



Lehrstuhl für Informatik 1
Friedrich-Alexander-Universität
Erlangen-Nürnberg



BACHELOR THESIS

Der Kompromiss zwischen Privatsphäre und Features am Beispiel Google

Richard Baumann

Erlangen, November 10, 2014

First Examiner:
Advisor:

Prof. Dr. Felix Freiling
Nadina Hintz

Eidesstattliche Erklärung / Statutory Declaration

Hiermit versichere ich eidesstattlich, dass die vorliegende Arbeit von mir selbstständig, ohne Hilfe Dritter und ausschließlich unter Verwendung der angegebenen Quellen angefertigt wurde. Alle Stellen, die wörtlich oder sinngemäß aus den Quellen entnommen sind, habe ich als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt.

I hereby declare formally that I have developed and written the enclosed thesis entirely by myself and have not used sources or means without declaration in the text. Any thoughts or quotations which were inferred from the sources are marked as such. This thesis was not submitted in the same or a substantially similar version to any other authority to achieve an academic grading.

Erlangen, November 10, 2014

Richard Baumann

Zusammenfassung

TODO

Abstract

TODO

CONTENTS

1	Einführung	1
1.1	Motivation	1
1.2	Zielsetzung und Forschungsfragen	1
1.3	Kategorien	2
1.4	Hypothesen	2
2	Ähnliche Arbeiten	3
3	Methodik	5
3.1	Fragebogenkonstruktion	5
3.2	Rekrutierung der Teilnehmer	6
4	Ergebnisse	7
4.1	Demographie	7
4.2	Hypothesen	9
4.3	Forschungsfragen	9
4.4	Sonstiges	9
5	Fazit	11
5.1	Resume	11
5.2	Ausblick	11
6	Anhang	13

EINFÜHRUNG

1.1 Motivation

Mit den vor kurzem aufgetretenen Veröffentlichungen durch Edward Snowden sind der Datenschutz und die Privatsphäre wieder in den Vordergrund geraten. So sagt zum Beispiel Hanspeter Thür, der Datenschutzbeauftragte der Schweiz: "Privatsphäre wird zu einem Privileg" [2]. Durch diese Debatten geraten auch große Unternehmen im Internet wieder in den Fokus. Da auch in Deutschland die Google Inc. eine sehr große Nutzerbasis hat (alleine in Deutschland nutzen ungefähr 38 Millionen Nutzer die Suchmaschine von Google [5]), ist es logisch, dass auch sie des öfteren mit Datenschützern in Konflikt gerät (siehe zum Beispiel "Datenschützer: Google verstößt gegen geltendes Recht" [3]). Die Google Suchmaschine ist dabei nicht der einzige Google Dienst der dabei in den Mittelpunkt von Diskussionen gerät. Spätestens seit Julian Assange Google eine "privatisierte NSA" genannt hat [4], ist es klar, dass man sich im Bezug auf persönliche Daten bei Google besonders Gedanken machen muss.

Dabei ist der wichtigste Aspekt der Mensch selbst - und zwar nicht die Mitarbeiter von Google oder die Datenschützer im Besonderen sondern die normalen Nutzer. Der Nutzer ist derjenige, dessen Daten von Google verwendet werden und somit derjenige der sich davor Schützen kann, oder damit leben muss dass Google die Daten besitzt.

1.2 Zielsetzung und Forschungsfragen

Diese Arbeit soll die Nutzer von Google näher betrachten und herausfinden wie diese Nutzer über ihren Datenschutz im Bezug auf Google denken und wie sie vorgehen um ihre Daten zu

schützen. Hierzu wird eine Umfrage durchgeführt die Nutzer über ihr Verhalten auf Google und ihre Gedanken über Google befragt. Die grundsätzlichen Forschungsfragen die behandelt werden sind "Wie viel wissen Nutzer von der Datensammlung durch Google?" und "Versuchen sie gegen dieses Sammeln von Daten vorzugehen?".

1.3 Kategorien

Zum Aufstellen der Hypothesen und zur späteren Einteilung der Fragen des Fragebogens wurden Kategorien aufgestellt, die wie folgt definiert wurden:

0. Nutzung von Google Diensten: Das Nutzungsverhalten der Google Dienste von Seiten der Nutzer. Dieser Teil fragt vor allem allgemeine Daten zum Nutzungsverhalten ab. Darunter fallen unter anderem die Information, welche Dienste genutzt werden und wie viele Accounts die Nutzer haben.
1. Kenntnisse über Googles Datenschutz: Angeeignetes Wissen über den Umgang mit personenbezogenen Daten bei Google, vor allem im Bezug auf die genutzten Dienste. Hierbei sind Fragen wie "Bietet Google auf Nutzer zugeschnittene Werbung an?" relevant.
2. Vertrauen in Google: Das Glauben an auftrende negative Konsequenzen im Zusammenhang mit dem Preisgeben der Daten (vgl. Kim et al., 2008).
3. Aufgeben der Privatsphäre: Das Bewusstsein über den Verlust der Privatsphäre bei der Verwendung von Google Diensten.
4. Schutzmaßnahmen: Verhalten des Nutzers zum Schutze der eigenen Privatsphäre

1.4 Hypothesen

Die Forschungsfragen werden beantwortet indem Hypothesen aufgestellt werden die den Zusammenhang der oben genannten Kategorien darstellen. Die zugehörigen zu testenden Hypothesen sind die Folgenden:

- H0** Nutzung von Google Diensten (0) → Kenntnisse über Googles Datenschutz (1): Wenn eine Person Google aktiver nutzt, bekommt sie mehr Erfahrung und Kenntnisse über Google Dienste und mögliche Probleme mit der Privatsphäre
- H1** Kenntnisse über Google (1) → Aufgeben der Privatsphäre (4): Je mehr eine Person über Googles Datenschutz weiß, desto seltener gibt sie Teile ihrer Privatsphäre auf
- H2** Vertrauen in Google (2) → Wahrgenommenes Risiko für Privatsphäre (3): Je höher das Vertrauen in Google ist, desto geringer wird das Risiko eingeschätzt.
- H3** Wahrgenommenes Risiko für Privatsphäre (3) → Schutzmaßnahmen (5): Je höher das Risiko eingeschätzt wird, desto mehr Schutzmaßnahmen werden unternommen.
- H4** Schutzmaßnahmen (5) → Aufgeben der Privatsphäre (4): Wenn eine Person mehr Schutzmaßnahmen verwendet, dann gibt sie weniger Teile ihrer Privatsphäre auf.

Durch die Umfrage sollen diese Hypothesen belegt oder widerlegt werden und somit die Zusammenhänge dargestellt und letzten Endes die Forschungsfrage beantwortet werden.

ÄHNLICHE ARBEITEN

Die Problematik der Privatsphäre bei der Kommunikation mit Google Diensten ist ein Thema zu der es bereits einige Artikel gibt. So beschreibt zum Beispiel Omer Tene in dem Artikel "What Google knows: Privacy and internet search engines" [6], dass die größten Probleme der Privatsphäre sich in 6 Kategorien einteilen lassen. Die Ansammlung der Daten, das heißt dass einzelne Daten die der Nutzer über sich abgibt zwar nicht zu Einschränkungen in der Privatsphäre führen, allerdings kommt es sehr schnell dazu wenn der Nutzer mehrmals Informationen herausgibt und diese zu einem Ganzen zusammengefügt werden. So wird als Beispiel angebracht, dass eine Suche nach "French Mountains" alleine nicht sehr aussagekräftig ist, wenn aber kurz darauf noch nach "ski vacation" und "gift to grandchild" und weiterem gesucht wird wird schnell klar worum es sich bei der ersten Frage gehandelt hat. Wenn danach zum Beispiel noch nach "disabled access" gesucht wird bildet sich ein sehr umfangreiches Bild über die Familie des Suchmaschinen-Nutzers aus. Der zweite Punkt ist die Verzerrung, das heißt dass Suchanfragen, wenn sie alleine gestellt sind sehr schnell ein falsches Bild liefern können. Als Beispiel wird hier die Suchanfrage "assassinate US president" gezeigt - Behörden könnten hierbei schnell aufmerksam werden, obwohl sich jemand nur über die Geschichte von früheren Präsidenten informieren wollte. Seine dritte Kategorie ist der Ausschluss. So dürfen sich Nutzer in Europa zwar ihre bei Google gespeicherten Informationen ausgeben lassen, da darüber aber nur wenige Menschen bescheid wissen gilt Googles Datenbank letzten Endes als "secret Database". Der vierte Punkt ist die "zweite Nutzung". So wird durch die Verwendung von Google der Datenschutzrichtlinie zugestimmt, die es Google erlaubt die Daten für weit mehr als der direkten Verwendung zu nutzen. Die letzten 2 Kategorien sind der Vertrauensbruch und die restlichen Probleme. So vertraut der Nutzer zum Beispiel darauf, dass Google ihre Daten vertraulich behandelt - und empfindet es als Vertrauensbruch wenn sie erfahren dass Google die Daten auch für andere Zwecke verwendet.

Als Analogon zu einer ähnlich tiefen Verbindung werden in "Facebook and online privacy: Attitudes, behaviors, and unintended consequences" [1] die Problematiken der Privatsphäre bei der Benutzung von Facebook untersucht. Genau wie auch Google ist Facebook ein Tool geworden das alltäglich von vielen Nutzern verwendet wird und dabei sehr tief in die Privatsphäre eingreifen kann. Wie auch in dieser Arbeit geht es bei der Untersuchung zu einem großen Teil um die Verhaltensweisen der Nutzer und darum wie viel sie über die Konsequenzen ihres Verhaltensmusters wissen. So wird zum Beispiel gefragt wie viele Nutzer der Meinung sind sich mit den Privatsphäreereinstellungen in Facebook auszukennen und ob sie sich in der Lage fühlen ihr Profil zumindest vor dem Zugriff Dritter zu schützen. Ähnliche Fragen werden auch in dieser Arbeit behandelt.

Es gibt vor allem im Bereich Social Network noch weitere Arbeiten die ähnliche Untersuchungen unternehmen, allerdings werden diese hier nicht dargestellt.

METHODIK

Die Studie wurde als Online-Umfrage entwickelt, mit dem direkten Ziel die Forschungsfragen zu beantworten. Dazu wurden die Forschungsfragen und die zugehörigen Hypothesen aufgestellt und mit diesen Mitteln der Fragebogen erstellt.

3.1 Fragebogenkonstruktion

Um die Umfrage zu erstellen wurde zuerst ein umfassender Fragenkatalog zum Thema Google erstellt. Dabei wurde noch nicht direkt auf die Hypothesen eingegangen und die Fragen deckten teilweise zu große oder zu kleine Bereiche ab. Aus diesem Fragenkatalog wurden danach die sinnvollen Fragen herausgefiltert und die restlichen verworfen. Die sinnvollen Fragen wurden überarbeitet und in Fragen umgewandelt die für eine Umfrage passender sind. Dieser neue Fragenkatalog wurde erweitert und belief sich am Ende auf ungefähr 60 Fragen. Diese Fragen wurden einer neuen Filterung unterzogen bei der diesmal vor allem die Relevanz für die Hypothesen im Vordergrund stand. Des weiteren wurden einige Fragen ausgefiltert die in einer Online-Umfrage nicht verwendbar gewesen wären, da für diese Fragen eine direkte Interaktion mit dem Benutzer nötig gewesen wäre (z.b. "Gehen Sie auf Google und geben sie den Begriff 'Restaurant' ein - was fällt ihnen dabei auf?", diese Frage hätte eventuell bei vielen Benutzern zum Abbruch der Umfrage geführt). Das Ergebniss war eine Umfrage mit 39 Fragen, die in Kategorien unterteilt waren, die ein schnelles und zielstrebiges durcharbeiten des Fragebogens vereinfachen sollte. Diese Kategorien sind nicht an spezielle Themen gebunden sondern eher so unterteilt, dass Fragen die eventuell die Meinung beeinflussen könnten, wie zum Beispiel "Durch die Nutzung von Google Diensten gebe ich zu viel von meiner Privatsphäre auf" erst gegen Ende der Umfrage

kommen, wohingegen am Anfang eher allgemeine Wissensfragen stehen. Der letzte Teil der Umfrage ist eine Liste an demografischen Fragen die eine Einordnung der Teilnehmer erleichtern soll. Am Ende dieses Teiles steht noch eine Frage nach Feedback zur Umfrage. Der Fragebogen war auf eine ungefähre Dauer von 10-15 Minuten ausgelegt. Nachdem dies bei den Testteilnehmern dem tatsächlichen Zeitverbrauch entsprochen hat (10min 58s, 13min 13s, 14min 43s, 11min 52s) wurde die Länge des Fragebogens nicht mehr verändert. TODO: UML Diagramm

3.2 Rekrutierung der Teilnehmer

Die Umfrage wurde zuerst über Facebook und über Familienmitglieder und Bekannte verbreitet. Nachdem so schon eine relativ große Anzahl an Antworten bekommen wurden wurde die Umfrage noch über den Email-Verteiler der Technischen Fakultät der Universität Erlangen-Nürnberg verbreitet. Insgesamt wurden somit innerhalb weniger Tage 856 vollständige Antworten erhalten, was zum vorzeitigen Abbruch führte. Eine zu große Anzahl an Antworten hätte die Auswertung nur gebremst und mit über 850 sollte bereits eine gute Auswertung möglich sein.

ERGEBNISSE

Die Umfrage ergab 856 vollständig ausgefüllte Fragebögen. Dazu 164 nicht vollständig ausgefüllte, die laut dem Feedback Feld teilweise auf das Fehlen eines Zurück-Knopfes zurückzuführen sind, aber wohl eventuell auch einfache Abbrüche waren.

4.1 Demographie

Von den 856 Teilnehmern waren 204 (23,83%) weiblich, 613 (71,61%) männlich und 39 haben keine Angabe zu ihrem Geschlecht abgegeben. Die hohe Anzahl an männlichen Teilnehmern liegt vermutlich daran, dass die Umfrage über den Email-Verteiler der Technischen Fakultät der Universität Erlangen-Nürnberg verteilt wurde, welche einen höheren Männer- als Frauenanteil hat. Das durchschnittliche Alter beträgt 24,99 Jahre mit einer Standardabweichung von 9,23 Jahren. Das Maximum ist 99 Jahre und das Minimum 0 Jahre, wobei dies wohl beides Fehler sind - das Minimum könnte durch ein nicht ausgefülltes Feld entstanden sein, wohingegen die Person die das Maximum eingegeben hat mir persönlich gesagt hat dass dies ein Fehler war. Das untere Quartil ist 20, das Mittlere 23 und das Obere 26. Interessant ist, dass trotz des niedrigen Wertes und der geringen Standardabweichung dennoch einige Fragebögen im Altersbereich von 45 bis 77 liegen. Diese müssen eventuell in der weiteren Auswertung gesondert betrachtet werden - denn selbst wenn ihre Anzahl nicht für statistisch exakte Aussagen reicht könnten hier dennoch interessante Ergebnisse gefunden werden. Die nächste Frage fragte nach dem höchsten bisher erreichten Schulabschluss. Hier zeigt sich dass ein Großteil der Teilnehmer entweder gerade studiert oder bereits einen Abschluss im Studium erreicht hat. Abitur und Bachelor/Master/Diplom zusammen geben über 90% der Teilnehmer ab. Dagegen sind Hauptschulabschluss und Realschulabschluss zusam-

men nur knapp über 5.73%, was die Relevanz der Umfrageergebnisse ein wenig einschränkt, da sie nicht als allgemeingültig angesehen werden können sondern nur für einen speziellen Teil der Bevölkerung gelten. In einer zukünftigen Arbeit zu diesem Thema sollten eventuell speziell die Gruppen "Kein Abschluss", "Hauptschulabschluss" und "Realschulabschluss" betrachtet werden.

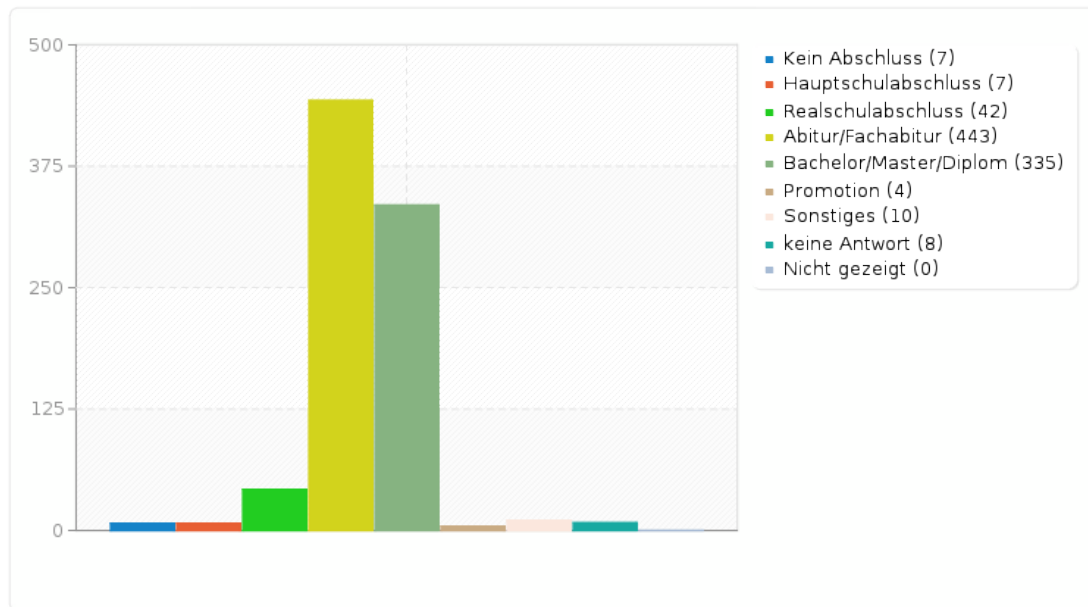


Figure 4.1: Höchster Schulabschluss

Auch die folgende Frage zeigt das selbe Problem auf - 76.99% der Teilnehmer (659) waren zum Zeitpunkt des Ausfüllens Studenten. Dahingegen sind nur 1.99% in Ausbildung und 14.84% berufstätig. Das arithmetische Mittel der selbst eingeschätzten Informatikkenntnisse liegt auf einer Skala von 1 (keine) bis 5 (sehr hohe) bei 3.26. Dass dieser Wert sehr nah an dem zu erwartenden Mittelwert von 3.0 liegt deutet darauf hin, dass die Umfrage trotz der fehlenden Gruppe an Realschülern und Hauptschülern dennoch Relevanz hat. Der größte Anteil hat mit 31.40% einen Wert von 3 angegeben, kurz darauf folgen Kenntnisse von 4 mit 28.22% und von 2 mit 24.20%. Ein größerer Unterschied von über 10% liegt zwischen denen die den höchsten Wert wählten (13.46%) und denen die den niedrigsten Wert wählten (2.72%). Diese Statistik verschiebt sich stark, wenn man die Frage nach den Kenntnissen in der IT-Sicherheit betrachtet. Hier liegt das arithmetische Mittel bei 2.63, was um 0.63 niedriger ist als das arithmetische Mittel der IT Kenntnisse. Hier liegt das Maximum mit 30.86% bei einem Wert von 2, Antwortmöglichkeit 3 und 1 haben 29.80% respektive 16.84%. Der Unterschied zwischen der Anzahl der Teilnehmer die mit 1 geantwortet haben und denen die mit 5 geantwortet haben liegt aber auch hier mit 12.13% nicht sehr weit von dem Abstand aus der letzten Frage weg. Die Ergebnisse dieser Frage korrelieren mit der nächsten Frage: 77.45% der Teilnehmer haben keinen Bildungsabschluss in der Informatik oder einem anderen IT-nahem Fachbereich.

4.2 Hypothesen

4.3 Forschungsfragen

4.4 Sonstiges

5.1 Resumee

5.2 Ausblick

ANHANG

First chapter

Bibliography

- [1] Bernhard Debatin, Jennette P Lovejoy, Ann-Kathrin Horn, and Brittany N Hughes. Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of Computer-Mediated Communication*, 15(1):83–108, 2009.
- [2] Jan Flückiger für NZZ. «Privatsphäre wird zu einem Privileg». <http://www.nzz.ch/aktuell/schweiz/privatsphaere-wird-zu-einem-privileg-1.18256915>.
- [3] Gulli News. Datenschützer: Google verstößt gegen geltendes Recht. <http://www.gulli.com/news/22876-datenschuetzer-google-verstoest-gegen-geltendes-recht-2013-12-02>, 12 2013.
- [4] Gulli News. Julian Assange über Google: Eine "privatisierte NSA", 9 2014.
- [5] Statista. Alle Statistiken und Zahlen zum Thema Google. <http://de.statista.com/themen/651/google/>.
- [6] Omer Tene. What Google knows: Privacy and internet search engines. *Utah Law Review*, *Forthcoming*, 2007.