abbildung 3.2: Darstellung des Ablaufs der durchgeführten SLR

3.5 Arbeiten aus weiteren Quellen

Zusätzlich zu der SLR wurden weitere Arbeiten aus zwei anderen Quellen bezogen. Die erste weitere Quelle für Arbeiten war durch solche Arbeiten gegeben, welche bereits bekannt waren. Diese wurden entweder durch vorherige wissenschaftliche Arbeit und das Studium an der Universität gefunden oder von Kommilitonen und Universitätsmitarbeitern empfohlen. Arbeiten aus dieser Quelle wurden gerade wegen ihrer nicht sehr strukturierten Herkunft jeweils kritisch betrachtet, bevor ihre Benutzung erwogen wurde. Die zweite weitere Quelle, aus der Literatur bezogen wurde, war eine Datenbanksuchen.

Eine Datenbanksuche wurde durchgeführt, da die während der SLR gefundenen Arbeiten zu wenig Bezug zur Nutzbarkeit hatten. Aus diesem Grund wurde mit folgendem Such-String gesucht:

allintitle: Usability Explainability OR Security -AI -ml -machine -„deep learning“ -xai -neural -study

Um aktuellere Forschungsergebnisse zu priorisieren, wurde der Suchbereich auf den Zeitraum seit 2019 begrenzt.

Auf diese Weise wurden 181 Arbeiten gefunden. Diese wurden zunächst oberflächlich, anhand des Titels, gesichtet und somit auf 15 für diese Arbeit relevante Arbeiten gekürzt. Nach einer erneuten, gründlicheren Sichtung dieser 15 Arbeiten wurde lediglich eine Arbeit als relevant eingestuft. Eine genaue Auflistung der in der vorliegenden Arbeit referenzierten Arbeiten und aus welcher der zuvor beschriebenen Quellen sie stammen ist im Anhang A zu finden.

Kapitel 4

Konzeptentwicklung

Vor Beginn der Studie wurde zuerst das folgende Konzept für erklärbare Registrierungsdialoge erarbeitet. Der erklärbare Registrierungsdialog soll neben dem eigentlichen Registrierungsdialog eine Passwortstärkeanzeige mit darunter stehenden Richtlinien besitzen. Dabei soll jede Richtlinie mit einer kurzen Erklärung des Zwecks dieser Richtlinie versehen sein. Auch die Stärkeanzeige soll mit einer Erklärung ihrer Funktionsweise versehen sein. Durch Zeigen mit der Maus auf einen „Mehr“-Knopf soll eine ausführliche Erklärung erreicht werden. Dieses Konzept erfüllt dabei die Kriterien, welche nach Zimmermann et al. ([33]) eine effektive Passwortstärkeanzeige ausmachen. So beinhaltet die Passwortstärkeanzeige sowohl einen visuellen Schubser als auch Informationen dazu, wie ein „besseres Passwort“ erzeugt werden kann. Somit kann die hier konzipierte Passwortstärkeanzeige als hybride Passwortstärkeanzeige (nach [33]) angesehen werden. Auch die „Mehr“-Knöpfe ähneln einem von Zimmermann et al. genutzten Konzept, wobei diese im Zuge dieser Arbeit unabhängig von ihrem Konzept erdacht wurden und sich nicht an diesen orientieren ([33]).

Dieses Konzept soll in der folgenden Studie einem Registrierungsdialog gegenübergestellt werden, welcher keinerlei Erklärungen aufweist. Die genaue Umsetzung dieses Konzepts ist im Folgenden zu finden (4.2).

4.1 Richtlinien und Passwortstärkealgorithmus

Als Algorithmus für das verwendete Passwortstärkemeter wurde „zxcvbn“ genutzt ([27]). Dieser wurde aufgrund der im Vergleich zu anderen Algorithmen relativ präzisen Einschätzung der Passwortstärke ([10]) und des geringen Implementierungsaufwandes ausgewählt. Außerdem beinhaltet er die Möglichkeit, einen Tipp für die Verbesserung eines Passwortes anzugeben, was für die Entwicklung des Prototyps hilfreich war und zur Entwicklung der Richtlinien verwendet wurde. Der Algorithmus wurde genauer im Kapitel 2.3 beschrieben.

Die Richtlinien ergeben sich aus den vom Algorithmus vorgeschlagenen Passwortverbesserungen. Diese wurden auf Deutsch übersetzt und teilweise zusammengefasst. Ein Beispiel für die zusammengefassten Tipps ist in der Tabelle 4.1 zu sehen.

Tipps von „zxcvbn“	**Richtlinie**
This is a top-100 common password	Dies ist ein häufig verwendetes Passwort und sollte vermieden werden.
This is a top-10 common password	
This is similar to a commonly used password	
This is a very common password	

Tabelle 4.1: Beispiel für das Zusammenfassen und Übersetzen von „zxcvbn“-Algorithmus-Tipps zu Richtlinien

Eine Tabelle mit allen zusammengefassten Tipps des „zxcvbn“-Algorithmus ist im Anhang B.3 zu finden. Aus den Tipps des „zxcvbn“-Algorithmus ergeben sich folgende Richtlinien, welche auf der Auswahl an überprüften Mustern des Algorithmus beruhen. Genaueres zu diesen kann in der entsprechenden Arbeit nachgelesen werden ([27]).

- Das Passwort sollte verlängert werden.
- Abfolgen von der Tastatur sollten vermieden werden.
- Großbuchstaben sind nicht besonders hilfreich.
- Vorhersehbare Ersetzungen helfen nicht sehr.
- Daten und Jahreszahlen sollten vermieden werden.
- Wiederholungen und direkte Abfolgen sollten vermieden werden.
- Dies ist ein häufig verwendetes Passwort und sollte vermieden werden.
- Vor- und Nachnamen sollten vermieden werden.

4.2 Prototypentwicklung

Der Prototyp wurde in Python 3.7 programmiert. Dabei wurde das „Pyside6“-Modul zusammen mit dem „qt-material“-package für die grafische Oberfläche genutzt. Grund für diese Entscheidungen war die so erreichte vergleichsweise einfache Umsetzung des Konzepts. Der fertige Prototyp besteht aus 2 Seiten: einer Seite mit Erklärungen zur Studie und einer Seite für Anmeldungs- und Registrierungsdialoge. Der genaue Ablauf der Studie wird in Kapitel 5 besprochen.

Die Abbildung 4.1 wurde dem Prototyp entnommen und zeigt einige relevante Elemente. Es sind die Eingabefelder für Benutzername, Passwort

und eine Wiederholung des Passworts zu sehen, wie es bei einer Registrierung heutzutage üblich ist. Auf der rechten Seite ist ganz oben zuerst die Passwortstärkeanzeige mit einer darunter stehenden Erklärung zu dieser zu erkennen. Darunter sind zwei Richtlinien mit jeweils einer Erklärung sichtbar. Es sind auch rechts neben den Erklärungen die farblich hervorgehobenen „Mehr“-Schriftzüge zu sehen, welche zu ausführlichen Erklärungen führen. Weitere Auszüge des Prototyps sind im Anhang B.3 zu finden.

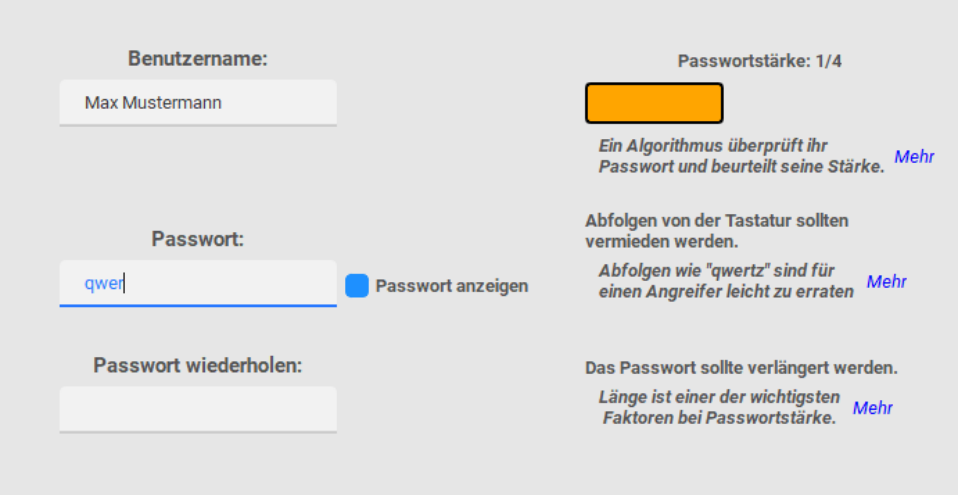

Abbildung 4.1: Ausschnitt aus dem Prototyp: Neuregistrierung mit Erklärungen

Kapitel 5

Studiendesign

Dieses Kapitel stellt zuerst Forschungsfragen auf, welche als Ziel dieser Arbeit beantwortet werden sollen. Dies soll mit Hilfe einer Studie und dem entwickelten Prototyp geschehen. Danach wird in diesem Kapitel genauer auf den Ablauf der durchgeführten Nutzerstudie eingegangen.

5.1 Forschungsfragen

RQ1: Wahrnehmung

RQ1: Welchen Einfluss haben Erklärungen in Registrierungsdialogen auf die Verständlichkeit und die Vertrauenswürdigkeit des Systems?

Um den Effekt von Erklärungen auf Registrierungsdialoge zu überprüfen, muss der Nutzen, den diese eventuell erbringen, überprüft werden. Dies soll mit der 1. Forschungsfrage durch die Wahrnehmung der Nutzer ermittelt werden. Als Metrik hierfür soll ein Fragebogen von den Nutzern nach der Benutzung des Prototyps ausgefüllt werden. Insbesondere werden Fragen zur wahrgenommenen Verständlichkeit und zum Vertrauen in die Sicherheit des Systems gestellt.

RQ2: Passwortsicherheit

RQ2: Wie beeinflussen Erklärungen in Registrierungsdialogen die Sicherheit der von Nutzern gewählten Passwörter?

Um die Effektivität von Erklärungen in Registrierungsdialogen nicht nur von der Wahrnehmung des Nutzers abhängig zu machen, wird diese in der 2. Forschungsfrage auch objektiv während der Nutzung des Systems untersucht. So soll bei dieser Frage geklärt werden, ob durch Erklärungen Nutzer tatsächlich sicherere Passwörter wählen. Als Metrik dafür sollen die Stärkeeinschätzungen des „zxcvbn"-Algorithmus, welcher in 2.3 genauer beschrieben wurde, genutzt werden.

RQ3: Nutzbarkeit

RQ3: Welche Auswirkungen haben Erklärungen in Registrierungsdialogen auf die Nutzbarkeit des Systems?

Erklärungen können neben positiven Auswirkungen auch negative Auswirkungen haben, besonders im Hinblick auf die Nutzbarkeit ([3]). Eventuelle negative Auswirkungen sollen mithilfe der 3. Forschungsfrage gefunden werden. Als Metrik zur Überprüfung der Nutzbarkeit soll eine Befragung mithilfe der „Usability Metric for User Experience" (UMUX) durchgeführt werden ([9]). Dies ist eine zuverlässige und leicht auszuwertende Metrik für Nutzbarkeit ([2]).

In Abbildung 5.1 ist ein Überblick über die für diese Arbeit aufgestellten Forschungsfragen in Form einer sogenannten „Goal-Question-Metric" zu sehen.

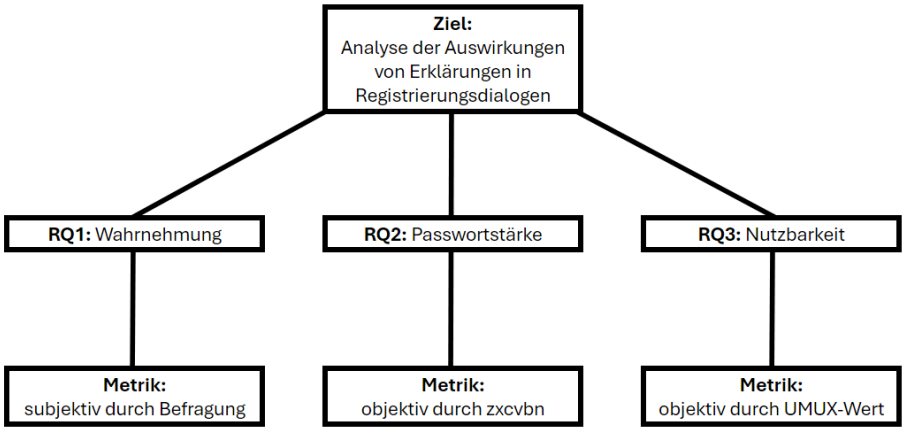

Abbildung 5.1: Schaubild einer „Goal-Question-Metric" zu dieser Arbeit

5.2 Forschungsmethodik

Die Struktur dieser Arbeit ist nicht gänzlich gleich dem für diese Arbeit durchgeführten Prozess. So wurde für diese Arbeit zuerst eine Literaturrecherche durchgeführt. Dabei wurde sowohl eine SLR als auch eine Datenbanksuche durchgeführt und darüber hinaus wurde auch unstrukturiert gesucht. Danach wurde mithilfe der gefundenen Literatur ein Konzept für erklärbare Registrierungsdialoge entwickelt. Als Nächstes wurde die Nutzerstudie geplant. Dazu wurden zuerst die Forschungsfragen formuliert und Metriken gewählt, um diese zu überprüfen. Außerdem wurde der allgemeine Ablauf der Studie geplant und es wurde sich um die Teilnehmergewinnung sowie kleinere Planungsaufgaben gekümmert. Nun wurde als nächstes der für die Studie erforderliche Prototyp programmiert, wobei einige wichtige Entscheidungen getroffen wurden, wie beispielsweise die Wahl des Algorithmus zur Bestimmung der Passwortstärke. Die zuvor geplante Studie wurde im Anschluss durchgeführt. Als Letztes mussten die bei der Durchführung der Studie gesammelten Daten ausgewertet und ihre Bedeutung analysiert werden. Der gesamte Prozess wird in Abbildung 5.2 grafisch dargestellt.

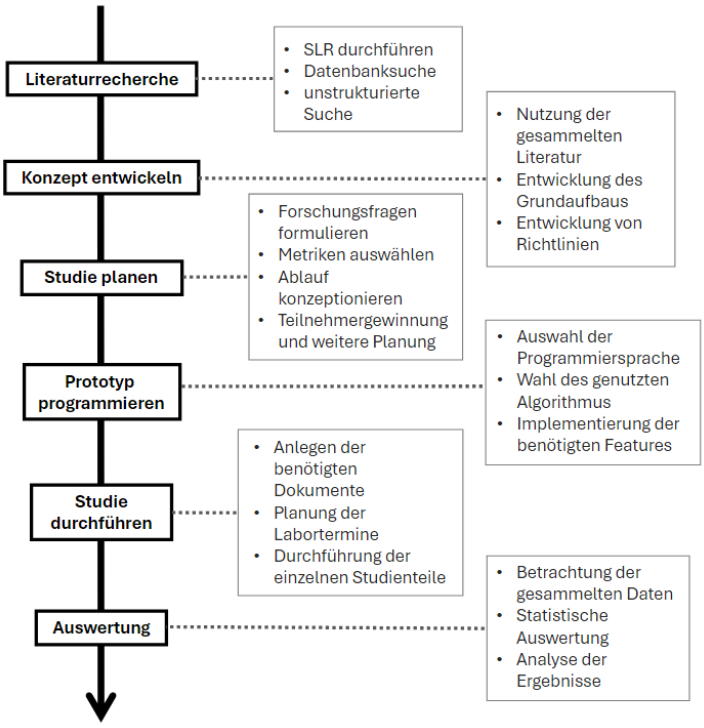

Abbildung 5.2: Darstellung des Arbeitsprozesses dieser Arbeit

5.3 Ablauf der Studie

Die Studie besteht aus 3 Teilen, welche im Folgenden genauer beschrieben werden. Der erste Teil ist eine Einleitung für die Probanden, in welcher der Ablauf der Studie und ihre Aufgabe den Probanden erklärt werden. Danach kommt die Benutzung des Prototyps durch die Probanden und der dritte Teil ist eine abschließende Befragung. Eine schematische Darstellung des hier beschriebenen Ablaufs ist in Abbildung 5.4 am Ende des Kapitels zu finden. Die Studie wurde mit 28 Probanden durchgeführt. Dabei wurde stets der gleiche Raum genutzt, damit die Ergebnisse besser miteinander verglichen werden konnten. Die Ergebnisse werden in Kapitel 6 präsentiert und in Kapitel 7 diskutiert.

Einleitung für die Probanden

Den Probanden wurde schriftlich ein Überblick über den Versuch sowie eine Einwilligungserklärung vorgelegt. Sie wurden angewiesen, sich beides gründlich durchzulesen, und wenn sie mit dem Inhalt einverstanden seien, die Einwilligungserklärung zu unterschreiben. In diesen zwei Dokumenten wurden die Probanden über den Zweck dieser Studie, den groben Ablauf und den Umgang mit ihren Daten informiert. Die Dokumente sind im Anhang B.1 zu finden.

Danach wurde den Probanden ein Laptop mit dem bereits gestarteten Prototyp gereicht, sodass für die Probanden zunächst nur der Erklärungstext des Prototyps sichtbar war. Der Einleitungstext teilte den Probanden mehrere wichtige Informationen mit. Zum einen wurde ihnen erneut mitgeteilt, dass es sich um eine Studie handelt, und zum anderen, wie der weitere Ablauf der Studie ist. Außerdem wurden die Probanden explizit darauf hingewiesen, dass keinerlei Zeitdruck besteht. So sollte dafür gesorgt werden, dass die Probanden sich möglichst viel mit dem Prototyp beschäftigen und eventuelle Effekte der Erklärungen auch mit wenig Probanden erkannt werden können. Weiterhin wurden sie darüber informiert, dass die von ihnen eingegebenen Passwörter weder sichtbar für andere sind noch gespeichert werden, sondern lediglich die Passwortstärken erhoben werden. Außerdem wurden die Probanden auch mündlich darauf hingewiesen, dass sie die von ihnen gespeicherten Daten nach der Studie einsehen können, wenn gewünscht. Dadurch sollte verhindert werden, dass die Passwortwahl der Probanden durch die Angst vor der Weitergabe ihrer Passwörter beeinflusst wird. Es wurde ebenfalls darauf hingewiesen, dass es keinerlei falsche Interaktion gibt, um so eine eventuelle Beeinflussung durch den Kontext, dass dies eine Studie ist, abzumildern. Der genaue Text, der in der Studie genutzt wurde, kann im Anhang B.2 gefunden werden.

Benutzung des Prototyps

Nach dem Erklärungstext wurden die Probanden durch einen Klick zur eigentlichen Durchführungsseite weitergeleitet. Neben den Informationen aus dem Einleitungstext wurden die Probanden bei der Benutzung des Prototyps lediglich verbal angewiesen, auf die Aufgabenstellung in der oberen linken Ecke des Prototyps zu achten. Bei der Benutzung des Prototyps war der Versuchsleiter so positioniert, dass er die Eingabe der Probanden nicht sehen konnte, um so eine Beeinflussung der Eingaben der Probanden zu verhindern.

Die Probanden sollten sich so verhalten, als ob sie sich bei drei verschiedenen vorgegebenen Webseiten mit ihren wirklichen Accounts anmelden würden. Die drei Webseiten waren dabei durch die Szenarien „Universitätsnetzwerk StudIP“, einen „Streamingdienst“ und „Onlinebanking“ vorgegeben, wobei sich die Probanden für einen genauen Anbieter eines solchen Dienstes selbst entscheiden sollten. Die Reihenfolge der drei Szenarien war zufällig, um so einer Beeinflussung durch die Reihenfolge und einem eventuellen Lerneffekt entgegenzuwirken. Nach der Durchführung der drei Anmeldedialoge, welche später als Vergleich dienen sollten, mussten die Probanden sich nun bei den gleichen drei Szenarien neu registrieren. Dabei waren die Probanden in zwei Gruppen eingeteilt. Eine Gruppe bekam Erklärungen zu der angegebenen Passwortstärke und den gezeigten Richtlinien und die andere Gruppe bekam keine solchen Erklärungen. In Abbildung 5.3 sind Ausschnitte aus dem Prototyp zu sehen, in welchen die Aufgabenstellungen zu sehen sind. Weitere Ausschnitte sind im Anhang B.3 zu finden.

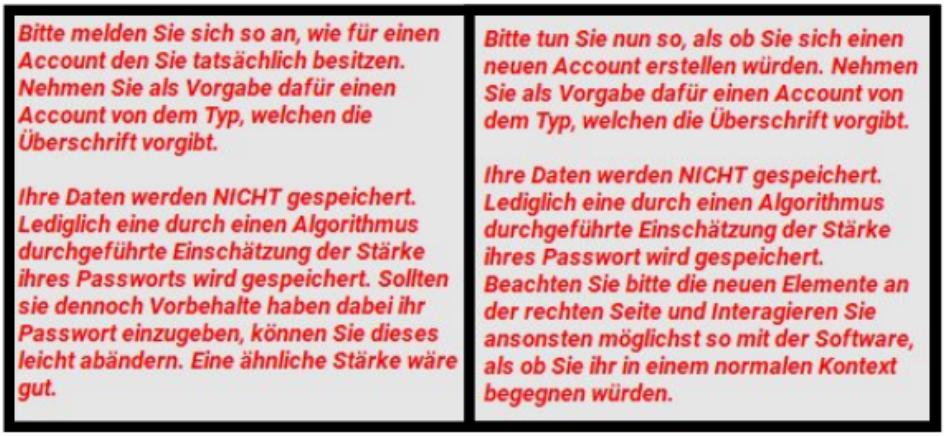

Abbildung 5.3: Im Prototyp genutzte Aufgabenstellungen, links: Anmeldung, rechts: Neuregistrierung

Abschließende Befragung

Nach der Durchführung des oben erklärten Ablaufs wurden die Probanden automatisch zu einer Umfrage in dem Tool „Limesurvey“ weitergeleitet. Diese wurde durchgeführt, um die im Prototyp gesammelten Erkenntnisse der Probanden zu erheben. Die Umfrage bestand dabei unter anderem aus den folgenden Fragen:

- **Wahrnehmung:**
  Bitte geben Sie Ihre Zustimmung zu den folgenden Aussagen an.
  Bitte beachten Sie, dass sich die folgenden Fragen sich nur auf den zweiten Teil der verwendeten Software, also die Neuregistrierungen, beziehen. Beantworten Sie die Fragen bitte dementsprechend.

  1. Ich halte die Anmeldung für verständlich. (7-Punkt-Likert)
  2. Ich finde die Anmeldung vertrauenswürdig. (7-Punkt-Likert)
  3. Ich hätte mir mehr Informationen zur Passwortstärke gewünscht. (7-Punkt-Likert)

- **UMUX:**
  Bitte geben Sie Ihre Zustimmung zu den folgenden Aussagen an.
  Bitte beachten Sie, dass sich die folgenden Fragen sich nur auf den zweiten Teil der verwendeten Software, also die Neuregistrierungen, beziehen. Beantworten Sie die Fragen bitte dementsprechend.

  1. Die Anmeldung erfüllt meine Anforderungen. (7-Punkt-Likert)
  2. Die Anmeldung zu benutzen ist eine frustrierende Erfahrung. (7-Punkt-Likert)
  3. Die Anmeldung ist leicht zu benutzen. (7-Punkt-Likert)
  4. Ich habe zu viel Zeit dazu verwenden müssen etwas bei der Anmeldung zu korrigieren. (7-Punkt-Likert)

- **Demografie**

  – Alter (Auswahl aus voreingestellten Optionen)
  – Geschlecht (Auswahl aus voreingestellten Optionen)
  – Ich halte mich selbst für technisch versiert. (7-Punkt-Likert)

Diese Fragen dienten der Erhebung von Ergebnissen für die Forschungsfragen 1 und 3 und der Erhebung der demografischen Daten. Dazu wurden zur Erhebung der Verständlichkeit und der Vertrauenswürdigkeit die drei oben unter „Wahrnehmung“ aufgeführten Fragen formuliert. Zur Bestimmung des Einflusses auf die Nutzbarkeit wurden die vier oben aufgeführten Fragen des UMUX ([9]) genutzt. Wie den Klammern zu entnehmen ist, wurde ein

Großteil dieser Fragen mit 7-Punkt-Likert-Skalen als Antwortmöglichkeiten erhoben. Die Fragen, welche für die Studie nicht weiter relevant waren und deswegen hier nicht behandelt wurden, können im Anhang B.4 gefunden werden.

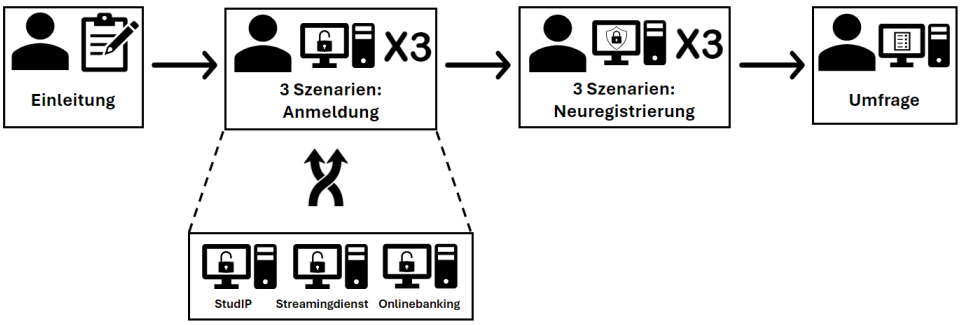

Abbildung 5.4: Schematische Darstellung des Studienablaufs

Kapitel 6

Ergebnisse

Im Folgenden wird zuerst die Demografie der Studienteilnehmer betrachtet. Dies beinhaltet generelle demografische Daten sowie die Selbsteinschätzung zur Versiertheit im Umgang mit Technik. Daraufhin werden die Ergebnisdaten präsentiert und statistisch analysiert. Ein genauer Ablauf der Studie wurde bereits in Kapitel 5 gegeben, während ein Überblick über den genutzten Prototypen in Kapitel 4 zu finden ist.

6.1 Demografie der Teilnehmer

Zur Teilnehmergewinnung wurden bekannte Personen und andere Studenten, welche ebenfalls eine Abschlussarbeit am Fachgebiet Software Engineering schreiben, kontaktiert. Insgesamt nahmen 28 Personen an der Studie teil. Mit ca. 89 % (25 Probanden) identifizierte sich der Großteil der Teilnehmer als männlich, ca. 11 % (3 Probanden) als weiblich und kein Teilnehmer als nicht binär. Der große Anteil männlicher Teilnehmer ist unter anderem auf den großen Anteil von männlichen Studenten in der Informatik zurückzuführen, aus welchem ein großer Teil der Teilnehmer stammte. Zudem hielten sich alle Teilnehmer für technisch versiert.

Das Alter der Probanden ist aufgrund der Art der Teilnehmergewinnung zu großen Teilen im gleichen Bereich. So waren alle Teilnehmer über 18 Jahre alt, da dies zur Teilnahme notwendig war. Weiterhin befanden sich ca. 86 % (24 Probanden) der Teilnehmer im Bereich 18 bis 30 Jahre. Eine graphische Darstellung des Alters der Probanden ist in Abbildung 6.1 zu sehen.

6.2 Auswertung der Studienergebnisse

Wie in Kapitel 5 bereits angemerkt, sollten drei Forschungsfragen mit dieser Studie untersucht werden (siehe 5.1). Für die erste Forschungsfrage RQ1 wurden drei Werte erhoben: Verständlichkeit, Vertrauen und der Wunsch nach mehr Informationen. Für die zweite Forschungsfrage RQ2 wurden die

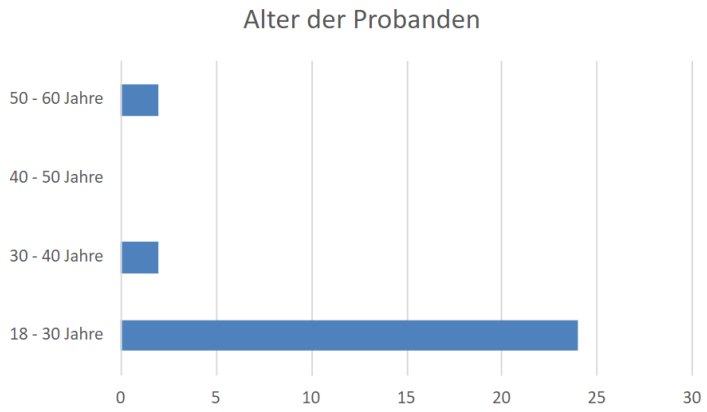

Abbildung 6.1: Alter der Probanden

von den Probanden bereits genutzten Passwörter sowie die im Zuge der Studie neu gewählten Passwörter betrachtet. Da pro Proband in je drei Kategorien sowohl bisher genutzte als auch neu zu wählende Passwörter erfragt wurden, ergeben sich hier über die Betrachtung der Differenz zwischen den zwei Passwörtern einer Kategorie und die Betrachtung der absoluten Werte der neu gewählten Passwörter insgesamt sechs auszuwertende Datenpunkte pro Proband. Für die dritte Forschungsfrage RQ3, welche sich mit der Nutzbarkeit beschäftigt, wurde mit Fragen die „Usability Metric for User Experience“ (UMUX) erhoben. Somit ergeben sich über alle drei Forschungsfragen hinweg insgesamt zehn auszuwertende Datenpunkte.

6.2.1 Verständlichkeit

Die Probanden teilten im Zuge der Studie ihre Meinung zu der Verständlichkeit des genutzten Prototyps mit, aufgeteilt in eine Gruppe ohne weitere Erklärungen und eine Gruppe mit weiteren Erklärungen. Diese Gruppeneinteilung wurde während der gesamten Studie beibehalten. Dazu wurde die Zustimmung zu der Aussage: „Ich halte die Registrierung für verständlich.“ mit einer 7-Punkt-Likert-Skala erhoben. Die dazugehörige einseitige Nullhypothese wurde wie folgt konstruiert:

H_0: Es gibt keine Verbesserung der Verständlichkeit von Registrierungsdialogen durch das Hinzufügen von Erklärungen.

Um die Nullhypothese zu überprüfen, wurde ein Mann-Whitney-U-Test mit 5 % Signifikanz durchgeführt. Dabei ergaben sich die folgenden U-Werte: $U_A = 121$ (A = Gruppe ohne Erklärungen), $U_B = 75$ (B = Gruppe mit

Erklärungen). Durch den Vergleich des kleineren U-Wertes mit dem der Tabelle für kritische Werte entnommenen Wert (Anhang C) ergibt sich folgender Vergleich: $U_{krit} = 61 > 75 = U_B$. Da dieser Vergleich offensichtlich nicht hält, kann die Nullhypothese nicht abgelehnt werden und es kann somit keine statistische Signifikanz nachgewiesen werden.

Bei der Berechnung des Pearson-Korrelationskoeffizienten ergibt sich ein r-Wert von ca. $r = 0,1997$, was nach der Interpretation von Cohen für einen schwachen Effekt spricht ([4]). In Abbildung 6.2 ist eine grafische Darstellung der von Probanden gegebenen Antworten zu sehen.

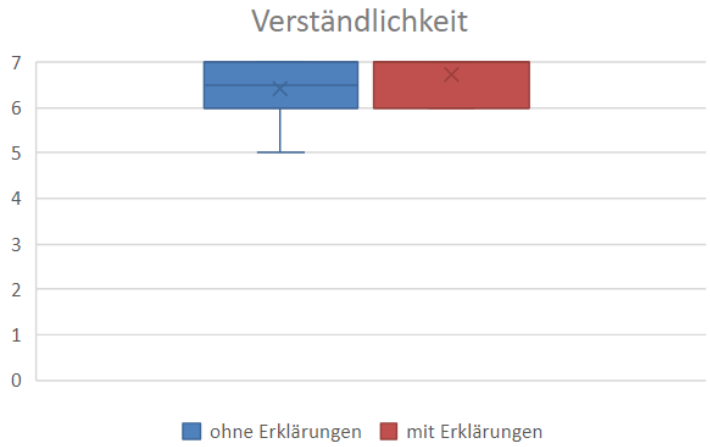

Abbildung 6.2: Box-and-Whisker-Plot der von Nutzern abgegebenen Zustimmung auf einer 7-Punkt-Likert-Skala zu der Aussage: „Ich halte die Registrierung für verständlich.“

6.2.2 Vertrauen

Auch teilten die Probanden im Zuge der Studie ihr Vertrauen in die genutzte Software mit, dabei immer noch aufgeteilt in eine Gruppe ohne weitere Erklärungen und eine Gruppe mit weiteren Erklärungen. Es wurde die Zustimmung zu der Aussage: „Ich finde die Registrierung vertrauenswürdig.“ mit einer 7-Punkt-Likert-Skala erhoben. Die passende einseitige Nullhypothese ist folgende:

H_0: Es gibt keine Verbesserung des Vertrauens der Nutzer gegenüber dem System durch das Hinzufügen von Erklärungen.

Um die Nullhypothese zu überprüfen, wurde auch hier ein Mann-Whitney-U-Test mit 5 % Signifikanz durchgeführt. Die sich ergebenen U-

Werte sind Folgende: $U_A = 97$ (A = Gruppe ohne Erklärungen), $U_B = 99$ (B = Gruppe mit Erklärungen). Auch hier wird wieder der kleinere Wert mit dem entsprechenden Wert aus der Tabelle verglichen (Anhang C): $U_{krit} = 61 > 97 = U_A$. Da dieser Vergleich erneut offensichtlich nicht hält, kann auch hier die Nullhypothese nicht abgelehnt werden und es kann somit keine statistische Signifikanz nachgewiesen werden.

Bei der Berechnung des Pearson-Korrelationskoeffizienten ergibt sich ein r-Wert von ca. $r = 0,0087$, was nach der Interpretation von Cohen für keinerlei Effekt spricht. Abbildung 6.3 ist eine grafische Darstellung der Antworten der Probanden.

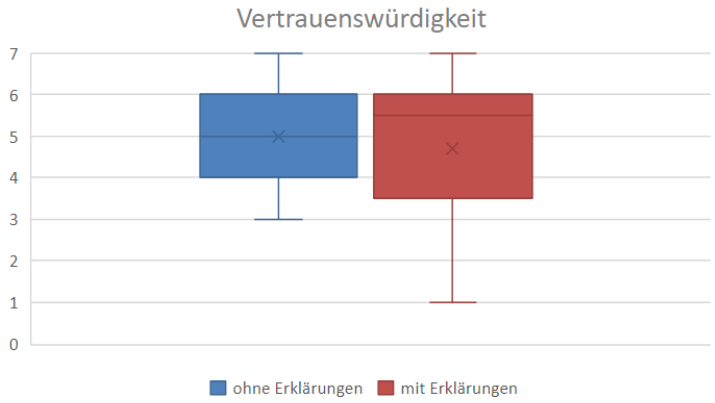

Abbildung 6.3: Box-and-Whisker-Plot der von Nutzern abgegebenen Zustimmung auf einer 7-Punkt-Likert-Skala zu der Aussage: „Ich finde die Registrierung vertrauenswürdig.“

6.2.3 Wunsch nach mehr Informationen

Weiterhin aufgeteilt in zwei Gruppen wurde auch der Wunsch der Probanden nach mehr Informationen erhoben. Dazu wurde ihre Zustimmung zu der Aussage: „Ich hätte mir mehr Informationen zur Passwortstärke gewünscht.“ mit einer 7-Punkt-Likert-Skala erhoben. Passend dazu wurde die Nullhypothese wie folgt konstruiert:

H_0: Es gibt keinen geringeren Wunsch nach mehr Informationen zur Passwortstärke durch das Hinzufügen von Erklärungen.

Um die Nullhypothese zu überprüfen, wurde erneut ein Mann-Whitney-U-Test mit 5 % Signifikanz durchgeführt. Es wurden folgende U-Werte berechnet: $U_A = 79,5$ (A = Gruppe ohne Erklärungen), $U_B = 116,5$ (B =

Gruppe mit Erklärungen). Durch den Vergleich des kleineren U-Wertes ergibt sich Folgendes: $U_{krit} = 61 > 79,5 = U_A$. Da dieser Vergleich offensichtlich nicht hält, kann die Nullhypothese nicht abgelehnt werden und es kann somit keine statistische Signifikanz nachgewiesen werden.

Erneut wird auch der Pearson-Korrelationskoeffizient berechnet und es ergibt sich ein r-Wert von ca. $r = 0,1606$, was nach der Interpretation von Cohen für einen kleinen Effekt spricht ([4]). Dieser Effekt tritt dabei in der erwarteten Richtung auf. Demnach sorgen Erklärungen für einen geringeren Wunsch nach mehr Informationen. Abbildung 6.4 können die abgegebenen Antworten der Probanden entnommen werden.

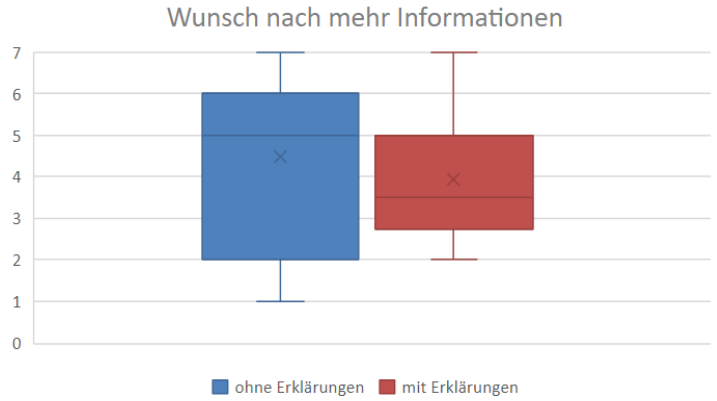

Abbildung 6.4: Box-and-Whisker-Plot der von Nutzern abgegebenen Zustimmung auf eine 7-Punkt-Likert-Skala zu der Aussage: „Ich hätte mir mehr Informationen zur Passwortstärke gewünscht.“

6.2.4 Passwortstärkedifferenz, Szenario: StudIP

Eine Aufgabe der Probanden während der Studie war es, sich sowohl mit einem bereits genutzten Passwort in dem Uni-Netzwerk StudIP anzumelden als auch sich dort neu zu registrieren, also ein neues Passwort mit Hilfe einer Passwortstärkeanzeige zu vergeben. Dabei waren die Probanden immer noch in zwei Gruppen aufgeteilt, in eine Gruppe ohne weitere Erklärungen und eine Gruppe mit weiteren Erklärungen. Es wurden jeweils die Stärken der von den Probanden gewählten Passwörter gespeichert, sodass hier die Differenz in der Passwortstärke zwischen zuvor genutztem Passwort und neu gewähltem Passwort ausgewertet werden kann. Die einseitige Nullhypothese dafür wurde wie folgt konstruiert:

H_0: Es gibt keine größere Verbesserung in der Stärke bei der Neuwahl eines Passworts für das System StudIP durch das Hinzufügen von Erklärungen.

Auch hier wurde, um die Nullhypothese zu überprüfen, ein Mann-Whitney-U-Test mit 5 % Signifikanz durchgeführt. Dabei ergaben sich folgende U-Werte: $U_A = 78$ (A = Gruppe ohne Erklärungen), $U_B = 118$ (B = Gruppe mit Erklärungen). Durch den Vergleich des kleineren U-Wertes mit dem der Tabelle für kritische Werte entnommenen Wert (Anhang C) ergibt sich folgender Vergleich: $U_{krit} = 61 > 78 = U_A$. Da dieser Vergleich offensichtlich nicht hält, kann die Nullhypothese nicht abgelehnt werden und es kann somit keine statistische Signifikanz nachgewiesen werden.

Bei der Berechnung des Pearson-Korrelationskoeffizienten ergibt sich ein r-Wert von ca. $r = 0,1737$, was nach der Interpretation von Cohen für einen kleinen Effekt spricht ([4]). Auffallend ist hierbei, dass der Effekt jedoch entgegen der erwarteten Richtung auftritt. Dies wird im Kapitel 7 besprochen. In der Abbildung 6.5 ist die Differenz der Passwortstärken grafisch dargestellt.

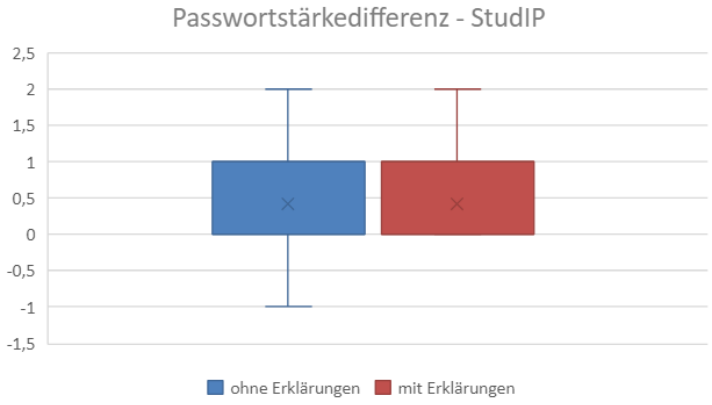

Abbildung 6.5: Box-and-Whisker-Plot der Differenzen der erfassten Passwortstärken für das Szenario StudIP.

6.2.5 Passwortstärkedifferenz, Szenario: Streamingdienst

Während der Studie sollte sich sowohl mit einem bereits genutzten Passwort als auch mit einem neuen Passwort bei einem Streamingdienst authentifiziert werden. Dabei waren die Probanden ebenfalls in zwei Gruppen aufgeteilt: in eine Gruppe ohne und eine Gruppe mit weiteren Erklärungen. Auch hier wurden jeweils die Stärken der von den Probanden gewählten Passwörter

gespeichert, sodass die Differenz in der Passwortstärke zwischen zuvor genutztem Passwort und neu gewähltem Passwort ausgewertet werden kann. Die einseitige Nullhypothese dazu ist die Folgende:

H_0: Es gibt keine größere Verbesserung in der Stärke bei der Neuwahl eines Passworts für einen Streamingdienst durch das Hinzufügen von Erklärungen.

Um die Nullhypothese zu überprüfen, wurde ein Mann-Whitney-U-Test mit 5 % Signifikanz durchgeführt. Dabei ergaben sich die folgenden U-Werte: $U_A = 102$ (A = Gruppe ohne Erklärungen), $U_B = 94$ (B = Gruppe mit Erklärungen). Durch den Vergleich des kleineren U-Wertes mit dem der Tabelle für kritische Werte entnommenen Wert (Anhang C) ergibt sich folgender Vergleich: $U_{krit} = 61 > 94 = U_B$. Da dieser Vergleich offensichtlich nicht hält, kann die Nullhypothese nicht abgelehnt werden und es kann somit keine statistische Signifikanz nachgewiesen werden.

Bei der Berechnung des Pearson-Korrelationskoeffizienten ergibt sich ein r-Wert von ca. $r = 0,0347$, was nach der Interpretation von Cohen für keinen Effekt spricht ([4]). Eine grafische Darstellung der erfassten Passwortstärkedifferenzen kann in Abbildung 6.6 gefunden werden.

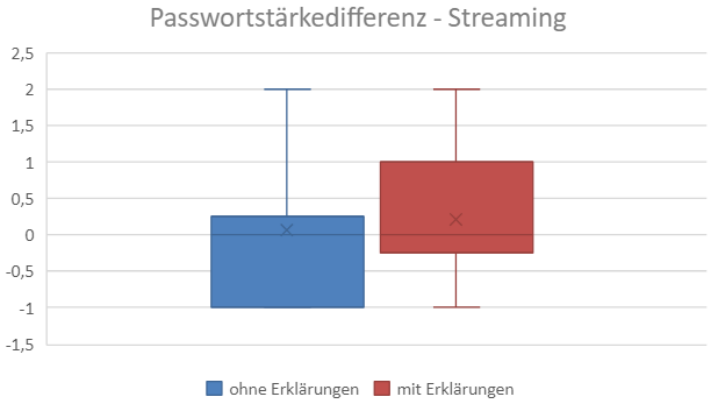

Abbildung 6.6: Box-and-Whisker-Plot der Differenzen der erfassten Passwortstärken für das Szenario Streaming.

6.2.6 Passwortstärkedifferenz, Szenario: Onlinebanking

Letztlich sollten die Probanden sich während der Studie auch sowohl mit einem bereits genutzten Passwort als auch mit einem neu gewählten Passwort beim Onlinebanking anmelden bzw. registrieren. Dabei waren die Probanden weiterhin in zwei Gruppen aufgeteilt: eine Gruppe ohne weitere Erklärungen

und eine Gruppe mit weiteren Erklärungen. Es wurden erneut jeweils die Stärken der von den Probanden gewählten Passwörter gespeichert, sodass hier die Differenz in der Passwortstärke zwischen zuvor genutztem Passwort und neu gewähltem Passwort ausgewertet werden kann. Dies ist die dazu konstruierte Nullhypothese:

H_0: Es gibt keine größere Verbesserung in der Stärke bei der Neuwahl eines Passworts für Onlinebanking durch das Hinzufügen von Erklärungen.

Erneut wurde zur Überprüfung der Nullhypothese ein Mann-Whitney-U-Test mit 5 % Signifikanz durchgeführt. Dabei ergaben sich folgende U-Werte: $U_A = 64,5$ (A = Gruppe ohne Erklärungen), $U_B = 131,5$ (B = Gruppe mit Erklärungen). Der Vergleich mit dem kritischen Wert aus der Tabelle ist folgender: $U_{krit} = 61 > 64,5 = U_A$. Dieser Vergleich hält offensichtlich nicht. Somit kann die Nullhypothese nicht abgelehnt werden und es kann keine statistische Signifikanz nachgewiesen werden.

Es ergibt sich ein Pearson-Korrelationskoeffizient von ca. $r = 0,2909$, was nach der Interpretation von Cohen für einen kleinen, fast schon mittleren Effekt spricht ([4]). Auffallend ist hierbei, dass der Effekt jedoch entgegen der erwarteten Richtung auftritt. Dies wird im nächsten Kapitel (7) besprochen. Eine grafische Darstellung der Differenzen der erfassten Passwortstärken ist in Abbildung 6.7 zu finden.

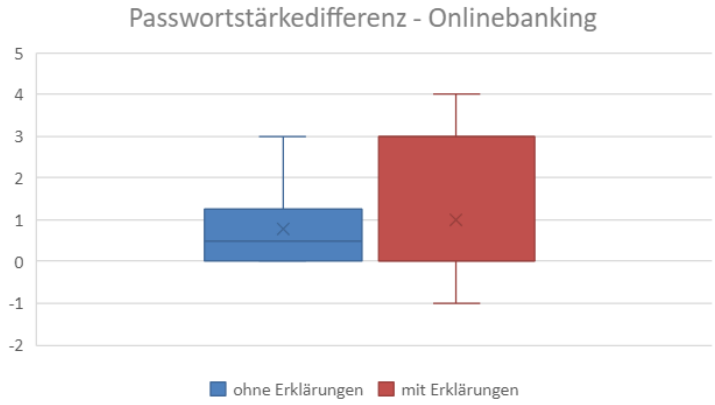

Abbildung 6.7: Box-and-Whisker-Plot der Differenzen der erfassten Passwortstärken für das Szenario Onlinebanking.

6.2.7 Absolute Passwortstärke, Szenario: StudIP

Bei der für die Neuregistrierung im Uninetzwerk StudIP gespeicherten Passwortstärke, kann nicht nur die Differenz zum zuvor verwendeten Passwort betrachtet werden, sondern auch die absolute Stärke des neuen Passwortes. Dies wird erneut, aufgeteilt in zwei Gruppen, im Folgenden durch Überprüfung dieser Nullhypothese gemacht:

H_0: Es werden keine stärkeren Passwörter bei der Neuwahl eines Passworts für das System StudIP durch das Hinzufügen von Erklärungen gewählt.

Um die Nullhypothese zu überprüfen, wurde ein Mann-Whitney-U-Test mit 5 % Signifikanz durchgeführt. Dabei ergaben sich die folgenden U-Werte: $U_A = 126$ (A = Gruppe ohne Erklärungen), $U_B = 70$ (B = Gruppe mit Erklärungen). Durch den Vergleich des kleineren U-Wertes mit dem der Tabelle für kritische Werte entnommenen Wert (Anhang C) ergibt sich folgender Vergleich: $U_{krit} = 61 > 70 = U_B$. Da dieser Vergleich offensichtlich nicht hält, kann die Nullhypothese nicht abgelehnt werden und es kann somit keine statistische Signifikanz nachgewiesen werden.

Bei der Berechnung des Pearson-Korrelationskoeffizienten ergibt sich ein r-Wert von ca. $r = 0,2431$, was nach der Interpretation von Cohen für einen kleinen, fast schon mittleren Effekt spricht ([4]). Eine Darstellung der absoluten Passwortstärken, welche bei der Neuregistrierung erfasst wurden, kann in Abbildung 6.8 gefunden werden.

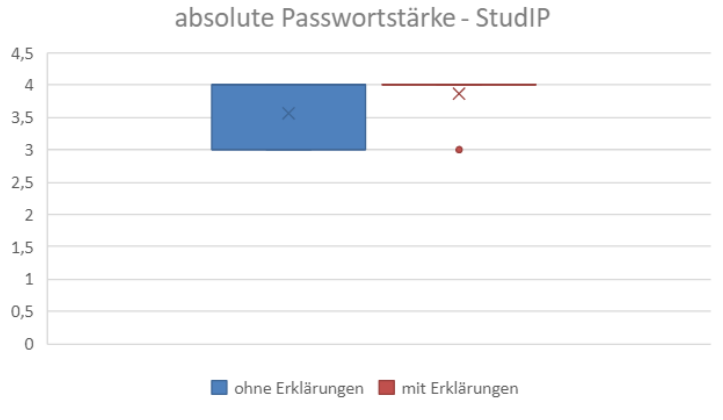

Abbildung 6.8: Box-and-Whisker-Plot der absoluten erfassten Passwortstärken für das Szenario StudIP.

6.2.8 Absolute Passwortstärke, Szenario: Streaming

Auch bei der Neuregistrierung bei einem Streamingdienst kann nicht nur die Differenz in der Passwortstärke zum zuvor verwendeten Passwort betrachtet werden, sondern auch die absolute Stärke des neuen Passworts. Erneut waren die Probanden in zwei Gruppen aufgeteilt, um die Auswirkung der hinzugefügten Erklärungen zu untersuchen. Dabei wurde folgende Nullhypothese betrachtet:

H_0: Es werden keine stärkeren Passwörter bei der Neuwahl eines Passworts für einen Streamingdienst durch das Hinzufügen von Erklärungen gewählt.

Es wurde zur Überprüfung der Nullhypothese erneut ein Mann-Whitney-U-Test mit 5 % Signifikanz durchgeführt. Dabei ergaben sich die folgenden U-Werte: $U_A = 107$ (A = Gruppe ohne Erklärungen), $U_B = 89$ (B = Gruppe mit Erklärungen). Der passende Vergleich mit dem kritischen Wert ist folgender: $U_{krit} = 61 > 89 = U_B$. Da dieser Vergleich offensichtlich nicht hält, kann die Nullhypothese nicht abgelehnt werden und es kann somit keine statistische Signifikanz nachgewiesen werden.

Der berechnete Pearson-Korrelationskoeffizient entspricht ca. $r = 0,0781$, was nach der Interpretation von Cohen für keinen oder einen nur sehr geringen Effekt spricht ([4]). Die Abbildung 6.9 zeigt die bei der Neuregistrierung erfassten Passwortstärken für dieses Szenario.

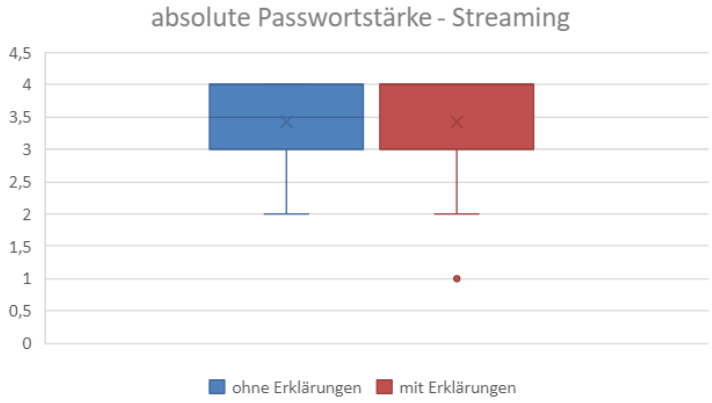

Abbildung 6.9: Box-and-Whisker-Plot der absoluten erfassten Passwortstärken für das Szenario Streaming.

6.2.9 Absolute Passwortstärke, Szenario: Onlinebanking

Auch für die Neuregistrierung beim Onlinebanking soll die absolute Passwortstärke ausgewertet werden. Neben der Aufteilung in zwei Gruppen soll dabei die folgende Nullhypothese dienlich sein:

> **H_0:** Es werden keine stärkeren Passwörter bei der Neuwahl eines Passworts für Onlinebanking durch das Hinzufügen von Erklärungen gewählt.

Auch hier wurde ein Mann-Whitney-U-Test mit 5 % Signifikanz durchgeführt. Die daraus resultierenden U-Werte sind diese: $U_A = 113,5$ (A = Gruppe ohne Erklärungen), $U_B = 82,5$ (B = ruppe mit Erklärungen). Es ergibt sich folgender Vergleich durch Entnahme des passenden kritischen Werts aus der Tabelle (Anhang C): $U_{krit} = 61 > 82,5 = U_B$. Da auch dieser Vergleich offensichtlich nicht hält, kann die Nullhypothese nicht abgelehnt werden und es kann somit keine statistische Signifikanz nachgewiesen werden.

Bei der Berechnung des Pearson-Korrelationskoeffizienten ergibt sich ein r-Wert von ca. $r = 0,1346$, was nach der Interpretation von Cohen für einen kleinen Effekt spricht ([4]). Im Folgenden ist in Abbildung 6.10 eine grafische Darstellung des Szenarios Onlinebanking für die bei der Neuregistrierung erhobenen Passwortstärken zu sehen.

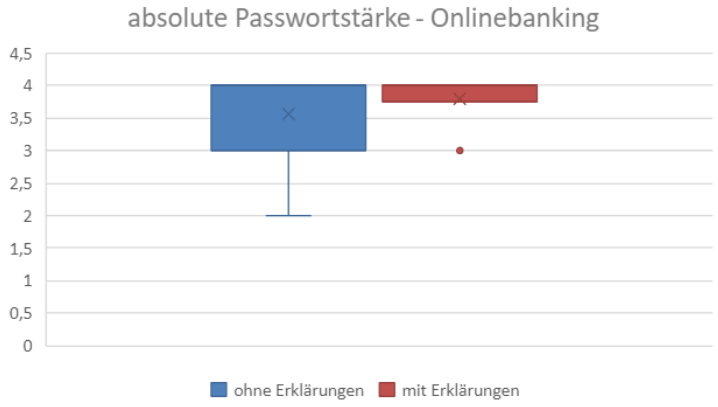

Abbildung 6.10: Box-and-Whisker-Plot der absoluten erfassten Passwortstärken für das Szenario Onlinebanking.

6.2.10 UMUX

Die Probanden teilten im Zuge der Studie auch ihre Meinung zu vier Fragen, welche sich mit Nutzbarkeit befassten, mit. Dabei waren sie weiterhin

aufgeteilt in eine Gruppe ohne weitere Erklärungen und eine Gruppe mit weiteren Erklärungen. Es wurde die Zustimmung zu vier Aussagen mit einer 7-Punkt-Likert-Skala erhoben (Genaueres in Kapitel 5). Aus diesen vier Fragen wurde dann der UMUX berechnet ([9]). Da die Richtung eines eventuellen Effekts auf die Nutzbarkeit unklar war, wurde hier eine zweiseitige Nullhypothese konstruiert:

H_0: Es gibt keinen Unterschied in der Nutzbarkeit eines Registrierungssystems durch das Hinzufügen von Erklärungen zur Passwortstärke.

Zur Überprüfung der Nullhypothese wurde ein weiterer Mann-Whitney-U-Test mit 5 % Signifikanz durchgeführt. Dabei ergaben sich die folgenden U-Werte: $U_A = 76$ (A = Gruppe ohne Erklärungen), $U_B = 120$ (B = Gruppe mit Erklärungen). Da es sich hier um eine zweiseitige Nullhypothese handelt, wurde die passende Tabelle genutzt, um den kritischen Wert für den Vergleich zu erhalten (Anhang C): $U_{krit} = 55 > 76 = U_A$. Da auch dieser Vergleich offensichtlich nicht hält, kann die Nullhypothese nicht abgelehnt werden und es kann somit keine statistische Signifikanz nachgewiesen werden.

Es ergibt sich ein Pearson-Korrelationskoeffizient von ca. $r = 0,1910$, was nach der Interpretation von Cohen für einen kleinen Effekt spricht ([4]). Die Richtung des Effekts zeigt dabei, dass die hinzugefügten Erklärungen eher für eine schlechtere Nutzbarkeit sorgen. Im Folgenden ist mit Abbildung 6.11 eine grafische Darstellung der berechneten UMUX-Werte zu sehen.

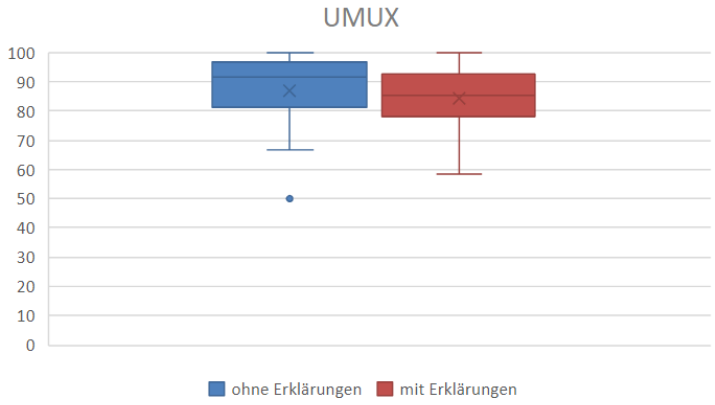

Abbildung 6.11: Box-and-Whisker-Plot des berechneten UMUX-Wertes

6.3 Weitere Anmerkungen der Probanden

Während der Studie gab es die Möglichkeit, schriftliche Anmerkungen zu hinterlassen. Zusätzlich gab es mehrere mündliche Anmerkungen von Probanden. Hier werden im Folgenden einige ausgewählte Anmerkungen hervorgehoben, während eine vollständige Liste der schriftlichen Anmerkungen im Anhang zu finden ist (B.5).

Mündlich wurden von verschiedenen Probanden insgesamt zwei verschiedene Anmerkungen gemacht. Zum einen wurde angemerkt, dass das Bestätigen der Anmeldung/Registrierung nicht durch die Enter-Taste möglich sei, wie es eigentlich der Standard wäre. Eine weitere Anmerkung war, dass das Durchschalten der Eingabefelder mit der Tabulator-Taste nicht richtig funktioniere. Beides sind kleine Fehler des Systems, welche jedoch eventuell die Ergebnisse beeinflusst haben.

Nachfolgend sind einige ausgewählte schriftliche Anmerkungen zu finden:

1. „Lange Passwörter waren nicht gut sichtbar, da das Fenster eine feste Größe hatte.“

2. „Man könnte es so machen, dass man bei der Registrierung das Passwort erneut eingeben muss, ohne dass man es copy-pasten kann.“

3. „Weitere Informationen, wie man die Passwort-Stärke verbessern könnte wären relativ hilfreich gewesen.“

4. „Tipps, welche Eingaben die Passwortstärke beeinflussen wären hilfreich gewesen.“

Die 1. Anmerkung sollte wahrscheinlich darauf hinweisen, dass sehr lange Passwörter nicht vollständig im Eingabefeld einsehbar waren. Dies hätte leicht verbessert werden können und stellt eine leichte Nutzbarkeitseinschränkung dar. Diese Einschrämkung war jedoch in beiden Gruppen vorhanden und sollte somit nur einen minimalen Effekt auf die Ergebnisse gehabt haben. Die 2. Anmerkung wurde in verschiedener Form noch drei weitere Male gemacht und bezieht sich auf das „Passwort wiederholen“-Feld. So umgeht das Kopieren des Passworts in dieses Feld den Nutzen des Feldes. Dies könnte das Vertrauen in das System oder die Nutzbarkeit beeinflusst haben. Auch dies war jedoch in beiden Gruppen vorhanden, wodurch eventuelle Effekte auf die Ergebnisse minimal ausfallen sollten. Die 3. Anmerkung wurde von einer Person aus Gruppe ohne weitere Erklärungen gemacht, welche sich eben diese Erklärungen zu wünschen scheint, die die andere Gruppe hatte und welche hier untersucht wurden. Ein ähnlicher Kommentar ist die 4. Anmerkung, welche jedoch von einer Person aus der Gruppe mit Erklärungen gemacht wurde. Während der Kommentar aus der Gruppe ohne Erklärungen (3.) die Relevanz von Erklärungen in diesem Kontext weiter zu bestätigen scheint, widerspricht der Kommentar aus der anderen

Gruppe (4.) diesem. Der Kommentar aus der Gruppe mit Erklärungen (4.) kann jedoch auch so interpretiert werden, dass der Proband die Richtlinien mit Erklärungen gar nicht bemerkt hat, da sie dynamisch erst nach einer Eingabe eingeblendet wurden. Demnach könnte dies auf die Notwendigkeit einer besseren Sichtbarkeit der Richtlinien hindeuten.

Kapitel 7

Diskussion

In diesem Kapitel wird die Bedeutung der im vorherigen Kapitel vorgestellten Ergebnisse diskutiert. Zunächst werden die in Kapitel 5 gestellten Forschungsfragen einzeln beantwortet und die Bedeutung dieser Antworten besprochen. Nachdem die Erkenntnisse noch einmal zusammengefasst wurden, wird noch auf eventuelle Gefahren für die Validität der Ergebnisse eingegangen und es werden Herausforderungen bei der Durchführung dieser Forschung besprochen.

7.1 Beantwortung der Forschungsfragen

7.1.1 Vertrauen und Verständlichkeit

Zwei wichtige in dieser Arbeit erforschte Aspekte sind die Verständlichkeit und die Vertrauenswürdigkeit von Registrierungsdialogen. Dafür sollte folgende Forschungsfrage untersucht werden:

> **RQ1:** Welchen Einfluss haben Erklärungen in Registrierungsdialogen auf die Verständlichkeit und die Vertrauenswürdigkeit des Systems?

Zu Beginn dieser Arbeit stand die Vermutung, dass Erklärungen in Registrierungsdialogen zu einer Erhöhung der Verständlichkeit und der Vertrauenswürdigkeit führen könnten. Dies wurde durch die Studie überprüft (siehe Kapitel 5) und die Ergebnisse wurden in Kapitel 6 präsentiert.

Die in der Studie gesammelten Ergebnisse zeigen weder bei der Verständlichkeit noch bei der Vertrauenswürdigkeit einen statistisch signifikanten Unterschied. Das bedeutet, dass nicht mit Sicherheit gesagt werden kann, dass die Ergebnisse nicht auf einem Zufall beruhen. Dennoch scheinen Tendenzen in positiver Richtung erkennbar zu sein. So scheint zwar das Hinzufügen von Erklärungen keine Auswirkungen auf das Vertrauen der

Nutzer gegenüber dem System zu haben. Bei der Auswirkung auf die Verständlichkeit und dem Wunsch nach mehr Informationen ist jedoch ein Effekt zu erkennen. Demnach scheint das Hinzufügen von Erklärungen zu Registrierungsdialogen eine leichte Verbesserung der Verständlichkeit und eine leichte Verminderung des Wunsches nach mehr Informationen zu bewirken.

7.1.2 Passwortstärke

Ein weiterer in dieser Arbeit untersuchter Aspekt ist die Auswirkung der hinzugefügten Erklärungen auf die Stärke der von Nutzern gewählten Passwörter. Dafür wurden innerhalb der Studie sowohl von den Nutzern bereits genutzte Passwörter als auch neu gewählte Passwörter erhoben. Wie in Kapitel 5 erläutert, wurden dazu drei Szenarien mit unterschiedlichen Software-Systemen verglichen. Die folgende Frage sollte in diesem Zusammenhang untersucht werden:

RQ2: Wie beeinflussen Erklärungen in Registrierungsdialogen die Sicherheit der von Nutzern gewählten Passwörter?

Um RQ2 zu beantworten, wurde im Kapitel 6 in jedem Szenario jeweils die Differenz zwischen den beiden erhobenen Passwortstärken und auch die absolute Stärke der neu gewählten Passwörter untersucht. Bei der Betrachtung der Differenz zwischen den beiden Passwörtern fällt auf, dass in keinem der drei Szenarien eine statistische Signifikanz nachgewiesen werden kann, jedoch sind Trends erkennbar. Zwar ist in dem Szenario Streamingdienst kein Effekt zu erkennen, aber in den Szenarien StudIP und Onlinebanking sind jeweils Effekte zu erkennen. Diese Effekte scheinen jedoch der ursprünglich angenommenen Richtung entgegenzuwirken. So könnte man auf Grundlage dieser Ergebnisse vermuten, dass Erklärungen eher für schwächere Passwörter in einigen Szenarien sorgen. Dies ist allerdings ein Trugschluss. Dies wird klar, wenn man auch die Ergebnisse zur absoluten Passwortstärke der neuen Passwörter betrachtet. Auch bei diesen können Effekte beobachtet werden, welche aber nicht statistisch signifikant sind. So ist auch bei den absoluten Passwortstärken im Szenario Streaming kein Effekt zu erkennen, jedoch ist sowohl bei dem Szenario StudIP als auch bei dem Szenario Onlinebanking ein schwacher Effekt zu erkennen. Dieser Effekt wirkt hier nun auch in der erwarteten Richtung und spricht somit für eine Erhöhung der Passwortstärke durch das Hinzufügen von Erklärungen.

Diese im ersten Moment widersprüchlichen Ergebnisse haben einen simplen Grund. So ist die Untersuchung der Differenz, auch wenn sie auf den ersten Blick sinnvoll erscheint, hier irreführend. Demnach haben bei

der zufälligen Zuweisung der Gruppen mehr Probanden in der Gruppe ohne Erklärungen verhältnismäßig schwächere Passwörter verwendet und haben somit auch mehr Raum gehabt, um diese Passwörter in der Stärke zu verbessern. Somit bilden, in der Art, wie die Studie hier durchgeführt wurde, die Differenzen keine geeignete Metrik, um die Forschungsfrage zu evaluieren.

Ebenfalls fällt auf, dass lediglich bei sicherheitskritischen Systemen eine Verbesserung zu beobachten ist. Bei dem vergleichsweise weniger kritischen Szenario Streamingdienst war keine Erhöhung der Passwortstärke, weder absolut noch in der Differenz, zu beobachten. Alles in allem scheinen Erklärungen in Registrierungsdialogen bei sicherheitskritischen Systemen zu leicht stärkeren Passwörtern zu führen.

7.1.3 Nutzbarkeit

Der letzte in dieser Arbeit untersuchte Aspekt ist die Nutzbarkeit von Registrierungsdialogen mit Erklärungen. Bei der Nutzbarkeit war vor der Durchführung der Studie noch nicht absehbar, in welche Richtung die Erklärungen die Nutzbarkeit eventuell beeinflussen würden. Sowohl ein negativer Einfluss als auch ein positiver Einfluss auf die Nutzbarkeit war denkbar (siehe Kapitel 5). Die dazugehörige Forschungsfrage war wie folgt formuliert:

RQ3: Welche Auswirkungen haben Erklärungen in Registrierungsdialogen auf die Nutzbarkeit des Systems?

Um diese Forschungsfrage zu überprüfen, wurde in der Studie der sogenannte UMUX, eine Nutzbarkeitsmetrik, erhoben. Bei der Auswertung dieser Metrik ergibt sich, wie in Kapitel 6 zu sehen, erneut keine statistische Signifikanz, jedoch ist ein Trend zu erkennen. Dieser verläuft in die Richtung, dass die Nutzbarkeit durch die Erklärungen negativ beeinflusst wird, jedoch ist nur eine geringe Effektstärke zu sehen. Somit kann gesagt werden, dass das Hinzufügen von Erklärungen zu Registrierungsdialogen eine leicht geringere Nutzbarkeit zur Folge hat.

7.1.4 Zusammenfassung der Erkenntnisse

Um die Ergebnisse noch einmal zusammenzufassen, kann gesagt werden, dass das Hinzufügen von Erklärungen zu Registrierungen die Verständlichkeit leicht erhöht, für eine leicht höhere Passwortstärke bei sicherheitskritischen Systemen sorgt und für eine leicht geringere Nutzbarkeit zur Folge hat. Auf die Vertrauenswürdigkeit und die Passwortstärke in weniger sicherheitskritischen Systemen hat das Hinzufügen von Erklärungen hingegen keine

Auswirkung. Diese Tendenzen wurden innerhalb der Studie beobachtet, sind jedoch nicht von statistischer Signifikanz gestützt.

7.2 Gefahren für die Validität

Es ergeben sich für die in dieser Arbeit gesammelten Erkenntnisse einige Limitierungen und Gefahren für ihre Validität. Zur Erhebung dieser wird die Einteilung in vier Kategorien nach Wolin et al. [29] genutzt.

Eine solche Kategorie ist die „construct validity“, welche sich mit Gefahren zur Verbindung von den Studienergebnissen mit der dazugehörigen, zu überprüfenden Theorie befasst. Eine solche Gefahr ist die Tatsache, dass es im Rahmen dieser Arbeit nicht möglich war, eine vollständige Literaturrecherche durchzuführen. So wurde das „Snowballing“ nach dem ersten Schritt als saturiert angenommen (nach [30]). Dies war wegen der Rahmenbedingungen dieser Arbeit notwendig, da für ein größeres „Snowballing“ der Zeitaufwand zu groß gewesen wäre. Auch wurde versucht, eventuelle Lücken in dieser Literaturrecherche mithilfe eine Datenbanksuche und auch mithilfe einer unstrukturierten Suche nach Literatur zu schließen. Jedoch kann so nicht sichergestellt werden, dass sämtliche für diese Arbeit relevante Literatur erfasst wurde. Somit könnten sich eventuell Lücken in der Literatur ergeben, welche der Grund dafür sein könnten, dass vielleicht wichtige Aspekte der Thematik übersehen wurden und somit auch die Forschungsfragen eventuell nicht voll umfassend bzw. nicht mit allen gängigen Methoden überprüft wurden.

Eine weitere Kategorie ist die „internal validity“ , welche sich mit weiteren Einflüssen auf die Ergebnisse beschäftigt. So wurde explizit eine zufällige Reihenfolge der verschiedenen Szenarien genutzt, um so dem Einfluss eines eventuellen Lerneffekts auf die Ergebnisse entgegenzuwirken. Die Art der Teilnehmergewinnung könnte jedoch einen Einfluss auf die Ergebnisse gehabt haben. So wurden zur Teilnehmergewinnung lediglich bekannte und leicht erreichbare Personen erwogen. Somit war die Auswahl an Probanden nicht zufällig, was die Ergebnisse beeinflusst haben könnte.

Auch die „external validity“ ist ein wichtiger Aspekt. Dabei geht es darum, inwiefern die Ergebnisse generalisierbar sind. Als Gefahr bei dieser Art der Validität ist die Demografie der Probanden anzuführen. Der Großteil der Probanden waren Studierende der Informatik, welche sich selbst als technisch versiert eingeschätzt haben. Dies könnte bedeuten, dass die beobachteten Effekte bei anderen Bevölkerungsgruppen in anderer Art auftreten, und dies ist somit eine Gefahr für die Validität.

Letztlich ist noch die „conclusion validity“ zu betrachten. Bei dieser geht es um die statistische Stärke der Ergebnisse. Weil gängige statistische Methoden genutzt wurden, sollte es hier nur eine geringe Gefährdung der Validität geben, jedoch ist eine eher geringe Teilnehmerzahl in gewissem

Maße eine Gefährdung. Es nahmen nur 28 Personen an der Studie teil. Es ist möglich, dass eventuelle Effekte bei einer höheren Teilnehmerzahl deutlicher bzw. stärker hervorgetreten wären. Außerdem wurden wie in Abschnitt 6.3 einige Einschränkungen im Prototypen angemerkt / entdeckt. Diese haben die gesehenen Effekte eventuell leicht beeinflusst, sodass die Effekte ohne diese Einschränkungen möglicherweise stärker zu sehen gewesen wären.

7.3 Herausforderungen

Zwar gibt es ausreichend Forschung zu Passwortstärkeanzeigen sowie auch zu Erklärbarkeit, jedoch nur sehr wenig zu beidem in Kombination. Darüber hinaus stammt ein großer Teil der Forschung zu Erklärbarkeit aus dem Bereich der künstlichen Intelligenz. So sind daraus resultierenden Ergebnisse nur teilweise oder sehr bedingt anwendbar, da sich wie in Kapitel 2.2 beschrieben die Erklärbarkeit im Zusammenhang mit der Softwaresicherheit und die aus der künstlichen Intelligenz unterscheiden. Dementsprechend war es die größte Herausforderung für diese Arbeit, genügend andere Arbeiten und somit genügend Wissen zu finden, welches sich sinnvoll verwenden lässt. Letztendlich wurde zwar eine ausreichende Menge erreicht, jedoch waren viele auf dem Weg dorthin gesichtete Arbeiten für diese Arbeit nicht hilfreich.

Kapitel 8

Verwandte Arbeiten

In diesem Kapitel wird zuerst diese Arbeit in den Forschungsbereich eingeordnet und ihre Bedeutung für diesen besprochen. Danach werden Arbeiten vorgestellt, welche in irgendeiner Art und Weise mit dieser Arbeit verwandt sind. Dies kann bedeuten, dass sie ein ähnliches Thema behandeln oder auch, dass sie Ergebnisse haben, welche sich mit denen dieser Arbeit decken.

8.1 Einordnung dieser Arbeit

Die in dieser Arbeit durchgeführte Forschung lässt sich in den Bereich der Sicherheit einordnen. Genauer gesagt in den Bereich von Passwörtern als Authentifizierungsmechanismus und den damit verbundenen Bereich der Passwortstärkeanzeigen. Während es in diesem Bereich bereits sehr viel Forschung gibt, ist dies im engsten Unterbereich dieser Arbeit nicht der Fall.

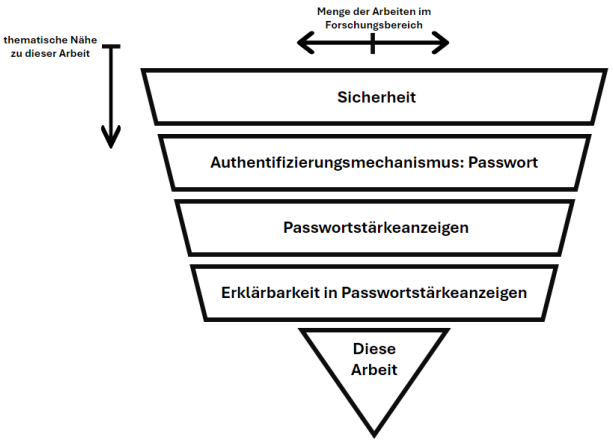

Abbildung 8.1: Schematische Darstellung zur Einordnung dieser Arbeit

So beschäftigt sich diese Arbeit mit der Erklärbarkeit im Zusammenhang mit Passwortstärkeanzeigen. Dadurch kann sie auch dem bislang nur wenig erforschten und relativ neuen Bereich der erklärbaren Sicherheit zugewiesen werden. Diese Arbeit zeigt tendenzielle Ergebnisse für Erklärungen in Registrierungsdialogen und bei Passwortstärkeanzeigen auf und macht so einen Schritt bei der Erforschung dieser Mechanismen. Diese Einordnung ist auch grafisch in der Abbildung 8.1 dargestellt.

8.2 Verwandte Arbeiten

Allgemein können alle Arbeiten, welche sich mit Passwörtern als Authentifizierungsmeachanismus beschäftigen, als verwandt zu dieser Arbeit angesehen werden. Da dieser Bereich jedoch enorm viel Forschung umfasst und er bereits in Kapitel 2 als Grundlage dieser Arbeit besprochen wurde, wird sich im Folgenden auf verwandte Arbeiten, welche sich mit Passwortstärkeanzeigen beschäftigen, und besonders auf solche, welche sich zusätzlich dazu mit Erklärbarkeit beschäftigen, fokussiert.

Allgemeine Arbeiten zur Passwortstärkeanzeigen

Bei den folgenden Arbeiten handelt es sich nur um eine kleine Auswahl von als besonders wichtig erachteten Arbeiten aus dem Forschungsbereich der Passwortstärkeanzeigen. Da dieser Bereich sehr viele Forschungsarbeiten umfasst, wird hier also kein Anspruch auf Vollständigkeit erhoben.

In „Does My Password Go up to Eleven? The Impact of Password Meters on Password Selection“ ([8]) untersuchen Egelman et al. die Wirksamkeit von Passwortstärkeanzeigen. Dazu wurden zwei Nutzerstudien durchgeführt. Bei einer wurden die Nutzer aufgefordert, das Passwort eines bereits genutzten Accounts neu zu wählen. Dabei hatte ein Teil der Probanden eine Passwortstärkeanzeige, während ein anderer Teil keine solche Anzeige hatte. Bei der zweiten Studie wurden die Probanden aufgefordert, sich bei einer bisher unbekannten Website neu zu registrieren. Diese Arbeit ist besonders relevant wegen ihrer Ergebnisse. Die Autoren kommen zu dem Schluss, dass Passwortstärkeanzeigen einen signifikanten Effekt auf die Passwortstärke bei der Neuwahl von Passwörtern für sicherheitskritische Konten haben, während bei eher unwichtigen Konten kein solcher Effekt erkennbar sei. In Anbetracht dieser Ergebnisse erscheint auch ein Teil der in Kapitel 7 besprochenen Ergebnisse schlüssig. So wurde festgestellt, dass die in dieser Arbeit untersuchten Erklärungen bei Registrierungen lediglich bei sicherheitskritischen Konten einen Effekt auf die Passwortstärke haben.

Es gibt bereits viel Forschung zu Passwortstärkeanzeigen, ihren Auswirkungen und wie man sie sinnvoll erweitern könnte. Ein gängiger Gedanke ist es dabei, das Feedback, welches die Passwortstärkeanzeige liefert, zu verbessern. Ein Beispiel für eine solche Arbeit ist „Enhancing Password

Security through Interactive Fear Appeals: A Web-based Field Experiment“ ([24]) von Vance et al. In dieser Arbeit wird die Erweiterung von Passwortstärkeanzeigen um sogenannte „Angst-Appelle“ untersucht. Dies sind Aussagen, die einen Nutzer über Angst dazu bringen sollen, ein stärkeres Passwort zu wählen. Ein Beispiel für einen solchen „Angst-Appell“ wäre folgende Aussage: „Ein Angreifer kann das Passwort, das sie eingegeben haben in ca. 8 Stunden erraten.“ Diese Arbeit kommt zu dem Ergebnis, dass „Angst-Appelle“ eine wirksame Erweiterung von Passwortstärkeanzeigen sind. „Angst-Appelle“ erscheinen ebenfalls als eine vielversprechende Erweiterung von Passwortstärkeanzeigen. Auch scheinen sie bei Betrachtung der Ergebnisse einen ähnlichen Effekt zu haben wie die in dieser Arbeit untersuchten Erklärungen. Es kann die Vermutung aufgestellt werden, dass der Effekt der untersuchten Erklärungen teilweise auf das Gleiche zurückzuführen ist, was auch die „Angst-Appelle“ wirksam macht. Es ist möglich, dass ein besseres Verständnis des Systems, das durch die Erklärungen erreicht wird, auch zu einem besseren Risikobewusstsein bei den Nutzern führt. Dies könnte dann dazu führen, dass die Erklärungen einen ähnlichen Effekt wie die „Angst-Appelle“ haben.

Arbeiten zu Erklärbarkeit in Passwortstärkeanzeigen

Es gibt einige Arbeiten, die sich bereits mit Erklärbarkeit und Passwortstärkeanzeigen beschäftigt haben. Auch an dieser Stelle muss erwähnt werden, dass die vorgestellten Arbeiten lediglich einen Auszug zu diesem Thema darstellen. Da jedoch in diesem spezifischen Forschungsbereich nur sehr wenige Arbeiten existieren, dürfte die Auswahl an vorgestellten Arbeiten dennoch einen großen Teil der relevanten Arbeiten abdecken.

Xu und Han stellen in ihrer Arbeit „An Explainable Password Strength Meter Addon via Textual Pattern Recognition“ ([31]) einen weiteren Algorithmus zur Bestimmung der Passwortstärke vor, welcher zur Erweiterung von Passwortstärkeanzeigen gedacht ist, um ihre Nutzbarkeit und Verständlichkeit zu erhöhen. Dieser Algorithmus benutzt auch Mustererkennung, ähnlich wie der „zxcvbn“-Algorithmus, versucht dabei aber, aufbauend auf diesem, eine genauere Bestimmung der Passwortstärke zu erreichen. Auch könnte der Algorithmus leicht seine Berechnungen offenlegen, um einem Nutzer Feedback zu gewähren. Dazu wird der neue Algorithmus mit den Algorithmen der Passwortstärkeanzeigen der Top 10 Webseiten verglichen. Die Schlussfolgerung ist, dass diese Webseiten nur sehr wenig Mustererkennung verwenden und die meisten keine Erklärungen liefern. Innerhalb einer Nutzerstudie sollte die Effektivität des Algorithmus überprüft werden, indem Probanden mehrere Fragen zu den gesuchten Mustern und ihren Passwörtern gestellt wurden. Das Ergebnis ist, dass viele Nutzer solche benutzen würden, wodurch der Algorithmus sinnvoll sei, und dass viele bereits bestehende Passwortstärkeanzeigen von weiteren Erklärungen

profitieren könnten. Der hier vorgestellte Algorithmus könnte eine gute Grundlage für weitere erklärbare Passwortstärkeanzeigen bieten.

In „An explainable online password strength estimator“ ([5]) stellen David und Wool einen Algorithmus zur Bestimmung von Passwortstärke vor, welcher nach eigener Aussage erklärbar sei. Dieser „PESrank“ genannte Algorithmus nutzt dabei für seine Stärkeeinschätzung (ähnlich wie „zxcvbn“), Mustererkennung und zum Training einen großen Korpus an bekannten Passwörtern. Der Algorithmus hat dabei die Möglichkeit, einem Nutzer Feedback zur Erhöhung der Passwortstärke in Form eines Tipps zu geben. Ein Beispiel für einen solchen Tipp ist in der Arbeit ([5]) wie folgt gegeben:

„Your password is sub-optimal, its guessability strength is 32 bits, based on 905 million leaked passwords. Your password is based on the leaked word: 'newyork' that was used by 129,023 people. It uses a suffix that was used by 17,631,940 people. It uses a capitalization pattern that was used by 592,568 people. It uses a l33t pattern that was used by 4,395,598 people.“

Wie in diesem Tipp zu erkennen ist, ist das gegebene Feedback zwar von Zahlen gestützt, jedoch für Laien eher unverständlich. So umfangreich wie der gegebene Tipp ist, ist er für den durchschnittlichen Nutzer komplett unverständlich. So ist davon auszugehen, dass er einen solchen Nutzer eher verwirren und eventuell sogar frustrieren würde. Dahingehend müssten die Tipps also noch angepasst werden, wenn „PESrank“ außerhalb einer Studie Anwendung finden würde. Außerdem wird in der Arbeit von David und Wool ([5]) die Eigenschaft der Erklärbarkeit offensichtlich anders verstanden als in dieser Arbeit. Die in Kapitel 4 erdachten Erklärungen sind auf technisch nicht besonders bewanderte Nutzer abgestimmt und haben das Ziel, diesem den Grund, warum etwas wichtig ist, zu erläutern. Die Erklärungen von „PESrank“ scheinen hingegen den Fokus auf ausführlicherem Feedback für Experten zu legen.

„Design and Evaluation of a Data-Driven Password Meter“ ([23]) ist eine Arbeit von Blase et al. In dieser schaffen sie eine Passwortstärkeanzeige, welche die Passwortstärke mithilfe einiger Heuristiken (orientiert am „zxcvbn“-Algorithmus) und einem neuronalen Netz beurteilt. Neben dem üblichen Balken, welcher die Passwortstärke anzeigt, beinhaltet die Anzeige Feedback zur Verbesserung in der Form von Richtlinien. Neben diesen Richtlinien ist ein Knopf zu finden, der eine ausführliche Erklärung sichtbar werden lässt. Außerdem kann auf Knopfdruck ein Vorschlag für ein besseres Passwort geliefert werden. Innerhalb einer groß angelegten Studie wurden viele verschiedene grafische Elemente der Passwortstärkeanzeige variiert. Dabei wurden die Probanden angewiesen, sich ein Passwort für ein wichtiges Konto zu erstellen. Dabei wurden die Passwortstärke, die Interaktion mit grafischen Elementen, die benötigte Zeit und die Erinnerungsfähigkeit (engl. memorability) an das gewählte Passwort gemessen. Es wird zu dem Schluss gekommen, dass ausführliches Text-Feedback (in diesem Fall Erklärungen) für eine höhere Passwortstärke bei den von Nutzern gewählten Passwörtern

sorge, während es kaum Einfluss auf die Erinnerungsfähigkeit habe und einige spezifische Nutzbarkeitsaspekte leicht negativ beeinflusse.

Die Arbeit von Blase et al. ([23]) ist dieser in einigen Punkten ähnlich. So wurde auch in dieser Arbeit eine Passwortstärkeanzeige erweitert und es wurde ein dem „zxcvbn"-Algorithmus ähnlicher, selbst gebauter Algorithmus verwendet. Auch waren in der entwickelten Passwortstärkeanzeige Erklärungen vorhanden, die jedoch erst nach aktiver Interaktion angezeigt wurden und nicht Hauptbestandteil der Forschung waren. Es wurde eine Nutzerstudie durchgeführt, um die Effekte dieser Erweiterung zu beurteilen. Dabei wurden Probanden dazu aufgefordert, so zu handeln, als ob sie ein neues Passwort für ihr wichtigstes E-Mail Konto wählen würden. So ähneln die entwickelten Passwortstärkeanzeigen von dieser Arbeit und die von der Arbeit von Blase et al. einander. Dabei unterscheiden sich Methodik und Forschungsziel. So lag der Fokus in der Arbeit von Blase et al. eher darauf, den Effekt einzelner Aspekte der Passwortstärkeanzeige zu testen, anstatt die Auswirkungen von Erklärungen. Auch die in der Studie gesammelten und untersuchten Daten unterscheiden sich demnach von dieser Arbeit. So wurden neben der Passwortstärke des zuvor beschriebenen Szenarios noch die Erinnerungsfähigkeit getestet. Des Weiteren wurden Fragen zur Erhebung einiger spezifischer Nutzbarkeits-Teilaspekte gestellt. Bei den Ergebnissen kommt die Arbeit von Blase et al. zu einem ähnlichen Ergebnis, was die Passwortstärke angeht. Demnach würde die Passwortstärke durch Erklärungen bei wichtigen Konten erhöht, was sich mit Ergebnissen dieser Arbeit deckt. Die Arbeit weist auch interessante, in dieser Arbeit nicht getestete Ergebnisse bei der Erinnerungsfähigkeit auf. Demnach sei diese kaum beeinflusst. Weiterhin gibt es teilweise überschneidende Ergebnisse bezüglich der Nutzbarkeit und zwar, dass Teile der Nutzbarkeit negativ beeinflusst würden, während in dieser Arbeit festgestellt wurde, dass die Nutzbarkeit als Ganzes negativ beeinflusst wurde.

Kapitel 9

Zusammenfassung und Ausblick

9.1 Zusammenfassung

Die heutzutage am häufigsten verwendete Authentifizierungsmethode ist die der Passwörter. Diese sind, wie bereits mehrmals gezeigt wurde, nicht ohne Fehler. So wird stetig an der Verbesserung von Passwörtern geforscht, wie auch in dieser Arbeit. Eine Möglichkeit, auf Passwörtern aufzubauen, sind Passwortstärkeanzeigen, welche über das Anzeigen der Passwortstärke Nutzer zu stärkeren Passwörtern führen sollen. Eine andere vielversprechende Thematik im Sicherheitsbereich allgemein ist die Erklärbarkeit. Hierbei steht im Fokus, dem Nutzer ein System oder einen Prozess verständlich zu machen. So wurden diese zwei Bereiche miteinander in Verbindung gebracht, indem Passwortstärkeanzeigen um Erklärungen erweitert wurden. Genauer gesagt war es Ziel dieser Arbeit, die Auswirkungen von Erklärungen in Registrierungsdialogen zu untersuchen, um womöglich den Prozess der Passwortwahl weiter zu verbessern.

Zu diesem Zweck wurde zuerst eine strukturierte Literaturrecherche durchgeführt, um die bereits bestehende Forschung in den relevanten Bereichen zu sichten. Ergänzend wurde auch mit einer Datenbanksuche und unstrukturiertem Suchen gearbeitet. Auf Basis der in der Literaturrecherche gesammelten Erkenntnisse wurde dann ein Konzept für erklärbare Registrierungsdialoge entwickelt. Dieses Konzept wurde im nächsten Schritt in Form eines Prototyps umgesetzt. Dieser wurde im Folgenden genutzt, um eine Nutzerstudie durchzuführen. Diese hatte insgesamt 28 Teilnehmer. Dabei wurden zwei Gruppen von Probanden Registrierungsdialoge gezeigt, wobei eine Gruppe ausführliche Erklärungen in ihren Registrierungen erhielt. Die andere Gruppe erhielt hingegen keinerlei Erklärungen und diente somit als Vergleich. In dieser Studie wurden drei zuvor formulierte Forschungsfragen untersucht. Diese befassten sich mit der Sicherheit der gewählten Passwörter

in unterschiedlichen wichtigen Szenarien und der Nutzbarkeit sowie der Vertrauenswürdigkeit und Verständlichkeit des Systems.

Bei der statistischen Auswertung der Ergebnisse stellte sich heraus, dass zwar keine statistische Signifikanz nachweisbar war, jedoch deutliche Tendenzen sichtbar wurden. Nach diesen Tendenzen sorgen Erklärungen in Registrierungsdialogen für ein besseres Verständnis des Systems, eine höhere Passwortstärke bei sicherheitskritischen Anwendungen und eine geringere Nutzbarkeit des Systems. Es könnte also besonders bei sicherheitskritischen Anwendungen sinnvoll sein, Registrierungsdialoge um Erklärungen zu erweitern.

9.2 Ausblick

Während in dieser Arbeit einige wichtige Erkenntnisse gesammelt wurden, gibt es noch weitere Fragen in Verbindung mit diesen, die erforscht werden sollten. Aus diesem Grund werden im Folgenden einige noch benötigte Forschungsschritte und auch weitere neue Forschungsthemen, welche im Zusammenhang mit dieser Arbeit stehen, beschrieben.

- **Statistische Signifikanz sicherstellen**
  Es sollte weitere Forschung durchgeführt werden, um die statistische Signifikanz der hier gesehenen Tendenzen zu sichern. Dabei könnten unter anderem eine größere Menge von Probanden sowie eine zufälligere Zusammenstellung dieser hilfreich sein. Ohne das Nachweisen der statistischen Signifikanz könnten die beobachteten Effekte eventuell nur auf Zufällen beruhen, auch wenn dies nicht besonders wahrscheinlich ist, ist es jedoch nicht gänzlich auszuschließen.

- **Technische Versiertheit berücksichtigen**
  Es sollte durch eine andere Zusammenstellung der Probanden der Einfluss von Erklärungen in Registrierungsdialogen auf weniger technisch versierte Personen überprüft werden. Da alle Teilnehmer der in dieser Arbeit durchgeführten Studie sich selbst als technisch versiert eingeschätzt haben, ist ein Effekt dieser Tatsache auf die gesammelten Daten noch unklar. Es wäre denkbar, dass weniger technisch versierte Personen stärker durch die Erklärungen beeinflusst würden.

- **Nutzungskontext des Systems untersuchen**
  Eine größere Untersuchung mit verschiedenen Systemen könnte interessante Ergebnisse liefern. So könnte genauer untersucht werden, in welchem Maße ein System sicherheitskritisch sein muss, um den Effekt auf die Passwortstärke zu beobachten, oder ob der genaue Nutzungskontext bzw. das Umfeld während der Nutzung des Systems irgendeinen Einfluss auf die Wirksamkeit der Erklärungen hat.

- **Konzept weiter verbessern**
  Es könnte versucht werden, das in dieser Arbeit entwickelte Konzept für erklärbare Registrierungen weiter zu verbessern. So könnten eine andere Art der Präsentation, ein anderer Weg der Erklärung oder auch eine andere größere Änderung für stärkere oder weitere, noch unentdeckte Effekte sorgen.

Anhang A

Literatur

Tabelle A.1 kann entnommen werden, auf welche Art welche Literatur gefunden wurde.

Snowballing	**Datenbanksuche**	**unstrukturierte Suche**
[1], [3], [6], [8], [13], [16], [19], [20], [26], [27], [33]	[14]	[2], [4], [5], [7], [9], [10], [11], [12], [15], [17], [18], [21], [22], [23], [24], [25], [29], [28], [30], [31], [32]

Tabelle A.1: Ursprung der einzelnen Literaturquellen

Anhang B

Studie

B.1 Datenschutzdokument

Der folgende Text ist der Text des Dokuments, welches jedem Probanden vor Beginn der Studie gegeben wurde. Dies umfasst auch eine Einwilligungserklärung, welche von jedem Teilnehmer an der Studie unterschrieben wurde.

Überblick über die Studie zu Registrierungsdialogen
Fachgebiet Software Engineering, Leibniz Universität Hannover

Bitte lesen Sie diesen Überblick sorgfältig durch. Dieser Überblick dient dazu, Ihnen die Studie vorzustellen und Sie auf Ihre Rechte als freiwillige*r Studienteilnehmer*in hinzuweisen. Bitte fragen Sie nach, wenn etwas unklar sein sollte. Vielen Dank, dass Sie an dieser Studie teilnehmen. Es werden Daten von mehreren Teilnehmenden erhoben. Die Studie dient zur Untersuchung, wie die Stärke von ihnen gewählter Passwörter beeinflusst werden kann. Es geht um die Evaluierung der Daten und nicht um Ihr individuelles Abschneiden im Experiment.

Ablauf und Dauer
Im Rahmen der Studie werden Sie mit einer selbst erstellten Software arbeiten und 6 kurze Aufgaben erledigen. Anschließend wird ein kurzer Fragebogen ausgefüllt. Die Studie wird insgesamt ca. 15 Minuten dauern. Sie haben das Recht, die Teilnahme an dieser Studie jederzeit und ohne Angabe von Gründen abzubrechen.

Erhobene personenbezogene Daten
Während der Studie wird nichts aufgezeichnet und auch in der Software angegebene Daten werden nicht gespeichert. Es wird lediglich eine von einem Algorithmus errechnete Einschätzung der Stärke der von ihnen angegebenen Passwörter gespeichert. Zusätzlich werden in unserer Studie mehrere Daten in Form von Antworten auf einem Fragebogen erhoben. Teile

davon sind demographische Daten, welche nicht einer Zuordnung, sondern einer Einordnung der Allgemeingültigkeit der Ergebnisse dienen.

Datenschutz und Datenspeicherung
Die von Ihnen zur Verfügung gestellten Daten werden ausschließlich anonym und ohne Rückschlüsse auf einzelne Personen ausgewertet. Die Daten aus den Fragebögen werden auf unserem internen SE-Server für einen begrenzten Zeitraum gespeichert. Die Daten sowie die darauf basierenden Auswertungen werden möglicherweise im Rahmen wissenschaftlicher Publikationen in anonymisierter Form veröffentlicht. Bitte lesen Sie sich nun sorgfältig die Einverständniserklärung und die dort beschriebenen Hinweise durch.

Einverständniserklärung zur Studie zu Registrierungsdialogen
Fachgebiet Software Engineering, Leibniz Universität Hannover

Diese Studie wird von Chris Burmeister (Fachgebiet Software Engineering) durchgeführt. Ich habe den Überblick über die Studie gelesen und verstanden. Ich nehme freiwillig an dieser Studie teil. Ich habe das Recht, die Teilnahme jederzeit und ohne Angabe von Gründen abzubrechen. Diese Einwilligung ist freiwillig. Ich kann sie ohne Angabe von Gründen verweigern, ohne dass ich deswegen Nachteile zu befürchten hätte.

1. Datenerfassung, -speicherung und -verwendung
Ich wurde darüber informiert, dass während der Studie einige personenbezogene Daten sowie die Angaben in dem Fragebogen erfasst, elektronisch auf internen SE-Servern gespeichert und zur Auswertung der Studie herangezogen werden. Die erfassten Daten werden nur für das Fachgebiet Software Engineering der Leibniz Universität Hannover zugänglich sein. Die erfassten Daten werden von uns alleinig für die wissenschaftliche Forschung genutzt und ausschließlich anonymisiert ausgewertet. Die erfassten Daten werden in wissenschaftlichen Publikationen nur in anonymisierter Form veröffentlicht. Ich habe gemäß Datenschutz gegenüber dem Informationsträger das Recht auf Auskunft, Berichtigung sowie Löschung meiner personenbezogenen Daten. Ich kann diese Einwilligungserklärung jederzeit schriftlich widerrufen. Nach erfolgtem Widerruf werden meine personenbezogenen Daten von den internen SE-Servern gelöscht und ab diesem Zeitpunkt für keine weiteren Publikationen mehr verwendet.

2. Unterschrift
Ich habe den Überblick über die Studie gelesen und verstanden. Ich bin mit den aufgeführten Punkten des Überblicks und der Einverständniserklärung einverstanden. Ich nehme freiwillig an dieser Studie teil.

B.2 Einleitungstext

Das Folgende ist der im Prototyp genutzte Einleitungstext.

„Dies ist eine Studie, die im Zuge der Masterarbeit von Chris Burmeister durchgeführt wird. Im Folgenden werden Sie zuerst mit einer Software interagieren und es werden danach noch einige wenige Fragen gestellt. Es besteht kein Zeitdruck, jedoch werden Sie voraussichtlich maximal 10 Minuten benötigen. Die Eingaben die Sie in der Folgenden Software tätigen werden nicht gespeichert und sind anderen nicht sichtbar. Schauen Sie sich die Software gerne gründlich an und interagieren Sie mit Dingen die Sie interessieren. Sie können hierbei keine Fehler machen. Interagieren Sie mit der Software möglichst so, als ob sie ihr in einem normalen Kontext begegnen würden. “

B.3 Prototyp & Ablauf

Tabelle B.1 ist zu entnehmen welche vom „zxcvbn“-Algorithmus ausgegebenen Tipps zu welchen im Prototyp dieser Arbeit verwendeten Richtlinien zusammengefasst wurden.

Tipps des „zxcvbn“-Algorithmus	**neu formulierte Richtlinie**
Add another word or two. Uncommon words are better.	Das Passwort sollte verlängert werden.
Use a longer keyboard pattern with more turns	Abfolgen von der Tastatur sollten vermieden werden.
Short keyboard patterns are easy to guess	
Straight rows of keys are easy to guess	
Capitalization doesn't help very much	Großbuchstaben sind nicht besonders hilfreich.
All-uppercase is almost as easy to guess as all-lowercase	
Predictable substitutions like '@' instead of 'a' don't help very much	Vorhersehbare Ersetzungen helfen nicht sehr.
Avoid dates and years that are associated with you	Daten und Jahreszahlen sollten vermieden werden.
Dates are often easy to guess	
Avoid repeated words and characters	Wiederholungen und direkte Abfolgen sollten vermieden werden.
Avoid sequences	
Sequences like abc or 6543 are easy to guess	
Repeats like „aaa“ are easy to guess	
This is a top-100 common password	Dies ist ein häufig verwendetes Passwort und sollte vermieden werden.
This is a top-10 common password	
This is similar to a commonly used password	
This is a very common password	
Common names and surnames are easy to guess	Vor- und Nachnamen sollten vermieden werden.
Names and surnames by themselves are easy to guess	

Tabelle B.1: Zu Richtlinien zusammengefasste und übersetzte „zxcvbn“-Algorithmus-Tipps

Im Folgenden sind Ausschnitte aus dem zur Umsetzung des Konzept programmierten und im Zuge der Nutzerstudie benutzten Prototyps zu finden.

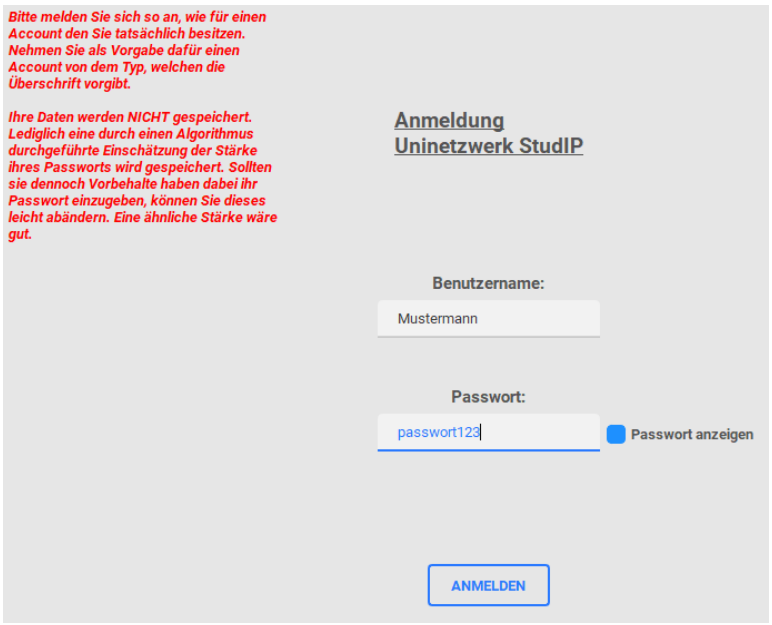

Abbildung B.1: Ausschnitt aus dem Prototyp: Anmeldungsdialog, Szenario StudIP

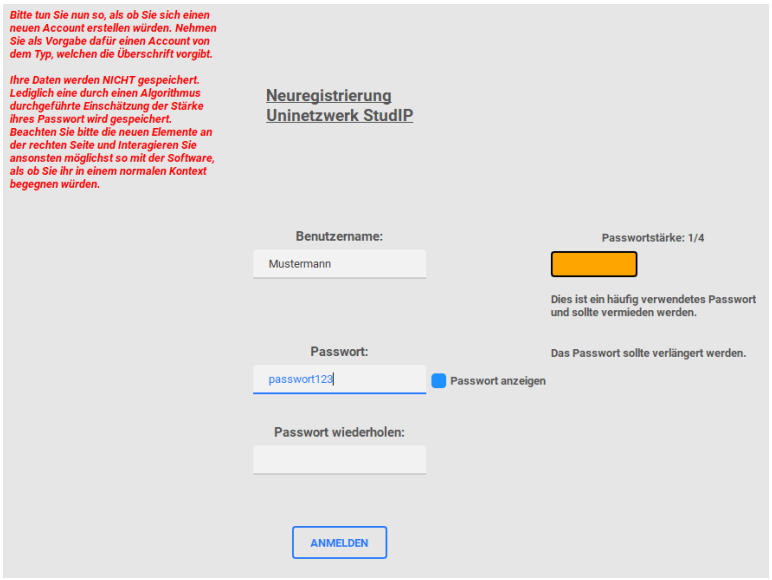

Abbildung B.2: Ausschnitt aus dem Prototyp: Registrierungsdialog der Gruppe ohne Erklärungen, Szenario StudIP

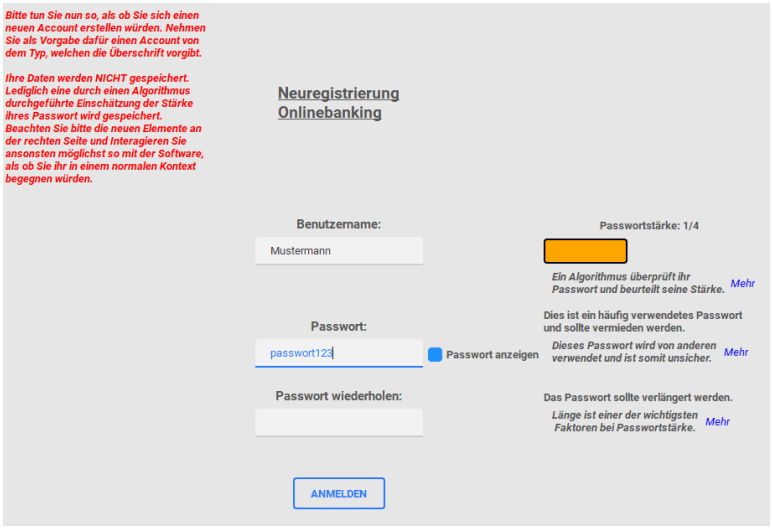

Abbildung B.3: Ausschnitt aus dem Prototyp: Registrierungsdialog der Gruppe mit Erklärungen, Szenario Onlinebanking

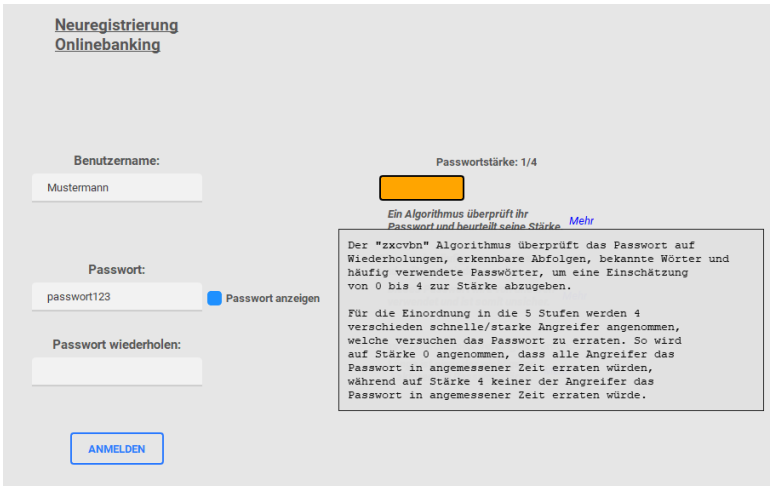

Abbildung B.4: Ausschnitt aus dem Prototyp: Registrierungsdialog der Gruppe mit Erklärungen, mit geöffneter ausführlicher Erklärung zum genutzten Algorithmus, Szenario Onlinebanking

B.4 Umfrage

Der Folgende Text Umfasst den Text der in der Nutzerstudie verwendeten Umfrage. Text innerhalb von ** ** beschreibt hier nicht dargestellte Elemente.

Chris Burmeister - Erklärbare Registrierungen

Diese Umfrage ist vollständig pseudonymisiert. Die Antworten auf die Fragen lassen keine Rückschlüsse auf Ihre Person zu. Die Daten sind nur dem Fachgebiet Software Engineering zugänglich und werden ausschließlich aggregiert für wissenschaftliche Veröffentlichungen verwendet. Bitte beachten Sie, dass sich die folgenden Fragen, welche sich auf die von Ihnen zuvor verwendete Software beziehen, sich nur auf den zweiten Teil, also die Neuregistrierungen, beziehen. Beantworten Sie die Fragen bitte dementsprechend.
In dieser Umfrage sind 15 Fragen enthalten.

Datenverknüpfung
Das Folgende dient dazu ihre Umfragedaten und die in der Software gepeicherten Passwortstärken miteinander verknüpfen zu können, was für eine sinnvolle Auswertung notwendig ist.
(Dies ist eine Pflichtfrage.)
Bitte geben sie eine beliebige ausgedachte 5 stellige Zahlenfolge ein und notieren Sie diese ebenfalls auf einen Zettel neben sich. Nach der Umfrage wird der Zettel genutzt, um die in der Software erhobenen Daten zu kennzeichnen und dann sofort vernichtet. (alles auf Wunsch noch in ihrer Anwesenheit)
In dieses Feld dürfen nur Zahlen eingegeben werden.
Eingabefeld

Wahrnehmung
Bitte geben Sie ihre Zustimmung zu den folgenden Aussagen an.
Bitte beachten Sie, dass sich die folgenden Fragen sich nur auf den zweiten Teil der verwendeten Software, also die Neuregistrierungen, beziehen. Beantworten Sie die Fragen bitte dementsprechend.

Ich halte die Registrierung für verständlich.
Bitte wählen Sie eine der folgenden Antworten:

1. stark ablehnend
2. ablehnend
3. eher ablehnend

4. weder ablehnend noch zustimmend
5. eher zustimmend
6. zustimmend
7. stark zustimmend

Ich finde die Registrierung vertrauenswürdig.
Bitte wählen Sie eine der folgenden Antworten:

1. stark ablehnend
2. ablehnend
3. eher ablehnend
4. weder ablehnend noch zustimmend
5. eher zustimmend
6. zustimmend
7. stark zustimmend

Ich hätte mir mehr Informationen zur Passwortstärke gewünscht.
Bitte wählen Sie eine der folgenden Antworten:

1. stark ablehnend
2. ablehnend
3. eher ablehnend
4. weder ablehnend noch zustimmend
5. eher zustimmend
6. zustimmend
7. stark zustimmend

Wahrnehmung
Bitte geben Sie ihre Zustimmung zu den folgenden Aussagen an.
Bitte beachten Sie, dass sich die folgenden Fragen sich nur auf den zweiten Teil der verwendeten Software, also die Neuregistrierungen, beziehen. Beantworten Sie die Fragen bitte dementsprechend.

Die Registrierung erfüllt meine Anforderungen.
Bitte wählen Sie eine der folgenden Antworten:

1. stark ablehnend
2. ablehnend
3. eher ablehnend
4. weder ablehnend noch zustimmend
5. eher zustimmend
6. zustimmend
7. stark zustimmend

Die Registrierung zu benutzen ist eine frustrierende Erfahrung.
Bitte wählen Sie eine der folgenden Antworten:

1. stark ablehnend
2. ablehnend
3. eher ablehnend
4. weder ablehnend noch zustimmend
5. eher zustimmend
6. zustimmend
7. stark zustimmend

Die Registrierung ist leicht zu benutzen.
Bitte wählen Sie eine der folgenden Antworten:

1. stark ablehnend
2. ablehnend
3. eher ablehnend
4. weder ablehnend noch zustimmend
5. eher zustimmend
6. zustimmend
7. stark zustimmend

Ich habe zu viel Zeit dazu verwenden müssen etwas bei der Registrierung zu korrigieren.
Bitte wählen Sie eine der folgenden Antworten:

1. stark ablehnend
2. ablehnend
3. eher ablehnend
4. weder ablehnend noch zustimmend
5. eher zustimmend
6. zustimmend
7. stark zustimmend

Haben Sie noch weitere Anmerkungen zur verwendeten Software?
Eingabefeld

Demographie
Geschlecht
Bitte wählen Sie eine der folgenden Antworten:

- Männlich
- Weiblich
- Nicht Binär
- Keine Angaben

Bitte geben Sie ihre Zustimmung zu der folgenden Aussagen an:
Ich halte mich selbst für technisch versiert.
Bitte wählen Sie eine der folgenden Antworten:

1. stark ablehnend
2. ablehnend
3. eher ablehnend
4. weder ablehnend noch zustimmend
5. eher zustimmend
6. zustimmend
7. stark zustimmend

Alter
Bitte wählen Sie eine der folgenden Antworten:
Dropdown-menü

Selbsteinschätzung
Bitte geben Sie eine Einschätzung ab:
Wie groß ist der Anteil von eigenständigen Softwareanwendungen mit eigener graphischer Benutzeroberfläche im Vergleich zu Softwareanwendungen allgemein?
Einstellbarer Slider, 0% bis 100%

Betrachten Sie Ihre Antwort auf die erste Frage (Anteil von eigenständigen Softwareanwendungen mit eigener graphischer Benutzeroberfläche) als Maximalwert des Sliders.
Bitte geben Sie eine Einschätzung ab:
Welchen Anteil eigenständiger Softwareanwendungen mit eigener graphischer Benutzeroberfläche könnten Sie mit minimalen Instruktionen bedienen?
Einstellbarer Slider, 0% bis 100%

Betrachten Sie Ihre Antwort auf die erste Frage (Anteil von eigenständigen Softwareanwendungen mit eigener graphischer Benutzeroberfläche) als Maximalwert des Sliders.
Bitte geben Sie eine Einschätzung ab:
Welchen Anteil eigenständiger Softwareanwendungen mit eigener graphischer Benutzeroberfläche könnte eine Person mit durchschnittlichen Fähigkeiten mit minimalen Instruktionen bedienen?
Einstellbarer Slider, 0% bis 100%

Vielen Dank!
Ihre Antworten wurden gespeichert.

B.5 Weitere Anmerkungen der Probanden

Im Folgenden sind die im Freitextfeld der Umfrage abgegebenen Anmerkungen der Probanden zu finden. Diese Anmerkungen sind aufgeteilt in die beiden Gruppen der Studie, da dieser Kontext für die Anmerkungen relevant ist.

Die Folgenden Anmerkungen wurden von Probanden der Gruppe ohne Erklärungen abgegeben.

- Ein typisches Anmelde-/Regestrierungsformular.
- Es wäre interessant zu wissen wie die Software Passwörter bewertet die stark den Namen ähneln.
- Anweisungen waren klar. Die Anzeige zur Passwortstärke war gut verständlich gestaltet und hat meinen Erwartungen entsprochen.
- Weitere Informationen, wie man die Passwort-Stärke verbessern könnte wären relativ hilfreich gewesen. - Das Wegklicken der „Passwörter stimmen nicht überein"-Fenster war ein wenig lästig. Eine Meldung, die direkt bei den Passwörtern angezeigt worden wär
- In dem Stärkemeter für das Passwort gab es 4 Punkte. Anfangs war ich davon ausgegangen, dass man je einen „Punkt " für: - Klein-/Großbuchstaben - Zahlen - Sonderzeichen - „lang genug" bekommt. Ich glaube ich habe bei den Anmeldeseiten auch 4/4 als stärke bekommen, wenn ich nicht alle der oben genannten Punkte erfüllt habe. Das hat mich überrascht. Generell denke ich auch, dass es durchaus hilfreich sein kann, wenn diese Information dem Nutzer mitgeteilt wird. (Also nicht genau wofür es die „Punkte" gibt, sondern dass generell mehrere unterschiedliche Zeichen verwendet werden sollten und das Passwort lang genug sein sollte.

Die Folgenden Anmerkungen wurden von Probanden der Gruppe mit Erklärungen abgegeben.

- Als Website würde ich das Ganze fischiger finden, weil man da bessere visuelle designs gewohnt ist. Bei installierten Desktop-Programmen hätte ich da weniger Probleme mit, weil ich mich darauf berufen würde, von welcher Webseite ich mir das Programm runtergeladen habe und ich bei desktop-programmen schlechteres visuelles Design mehr gewohnt bin.
- Es wäre schon gewesen, dass beim Passwort erstellen auch erwähnt werden wurde z.B. es sollen nicht nur Buchstaben verwenden werden, sondern auch Nummern, sowie Sonderzeichen um die stärke des Passwortes zu erhöhen.

- Tipps, welche Eingaben die Passwortstärke beeinflussen wären hilfreich gewesen. Ansonsten war es ja ein Standardregistrierungsscreen. Ich weiß nicht, ob das gewollt war/ist aber ich habe, so wie ich es bei Neuregistrierungen immer mache, einmal das Passwort getippt und in das Feld für die Wiederholung kopiert.
- Sehr intuitiv gehalten, Password-Meter ist selbsterklärend. Ein wenig ist dem Prototyp geschuldet, dass man (fiktive oder nicht) Anmeldedaten eingibt, und typische Elemente wie Firmenlogo, Uni-Logo oder Bank-Logo fehlen und ohne entsprechende Einweisung vom Durchführenden das gesamte (logischerweise) sehr schlicht gehaltene UI wie ein schlechter Scam-Versuch aussehen würde.
- Die Enter-Taste funktioniert nicht
- Lange Passwörter waren nicht gut sichtbar, da das Fenster eine feste Größe hatte. Farben wie Grün bei Bestätigung des neuen Passworts hätten das Klicken auf das Feld zum Überprüfen vereinfacht.
- Es war mir möglich das Passwort in das Passwort bestätigen Feld zu kopieren. Das umgeht ein wenig den Sinn dieses Feldes, da die Idee ja ist, dass der Nutzer das Feld manuell erneut ausfüllen muss. So ist es nur ein unnötiger Extra schritt wo ich das Passwort mit Strg + C und Strg + V kurz in ein neues Feld kopiere ohne wirklichen Vorteil.
- @ = a und 1337 hinweis ist sehr praktisch gewählt. finde ich gut
- Nein
- Generell sollte auch bedacht werden, dass z.B. im Falle von Onlinebanking häufig auch noch eine 2FA mit genutzt wird. Unter der Prämisse habe ich auch mein Passwort gewählt.
- Man könnte es so machen, dass man bei der Registrierung das Passwort erneut eingeben muss, ohne dass man es copy-pasten kann.

Anhang C

Kritische Werte, Mann-Whitney-U-Tests

Die folgenden Tabellen sind die in dieser Arbeit für die Auswertung genutzten Tabellen, der kritischen Werte des Mann-Whitney-U-Tests ([32]).

n1\n2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2				0	0	0	1	1	1	1	2	2	2	3	3	3	4	4	4
3			0	1	2	2	3	3	4	5	5	6	7	7	8	9	9	10	11
4		0	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	17	18
5	0	1	2	4	5	6	8	9	11	12	13	15	16	18	19	20	22	23	25
6	0	2	3	5	7	8	10	12	14	16	17	19	21	23	25	26	28	30	32
7	0	2	4	6	8	11	13	15	17	19	21	24	26	28	30	33	35	37	39
8	1	3	5	8	10	13	15	18	20	23	26	28	31	33	36	39	41	44	47
9	1	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54
10	1	4	7	11	14	17	20	24	27	31	34	37	41	44	48	51	55	58	62
11	1	5	8	12	16	19	23	27	31	34	38	42	46	50	54	57	61	65	69
12	2	5	9	13	17	21	26	30	34	38	42	47	51	55	60	64	68	72	77
13	2	6	10	15	19	24	28	33	37	42	47	51	56	61	65	70	75	80	84
14	2	7	11	16	21	26	31	36	41	46	51	56	61	66	71	77	82	87	92
15	3	7	12	18	23	28	33	39	44	50	55	61	66	72	77	83	88	94	100
16	3	8	14	19	25	30	36	42	48	54	60	65	71	77	83	89	95	101	107
17	3	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96	102	109	115
18	4	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109	116	123
19	4	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	130
20	4	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138

Abbildung C.1: Tabelle der kritischen Werte für den Mann-Whitney-U-Tests mit einer Signifikanz von 5% für eine einseitige Hypothese (zweiseitig 10% Signifikanz)

n1\n2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2							0	0	0	0	1	1	1	1	1	2	2	2	2
3				0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4			0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	14
5		0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6		1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7		1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	69	74	78	83
15	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	2	6	11	17	22	28	34	39	45	51	57	63	69	75	81	87	93	99	105
18	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	2	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

Abbildung C.2: Tabelle der kritischen Werte für den Mann-Whitney-U-Tests mit einer Signifikanz von 5% für eine zweiseitige Hypothese

Literaturverzeichnis

[1] G. Bella, J. Ophoff, K. Renaud, D. Sempreboni, and L. Viganò. Perceptions of beauty in security ceremonies. *Philosophy & Technology*, 35(3):72, 2022.

[2] M. I. Berkman and D. Karahoca. Re-assessing the usability metric for user experience (umux) scale. *Journal of Usability Studies*, 11(3), 2016.

[3] L. Chazette. Requirements engineering for explainable systems. *Hannover: Institutionelles Repositorium der Leibniz Universität Hannover*, 2023.

[4] J. Cohen. *Statistical power analysis for the behavioral sciences*. Academic press, 2013.

[5] L. David and A. Wool. An explainable online password strength estimator. In *Computer Security–ESORICS 2021: 26th European Symposium on Research in Computer Security, Darmstadt, Germany, October 4–8, 2021, Proceedings, Part I 26*, pages 285–304. Springer, 2021.

[6] X. de Carné de Carnavalet and M. Mannan. From very weak to very strong: Analyzing password-strength meters. In *Network and Distributed System Security Symposium (NDSS 2014)*. Internet Society, 2014.

[7] M. Dupuis and F. Khan. Effects of peer feedback on password strength. In *2018 APWG Symposium on Electronic Crime Research (eCrime)*, pages 1–9. IEEE, 2018.

[8] S. Egelman, A. Sotirakopoulos, I. Muslukhov, K. Beznosov, and C. Herley. Does my password go up to eleven? the impact of password meters on password selection. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2379–2388, 2013.

[9] K. Finstad. The usability metric for user experience. *Interacting with computers*, 22(5):323–327, 2010.

[10] M. Golla and M. Dürmuth. On the accuracy of password strength meters. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 1567–1582, 2018.

[11] M. Golla, B. Hahn, K. M. zu Selhausen, H. Hosseini, and M. Dürmuth. Bars, badges, and high scores: On the impact of password strength visualizations. *Who Are You*, 2018.

[12] International Organization for Standardization. Iso 9241-11:2018, ergonomics of human-system interaction, part 11: Usability: Definitions and concepts, 2018. `https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en` [Accessed: 3.4.2024].

[13] W. Khern-am nuai, M. J. Hashim, A. Pinsonneault, W. Yang, and N. Li. Augmenting password strength meter design using the elaboration likelihood model: Evidence from randomized experiments. *Information Systems Research*, 34(1):157–177, 2023.

[14] B. Naqvi and A. Seffah. Interdependencies, conflicts and trade-offs between security and usability: why and how should we engineer them? In *HCI for Cybersecurity, Privacy and Trust: First International Conference, HCI-CPT 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings 21*, pages 314–324. Springer, 2019.

[15] National Institute of Standards and Technology. Nist special publication 800-63b, 2017. `https://pages.nist.gov/800-63-3/sp800-63b.htmln` [Accessed: 22.4.2024].

[16] B. D. Payne and W. K. Edwards. A brief introduction to usable security. *IEEE Internet Computing*, 12(3):13–21, 2008.

[17] W. Pieters. Explanation and trust: what to tell the user in security and ai? *Ethics and information technology*, 13:53–64, 2011.

[18] F. Raczkowski and N. Schrape. Gamification. *Game studies*, pages 313–329, 2018.

[19] A. Rosenfeld and A. Richardson. Explainability in human–agent systems. *Autonomous agents and multi-agent systems*, 33:673–705, 2019.

[20] R. Shay, L. Bauer, N. Christin, L. F. Cranor, A. Forget, S. Komanduri, M. L. Mazurek, W. Melicher, S. M. Segreti, and B. Ur. A spoonful of sugar? the impact of guidance and feedback on password-creation behavior. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 2903–2912, 2015.

[21] Y. Shin and S. S. Woo. What is in your password? analyzing memorable and secure passwords using a tensor decomposition. In *The World Wide Web Conference*, pages 3230–3236, 2019.

[22] R. H. Thaler and C. R. Sunstein. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin (publisher), 2009.

[23] B. Ur, F. Alfieri, M. Aung, L. Bauer, N. Christin, J. Colnago, L. F. Cranor, H. Dixon, P. Emami Naeini, H. Habib, et al. Design and evaluation of a data-driven password meter. In *Proceedings of the 2017 chi conference on human factors in computing systems*, pages 3775–3786, 2017.

[24] A. Vance, D. Eargle, K. Ouimet, and D. Straub. Enhancing password security through interactive fear appeals: A web-based field experiment. In *2013 46th Hawaii International Conference on System Sciences*, pages 2988–2997. IEEE, 2013.

[25] L. Vigano and D. Magazzeni. Explainable security. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 293–300. IEEE, 2020.

[26] R. Wash and M. E. Zurko. Usable security. *IEEE Internet Computing*, 21(3):19–21, 2017.

[27] D. L. Wheeler. zxcvbn:{Low-Budget} password strength estimation. In *25th USENIX Security Symposium (USENIX Security 16)*, pages 157–173, 2016.

[28] C. Wohlin. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, pages 1–10, 2014.

[29] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén. *Experimentation in software engineering*. Springer Science & Business Media, 2012.

[30] J. F. Wolfswinkel, E. Furtmueller, and C. P. Wilderom. Using grounded theory as a method for rigorously reviewing literature. *European journal of information systems*, 22(1):45–55, 2013.

[31] M. Xu, W. Han, et al. An explainable password strength meter addon via textual pattern recognition. *Security and Communication Networks*, 2019, 2019.

[32] C. Zaiontz. Mann-whitney table, 2024. `https://real-statistics.com/statistics-tables/mann-whitney-table/` [Accessed: 11.4.2024].

[33] V. Zimmermann, K. Marky, and K. Renaud. Hybrid password meters for more secure passwords–a comprehensive study of password meters including nudges and password information. *Behaviour & Information Technology*, 42(6):700–743, 2023.