Der Wert von Privatsphäre bei Websuchen und Personalisierung von Websuchen

Sebastian Vogl
Betreuer: Heiko Niedermayer
Seminar: Innovative Internettechnologien und Mobilkommunikation SS2014
Lehrstuhl Netzarchitekturen und Netzdienste
Fakultät für Informatik, Technische Universität München
Email: s.vogl@tum.de

KURZFASSUNG

In dieser Arbeit wird der Wert von Privatsphäre bei Websuchen unter Berücksichtigung von Personalisierung besprochen. Zu Beginn wird eine kurze Einführung in Anonymisierung und Personalisierung gegeben. Diese beinhaltet jeweils verschiedene Arten und deren Voraussetzungen. Danach wird auf das Hauptthema „Was ist der Wert von Privatsphäre bei Websuchen?“ eingegangen. Dazu werden ein Experiment und eine Umfrage herangezogen. Für das Experiment sollten Probanden verschiedene Websuchen ausführen. Dabei wurden deren Verhalten und Privatsphäre Einstellungen beobachtet und ausgewertet. Im weiteren Verlauf der Arbeit werden der Ablauf des Experiments und dessen Ergebnisse aufgezeigt. Nachfolgend wird der Aufbau der erstellten Umfrage erklärt und zur Bewertung der Ergebnisse des Experiments herangezogen.

Schlüsselworte

Websuche, Suchmaschine, Anonymisierung, Personalisierung, Profile, Privatsphäre, Datenschutz, Arten der Personalisierung, Wert von Datenschutz, Personalisierte Suche

1. EINLEITUNG

„Wie wichtig ist der heutigen Gesellschaft Ihre Privatsphäre?“ beziehungsweise „Wie wichtig ist mir meine Privatsphäre?“. Durch das Preisgeben persönlicher Angaben und eigener Interessen wird der Umgang auf Webseiten sehr vereinfacht. Jedoch den Zugriff dieser Daten wieder einzuschränken, oder sie gar zu löschen ist weitaus schwieriger als sie vorher zu erstellen. Ein Beispiel zu diesem Thema ist „Das Recht auf Vergessen“, wie es in den Medien genannt wird. Dabei wird Google nach dem Verlorenen Präzedenzfall [4] am Europäischen Gerichtshof dazu veranlasst, innerhalb der EU die Möglichkeit zur Löschung von eigenen Suchanfragenergebnissen zu geben. Zwar ist die grundsätzliche Idee nachvollziehbar, aber man kann im Moment davon ausgehen, dass, meiner Meinung nach, dieses Recht auf längere Zeit nur von wenigen genutzt werden wird. Zum einen liegt dies daran, dass der Vorgang aufwendig durchzuführen ist, indem man ein Formular ausfüllen und jede einzelne Anfrage explizit benennen muss, zum Anderen gilt dieses Recht nur für Google-Domänen in der Europäischen Union. [5] Die Hauptwebseite google.com ist davon nicht betroffen, da diese nicht innerhalb der EU betrieben wird. Weiter kommt hinzu, dass eine Kopie des Personalausweises der Privatperson hochgeladen werden muss, welche eine eindeutige Identifizierung durch Google ermöglichen würde. Dies wird von mehreren Seiten bemängelt. Ein Beispiel dazu findet sich unter [14]. Somit kommen folgende Überlegungen auf: „Was ist uns eigentlich die Privatsphäre bei Websuchen wert?“ und „Wie sieht Personalisierung heute aus?“

2. ANONYMISIERUNG UND PERSONALISIERUNG

Nach §3 Abs. 6 Bundesdatenschutzgesetzt [13] ist Anonymisierung definiert als „[…] das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können.“. Somit zusammengefasst, das Verändern von Daten, dass sie nicht mehr einer Person zugeordnet werden können. Im Gegensatz dazu ist Personalisierung definiert als das Erheben und Verarbeiten von Daten einer Person, dabei ist das „Erheben […] das Beschaffen von Daten über den Betroffenen (Person).“ und das Verarbeiten „[…] das Speichern, Verändern, Übermitteln, Sperren und Löschen personenbezogener Daten.“

2.1 Möglichkeiten der Anonymisierung

Informationsbeschaffung war für die Menschen schon immer wichtig, doch nicht jedes Thema, das man im Internet anspricht, möchte man in sein öffentliches Profil mit protokolliert haben. Vor allem, wenn diese Protokolle eine Person, dessen Umfeld mit Freunden, Verwanden, dessen Kaufverhalten und Interessen bis ins kleinste Detail miteinschließen. Wie viel man mit solchen Datensätzen anfangen kann und, wie man diese wieder Vermarktet sieht man zum Beispiel an der Google Trend-Funktion [17], bei der die meist gesuchten Begriffe aufgelistet und Trend-Tendenzen ausgewertet werden. Deswegen wird gerne auch auf die Möglichkeit der Anonymisierung zurückgegriffen.

Es gibt verschiedene Arten seine Identität im Internet oder bei Websuchen zu verschleiern. Zum einen gibt es die Möglichkeit falsche Angaben (z.B. Alter, Geschlecht, Wohnort) bei der Anmeldung zu machen. Dies ist für den Nutzer die einfachste Möglichkeit von sich abzulenken. Andererseits hilft dies nicht

doi: 10.2313/NET-2014-08-1_17

bei dem Problem der Personalisierung, da bisher immer noch eine, wenn auch nur virtuelle, Person vorhanden ist. Auch bei der Verwendung von verschiedenen „Pseudo“-Profilen, besteht die Möglichkeit, dass mittels Verhaltensmuster-Analysen eine Personalisierung durchgeführt wird. Deswegen werden oft weitere externe Dienste hinzugezogen (z.B. Tor [6], JonDonym [7] und verschiedene Proxy oder VPN Anbieter), die nicht nur den eigenen Zugangsstandort, sondern auch die Identität einer jeweiligen Person durch eine öffentliche, von Mehreren genutzte, Identität ersetzen. Somit kann zum Beispiel im Falle von Websuchen, keine Personalisierung nach Interessen durchgeführt werden, da über die gleiche IP-Adresse unterschiedliche Nutzer mit ihren jeweiligen Interessen auf die Websuchmaschinen gleichzeitig zugreifen. Bei der Nutzung von externen Diensten besteht dennoch weiterhin die Gefahr der Datenspeicherung. Denn diese Dienste müssen, wenn auch nur temporär, die korrekten Daten der zu verschleiernden Personen abspeichern, damit der Netzwerkverkehr an die richtigen Personen weitergeleitet werden kann. Eine weitere Option um die Identifizierung von Personen aufgrund dieser Datenspeicherung zu vermeiden wäre die Nutzung von öffentlichen Internetzugängen (z.B. die Nutzung öffentlicher Hotspots an Bahnhöfen oder Flughäfen). Außerdem sollte man beachten, sobald man persönliche Daten, auch über eine anonymisierte Verbindung, versendet, kann man schnell wieder enttarnt werden. Eine letzte Möglichkeit, die verwendet wird und in dieser Arbeit genannt werden soll, ist das Verwenden einer widerrechtlichen erlangten Identität (z.B. das Nutzen eines privaten WLan- Anschlusses einer fremden Person). So wird zwar die eigene Identität verschleiert, aber dieses Vorgehen kann strafrechtliche Konsequenzen mit sich ziehen.

Da dies nicht Thema dieser Arbeit ist, wird nicht weiter darauf eingegangen. Sofern weiteres Interesse besteht, wird auf [15] und [16] verwiesen.

2.2 Arten der Personalisierung

Genauso wie es verschiedene Arten der Anonymisierung gibt, existieren auch verschiedene Arten der Personalisierung. Die folgenden Arten werden als Grundlage für Websuchen verwendet.

Die erste Art wird als „Relevance Feedback and Query Modification“ bezeichnet. Zwar dienen die Informationen davon nicht direkt der Personalisierung, aber im Datensammlungs-prozess ist es eine wichtige Grundlage. Das eigentliche Ziel hierbei ist ein besseres Suchergebnis zu finden, das auf Basis der eingegeben Informationen und bestimmten Ergebnis-gewichtungen erstellt wird. Dabei wird die ursprüngliche Suchanfrage intern durch andere Ergebnisse und Daten (z.B. Standortrelevanz) angepasst. Man kann das Vorgehen mit einer Personalisierung einer einzelnen Suchanfrage vergleichen.

Die zweite Art basiert auf der Analyse des Inhalts der Webseiten. Dabei werden Inhalte von verschiedenen Webseiten verglichen und auf Basis der Relevanz und den bisherigen Gewohnheiten beziehungsweise Interessen des Nutzers angepasst ausgegeben. Das Profil des Nutzers wurde bei dieser Methode bereits angelegt. Im Gegensatz zum vorher genannten „Relevance-System“ wird hier ein größerer Raum betrachtet als eine einzelne Suchanfrage. Sobald ein Nutzer eine Suchanfrage stellt, wird mithilfe des „Relevance-Systems“ eine Sortierung von Suchergebnissen zurückgegeben. Diese wird jedoch nochmal durch das zuvor erstellte Profil des Nutzers neu angeordnet.

Die bisherigen Arten sind relativ statisch und arbeiten auf einer bestimmten Vorgabe. Da es sehr aufwendig ist, ein gutes Profil zu erstellen, sollte dies das Analysesystem dynamisch selbst durchführen können. Dies ist nun Aufgabe der „Recommender Systeme“. Hierbei wird mithilfe von Algorithmen, die als Datenbasis Langzeitinformationen nutzen, die Analyse von Inhalten der Webseiten weiter an den Nutzer angepasst. Da dieses gesamte Konstrukt auf dem Relevance System basiert und dieses mit verschieden passenden Anfragen erweitert, um das Ergebnis zu strukturieren, muss das Recommender System abschätzen können, welche Suchanfragen der Nutzer an dessen Stelle geben würde.

Im Gegensatz zu den drei vorher genannten Systemen gibt es auch ein System, das auf der Analyse der Links basiert, und somit „Link-Analysis“ genannt wird. Bei diesem Ansatz, wird die Topologie des Internets zur Relevanzbestimmung genutzt. So muss nicht der Inhalt selbst auf Relevanz überprüft werden, sondern die Verlinkungen der Webseiten werden analysiert. Beim sogenannten topic-sensitive-PageRank-Algorithmus werden, basierend auf den Verlinkungen unterschiedlicher Webseiten und der Thematik selbst, verschiedene Kategorien und eine Wertung für Webseiten erstellt. Der personalisierte topic-sensitive-PageRank -Algorithmus ist eine Erweiterung des allgemeinen topic-sensitive-PageRank. Hierbei werden Nutzer Profile beziehungsweise deren Webseiten-Vorlieben in die Bewertung der Webseiten mit einbezogen.

Zuletzt soll noch ein weiteres System genannt werden, welches nicht nur auf einen Nutzer fixiert ist. Beim sogenannten „Social Search Engine“ wird das Verhalten von verschiedenen Nutzergruppen in die Wertung für Webseiten mit einbezogen. Dieser Ansatz entstand durch die Beobachtung der Gesellschaft, da zum Beispiel Bücher- oder Filmkäufe meist aufgrund von Empfehlungen Anderer durchgeführt werden. Die „Social-Search-Engine“ wird zu den genannten Arten verhältnismäßig öfter bei den „Recommender Systems“ verwendet, indem man Nutzer, die ein ähnliches Profil besitzen zusammen betrachtet. Die Seitenwertung wird damit unter dem Leitgedanken „Das Kollektiv hat (meistens) recht“ ausgeführt.

Schlussendlich kann gesagt werden, dass heutige Suchmaschinen auf einer Mischung der genannten Arten basieren. Außerdem fällt auf, dass für ein gutes Suchergebnis ein gutes Profil des Nutzers benötig wird, welches durch die Verwendung von dynamischen Systemen immer weiter verbessert wird und somit eine kontinuierliche Analyse des Nutzers voraussetzt. In [2] findet man einen tieferen Einblick in diese Thematik.

2.3 Voraussetzungen der Personalisierung

Um eine Personalisierung von Nutzern durchzuführen, reicht es nicht aus, verschiedene Daten nur in einer großen Liste zu speichern. Dazu muss man sich verschiedene Fragen stellen, die speziell für die Art und deren Ziel zu beantworten sind. Da für jedes Verwendungsgebiet jeweils andere Anforderungen vorhanden sind, werden hier nur Fragen angeführt, die im Fall einer Suchmaschinenpersonalisierung zu beantwortet sind. Im Folgenden wird hauptsächlich auf das Thema Personalisierung

von Websuchen eingegangen und erklärt, wie diese optimal verwendet wird.

Angefangen wird mit der Frage „Wie beschafft man sich die Daten?“, beziehungsweise „Welche Methode soll man zu Speicherung verwenden?“ Hierbei wird zwischen direkter und indirekter Datenbeschaffung unterschieden. Da es für den Nutzer unangenehm ist, bei jeder Systemverwendung ein direktes Feedback zu geben, wird meist die indirekte Form verwendet. Hierbei wird das Verhalten von Nutzer und System beobachtet und teilweise noch mit der Möglichkeit der optionalen direkten Feedbackgabe erweitert. Für die Websuche bietet sich eine indirekte Datensammlung sehr gut an, da eine nahtlose Kommunikation zwischen Nutzer und System vorhanden ist und Suchanfrage eine gute Basis zur Personalisierung bieten.

Die nächste Frage ist „Wie oder wo sollen bereits erstellte oder noch zu erstellende Profile gespeichert werden?“ Auch hier gibt es zwei Möglichkeiten, die serverseitige und die clientseitige Speicherung. Die Vorteile der clientseitigen Datenspeicherung liegen in der erhöhten Privatsphäre und Sicherheit des Nutzers. So verbleiben private Daten auf dem Nutzersystem und können hier auch jederzeit gelöscht werden. Die serverseitige Datenspeicherung würde aber die Auswertung von Verhaltensmustern der Nutzer besser unterstützen, außerdem spielt das verwendete Clientsystem (z.B. verschiedene Arten von Betriebssystemen) keine Rolle. Andererseits erzeugt eine serverseitige Nutzung eine höhere Auslastung des Servers. Somit sollte eine Lösung gewählt werden, die beide Seiten, Client wie auch Server, optimal ausnutzt. Hierbei ist es zu beachten, wie oft ein Austausch von Informationen der beiden Seiten stattfindet. Eine durchgehende Verbindung ist dabei empfehlenswert, da Informationen unmittelbar auf beiden Seiten synchronisiert und somit aktuell gehalten werden können.

Eine weitere wichtige Frage ist „Soll sich das System über die Dauer der Verwendung an den Nutzer anpassen?“ Bei Websuchen ist diese Frage mit einem eindeutigen „Ja“ zu beantworten, da Websuchen zumeist auf dem Themen basiert, die für eine Person in diesem Zeitraum interessant sind. Da sich Interessen aber schnell ändern können, sollte die Möglichkeit der Anpassbarkeit gegeben sein. Empfehlenswert ist es, dem Nutzer selbst eine Anpassung des Systems zu erlauben. Diese Form findet man zum Beispiel oft bei werbefinanzierten Streaming-Portalen, die eine Grundeinstellung der Werbeinteressen ermöglichen, aber auch Suchmaschinenbetreiber stellen individuelle Konfigurationsmöglichkeiten bereit.

Wenn man sich überlegt, wie man die Profile an Nutzer anpasst, sollte auch die Frage „Wie werden die Profile aufgebaut und was sollen sie enthalten?“ gestellt werden. Um diese aber zu beantworten, sollte die Art und Methode der Personalisierung beziehungsweise deren Algorithmus festgelegt sein. Dabei kann man entweder auf verschiedene Drittanbieter zurückgreifen, an die nur die reinen Informationen gesendet werden, welche dann ausgewertet werden, oder man verwendet ein eigenes System, die die Relevanz der Suchergebnisse mit berücksichtigt und bewertet. Da die Entwicklung von korrekten und vor allem schnellen Algorithmen mit viel Aufwand verbunden ist, wird ein solcher Algorithmus entweder nicht oder nur mit hohen Lizenzgebühren veröffentlicht. Aus der Struktur und der Funktionsweise ergibt sich eine Vorgabe für den Aufbau der Profile, zum Beispiel wie verschiedene Themen gewichtet und wie die Informationen gespeichert werden (Bsp.: Liste, Tabelle, Cluster).

Letztendlich muss man sich noch überlegen: „Wie sieht die Nutzerschnittstelle aus?“ Dabei ist vor allem zu beachten, welche Arten von Clients verwendet werden. So verhalten sich Desktop-Webbrowser anders als mobile Webbrowser. Außerdem kann eine spezielle Nutzer-Applikation auf der Client Seite vorliegen, die ein bestimmtes Format vorrausetzen, zum Beispiel eine beschränkte Zeichenmenge, aufgrund von Ressourcenmangel.

Unter [2] werden die hier genannten Themen gezielter ausgeführt.

3. EXPERIMENT UND UMFRAGE

In den folgenden Kapiteln werden ein Experiment und eine Umfrage jeweils mit deren Ergebnissen zum Thema „Der Wert von Datenschutz bei Websuchen“ vorgestellt und verglichen.

3.1 Vorbereitung und Ablauf des Experiments

3.1.1 Durchführung

Das Ziel des Experiments war es herauszufinden, wie viel den Nutzern einer Suchmaschine die Privatsphäre wert sei. Dazu wurden zu verschiedenen Zeitpunkten am University College London (UCL) 189 Testpersonen eingeladen. Diese wurden vorher innerhalb des Universitätsbereichs angesprochen, ob sie an einem Test einer neuen Suchmaschine, welche „Find Fever“ genannt wurde, teilnehmen würden. Die Absicht das Verhalten und Sucheinstellungen zu beobachten wurde nicht explizit genannt, aber im Verlauf des Experiments wurde gefragt, welche Einstellung die jeweilige Person zum Thema Datenschutz hat.

Nachdem sich die Testpersonen registriert und angemeldet hatten, sollten sie verschiedene Fragen beantworten, wobei sie das Internet frei verwenden konnten, jedoch alle Suchmaschinenanfragen auf die „Find Fever“-Seite verlinkt wurden. Es wurden verschiedene Themen behandelt, welche zum Teil einen allgemeinen Charakter hatten, zum anderen Teil sensiblere Informationen abfragten. Folgend ein kurzer Ausschnitt aus dem Fragenkatalog [1]:

- „What is the population size of Little Shelford, England? “
- „What is the address of the tourist information in York, England? “
- „Where can you buy lingerie in Chelsea? “
- „What is the maximum penalty for attempted sexual intercourse with girl under 13? “
- „Which airport is closest to you? “
- „What is the proportion of water in human faces? “

Das Experiment selbst war in zwei Teile gegliedert. Im ersten Durchlauf waren die in Tabelle 1 gegebenen Einstellungs-möglichkeiten kostenlos verfügbar. Im zweiten Teil konnte man, wie in Tabelle 1 gezeigt, Credits für bestimmte Einstellungen ausgeben und für andere zurückgewinnen. Die vergebenen Credits wurden am Ende mit der gegebenen Entlohnung der

doi: 10.2313/NET-2014-08-1_17

Teilnahme an dem Experiment abgeglichen. (Ein Credit war ungefähr 0,30€ wert.) [1]

Tabelle 1. Einstellungsmöglichkeiten und deren Kosten beziehungsweise Gewinne während des Experiments

Einstellung	Credits
Speichere meinen Suchverlauf nicht.	-2
Entferne Werbung auf der Ergebnisseite.	-2
Gib meine Suchanfragen nicht an Dritte weiter.	-2
Verbessere das Suchergebnis.	-1
Hebe die Suchanfrage in den Ergebnissen hervor.	-1
Kopple die Suchergebnisse mit meinem Standort.	kostenlos
Verwende die sichere Suche Funktion.	kostenlos
Veröffentliche meine Suche auf Twitter.	+2
Speichere, auf welche Ergebnisse ich klicke.	+2

3.1.2 Aufgestellte Hypothesen

Vor Beginn des Experimentes wurden andere Versuche betrachtet. Die Versuche sind unter [8] und [9] beschrieben. Dabei fand man heraus, dass Privatsphäre schon ein wichtiger Teil in Online-Shops ist. So sind Kunden bereit, mehr Geld für Waren auszugeben, wenn ein sensiblerer Umgang mit ihren Daten gegeben ist. Dennoch gibt es Abstufungen, wie viel mehr ein Produkt kosten darf und auf welche Art in die Privatsphäre eingedrungen wird. Somit zeigten die Ergebnisse, dass, sobald der Rabatt verlockend genug war, auf in die Privatsphäre eingreifende Webseiten gewechselt wurde. Des Weiteren wurde festgestellt, dass zum Beispiel kaum mehr Geld ausgegeben wurde, wenn es um das Versenden von Werbung an die E-Mailadressen der Nutzer ging. Siehe hierzu auch [10]. Außerdem fand man heraus, dass Kunden, sobald es sich um sensiblere Produkte oder Suchanfragen handelt, ihre Privatsphäre ausgeprägter schützen wollen, als bei Alltäglichem. Wie sich die Person dabei genau verhalten hat, war je nach Land und Person zwar unterschiedlich, es konnten jedoch Tendenzen festgestellt werden.

Aus diesem Vorwissen wurden in [1] diese fünf Hypothesen aufgestellt, die es zu bestätigen oder zu widerlegen galt:

H1 – The price of privacy-enhancing features and the proportion of users enabling them are negatively associated.

H2 – The more sensitive the search task, the more likely users will enable privacy-enhancing features.

H3 – The more sensitive the search task, the less likely users will enable privacy-invasive features.

H4 – Users who are more concerned about privacy will enable privacy-enhancing features more often.

H5 – Users who consider privacy-enhancing features more important will enable them more often.

3.2 Ergebnisse des Experiments

Bevor die Ergebnisse des Experiments dargestellt werden, soll hiermit noch angemerkt werden, dass bei der Durchführung verschiedene Punkte aufgefallen sind, die in zukünftigen Feldversuchen verändert werden sollten. So sollte man zum Beispiel beachten, dass Probanden im Verlauf eines Experiments ihr Verhalten ändern können und unterschiedliche Arten für Einstellungen zu Bezahlen auch ein unterschiedliches Verhalten gegenüber der Auswahl von Optionen mit sich bringt (z.B. ein Abonnement). Siehe hierzu Seite 12 unter [1].

3.2.1 Bestätigte Hypothesen

Nachdem das Experiment durchgeführt und ausgewertet war, ist ein Ergebnis stark aufgefallen. Viele der Einstellungs-möglichkeiten, die eine positive Auswirkung auf die Privatsphäre ausübten, wurden, solang sie kostenlos waren, von sehr vielen Probanden genutzt, jedoch verringerte sich die Anzahl der Nutzer drastisch, sobald man dafür bezahlen musste. So fiel zum Beispiel die Anzahl der Personen, die die „Verlauf nicht speichern“-Funktion nutzten, auf ein Viertel, sobald 2 Credits dafür verlangt wurden. Somit zeigte sich, dass die Einführung von Kosten die Nutzer herausfiltert, die Wert auf diese Funktionen legen. Zum Beispiel fiel die Nutzung der „Kein Verlauf“-Funktion von 58% auf 15% und die Nutzung der „Keine Weitergabe an Dritte“-Funktion von 79% auf 16%, als diese mit Kosten verbunden waren. Durch dieses Ergebnis konnte Hypothese 1 bestätigt werden.

Weiterhin konnte Hypothese 2 und Hypothese 3 ebenfalls voll bestätigt werden. So bemerkte man einen deutlichen Anstieg der Nutzung von den Funktionen „Verlauf nicht Speichern“ und „Keine Weitergabe an Dritte“, sobald es um sensible Suchanfragen ging. Bei der Veröffentlichung von Suchanfragen auf Twitter, war das Ergebnis genau umgedreht. Umso sensibler die Suchanfrage war, desto seltener wurde die „Auf Twitter veröffentlichen“-Funktion genutzt. Während des Experiments war auch aufgefallen, dass sobald die Funktionen nicht mehr kostenlos verfügbar waren, diese auch bei sensibleren Anfragen nicht mehr aktiviert wurden.

Hypothese 5 konnte hingegen nur teilweise bestätigt werden. Die Nutzer für die eine „Verlauf nicht Speichern“-Funktion laut ihrer Angabe wichtig sei, haben diese nicht häufiger oder weniger häufig eingesetzt, als andere Teilnehmer. Eine Bestätigung der Hypothese konnte man aber bei der „Nicht an Dritte weitergeben“-Funktion registrieren. Hier war der Anteil der Personen, die für die Wichtigkeit gestimmt haben, höher als bei der anderen Gruppe, die nicht so viel Bedeutung in diese Funktion legen. Dennoch wurde ersichtlich, dass Nutzer ungern Geld für ihre Sicherheitseinstellungen ausgeben, auch wenn sie vorher angaben, diese Funktionen seien ihnen wichtig. [1]

3.2.2 Widerlegte Hypothesen

Neben der teilweise widerlegten beziehungsweise bestätigten Hypothese 5 (dessen Ergebnis in Kapitel 3.2.1 beschrieben wird), konnte Hypothese 4 vollständig widerlegt werden. So haben die Nutzer zwar angegeben, dass großer Wert auf Datenschutz gelegt wird, jedoch wurden, unabhängig von den Creditkosten, die datenschützenden Funktionen, wie zum Beispiel „Verlauf nicht speichern“ und „Keine Weitergabe an

Dritte“, nicht häufiger genutzt. Bei weiterem Nachfragen fiel aber auf, dass die Probanden keine genaue Auskunft darüber geben konnten, welche Optionen zum Datenschutz beitragen und welche nicht. So wurde das Deaktivieren von Werbung auf der Ergebnisseite der Suchanfrage schon von vielen Nutzern als eine Erhöhung des Datenschutzes gesehen. Außerdem ist bei der Auswertung aufgefallen, dass mehrere Nutzer zwar angaben, ihnen sei Datenschutz wichtig, jedoch wurden ihre Aussagen von ihren Handlungen und Einstellungen widerlegt. [1]

3.3 Umfrage

Wie auch bei Kapitel 3.2 wird hiermit verdeutlicht, dass die hier genannte Umfrage aufgrund der geringen Anzahl an Teilnehmern, selbst nicht repräsentativ, jedoch soll damit ein Vergleich zum Experiment zur Verfügung gestellt und ein Bezug zwischen Aussagen und Handeln verschiedener Personen dargestellt werden.

3.3.1 Aufbau

Um die Hypothesen und Ergebnisse aus dem im Absatz 3.1. und 3.2 vorgestellten Experiment vergleichen zu können, wurde eine, an das Experiment angepasste, Internetumfrage erstellt, die sich hauptsächlich an den Thesen und Einstellungsmöglichkeiten des Experiments orientiert. Diese Umfrage wurde an 19 Personen, mit unterschiedlichem Alter und Lebenslage verteilt, welche diese vollständig ausgefüllt und zurückgesendet haben. In der Umfrage wurden verschiedene Fragen zu den verwendeten Sucheinstellungen gestellt. Die Fragestellung wurde teilweise so angepasst, dass sich eine Vergleichbarkeit zum Experiment ergab. Dennoch konnte man eigene Kommentare zu den jeweiligen Fragestellungen abgeben, welche auch oft genutzt wurden und so das Gesamtbild der Ergebnisse verbesserten. Am Anfang der Umfrage wurde nach der verwendeten Suchmaschine, der Nutzung der „Verlauf speichern“-Funktion und den gewünschten Einstellungsmöglichkeiten gefragt. Man konnte jeweils angeben, ob die in Tabelle 3 angegebenen Möglichkeiten genutzt werden oder nicht. Diese Frage wurde in zwei Teile aufgespalten. Im ersten Teil wurde angenommen, dass die Nutzung der Einstellungsmöglichkeiten kostenfrei zur Verfügung gestellt wird. Im zweiten Teil hingegen, sollte man für die Aktivierung der Einstellungsmöglichkeiten bezahlen. Ein Beispiel für eine Fragestellung ist: „Würden Sie eine No-Ads-Funktion verwenden?“, diese hatte die Antwortmöglichkeit „Ja, ich würde No-Ads verwenden.“, beziehungsweise „Nein, oder es ist mir egal.“. Der nächste Teil der Umfrage beschäftigte sich mit der Thematik der Seitenfinanzierung, welche Optionen die Befragten akzeptieren würden. Die möglichen Themen waren: „Werbung auf der Ergebnisseite“, „Sponsored Links“, „Die Suche würde auf Twitter veröffentlicht“ und „Die Suchergebnisse, auf die Sie klicken, werden gespeichert.“ Zum Schluss wurden noch nach der persönlichen Einstellung gegenüber Datenschutz und der Häufigkeit der Verwendung von Anonymisierungsdiensten und VPN-Verbindungen gefragt. [3]

3.3.2 Ergebnis und Vergleich mit dem Experiment

Sowohl im Experiment, als auch in der Umfrage stellte sich heraus, dass den Nutzern, laut eigener Angabe, Datenschutz sehr wichtig ist, aber das Handeln diese Aussage nicht unterstützt. So werden persönliche Daten eher preisgegeben, sobald der Schutz entweder etwas kostet oder mit Aufwand verbunden ist.

In Tabelle 2 und 3 werden jeweils die Prozentzahlen der Nutzer einer bestimmten Einstellung gegeben. Dabei steht ein + für die Aktivierung beziehungsweise Nutzung des Features und ein – für die Deaktivierung beziehungsweise nicht Nutzung eines Features, jeweils im Vergleich, ob es kostenlos oder kostenpflichtig angeboten wurde. Ein großer Unterschied zwischen dem Experiment und der Umfrage ist beim Thema „Weitergabe an Dritte“. Dabei ist der Prozentsatz im Experiment von 79% auf 16% gefallen, sobald dafür Geld verlangt wurde. In der Umfrage jedoch sank der Prozentsatz von 100% nur auf 74%. Somit würden die Nutzer auch in einer gewissen Form dafür bezahlen, dass ihre Daten unter Verschluss bleiben. Auch fällt hierbei auf, dass die Nutzer ihre Daten nur an Personen oder Firmen weitergeben wollen, denen sie persönlich vertrauen. In den Ergebnissen lässt sich auch erkennen, dass das Entfernen von Werbung und Sponsored-Links, beziehungsweise das verbessern des Suchergebnisses und das Hervorheben von Suchanfragen im Ergebnis, zwar ein gern gesehenes Feature ist, jedoch vernachlässigt wird, sobald dafür Kosten entstanden wären. Beim Thema „Miteinbeziehen des Standorts in der Suchanfrage“ fällt auf, dass dieses Feature im Experiment nach den Zahlen in der Tabelle von fast allen genutzt wurde und bei der Umfrage jedoch kaum. Dies lässt sich aber leicht durch das Entstehen der Zahlen erklären. So wurde im Experiment nur berücksichtigt, ob diese Funktion min. einmal oder häufiger genutzt wurde. Somit, wurde jede Person gezählt, die auch nur ein einziges Mal, diese Funktion verwendet hat. In der Umfrage ist dies umgedreht. Hierbei wurde die Ortung bei der Ergebnissuche im Allgemeinen abgelehnt, jedoch wenn zum Beispiel Restaurants in der Nähe gesucht werden sollten, würden die meisten Befragten diese Funktion aktivieren, sofern sie kostenlos zur Verfügung gestellt wird.

Des Weiteren kann man an den Ergebnissen sehen, dass die Nutzer ihre Sucheinstellungen nach ihrer Suchanfrage anpassen. Das erkennt man sowohl an den Einstellungen bei empfindlichen Suchanfragen im Experiment, als auch in den Feedbacktexten der Umfrage. Ein Beispiel dafür ist die Safe-Search-Funktion. Diese wurde im Experiment etwa gleich oft genutzt, egal ob dafür Geld verlangt wurde oder nicht. Laut Umfrage wurde diese Funktion, solang Sie kostenlos war, von 21% immer und von 53%, wenn Kinder und Jugendliche geschützt werden sollten, genutzt. Sobald Geld dafür verlangt wurde verschob sich dieser Bereich zu Nein, jedoch war das Thema Jugendschutz immer noch wichtig, wie man im Feedback lesen konnte.

Zum Thema Seitenfinanzierung, verhielten sich Experiment und Umfrage ähnlich, bis auf eine Einstellungsmöglichkeit. Im Experiment stimmten 75% einer Veröffentlichung ihrer Suche auf Twitter zu. Nach der Umfrage wurde die Veröffentlichung strikt abgelehnt, siehe Tabelle 4. Die restlichen 16% würden nur zustimmen, sofern die Veröffentlichung anonym oder mit großen Vorteilen verbunden wäre. [1][3]

doi: 10.2313/NET-2014-08-1_17

Tabelle 2. Nutzung von Einstellungen im Experiment (in %)

	Kostenlos		Mit Credits	
Einstellung	+	-	+	-
Speichere meinen Suchverlauf nicht.	58	42	15	85
Entferne Werbung auf der Ergebnisseite.	94	6	19	81
Gib meine Suchanfragen nicht an Dritte weiter.	79	21	16	84
Verbessere das Suchergebnis.	~61	~39	~52	~48
Hebe die Suchanfrage in den Ergebnissen hervor.	~50	~50	~45	~55
Kopple die Suchergebnisse an meinen Standort.	~90	~10	~80	~20
Verwende die sichere Suche Funktion.	~69	~31	~69	~31
Veröffentliche meine Suche auf Twitter.	~75	~25	~75	~25
Speichere, auf welche Ergebnisse ich klicke.	~88	~12	~80	~20

Tabelle 3. Nutzung von Einstellungen nach der Umfrage (in %)

	Kostenlos		Kostenplichtig	
Einstellung	+	-	+	-
Entferne Werbung auf der Ergebnisseite.	89	11	42	58
Sponsored-Links verbieten	47	53	26	74
Gib meine Suchanfragen nicht an Dritte weiter.	100	0	74	26
Verbessere das Suchergebnis.	74	26	32	68
Hebe die Suchanfrage in den Ergebnissen hervor.	74	26	21	79
Kopple die Suchergebnisse an meinen Standort.	11	89	5	95
Verwende die sichere Suche Funktion.	74	26	49	53

Tabelle 4. Möglichkeiten der Seitenfinanzierung in der Umfrage (in %)

Finanzierungsmöglichkeit	+	-
Werbung auf der Ergebnisseite	84	16
Sponsored Links auf der Ergebnisseite	74	26
Suchergebnisse auf Twitter veröffentlichen	16	84
Suchergebnisse, auf die Sie klicken werden gespeichert	64	37

Die Umfrage ging noch auf zwei weitere Themen ein, die so nicht im Experiment behandelt wurden. Zum einen, welche Suchmaschinen verwendet werden, zum Anderen wie wichtig Datenschutz ist und ob Schutzmaßnahmen ergriffen werden.
Dabei stellte sich, wie in Abbildung 1 gegeben, heraus, dass der Hauptanteil eindeutig bei Google liegt. Obwohl es in der Datenschutzerklärung von Google genau beschrieben ist, dass Google alle Suchanfragen detailreich abspeichert [11], verwenden nur 3 Personen Startpage beziehungsweise nur 2 DuckDuckGo, welche die Google Search Engine zur Suchergebnisverbesserung verwenden, aber keine Datenspeicherung vornehmen. [3]

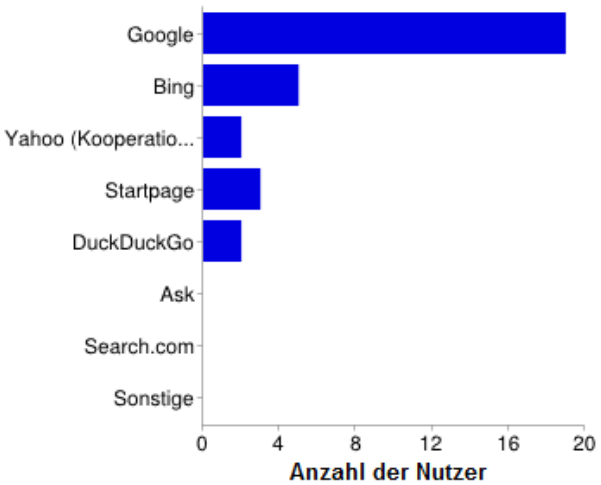

Abbildung 1: Auswertung zur Frage: „Welche Suchmaschinen verwenden Sie.

Außerdem wurde nach der Bedeutung von Datenschutz gefragt und der Häufigkeit der Verwendung von Anonymisierungssoftware, gegeben in Abbildung 2 und 3. Beim Vergleich der beiden Diagramme fällt auf, dass diese Spiegelungen voneinander sind. So ist den Befragten zwar ihr Datenschutz sehr wichtig, aber sie nutzen meistens keine Software zum Verschleiern ihrer Identität. Genauso sieht es bei der Verwendung von VPN-Verbindungen aus. So verwenden 64% entweder gar keine gesicherte Verbindung oder nur, wenn sie müssen, zum Beispiel das Einwählen in das Uni- oder Firmennetz. Nur 37% nutzen eine gesicherte Verbindung, sobald sie mit einem öffentlichen, für alle zugänglichen Netz verbunden

doi: 10.2313/NET-2014-08-1_17

sind. Die kaum vorhandene Sicherheit dieser öffentlichen Verbindungen wurde in den Medien bereits ausgiebig diskutiert. Als Beispiel siehe hierzu [12]. Dennoch wird die Möglichkeit einer VPN-Verbindung nicht häufig genutzt, da dies entweder mit Kosten oder mit Aufwand verbunden ist. [3]

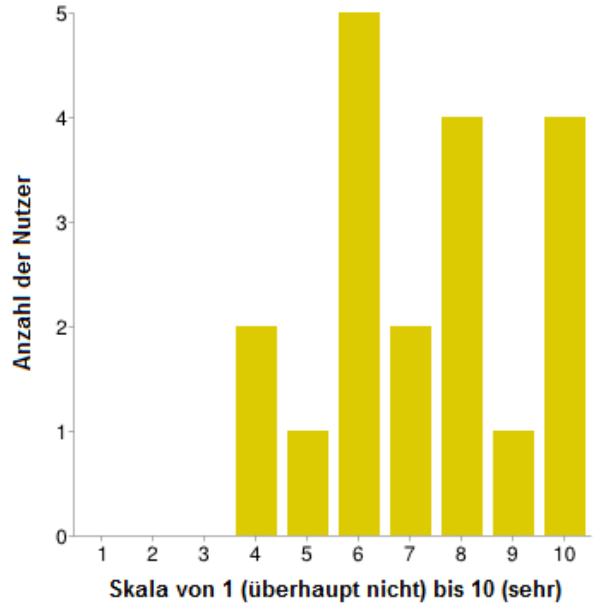

Abbildung 2: Auswertung zur Frage: „Wie wichtig ist Ihnen Datenschutz?“

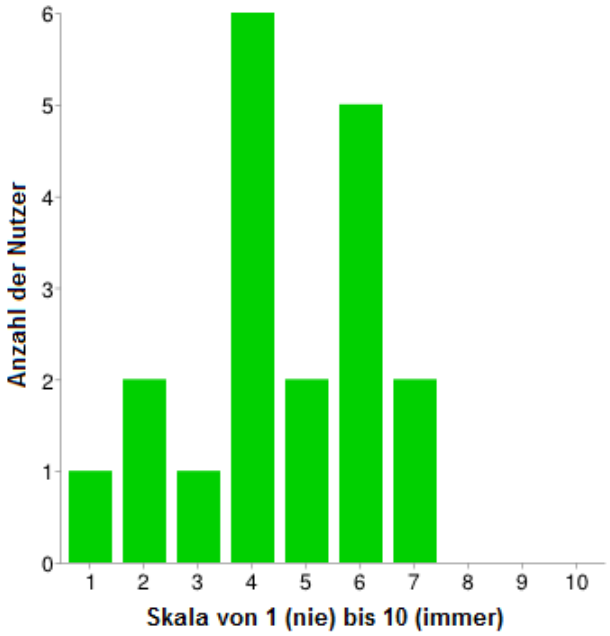

Abbildung 3: Auswertung zur Frage: „Wie oft verwenden Sie Anonymisierungssoftware?“

4. ZUSAMMENFASSUNG

Der Begriff Datenschutz beziehungsweise Schutz der persönlichen Daten im Internet ist in den Medien bereits seit längerer Zeit ein fester Bestandteil. Man müsste somit annehmen, dass ein Mindestmaß an Anonymisierung und somit dem Verschleiern von eigenen persönlichen Daten standardmäßig bei der Nutzung des Internets und hier vor allem auch bei der Suche im World Wide Web vorhanden sein müsste.

Sowohl nach dem Experiment, als auch in der Umfrage, stellte sich heraus, dass den Nutzern von Suchmaschinen, laut eigener Aussagen, sehr viel an Datenschutz liegt. Aber sobald Kosten oder ein gewisser Aufwand entstehen, tritt der Schutz persönlicher Daten oft in den Hintergrund. Außerdem ist aufgefallen, dass anonymisierte Suchmaschinen eher unbekannt sind und die bekannten Firmen, wie Google und Microsoft (Bing), bisher keine Möglichkeit bieten für die Anonymisierung eigener Daten zu bezahlen. Somit gibt es für den normalen Internetnutzer nur die Möglichkeit, durch die Preisgabe seiner Daten, ein akkurates Suchergebnis zu erhalten.

doi: 10.2313/NET-2014-08-1_17

5. REFERENZEN

[1] Sören, Preibusch. 2013. The value of privacy in Web search. Microsoft Research Cambridge UK. The twelfth workshop on the economics of information security in WEIS 2013. http://weis2013.econinfosec.org/papers/PreibuschWEIS2013.pdf. Zuletzt aufgerufen: 30.07.2014

[2] K. Keenoy und M. Levene. Personalisation of Web search. In B. Mobasher und S. S. Anand, Editoren, Intelligent Techniques for Web Personalization, Volume 3169 of Lecture Notes in Computer Science, Seiten 201-228. Springer Berlin Heidelberg, 2005. http://link.springer.com/chapter/10.1007%2F11577935_11

[3] Eigene durchgeführte Umfrage Mai 2014

[4] Urteil des Europäischen Gerichtshofs in der Rechtssache C-131/12 vom 13. Mai 2014. http://curia.europa.eu/juris/document/document.jsf?text=&docid=152065&pageIndex=0&doclang=DE&mode=req&dir=&occ=first&part=1&cid=258738, Zuletzt aufgerufen: 10.06.2014

[5] Google Formular zur Löschung von Suchanfragen. 2014. https://support.google.com/legal/contact/lr_eudpa?product=websearch&hl=de. Zuletzt aufgerufen: 30.07.2014

[6] Tor-Project. Juli 2014. https://www.torproject.org/about/overview.html.en. Zuletzt aufgerufen: 30.07.2014

[7] JonDoNym. Juli 2014. http://www.anonym-surfen.de/prinzip.html. Zuletzt aufgerufen: 30.07.2014

[8] Janice Tsai, Serge Egelman, Lorrie Cranor, and Alessandro Acquisti, "The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study," in WEIS 2007.

[9] Alastair R. Beresford, Dorothea Kübler, and Sören Preibusch, "Unwillingness to pay for privacy: A field experiment," Economics Letters, vol. 117, no. 1, Seiten 25-27, 2012.

[10] Nicola Jentzsch, Sören Preibusch, and Andreas Harasser, "Study on monetising privacy. An economic model for pricing personal information," European Network and information Security Agency (ENISA), 2012.

[11] Datenschutzerklärung Google. Stand 31. März 2014. Absatz: Von uns erhobene Informationen. https://www.google.de/intl/de/policies/privacy/. Zuletzt aufgerufen: 30.07.2014

[12] Artikel: Die Hotspot Falle. Heft C't 1/2012. http://heise.de/-1394646. Zuletzt aufgerufen: 30.07.2014

[13] Bundesdatenschutzgesetz §3. Stand 2014. http://www.gesetze-im-internet.de/bdsg_1990/__3.html. Zuletzt aufgerufen: 31.07.2014

[14] Artikel: Gelöscht und Gut? von Alexander Drößler. http://www.tagesschau.de/wirtschaft/google-urteil-100.html. Zuletzt aufgerufen: 31.07.2014

[15] TUM Vorlesung „IT-Security". Prof. Dr. Eckert, Dipl.-Inf. Thomas Kittel. Wintersemeseter 13/14

[16] TUM Vorlesung „Network Security". Prof. Dr.Ing. Georg Carle, Dr. Heiko Niedermayer, Ralph Holz. Wintersemester 13/14

[17] Google Trends. Juni 2014. https://www.google.de/trends/. Zuletzt auftgerufen: 01.08.2014