Query Suggestion mit Siri

Bachelorarbeit zur Erlangung des Bachelor-Grades
Bachelor of Science im Studiengang Angewandte Informationswissenschaft
an der Fakultät für Informations- und Kommunikationswissenschaften
der Technischen Hochschule Köln

vorgelegt von: Birte Langkammerer

eingereicht bei: Prof. Dr. Philipp Schaer
Zweitgutachter/in: Malte Bonart M.Sc.

Köln, 07.08.2019

Kurzfassung/Abstract

Suchanfragen und automatische Vorschläge zu diesen, wenn der Nutzer die Anfrage gerade noch eintippt, gehören heutzutage zum Standard. Das nicht nur bei Suchen im Internet, sondern auch mithilfe von integrierten Assistenten an PC oder auf Mobilgeräten, wie Smartphones oder Tablets. Einer dieser persönlichen Assistenten ist Siri, eine Software auf iOS-Geräten des Technologiekonzerns Apple. Siri ist hauptsächlich bekannt dafür, als Sprachassistent auf gesprochene Anfragen zu reagieren. Allerdings bietet Siri auch eine Suchfunktion auf dem Homescreen des Geräts an, in die Suchen eingetippt werden können. Auch hier werden Vorschläge gemacht, die die Eingabe während des Schreibens automatisch vervollständigen. Dabei ist aber nicht klar, woher diese stammen. Gut denkbar ist eine Kooperation mit einem etablierten Anbieter einer Websuchmaschinen. Aber klare Aussagen, von Apple selbst, finden sich nicht. Lässt sich dies eventuell auf experimentellem Weg ermitteln? Um sich der Lösung dieser Fragestellung zu nähern, stelle diese Bachelorarbeit die Umsetzung eines Versuchsaufbaus dar, bei dem über einen vierwöchigen Zeitraum definierte Suchanfragen an die Websuchmaschinen Google, Bing, DuckDuckGo und an die Siri-Suche gestellt wurden. Durch Analysemethoden, wie unter anderem Rank-biased overlap (RBO), sollten so Gemeinsamkeiten ermittelt werden, die gegebenenfalls auf einen konkreten Partner schließen lassen. Zwar zeigten die Vergleichsmethoden durchaus Unterschiede auf, ein klares Ergebnis in Bezug auf eine der betrachteten Suchmaschinen, konnte allerdings nicht erzielt werden.

Query Suggestion; Query Autocompletion; Suchmaschine; Siri; Apple

Search queries and automatic suggestions for these, when the user is just typing in the query, are standard nowadays. This not only applies to searches on the Internet, but also with the help of integrated assistants on PCs or mobile devices such as smartphones or tablets. One of these personal assistants is Siri, software on iOS devices from the Apple technology group. Siri is mainly known as a speech assistant for responding to spoken requests. However, Siri also offers a search function on the device's home screen, where searches can be typed in. Here, too, suggestions are made that automatically complete the input as you type. However, it is not clear where they come from. A cooperation with an established provider of web search engines is conceivable. But clear statements, from Apple itself, cannot be found. Can this possibly be determined experimentally? In order to approach the solution to this question, this bachelor thesis represents the implementation of an experimental setup in which defined search queries were made to the web search engines Google, Bing, DuckDuckGo and Siri search over a four-week period. Analysis methods such as Rank-biased overlap (RBO) were used to identify commonalities that might lead to the conclusion that a specific partner is involved. Although the comparison methods did reveal differences, a clear result in relation to one of the search engines considered, could not be achieved.

Query Suggestion; Query Autocompletion; Search Engine; Siri; Apple

Inhalt

1 Einleitung

Suchanfragen werden längst nicht mehr nur über klassische Suchen im Web und mit den bekannten Suchmaschinen wie Google, Bing oder DuckDuckGo gestellt. Inzwischen gibt es auch die Möglichkeit, in sogenannten integrierten Assistenten im Internet zu suchen. Beispiele hierfür sind Cortana in der Microsoft-Umgebung oder Siri bei Apple-Produkten. Bekannt sind diese Assistenten vor allem für die Fähigkeit, auf eine gesprochene Anfrage eines Users zu reagieren.[1]
Gleichzeitig bieten diese Technologien auch Desktopsuchen direkt auf den Geräten an. Neben dem Durchsuchen von lokalen Daten wie E-Mails oder Kontakten tauchen dabei dann zusätzlich Inhalte aus dem Internet in den Ergebnissen auf.[2]

Wie in den bekannten Websuchen werden auch hierbei die gestellten Suchanfragen durch sogenannte Query Suggestion automatisch vervollständigt. Wo diese Vorschläge bei Cortana denen der Suchmaschine Bing entsprechen[3], bleibt die Quelle solcher Vorschläge bei Siri unklar. Apple selbst bestätigt die Weitergabe vereinzelter Anfragen in einem iOS Security Guide aus dem Mai 2019:

> „In some cases, Suggestions may forward queries for common words and phrases to a qualified partner in order to receive an display the partner's search results."[4]

Zudem gibt es Berichte darüber, dass Webanfragen innerhalb von iOS Google-Ergebnisse enthalten, es herrscht aber Unklarheit darüber, ob genau diese Quelle bei Siri genutzt wird, um die Suchen der Nutzer durch Vorschläge zu vervollständigen.[5] Erste eigene Tests konnten hierzu allerdings keine eindeutige Deckungsgleichheit aufzeigen.

Dementsprechend lautet die zentrale Fragestellung dieser Bachelorarbeit: Ist es möglich, durch einen Vergleich von zurückgelieferten Vorschlägen zu definierten Anfragen über einen festgesetzten Zeitraum konkrete Quellen zu bestimmen, auf denen die Query Suggestions in der Siri Suche beruhen?

Hierfür bildet ein Vergleich der Query Suggestions von Siri mit denen der genannten Websuchmaschinen Google, Bing oder DuckDuckGo die Basis. Dabei bietet es sich an, über einen definierten Zeitraum, regelmäßig und möglichst standardisiert die gleichen Anfragen an alle Suchmaschinen abzusenden. Dahinter steht die Idee, die so entstehenden Listen von Vorschlägen miteinander vergleichen zu können.

[1] vgl. Kerkmann, C. / Scheuer S. (2018): Elektronikmesse IFA: Siri, Alexa und Google Home – wie Sprachassistenten die Technikwelt verändern.

[2] vgl. Apple Inc. (20. Mai 2019): Suchfunktion auf dem iPhone, iPad oder iPod touch verwenden.

[3] vgl. Gavin, R. (2018): Delivering Personalized Search Experiences in Windows 10 through Cortana.

[4] Apple Inc. (Mai 2019): iOS Security: iOS 12.3, May 2019, S. 70.

[5] vgl. Panzarino, M (2017): Apple switches from Bing to Google for Siri web search results on iOS and Spotlight on Mac.

Ausgehend von der Annahme, dass es sich bei den Query Suggestions um ein unvollständiges Ranking handelt, wie im Bereich der Suchmaschinen auch die Sortierung der Suchergebnisse eines ist, kommt zum Vergleich der Vorschläge der genannten Suchmaschinen die Methode des Rank-Biased-Overlap (RBO) in Frage. Diese Methode eignet sich, um die Ähnlichkeit von solchen Rankings miteinander zu vergleichen. [6]

Abgesehen von dieser Methode, mit der ein Vergleich von unterschiedlichen Listen möglich ist, werden als weitere Methoden die Frequenz- und Schnittmengenanalyse, sowie eine intellektuelle Clusterbildung mit Berechnung eines Scores zur Ermittlung von Häufigkeiten der festgelegten Kategorien des Clusters, angewendet. Diese voneinander unabhängigen Methoden dienen dazu, sich der Beantwortung der oben genannten Fragestellung zu nähern.

Im Rahmen dieser Arbeit soll eine programmatische Umsetzung erfolgen, um den Prozess der täglichen Datenerhebung zu automatisieren.

[6] vgl. Webber et al. (2010): A similarity measure for indefinite rankings, S. 1.

2 Hintergründe zur Software Siri

In diesem Kapitel wird einleitend erläutert, um was für eine Software es sich bei Siri handelt und welcher Teil der von Siri bereitgestellten Daten im Rahmen dieser Arbeit genau analysiert werden.

2.1 Siri

Siri ist eine Software, die das Technologie Unternehmen Apple Inc. zunächst in die Software iOS seiner Mobilgeräte integriert hat. Dabei ist Siri ein sogenannter persönlicher Assistent, der in erster Linie auf gesprochene Kommandos reagiert. [7] Die Software dient zur Erledigung von Aufgaben, wie Anrufe tätigen, Textnachrichten schreiben oder Internetsuchen auf Basis der gesprochenen Anfragen durchzuführen [8] Siri wird seit dem iPhone 4s unter iOS mitgeliefert.[9]

Neben dieser Funktion als Sprachassistent bieten iOS Geräte auch eine Suchfunktion an, für die ebenfalls Siri verwendet wird. [10] Diese funktioniert mit einem Suchschlitz auf dem Home-Bildschirm der Geräte. In diesen Suchschlitz kann eine Suchanfrage eingegeben werden kann. Diesen Bereich bezeichnet Apple im bereits vorgestellten iOS Security Guide als „Search“.[11]

Dabei werden schon während der Eingabe Vorschläge generiert, die textuell oder in abgesetzten Boxen unter dem Suchschlitz erscheinen. Diese Vorschläge stammen ebenfalls von Siri und laufen unter dem Titel Siri-Vorschläge in Suchen. Diese beziehen dabei Inhalte aus lokalen Apps, wie E-Mails, aber eben auch Internet-Inhalte mit ein. [12]

Die Vorschläge werden individuell zur Suchanfrage generiert. Dabei sind in einem explorativen Test im Zuge dieser Arbeit folgende gängige Inhalte ermittelt worden. In den beschriebenen Boxen werden unter anderem aktuelle „News“ und „Siri-Website-Vorschläge“ angeboten. Daneben wird je nach Eingabe „Siri-Wissen“ angezeigt. Dieses beinhaltet kurze Informationstexte unter Einbeziehung von Internetquellen. Bei erwähntem explorativem Test hat sich ergeben, dass zur Anzeige dieser Informationen meist auf die Internet-Enzyklopädie Wikipedia zurückgriffen wird. Dies ist auch im iOS-Security Guide belegt.[13] Weitere Quellen sind jedoch ebenfalls nicht auszuschließen. Neben dem Genannten, werden teilweise auch passende „Webvideos“ oder Vorschläge zu „Filmen“ unterbreitet. Darüber hinaus erfolgen oftmals auch passende Ortsvorschläge innerhalb der systeminternen App „Karten“ oder zu sozialen Netzwerken wie Twitter. Exemplarisch

[7] vgl. Anderson B. / Galvez K. (2018): Apple, S. 1.

[8] vgl. ebd.

[9] vgl. Apple Inc. (2019): iPhone Benutzerhandbuch, S 41.

[10] vgl. Apple Inc. (20. Mai 2019): Suchfunktion auf dem iPhone, iPad oder iPod touch verwenden.

[11] Apple Inc. (Mai 2019): iOS Security: iOS 12.3, May 2019, S. 70.

[12] vgl. ebd.

[13] vgl. ebd.

zeigen Abbildung 1 und 2 einen Screenshot der auf dem Home-Screen generierten Vorschläge bei Eingabe des Suchbegriffs „merkel“ am 19.06.2019.

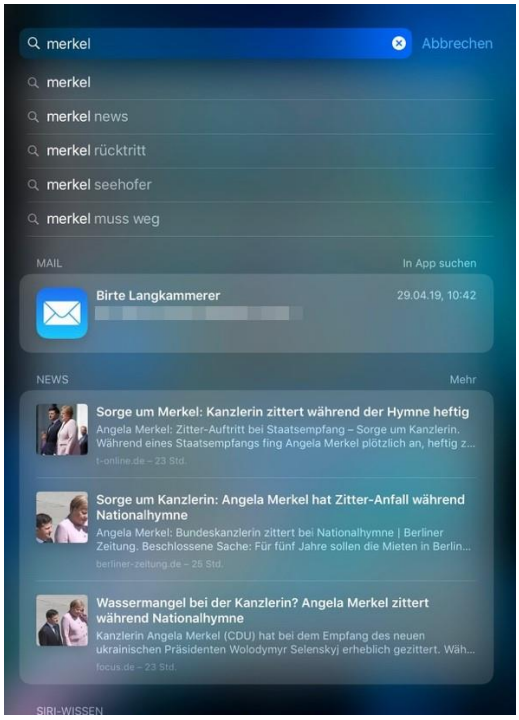


Abbildung 1: Screenshot am Apple iPad von Siri-Vorschlägen in „Suchen“ zur Suchanfrage nach „merkel“ mit Autocompletions, Anzeige lokalen Inhalts aus der App „Mail“ (Betreff der Mail unkenntlich gemacht) und Vorschlägen zu „News“

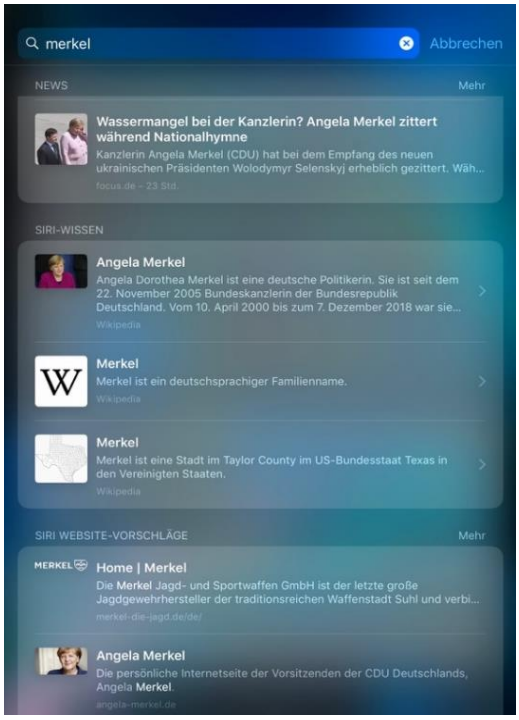


Abbildung 2: Screenshot am Apple iPad von Siri-Vorschlägen in „Suchen“ zur Suchanfrage nach „merkel“ Vorschlägen zu „News“, „Siri-Wissen“ und „Siri-Website-Vorschläge“

Im nachfolgend in Kapitel 3 beschriebenen Versuchsaufbau, auf den sich diese Bachelorarbeit in der Hauptsache stützt, erfolgt allerdings nicht die Untersuchung der oben beschriebenen thematisch untergliederten Felder, die bei der Suche erscheinen. Vielmehr werden die Autocompletions, untersucht, die unmittelbar unter der Suchleiste generiert werden (s. Abbildung 1).

3 Versuchsaufbau

3.1 Vorstellung der für den Vergleich verwendeten Websuchmaschinen

Um Ähnlichkeiten zwischen den Autocompletions der Suchmaschinen zu ermitteln und so idealerweise auch Hinweise darauf aufzudecken, wie die Query Suggestion in der Siri Suche unter iOS entstehen, wird für diese Arbeit ein Versuchsaufbau konstruiert.

In diesen Versuch fließen neben den Ergebnissen der Siri-Suche auch die Vorschläge der Suchmaschinen Google, Bing und DuckDuckGo ein.

3.1.1 DuckDuckGo

DuckDuckGo ist eine Suchmaschine, angibt, keine Daten zu sammeln und diese nicht mit dritten Personen teilt. In den Datenschutzbestimmungen vom Anbieter wird verdeutlicht, wie das sogenannte „search leakage“, also das Sammeln von Informationen über den Anfragenden und weiterleiten an aufrufende Seiten verhindert wird.[14]

3.1.2 Google

Für einen Suchmaschinenvergleich wird ebenfalls Google verwendet. Google hat im weltweiten Vergleich einen Marktanteil von fast 74 Prozent.[15] Da dies die bekannteste und am meisten verwendete Suchmaschine darstellt, ist hier ein Vergleich mit der Siri-Suche interessant.

3.1.3 Bing

Als Suchmaschine mit dem zweithöchsten Marktanteil weltweit hinter Google, wird Bing ebenfalls zum Vergleich herangezogen.[16] Bing ist eine Websuchmaschine von Microsoft.

3.1.4 Yahoo

Die Suchmaschine Yahoo ist zeitweise Teil der Datenerhebung (12.06.-23.06.2019). Da zwischen DuckDuckGo und Yahoo eine bestätigte Kooperation besteht und sich die Erhebungsergebnisse nachgewiesenermaßen gleich, gehen die Vorschläge von Yahoo abgesehen von der Schnittmengen- und Frequenzanalyse nicht in die weitere Auswertung ein.[17]

[14] vgl. DuckDuckGo (o.D): We don’t collect or share personal information: That’s our privacy policy in a nutshell.

[15] vgl. NetMarketShare (2019) Marktanteile der Suchmaschinen - Mobil und stationär 2019.

[16] vgl. ebd.

[17] vgl. DuckDuckGo (2019) Result Sources.

3.2 Vorgehen bei Datensammlung und Datenspeicherung

Das folgende Kapitel thematisiert die programmatische Umsetzung der automatischen Datenerhebung und die manuell durchgeführte Datensammlung unter iOS.

3.2.1 Automatisierte Datensammlung bei Websuchmaschinen

Zugriff auf die Autocompletions der Websuchmaschinen erhält man über eine API (Application Programming Interface). Eine API ist eine Programmierschnittstelle, die die Kommunikation zwischen zwei Programmen möglich macht. Hierbei werden Daten nach einer standardisierten Struktur ausgetauscht. [18]

Für alle für den Vergleich herangezogenen Websuchmaschinen existieren sogenannte Autocompletion APIs. DuckDuckGo stellt eine offizielle API zur Verfügung.[19] Die API URLs von Google[20] und Bing[21] wurden für diese Arbeit von Malte Bonart zur Verfügung gestellt.

Um die Durchführung des täglichen Sammelns der Daten zu erleichtern, besteht die Möglichkeit, das Absenden der Suchbegriffe und das Dokumentieren der Antworten in ein Skript auszulagern.

Hierfür wurde ein JavaScript Skript geschrieben, das eine Datei im CSV-Format zeilenweise ausliest. Jeder Eintrag in einer Zeile repräsentiert einen Suchbegriff. Die jeweilige Anfrage erfolgt dabei über einen HTTP Request, der für jeden Suchbegriff abgesendet wird.

Die Antworten der APIs erfolgen dann im Format JSON (JavaScript Object Notation). Dabei handelt es eine leichte, textbasierte, von einer bestimmten Programmiersprache unabhängige Syntax zur Definition von Datenaustauschformaten.[22]JSON ist „für Menschen einfach zu lesen und zu schreiben und für Maschinen einfach zu parsen“ [23]

Eine Hürde bei der Erstellung des Skripts entsteht dabei dadurch, dass zwar alle Antworten im JSON Format vorliegen, die Struktur der Antwort dabei aber jeweils unterschiedlich ist, wie Abbildungen 3-5 verdeutlichen.

[18] vgl. Geißler, O./ Ostler, U. (2018): Was ist ein Application-Programming-Interface (API)?.

[19] https://api.duckduckgo.com/ac/?q=beispiel&kl=de-de&format=json&pretty=1

[20] http://www.google.com/complete/search?q=beispiel&client=psy-ab

[21] http://api.bing.net/osjson.aspx?q=beispiel

[22] vgl. Ecma International (2017): Standard ECMA-404. The JSON Data Interchange Syntax.

[23] Ohne Autor (o.D.): Einführung in JSON.

```
[ 'cristiano ronaldo',
  [ 'cristiano ronaldo',
    'cristiano ronaldo instagram',
    'cristiano ronaldo steckbrief',
    'cristiano ronaldo vermögen',
    'cristiano ronaldo wikipedia',
    'cristiano ronaldo 7 live stream',
    'cristiano ronaldo kinder',
    'cristiano ronaldo freundin',
    'cristiano ronaldo transfermarkt',
    'cristiano ronaldo schlafrhythmus',
    'cristiano ronaldo bugatti',
    'cristiano ronaldo erfolge' ] ]
```

Abbildung 3: Antwortbeispiel über die Bing-API

```
Cristiano Ronaldo
[ { phrase: 'cristiano ronaldo freundin' },
  { phrase: 'cristiano ronaldo sohn' },
  { phrase: 'cristiano ronaldo instagram' },
  { phrase: 'cristiano ronaldo bilder' },
  { phrase: 'cristiano ronaldo wallpaper' },
  { phrase: 'cristiano ronaldo girlfriend' },
  { phrase: 'cristiano ronaldo wikipedia' },
  { phrase: 'cristiano ronaldo news' },
  { phrase: 'cristiano ronaldo frisur' } ]
```

Abbildung 4: Antwortbeispiel über die DuckDuckGo-API

```
[ 'Cristiano Ronaldo',
  [ [ 'cristiano ronaldo', 0 ],
    [ 'cristiano ronaldo<b> jr</b>', 0 ],
    [ 'cristiano ronaldo<b> kinder</b>', 0 ],
    [ 'cristiano ronaldo<b> alter</b>', 0 ],
    [ 'cristiano ronaldo<b> gehalt</b>', 0 ],
    [ 'cristiano ronaldo<b> auto</b>', 0 ],
    [ 'cristiano ronaldo<b> frau</b>', 0 ],
    [ 'cristiano ronaldo<b> statue</b>', 0 ],
    [ 'cristiano ronaldo<b> gewicht</b>', 0 ],
    [ 'cristiano ronaldo<b> insta</b>', 0 ] ],
  { q: 'gqf5vK6XFk2-pEoc0xwXLS9KKl8',
    t: { bpc: false, phi: 0, tlw: false } } ]
```

Abbildung 5: Antwortbeispiel über die Google-API

Die Unterschiede in der Struktur haben Auswirkungen auf das Parsen der Antwort-Arrays. Diese werden so weiterverarbeitet, dass aus den JSON-Antworten nur die entsprechenden Autocompletions in eine CSV-Datei geschrieben werden. Das gewählte CSV-Format bietet die Möglichkeit, die gesammelten Daten durch Trennzeichen separiert abzuspeichern und für die spätere Analyse flexibel wieder einzulesen. Die Datei hat dabei die in Abbildung 6 dargestellte Struktur. Mit der Plattform wird die Quelle der jeweiligen Vorschläge dokumentiert. Daneben der Suchbegriff, zu dem die Vorschläge erfolgen und zur Dokumentation des Datums ein Zeitstempel im ISO-Format. Dahinter folgen dann die Autocompletions und die Position, um für die Weiterverarbeitung den Rang des Vorschlags vorliegen zu haben.

```
Plattform, Suchbegriff, Datum, Vorschlag, Position
Bing, Instagram, 2019-05-26T11:49:16.879Z,  login, 1
Bing, Instagram, 2019-05-26T11:49:16.879Z,  suche, 2
Bing, Instagram, 2019-05-26T11:49:16.879Z,  heidi klum, 3
Bing, Instagram, 2019-05-26T11:49:16.879Z,  anmelden, 4
Bing, Instagram, 2019-05-26T11:49:16.879Z,  account löschen, 5
Bing, Instagram, 2019-05-26T11:49:16.879Z,  dolunay, 6
Bing, Instagram, 2019-05-26T11:49:16.879Z,  download, 7
Bing, Instagram, 2019-05-26T11:49:16.879Z,  friesennerz, 8
Bing, Instagram, 2019-05-26T11:49:16.879Z,  daniela katzenberger, 9
Bing, Instagram, 2019-05-26T11:49:16.879Z,  lena meyer landrut, 10
Bing, Instagram, 2019-05-26T11:49:16.879Z,  bibisbeautypalace, 11
Bing, Cristiano Ronaldo, 2019-05-26T11:49:16.879Z,  instagram, 1
Bing, Cristiano Ronaldo, 2019-05-26T11:49:16.879Z,  steckbrief, 2
Bing, Cristiano Ronaldo, 2019-05-26T11:49:16.879Z,  kinder, 3
Bing, Cristiano Ronaldo, 2019-05-26T11:49:16.879Z,  wikipedia, 4
Bing, Cristiano Ronaldo, 2019-05-26T11:49:16.879Z,  autos, 5
Bing, Cristiano Ronaldo, 2019-05-26T11:49:16.879Z,  vermögen, 6
Bing, Cristiano Ronaldo, 2019-05-26T11:49:16.879Z,  bugatti, 7
Bing, Cristiano Ronaldo, 2019-05-26T11:49:16.879Z,  rekorde, 8
Bing, Cristiano Ronaldo, 2019-05-26T11:49:16.879Z,  transfermarkt, 9
Bing, Cristiano Ronaldo, 2019-05-26T11:49:16.879Z,  familie, 10
Bing, Cristiano Ronaldo, 2019-05-26T11:49:16.879Z,  stream, 11
```

Abbildung 6: Auszug von Teilergebnissen eines Erhebungstages zur Suchmaschine Bing

Neben der in Abbildung 6 gezeigten Plattform „Bing“ wurden für die übrigen Plattformen für die Kennzeichnung innerhalb der CSV folgende Benennungen gewählt „Google“, „DuckDuckGo“, „siri_simulator“ für die Suche innerhalb der iOS-Simulation und „siri_tab“ für die Suche am Tablet.

3.2.2 Händisches Vorgehen unter iOS

Eine Automatisierung, wie zuvor für die Websuchmaschinen beschrieben, ist für den Zugriff auf die Autocompletions in der Siri-Suche so nicht möglich. Apple stellt hierfür keine API zur Verfügung, mit der man eine Liste der Vorschläge für einen Suchbegriff erhalten würde.

Somit ist das händische dokumentieren der Vorschläge, die die Siri-Suche zu den gewählten Suchbegriffen liefert, für diese Bachelorarbeit unumgänglich.

Die Suggestions der Suche in Siri werden für den Versuch an zwei Stellen erhoben. Für diese gewählte Aufteilung wird ein Apple-Gerät mit aktuellem MacOS und ein weiteres Gerät mit dem Betriebssystem iOS benötigt.

Zum einen kommt zur Umsetzung der Datensammlung Xcode 10.2.1 zum Einsatz, welche die integrierte Entwicklungsumgebung (IDE) von Apple ist, die für die App-Entwicklung für die Apple-Betriebssysteme MacOS, iOS konzipiert ist.[24] Innerhalb von Xcode steht ein Simulator zur Verfügung, in dem sämtliche Geräte aus Apple-Portfolio simuliert werden können. Das für den Versuchsaufbau gewählte Gerät im Simulator ist ein Apple iPad Air 2019. Üblicherweise dient dieser Simulator dazu, entwickelte Apps zu testen.

Innerhalb des im Simulator gestarteten Geräts lässt sich allerdings auch auf das Widget Fenster zugreifen, in dem sich die Siri-Suche befindet. An dieser Stelle werden die Suchbegriffe eingegeben. Die Einstellungen des Apple-Geräts im Simulator können

[24] Vgl. Brunsmann et al. (2017): Apps prorgrammieren mit Swift, S. 32.

durch einen simulierten Werksreset auf die Grundeinstellung zurückgesetzt werden. Somit ist durch die Verwendung des Simulators gewährleistet, dass täglich in einer „sauberen“ Umgebung gesucht wird.

Neben dem Simulator werden die Abfragen ebenfalls an einem iPad Air 2019 durchgeführt. Das Gerät ist regelmäßig in Gebrauch mit einer Apple-ID, also einem Benutzerkonto für Apple-Geräte verknüpft, die seit mehreren Jahren verwendet wird. Durch diese Apple-ID sollte auf diesem Gerät also eine Nutzerhistorie vorhanden sein, die sich gegebenenfalls auf die Vorschläge auswirken.

Somit gib es für die Siri Suche zwei Szenarien, zu denen die Autocompletions untereinander und jeweils auch einzeln mit den Ergebnissen der Websuchmaschinen verglichen werden können.

3.3 Verwendete Suchbegriffe

Um einen Vergleich über die verschiedenen Suchmaschinen zu gewährleisten, werden über einen Zeitraum von vier Wochen täglich die gleichen, zuvor definierten Anfragen an alle Suchmaschinen bzw. in der Siri-Suche abgesetzt. Hierfür wurden die Profile des sozialen Netzwerks Instagram, genauer die 20 Accounts mit den weltweit meisten Followern gewählt.[25] In diesem Ranking tauchen sowohl Namen von Prominenten als auch von Marken und Firmen auf.

Teilweise handelt es sich bei den Bezeichnungen der Instagram-Accounts um Künstlernamen oder Kunstformen der Namen der Personen. Um hier eine bessere Grundlage für einen Vergleich erzielen zu können und den damit entstehenden Bezug der Namen zu den Instagram-Profilen aufzulösen, wurde in solchen Fällen der Name des Instagram-Profils gegen geläufige Namen der betrachteten Personen ersetzt (s. Tabelle 1). Für das Instagram-Profil „Barbie“, hinter dem eigentlich die Sängerin Nicki Minaj steht, wurde diese Anpassung versäumt.[26] Diese Profilbezeichnung und folglich auch der als Suchbegriff verwendetet Term wurde fälschlicherweise als die Marke „Barbie“ der Firma Mattel interpretiert.[27] Daraus folgend sind Vorschläge zu diesem Suchbegriff auch in allen Analysemethode als Vorschlag zu einem Suchbegriff der Gruppe der Marken und Firmen zugeordnet. Dies stellt allerdings für die angewendeten Vergleichsmethoden keine Verfälschung dar.

[25] vgl. Trackalytics.com (2019): The Most Followed Instagram Profiles.

[26] vgl. Minaj, Nicki (2019): Barbie (@nickiminaj).

[27] vgl. Mattel (2019): Barbie – Lustige Spiele, Videos und Aktivitäten für Mädchen.

Bezeichnung des Instagram-Profils[28]	**Verwendete Bezeichnung zur Datenerhebung**
therock	The Rock
Kylie	Kylie Jenner
Beyonc	Beyoncé
3n310ta neymarjr	Neymar jr
Kendall	Kendall Jenner
Khlo	Khloe Kardashian

Tabelle 1: Name einzelner im Versuch betrachteter Instagram-Profile, die für die Datenerhebung ersetzt oder ergänzt werden

Zu den 16 in Tabelle 2 aufgelisteten Personen kommen noch folgende vier Instagram-Profile dazu, die Firmen bzw. Marken repräsentieren:

- Instagram
- National Geographic
- Nike
- Barbie

Person	**Geschlecht**	**Beruf**
Cristiano Ronaldo	männlich	Sportler
Ariana Grande	weiblich	Sängerin/Schauspielerin
Selena Gomez	weiblich	Schauspielerin/Sängerin
The Rock	männlich	Schauspieler
Kim Kardashian West	weiblich	Kardashians/Model
Kylie Jenner	weiblich	Kardashians/Unternehmerin
Beyoncé	weiblich	Sängerin
Taylor Swift	weiblich	Sängerin
Leo Messi	männlich	Sportler
Neymar jr	männlich	Sportler
Kendall Jenner	weiblich	Kardashians/Model
Justin Bieber	männlich	Sänger
Khloe Kardashian	weiblich	Kardashians/Model
Jennifer Lopez	weiblich	Sängerin/Schauspielerin
Miley Cyrus	weiblich	Sängerin/Schauspielerin
Katy Perry	weiblich	Sängerin

Tabelle 2: Namen der natürlichen Personen unter den für den Versuch ausgewählten Instagram Profilen, mit Zuordnung des Geschlechts und dem Beruf der Person

Hinter der Auswahl dieser Accounts als Suchbegriffe für den Versuchsaufbau während steht der Gedanke, dass es sich dabei vornehmlich um prominente Personen und große Firmen bzw. Marken handelt. So ergibt sich zum einen durch die vorherige Definition die Möglichkeit, die Ergebnisse im zeitlichen Verlauf zu betrachten und zu vergleichen,

[28] vgl. Trackalytics.com (2019): The Most Followed Instagram Profiles.

andererseits ist es realistisch, dass sich im Zeitverlauf Veränderungen in den Vorschlägen ergeben können. Dies könnte dann zum beispielsweise, bedingt durch aktuelle Berichterstattung, Auswirkungen zeigen.

4 Methoden zur Auswertung der gesammelten Daten

In diesem Kapitel werden die drei Methoden vorgestellt, die für diese Bachelorarbeit zu einem analytischen Vergleich der erhobenen Query Suggestions der einbezogenen Suchmaschinen, eingesetzt werden. Nach einer kurzen Erläuterung zu den eingesetzten Programmen zur Analyse wird die angewendete Frequenz- und Schnittmengenanalyse erklärt. Anschließend wird das angewendete Verfahren des Rank-biased Overlap erläutert, wonach eine Erklärung des vorgenommenen Clusterings und der darauf aufbauenden Score Berechnung erfolgt.

Die verwendete Software zur Auswertung ist Microsoft Excel und R. Excel bietet sich daher an, da mit dem dort enthaltenen Tool Pivot-Tabellen aus der Gesamtheit der Datensammlung gewünschte Daten herauszugreifen und für weitere Analyse gegenüberzustellen. „PivotTable ist ein leistungsfähiges Tool zum Berechnen, Zusammenfassen und Analysieren von Daten, mit dem Sie Vergleiche vornehmen sowie Muster und Trends in Ihren Daten erkennen können.“[29]

Die Programmiersprache R ist Open Source und bietet sich für den Einsatz von statistischen Berechnungen und anschließender Visualisierung an. Da in dieser Arbeit mit einer großen Datenmenge gearbeitet wird, können diese mit R eingelesen, verarbeitet und auch grafisch dargestellt werden. [30]

4.1 Frequenzanalyse + Schnittmengenanalyse

Für eine erste Annäherung an die Autocompletions, die während des Versuchs gesammelt werden, wird eine Frequenzanalyse durchgeführt.

Dabei werden die in der Ergebnismenge vorgeschlagenen Terme zunächst gezählt. Hier erfolgt eine Aufschlüsselung nach der Anzahl in den Vorschlägen der einzelnen Suchmaschinen und das Vorkommen in der Summe.

Daneben wird aus der Ergebnismenge die Anzahl der Unique Terms ermittelt. Um diese zu erhalten, wird jeder Begriff in der Liste nur einmal in die Zählung aufgenommen. Jedes weitere Vorkommen wird in der Rechnung nicht berücksichtigt. Die Anzahl der Unique Terms ist somit ein Maßstab, wie viele unterschiedliche Vorschläge in die Ergebnisse einfließen.

Die Unique Terms bilden die Basis für eine anschließende Schnittmengenanalyse zwischen den betrachteten Suchmaschinen. Dabei wird die Schnittmenge von Unique Terms zwischen den Suchmaschinen berechnet, was erste Hinweise darauf geben kann, welche Suchmaschinen grundsätzlich ähnliche Query Autocompletions liefern.

Auch wenn die Schnittmengen der Unique Terms einen ersten Hinweis auf Ähnlichkeiten bieten können, so ist dennoch nur eine vorsichtige Betrachtung der Ergebnisse

[29] Microsoft (o.D.): Erstellen einer PivotTable zum Analysieren von Arbeitsblattdaten.

[30] vgl. The R Foundation (o.D.) What is R?.

angeraten. Denn bei dieser Herangehensweise werden zwar Überschneidungen sichtbar, allerdings geht durch das Herunterbrechen auf das Vorkommen eines Terms die Information über die Häufigkeit des Auftretens bei der Schnittmengenanalyse verloren.

4.2 Rank-biased overlap (RBO)

In der Arbeit „A Similarity Measure for Indefinite Rankings" von William Webber et. al. wird das Maß rank-biased overlap (RBO) zur Bestimmung der Ähnlichkeit von unvollständigen Rankings vorgestellt. Es vergleicht die Überschneidung von zwei unvollständigen Ranglisten in zunehmender Tiefe.[31]

Klassischerweise sind die Ergebnislisten von Suchmaschinen - und darauf übertragen, auch die Vorschlaglisten von Suchmaschinen - eben solche unvollständigen Rankings. Das bedeutet in diesem Zusammenhang beispielsweise, dass die Liste nicht alle möglichen Vorschläge führt, die es zum abgefragten Suchbegriff geben kann. Neben dem Charakteristikum der Unvollständigkeit weisen solchen Listen die Eigenschaft der Top-Gewichtung auf, also der obere Teil der Liste bedeutender ist, als das Ende. Ein drittes Merkmal ist das der Unbestimmtheit. Eine Liste in einer bestimmten Tiefe abzuschneiden ist in erster Linie eine willkürliche Entscheidung, ist allerdings zu rechtfertigen, da der Wichtigkeit der Listenelemente durch das Prinzip der Top-Gewichtung abnimmt.[32]

RBO kann Werte zwischen 0 und 1 annehmen. Dabei bedeutet ein Wert von 0, dass die Listen keine Überschneidungen aufweisen. 1 wiederum zeigt vollkommene Gleichheit an.[33]

Mittels eines Parameters p, der zwischen 0 und 1 liegen kann, lässt sich bestimmen, wie stark die Top-Gewichtung des Verfahrens ausgeprägt ist. Je kleiner das p ist, desto deutlicher ist diese ausgeprägt. Das bedeutet, wenn p=0, so wird nur das erste Element einer Liste betrachtet. Je näher p umgekehrt an 1 liegt, desto weniger Bedeutung erhält diese Gewichtung der Spitze einer Liste. [34]

Um den Parameter p passend auf den Anwendungsfall bestimmen zu können, wird im Vorfeld eine Betrachtungstiefe d festgelegt. Dieser Parameter d repräsentiert dabei die Anzahl der Elemente einer Liste, die in Betrachtung einbezogen werden.

Die Berechnung von p erfolgt dann mit folgender Formel, wobei die Bezeichnung $W_{RBO}(d)$ die Gewichtung von Rang *d* repräsentiert:

$$p = 1 - W_{RBO}\,(1{:}\,d)$$

[31] vgl. Webber et al. (2010): A similarity measure for indefinite rankings, S. 2.

[32] vgl. ebd,, S. 1f.

[33] vgl. ebd,, S. 15

[34] ebd.

Dieser Berechnung folgend, liegen bei einem p=0,9 86% des Gewichts des Verfahrens auf den ersten 10 betrachteten Positionen eines Rankings. Bei einem Ranking, bei dem die ersten 50 Positionen mit gleicher Gewichtung betrachtet werden sollen, wäre Parameter p entsprechend 0,98.[35]

Um der ungleichen Länge der zu vergleichenden Listen Rechnung zu tragen, wird in dieser Bachelorarbeit die extrapolierende Version des RBO zur Anwendung kommen. Dieser rechnet den Grad der Übereinstimmung in einer sichtbaren Tiefe hoch. Dies geschieht unter der Annahme, dass sich die bis dahin berechnete Übereinstimmung weiter fortsetzt.[36]

4.3 Clustering

In Christopher D. Mannings „Introduction to Information Retrieval“ wird die grundsätzliche Idee des Clusterings so beschrieben, dass Algorithmen eine Reihe von Dokumenten in Untergruppen bzw. Cluster einteilen.[37] Das Ziel des Clusterings ist demnach das Folgende:

> „The algorithms´goal is to create clusters that are coherent internally, but clearly different from each other.”[38]

Im Rahmen dieser Arbeit bietet es sich an, ein solches Clustering durchzuführen, um aus den gesammelten Vorschlägen thematische Gruppierungen herauszuarbeiten. Anstelle der Anwendung eines Algorithmus wird für diese Arbeit allerdings ein rein intellektuelles, also manuelles Verfahren zur Bildung der Cluster durchgeführt. Die Umsetzung dieses Verfahrens ist in Kapitel 5.5 beschrieben. Dieses Clustergruppen können dann die Basis zur Berechnung eines Scores herangezogen werde, um zu ermitteln, zu welchen Clusterkategorien besonders häufig Vorschläge geliefert werden

[35] vgl. Webber et al. (2010): A similarity measure for indefinite rankings, S. 15.

[36] ebd.

[37] vgl. Manning et al. (2010): Introduction to Information Retrieval, S. 321

[38] ebd.

4.4 Berechnung eines Scores zur Darstellung der durchschnittlichen Precision zu einer Kategorie

Die Precision ist ein Standardmaß zur Evaluierung der Qualität der Ergebnisse eines Retrieval Systems. Dabei repräsentiert sie den Anteil der Dokumente der für eine Anfrage relevant ist.[39]

Die entsprechende Formel dazu lautet:

$$Precision = p = \frac{|\, R \cap A \,|}{|\, A \,|}$$

Dabei repräsentiert $|\, R \cap A \,|$ die Schnittmenge von relevanten Dokumenten und der Anzahl der Dokumente, die zu einer Anfrage zurückgeliefert wurden. $|\, A \,|$ repräsentiert die zurückgelieferte Anzahl der Dokumente zu einer Anfrage.[40]

Diese Berechnung wird auf den Anwendungsfall in dieser Arbeit übertragen. Es soll ermittelt werden, wie oft ein Vorschlag aus einer betrachteten und somit relevanten Kategorie in einer Antwortliste vorkommt. Dies geschieht für jede Kategorie an jedem Erhebungstag.

Um einen durchschnittlichen Score über den gesamten Erhebungszeitraum für jede Kategorie zu ermitteln, werden die einzelnen errechneten Precision-Werte aufsummiert und ein Mittel über die Anzahl der Erhebungstage gebildet.

Dieser Score dient dazu in den Rankings der Vorschläge auf das reine Vorkommen von Kategorien zu prüfen, um sich der Fragestellung zu nähern mit welcher Wahrscheinlichkeit ein Vorschlag X zu einer Anfrage Y aus Kategorie Z stammt.

[39] vgl. Baeza-Yates R. / Ribeiro-Neto, B. (2011): Modern Information Retrieval, S.134 f.

[40] ebd.

5 Durchführung der gewählten Methoden

5.1 Datenerhebung

In diesem Kapitel wird exemplarisch der Ablauf der Datenerhebung beschrieben.

Die Datenerhebung erfolgte vom 26.05.2019 bis 23.06.2019 über 29 Tage. Die Anfragen wurden dabei in der Regel aus Köln abgesetzt. Ausnahme bildete ein Zeitraum vom 06.06.2019 bis 11.06. 2019, in dem die Erhebung von Simmern/Hunsrück in Rheinland-Pfalz aus, durchgeführt wurde.

Das Ausführen des Skripts für das Erfassen der Autocompletions erfolgt über die Kommandozeile (Konsole). Das Skript ist so aufgebaut, dass Abfragen getrennt nach Suchmaschinen durchgeführt werden können. Beispielhaft ist dies in Abbildung 7 für DuckDuckGo dargestellt. Das Skript wird mittels des Befehls „node suchmaschinenAbfrage.js“ gestartet. Die Ergänzung dauer.txt markiert die Mitgabe der Bedingung, dass auf die Textdatei dauer.txt zugegriffen werden soll, aus der die Queryterms ausgelesen werden sollen. Mit der Angabe „ddg“ in der Kommandozeile wird bestimmt, dass der http-Request nur an die API von DuckDuckGo gesendet wird. Ohne eine Bestimmung an dieser Stelle würde eine Anfrage an alle hinterlegten APIs erfolgen.

```
Christians-MBP:skript birtelangkammerer$ node suchmaschinenAbfrage.js dauer.txt ddg
DUCKDUCKGO
https://api.duckduckgo.com/ac/?q=instagram&kl=de-de&format=json&pretty=1
https://api.duckduckgo.com/ac/?q=cristiano ronaldo&kl=de-de&format=json&pretty=1
https://api.duckduckgo.com/ac/?q=ariana grande&kl=de-de&format=json&pretty=1
https://api.duckduckgo.com/ac/?q=selena gomez&kl=de-de&format=json&pretty=1
https://api.duckduckgo.com/ac/?q=the rock&kl=de-de&format=json&pretty=1
https://api.duckduckgo.com/ac/?q=kim kardashian west&kl=de-de&format=json&pretty=1
```

Abbildung 7: Kommando in der Konsole zur Ausführung des Skripts. Dabei das Auslesen der Suchbegriffe aus der Datei „dauer.txt“ und Absenden des https-Requests an die DuckDuckGo-API

Die Ergebnisse werden hierbei für jede Suchmaschine in eine eigene CSV-Datei geschrieben. Dieses Vorgehen zeigt sich insofern praktikabel, als dass es sofort die tägliche Bereinigung der Daten erleichtert. Diese ist täglich notwendig, da es vorkommt, dass das Muster einer Antwort bei den Suchmaschinen bei DuckDuckGo und Bing nicht immer „Suchbegriff – Vorschlag“ entspricht. In diesem Fall kann das Skript zur Erhebung der Daten diese Antwort nicht korrekt verarbeiten und schreibt ein „undefined“ zum Suchbegriff in der CSV-Datei als Vorschlag. Zudem macht eine fehlerhafte Darstellung von Sonderzeichen bedingt durch die Codierung bei Google macht eine Bereinigung notwendig.

Um die Autocompletions zu erheben, die die Siri Suche liefert, gibt es, wie zuvor beschrieben, zwei Setups. Zum einen von einem privat genutzten iPad, mit einer verknüpften Apple ID und zum anderen in einem Simulator in Xcode (Abbildung 8).

Jeder Suchbegriff wird sowohl am iPad als auch in der Simulation des iPads innerhalb manuell eingegeben und die Autocompletions äquivalent zu den automatisch herangezogenen Autocompletions in einer CSV-Datei abgelegt.

Dabei unterscheidet sich das Vorgehen an iPad und im Simulator dadurch, dass die Simulation am MacBook täglich neu gestartet wird und bei gestartetem Simulator über den Menüpunkt „Hardware“ > „Erase all Content and Settings“ auf „Werkseinstellung“ zurückgesetzt wird. Dies ermöglicht eine Umgebung, in der, so die Annahme, Vorschläge geliefert werden, die nicht von bisher gesammelten Erkenntnissen über Suchpräferenzen eines Nutzers oder geografische Daten beeinflusst werden. Zumindest, was lokale Einflüsse betrifft. Denn um in der Simulation deutschsprachige und für den deutschsprachigen Raum relevante Ergebnisse zu erhalten, werden die Sprache und die Region des simulierten Geräts auf Deutsch bzw. Deutschland eingestellt.

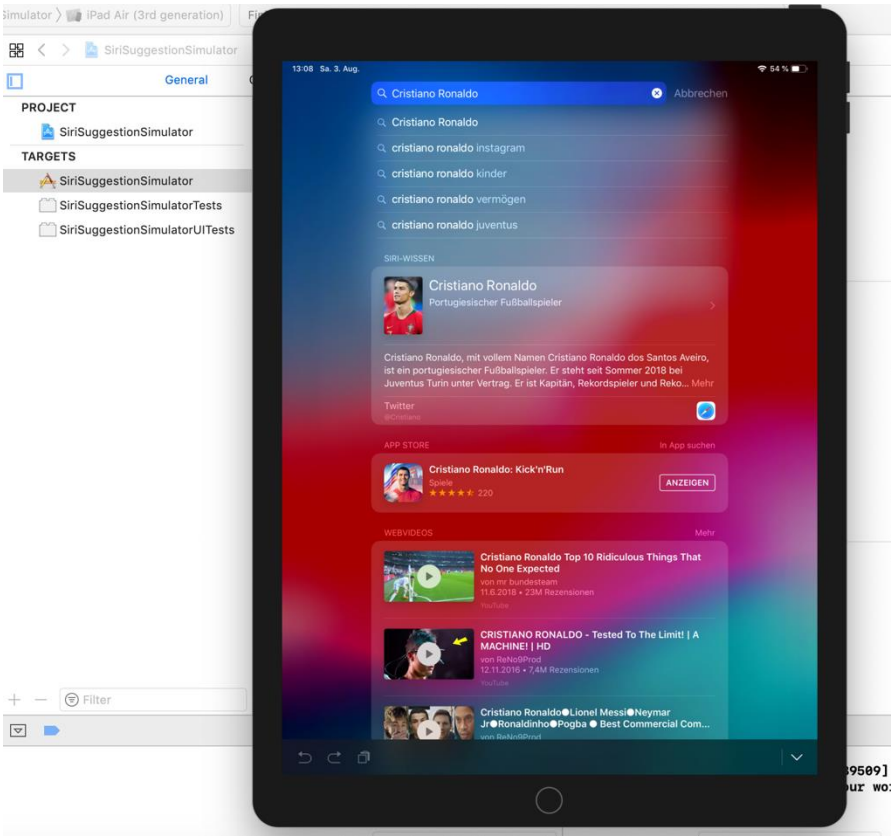


Abbildung 8: Screenshot mit Beispieldarstellung beim Verwenden des Xcode-Simulators zur Datenerhebung

Ein erster Unterschied zwischen den Ergebnislisten bei der Siri-Suche und den weiteren Vergleichssuchmaschinen ist, dass sowohl bei Siri_Simulator als auch beim Siri_Tab in der Regel maximal vier Ergebnisse zu einem Suchbegriff geliefert werden. Die Anzahl der Vorschläge bei den genannten Websuchmaschinen im Vergleich liegt dagegen bei etwa zehn Vorschlägen pro Anfrage.

5.2 Datenbereinigung als Vorbereitung für die Clusterbildung

Dieses Kapitel erklärt das Vorgehen während der Datenbereinigung, die als vorgelagerter Schritt zur Anwendung der beschriebenen Analysemethoden notwendig ist.

In der Phase der Datenerfassung wurde die Daten bis auf die oben beschriebene Bereinigung der Ergebnislisten in den CSV-Dateien in Bezug auf „undefined“-Fälle und falsch codierte Sonderzeichen, ohne weitere Vorverarbeitung übernommen.

Um bei der Clusterbildung Redundanzen in den Kategorien zu vermeiden, müssen allerdings noch Bereinigungsschritte vorgenommen werden. Zu einigen Querytermen tauchen Autocompletions auf, die den Term zunächst in anderer Schreibweise wiederholen und ein Zusatz dazu. Dies ist besonders bei den Begriffen „Beyoncé“ und „Kim Kardashian West“ auffällig.

Für das Clustering werden diese vorangestellten Begriffe, die in den allermeisten Fällen den Suchterm in einer abgeänderten Schreibweise darstellen, entfernt. Somit gehen diese Terme meist in anderen, als Einzelbegriff vorkommenden, Vorschlägen auf. Mit diesem Vorgehen wird zumindest größtenteils verhindert, dass es in den Clustern zu Redundanzen kommt. Dies trifft auf die meisten Clusterkategorien zu. Der Kategorie „Profession“ sind beispielsweise auch Songtitel oder Filmtitel von Künstlern in unterschiedlichen Schreibweisen oder mit unterschiedlichen Zusätzen zugeordnet, wie „lyrics“ oder „übersetzung“. Diese Zusätze werden während der Bereinigung, anders als die vorgelagerten Teile der Suggestions, nicht aufgelöst. Eine Liste der bereinigten Terme findet sich in Anhang 2.

5.3 Durchführung Schnittmengen und Frequenzanalyse

Diese Auswertung funktioniert mittels einer Pivot Tabelle in Microsoft Office Excel. Das bietet den Vorteil, dass die erstellten CSV-Dateien unkompliziert in Excel importiert werden können und dort als Datenbasis für Verarbeitung in einer Pivot Tabelle verwendet werden können.

Die Basis zur Durchführung dieser Analyse bildet der, wie im vorherigen Kapitel 5.2 beschrieben, bereinigte Datensatz mit der gesamten Ergebnisliste. Um für diese großen Datenbasis Ergebnisse darstellen zu können, wird sich die Möglichkeit der individuellen Filterkonfiguration einer Pivot-Tabelle zu Nutze gemacht.

Zur Bestimmung der Anzahl der Unique Terms, in der Gesamtheit und heruntergebrochen auf einzelne Suchmaschinen, wurde Pivot-Tabelle mit folgender Konfiguration umgesetzt:

- Filtern über Suchbegriffe, um eine Auswahl von Suchbegriffen gegebenenfalls einzugrenzen
- in den Tabellenzeilen werden die Vorschläge gelistet

- in den Spalten der Tabelle sind alle Plattformen, also Suchmaschinen, des Vergleichs dargestellt
- für das Attribut „Werte“ wird Anzahl von Vorschlag gewählt

Mit dieser Konfiguration ist die Möglichkeit geschaffen, eine übersichtliche Darstellung über die Anzahl des Vorkommens der Vorschlagsterme zu erhalten.

Zudem dient das Zählen der in der Auflistung enthaltenen Vorschläge zur Ermittlung der sogenannten Unique Terms.

Darüber hinaus können auf Basis der Unique Terms zwei oder mehr Suchmaschinen gegenüberstellt werden, um gemeinsame Schnittmenge zu ermitteln.

5.4 Durchführung RBO

Die Umsetzung und Durchführung der in Kapitel 4.4 beschriebenen Methode Rank-biased Overlap erfolgt in R. Als Grundlage des hierfür verwendeten Skripts dient Code von von Malte Bonart zur Umsetzung von RBO in R.[41]

Die dafür benötigten Daten werden zunächst aus einer CSV eingelesen, in der sich die gesammelten Ergebnisse zu allen Suchmaschinen sowie Suchbegriffen über alle Erhebungstage befinden. Diese Ergebnis-CSV enthielt zunächst einen Zeitstempel mit Datum und Uhrzeit im ISO-Format. Zugunsten einer erleichterten Weiterverarbeitung wurde dieser unter Anwendung einer Regular Expression auf das Datum im Format JJJJ-MM-DD eingekürzt. Das Skript ist nun so aufgebaut, dass es über alle 20 Suchbegriffe iteriert. Zudem erfolgt dies außerdem für jeden Erhebungstag. Bedingt durch die im Skript eingesetzte Schleife wird so für jeden Tag einzeln ein RBO errechnet.

Dies geschieht, indem aus den zuvor eingelesenen Daten je zwei Rankings erstellt werden, die mittels der RBO-Berechnung miteinander verglichen werden. Die Erstellung dieser Rankings aus den zur Verfügung stehenden Gesamtdaten erfolgt über einen Filter, in dem konfiguriert wird, zu welchen Plattformen die Vorschläge selektiert werden sollen. Mittels dieses Filters werden dann die entsprechende Spalte mit Vorschlägen extrahiert. Mit den in diesen Spalten enthaltenen Vorschlägen werden dann die zwei Rankings gebildet, auf denen die Berechnung des RBO stattfindet.

Für diese Berechnung wird zudem ein Parameter p bestimmt. Für den Vergleich im Rahmen dieser Arbeit werden die Berechnungen des RBO mit zwei verschiedenen p-Werten durchgeführt. Zunächst erfolgt sie mit einem auf 1.0 festgesetzten Parameter p. Diese Konfiguration ermöglicht einen Vergleich der beiden in die Berechnung eingehenden Listen, bei dem auf das reine Vorkommen von Vorschlägen in den Listen geprüft wird. Dabei spielt die Rangfolge der Vorschläge keine Rolle.

[41] Bonart, M. (2019): rbo in r.

Des Weiteren wird p=0.75 gewählt. Gemäß der in Kapitel 4.2 vorgestellten Formel ergibt sich die Festsetzung des p dadurch, dass die Evaluationstiefe (d) auf 4 festgelegt werden soll. Somit entfallen 86% des Gewichts auf die ersten Vorschläge eines Rankings. Diese Zahl Vier entspricht der Anzahl der Vorschläge, die in der Regel bei der Suche mit Siri am Tablet und äquivalent auch in der Simulation zurückgegeben werden.

Die errechneten RBO-Scores für jeden Tag werden in eine CSV-Datei geschrieben. Für jede beim RBO betrachtete Suchmaschinen bzw. p-Wert-Kombination existiert danach eine solche Datei. Bei sieben Vergleichspaarungen in Bezug auf die verwendeten Suchmaschinen und unter Betrachtung der zwei erläuterten Parameter p, werden insgesamt 14 CSV-Dateien erstellt.

Um die berechneten RBO in ihrem Verlauf über den 29-tägigen Erhebungszeitraum zu visualisieren, werden mit Hilfe eines weiteren R-Skripts Plots erstellt. Dazu werden zuvor erzeugten CSV-Dateien eingelesen und diese zeilenweise (über jeden Suchbegriff) durchlaufen. Für jeden Suchbegriff wird damit eine graphische Darstellung generiert. In dieser werden die Erhebungstage auf der X-Achse des Koordinatensystems dargestellt. Die Höhe des jeweiligen RBO-Scores erfolgt auf der Y-Achse. Somit wird die Darstellung von eventuell gegebenen Veränderungen im Zeitverlauf erreicht.

Das vollständige Skript zur Berechnung des RBO, sowie zur Erstellung der Plots ist der CD zur Bachelorarbeit beigefügt.

5.5 Durchführung Clustering und Score Berechnung

Das Clustering, das im Rahmen dieser Arbeit angewendet wird, erfolgt durch eine intellektuelle Zuordnung der identifizierten Unique Terms aus den Autocompletions in Kategorien. Die Entscheidung zu Anzahl und Benennung der Kategorien passiert ebenfalls intellektuell auf Basis von Beobachtung, welche Themen sich oft in der Liste der Unique Terms finden und sich als Oberbegriff zur Kategorie Benennung eignen.

Die Einteilung erfolgt in die in Tabelle 3 aufgelisteten Kategorien. Dabei werden den Kategorien 1, 2, 3, 4, 5 und 6 solche Suggestterms zugeordnet, denen als Query Term eine natürliche Person zu Grunde liegt. Denen Kategorien 9 und 10 werden Terme zugeordnet, die auf eine Firma bzw. eine Marke als Suchbegriff zurückzuführen sind. Die Kategorien „Orte/Sprache“ (7) und „Sonstiges“ (8) teilen sich beide Gruppen.

Kategorie Name	Kategorie Nummer
Social Media/Information	1
Körpermerkmale	2
Profession	3
Statussymbole	4
Beziehung/Familie	5
Nacktheit/Tod	6
Orte/Sprache	7
Sonstiges	8
Produkte	9
Services	10

Tabelle 3: Kategorie Name und zugehöriger Nummer im erstellten Cluster

In der Kategorie „Social Media/Information“ werden Autocompletions zu den gängigen Social Media Plattformen, wie „instagram“, „twitter“, „snapchat“ und „facebook“ eingeordnet. Darüber hinaus aber auch Vorschläge, die Informationsbedürfnis über die jeweilige Person thematisieren, wie beispielsweise „wikipedia“, „news“, „bilder“, „website“ oder das Forum „superiorpics“.

Die Kategorie Körpermerkmale wird mit Begriffen, wie „tattoo“, „ungeschminkt“, „lippen“ oder „fat“ befüllt. Auch Vorschläge, wie „diät“ oder „op“ fließen hier mit ein.

In der dritten Kategorie „Profession“ werden solche Vorschläge abgebildet, die im engeren und im weiteren Sinne Informationen oder Fragestellungen zum Beruf der Personen abbilden. Dies können allgemeiner gehaltene Begriffe, wie „trikot“ zum Suchbegriff Neymar Jr. (Fußballer von Beruf) sein. Andererseits werden hier auch spezifischer Suggestterms eingeordnet. So zum Beispiel ein Songtitel „wrecking ball“ von Sängerin Miley Cyrus. Diese Kategorie beinhaltet mit einer Anzahl von 120 die meisten zugeordneten Vorschlägen. Von diesen 120 Vorschlägen fallen ein Großteil auf Songtitel, Chords und Lycrics.

Des Weiteren sind in Kategorie 4 „Statussymbole“ solche Vorschläge zugeordnet, die Status oder Reichtum thematisieren. Darunter fallen unter anderem „vermögen“, „gehalt“ oder „bugatti“ als bekanntermaßen edle Automarke.

Begriffe, wie „freundin“, „tochter“ oder „baby“ sind Kategorie „Beziehung/Familie“ zugeordnet. Neben den neutral gehaltenen Termen, fallen aber auch Namen in diese Kategorie, die nur durch einen intellektuellen Rückschluss auf eine aktuelle oder vergangene Liebsbeziehung hinweisen. So zum Beispiel der Vorschlag „justin bieber“ zum Suchterm „selena gomez“, die in der Vergangenheit ein Paar gewesen sind.

In die Kategorie „Nacktheit/Tod“ fallen solche Vorschläge, die intime Körperstellen thematisieren, wie zum Beispiel „po“, aber auch Vorschläge, die Tod oder Krankheit zum Thema habe. Also Begriffe, bei denen eine latente Sensationslust mitschwingt.

Die Kategorie „Orte/Sprache“ werden von Suggestions belegt, die sowohl ihren Ursprung bei natürlichen Personen als auch bei den Marken bzw. Firmen innerhalb der Anfrageterme haben. In diese Kategorien fallen Vorschläge, die den Suchterm um eine geografische Angabe, wie einen Ort oder aber eine Sprache ergänzen.

In Kategorie „Sonstige“ befinden sich ebenfalls Suggestions, die aus Anfragen an beide Gruppen, natürliche Personen und Marken bzw. Firmen hervorgegangen sind. In dieser Kategorie befinden sich diese Vorschläge, die nicht eindeutig einer der festgelegten Kategorien zugeordnet werden können.

Ausschließlich Vorschläge zu den Marken und Firmen aus den Suchbegriffen gehen in die Kategorie „Produkte“ (9) und „Services“ (10) ein. Dabei findet diese Unterteilung statt, da in Kategorie 9 Vorschläge, wie „air max“ als Vorschlag zu einer Suche mit dem Queryterm „Nike“, eingehen. Wobei es sich klar um ein Modell, also Produkt der Schuhmarke „Nike“ handelt. In Abgrenzung dazu fallen in die Kategorie 10 Vorschläge, wie „store“, was nicht unmittelbar ein Produkt der Marke „Nike“ repräsentiert, aber einen Kontaktpunkt für einen Kunden. Dieser Kategorie sind ebenso Vorschläge, wie „account löschen“ als Vervollständigung einer Suche nach Instagram, zu geordnet.

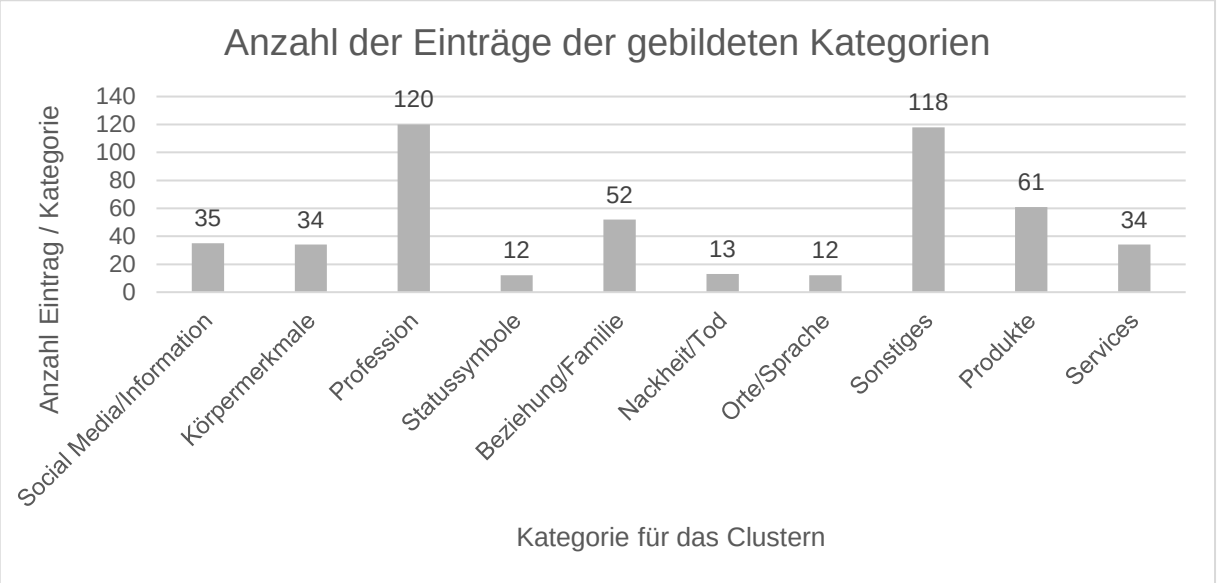


Abbildung 9: Anzahl von Unique Terms aufgeteilt nach Kategorien der Clusterbildung

Die Verteilung von Vorschlägen in die Kategorien ist in Abbildung 9 dargestellt. Hier zeigt sich, dass die die Anzahl von Begriffen in den Kategorien sehr unterschiedlich verteilt ist. Die zahlenmäßig größte Gruppe ist Kategorie „Profession“ mit 120 Einträgen, davon sind etwa 70 Einträge Song oder Albentitel von den unter den Suchbegriffen vorkommenden Sängerinnen.

Die zweitgrößte Kategorie stellt „Sonstige“ dar. Allerdings sind hier die Einträge sowohl aus Suchanfragen aus der Gruppe der natürlichen Personen als auch aus Anfragen aus der Gruppe der Firmen bzw. Marken generiert.

Die Kategorien „Produkte“ und „Services“ sind mit 61 bzw. 34 Zuordnungen im Vergleich über alle Kategorien hinweg eine hohe Anzahl an Vorschlägen zugeordnet. Dies liegt wiederum darin begründet, dass diese beiden Kategorien mit Vorschlägen zu Marken und Firmen belegt sind. Für diese Gruppe der Query Terms ist die Anzahl der möglichen vier Kategorien, auf die die Verteilung stattfindet, kleiner. Dem stehen acht Kategorien für Vorschläge zu der Gruppe der Personen gegenüber. Aus diesem Grund spaltet sich die Verteilung dieser deutlicher in kleinere Gruppen auf.

In den übrigen Kategorien variiert die Anzahl der eingeordneten Suggestions zwischen zwölf bei „Statussymbole“ bis 52 in der Kategorie „Beziehung/Familie“.

Über alle Kategorien ist die Anzahl der zugeordneten Suggestterms in der Summe 491. Das sind 7 mehr, als in der zugrundeliegenden Liste der Unique Terms mit einer Anzahl von 484 Einträgen. Das ist auf die Vorgehensweise bei der Bildung der Cluster zurückzuführen. Denn dabei werden in der Liste der Unique Terms zunächst nur die 16 Suchbegriffe der Gruppe der natürlichen Personen gefiltert und auf die genannten acht Kategorien verteilt. In einem zweiten Schritt wird dann nur auf die 4 Marken und Firmen innerhalb der Suchbegriffe reduziert und auf die vier hierfür festgelegten Kategorien verteilt. Hierbei entstehen Dopplungen, bei den folgenden Termen:

- bilder → Vorkommen in Kategorie „Social Media/Information“ und „Services“
- filme → Vorkommen in Kategorie „Profession“ und „Produkte“
- schuhe → Vorkommen in Kategorie „Statussymbole“ und „Produkte“
- auto → Vorkommen in Kategorie „Statussymbole“ und „Produkte“
- haus → Vorkommen in Kategorie „Statussymbole“ und „Produkte“
- spiele → Vorkommen in Kategorie „Produkte“ und „Services“
- shop → Vorkommen in Kategorie „Sonstige“ und „Services“

Das gesamte Clustering, mit allen Kategorien und allen Eintragungen ist in Anhang 3 einsehbar.

Das Cluster ist innerhalb von Microsoft Excel erstellt und hierfür dient, wie erläutert, die ebenfalls dort erarbeitete Pivot-Tabelle mit den Unique Terms als Grundlage. Deshalb findet in Excel ein Abgleich statt, mit welchem auf das Vorkommen eines Begriffs in einem Cluster geprüft wird. Diese Prüfung erfolgt durch Anwendung einer Formel, die im ersten Schritt das Cluster mit Vorschlägen zu der Gruppe der natürlichen Personen vergleicht. In einem zweiten Schritt wird die Formel auf das Cluster für die Firmen und Marken innerhalb der Suchbegriffe angepasst und ebenfalls geprüft. Durch diesen Abgleich wird in der CSV-Datei (Abbildung 10) mit den gesamten Erhebungsergebnissen zu den Dauerthemen eine Spalte ergänzt, in der mit Hilfe von zwei verketteten Excel-Funktionen „SVerweis“, zuerst die Zuordnungen zu den Kategorien von 1 bis 8 für natürliche Personen und danach für Suchbegriff zu Marken und Firmen die Kategorien 8 bis 10 abgebildet werden. Die Funktion „SVerweis“ wird verwendet, um tabellenübergreifend zu prüfen, in welchem Cluster der jeweilige Vorschlag eines Suchbegriffs auftaucht.

```
Plattform;Suchbegriff;Datum;Vorschlag;Position;Kategorie
Bing;Instagram;2019-05-26;login;1;10
Bing;Instagram;2019-05-26;suche;2;10
Bing;Instagram;2019-05-26;heidi klum;3;8
Bing;Instagram;2019-05-26;anmelden;4;10
Bing;Instagram;2019-05-26;account löschen;5;9
Bing;Instagram;2019-05-26;dolunay;6;8
Bing;Instagram;2019-05-26;download;7;10
Bing;Instagram;2019-05-26;friesennerz;8;8
Bing;Instagram;2019-05-26;daniela katzenberger;9;8
Bing;Instagram;2019-05-26;lena meyer landrut;10;8
Bing;Instagram;2019-05-26;bibisbeautypalace;11;8
Bing;Cristiano Ronaldo;2019-05-26;instagram;1;1
Bing;Cristiano Ronaldo;2019-05-26;steckbrief;2;1
Bing;Cristiano Ronaldo;2019-05-26;kinder;3;5
Bing;Cristiano Ronaldo;2019-05-26;wikipedia;4;1
Bing;Cristiano Ronaldo;2019-05-26;autos;5;4
Bing;Cristiano Ronaldo;2019-05-26;vermögen;6;4
Bing;Cristiano Ronaldo;2019-05-26;bugatti;7;4
Bing;Cristiano Ronaldo;2019-05-26;rekorde;8;8
Bing;Cristiano Ronaldo;2019-05-26;transfermarkt;9;3
Bing;Cristiano Ronaldo;2019-05-26;familie;10;5
Bing;Cristiano Ronaldo;2019-05-26;stream;11;8
```

Abbildung 10: Ausschnitt der um die Kategorie ergänzte CSV, die die Gesamtheit aller Erhebungsergbenisse darstellt

Mit der vorgestellten Einteilung in Cluster und der Abbildung dieser in einer CSV Datei ist eine Basis geschaffen, um in einem nächsten Schritt für das Vorkommen von Kategorien einen Score zu berechnen. Dieser Score ist, wie in Kapitel 4.4 bereits erläutert, angelehnt an das Maß der Precision.

Um die Berechnung durchzuführen wird diese Methode in ein R-Skript überführt, worin folgendermaßen vorgegangen wird, um einen durchschnittlichen Score für alle Kategorien über den Erhebungszeitraum zu erhalten. Die Summe aller Scores über die Kategorien hinweg bezogen auf einen Suchbegriff ist dabei insgesamt immer 1.

Die zuvor beschriebene CSV-Datei mit der ergänzten Spalte der Kategorienzuweisung wird in R eingelesen.

Besagtes Skript iteriert dann über die in einem Array abgelegten Benennungen der Suchmaschinen und die sich ebenfalls in einem Array befindenden Suchbegriffe. Genauso wiederholt es die Schleife über alle Kategorien und für jedes Datum im Erhebungszeitraum. Somit kann der Summe aller ermittelten (Tages-)scores zu jeder Kategorie ein Durchschnittsscore errechnet werden kann.

Über die Filterkriterien „Plattform“, „Suchbegriff“ und „Datum“, wie in Abbildung 11 gezeigt, werden die zugeordneten Kategorien aus der entsprechenden Spalte ausgelesen und in einen Vektor „kategorien2tag“ geschrieben.

```
# Rückgabe der Kategorien für einen Tag, Suchmaschine, Suchbegriff
kategorien2tag <- select(filter(dat, Plattform == suchmaschine, Suchbegriff == suchbegriff, Datum == datum), c(Kategorie))
```

Abbildung 11: Filterfunktion für einen Zugriff auf zugeordnete Clusterkategorien nach Suchmaschine, Suchbegriff und Datum

Bei einer beispielhaften Betrachtung der Plattform „Google“ und dem „Suchbegriff „Cristiano Ronaldo“ am Datum 2019-05-26 gestaltet sich das Array mit den ausgelesenen Kategorien, wie in Abbildung 12 dargestellt, von Rang 1 bis 9.

Kategorie
5
5
4
2
5
3
1
4
1

Abbildung 12: Beispielarray „kategorien2tag“ mit der Suchmaschine Google, dem Suchbegriff Cristiano Ronaldo am 2019-05-26

Für eine beispielhafte Score Berechnung zu Kategorie 5, wird nun über den zuvor gefilterten Vektor iteriert. Die Prüfung, ob die gerade betrachtete Kategorie zutrifft, erfolgt über eine if-Bedingung. Wenn diese Bedingung erfüllt ist, wird eine Variable, die zu Beginn als 0 definiert ist und als ein Laufzähler für die Bedingung fungiert, um 1 erhöht.

```
# katScore repräsentiert einen Laufzähler zur Ermittlung ob betrachtete Kategorie zutrifft
katScore = 0
for (rang in 1:nrow(kategorien2tag)) {
    # Falls Kategorie zutrifft, wird der Laufzähler um 1 erhöht
    if (kategorien2tag[rang, 1] == katNr) {
        katScore = katScore + 1
    }
}
```

Abbildung 13: Prüfung zum Erhöhen des Laufzählers bei einer erfüllten if-Bedingung, wenn Kategorie aus kategorie2tag mit aktueller betrachteter Kategorie übereinstimmt

Das hier ermittelte Ergebnis Anzahl 3 (dreimal Vorkommen der Kategorie 5), wird durch die Zahl der im Vektor „Kategorie“ enthaltenen Elemente, also der Anzahl der zurückgelieferten Vorschläge, geteilt (Abbildung 12). Damit wird ein Score zu der betrachteten Kategorie für einen Tag der Erhebung errechnet. Somit lautet das Ergebnis zu diesem für einen Tagesscore zu Kategorie 5 *3/9=0,33*. Dieser Rechenschritt wird, bedingt durch das Iterieren über eine Schleife, für alle Erhebungstage wiederholt und diese Scores summiert. Zuletzt wird die erhaltene Summe der Scores nochmals durch 29, also die Gesamtzahl der Erhebungstage, dividiert. Das Ergebnis dieser Berechnungen ist dann ein Durchschnittsscore für eine Kategorie zu einer Suchplattform und zu einem Suchbegriff.

Mit Hilfe des Skripts werden so für fünf Suchmaschinen, 20 Suchbegriffe und 10 Kategorien insgesamt 1000 Scores berechnet, die für einen Vergleich dienen können.

```
Plattform;Suchbegriff;Kat1;Kat2;Kat3;Kat4;Kat5;Kat6;Kat7;Kat8;Kat9;Kat10
siri_simulator;Instagram;0,000;0,000;0,000;0,000;0,000;0,000;0,000;0,000;0,009;0,250;0,741
siri_simulator;Cristiano Ronaldo;0,250;0,000;0,250;0,250;0,250;0,000;0,000;0,000;0,000;0,000;0,000
siri_simulator;Ariana Grande;0,241;0,250;0,362;0,000;0,138;0,000;0,000;0,009;0,000;0,000
siri_simulator;Selena Gomez;0,250;0,000;0,181;0,000;0,319;0,250;0,000;0,000;0,000;0,000
siri_simulator;The Rock;0,000;0,491;0,000;0,000;0,000;0,000;0,000;0,509;0,000;0,000
siri_simulator;Kim Kardashian West;0,500;0,000;0,250;0,000;0,000;0,000;0,000;0,000;0,250;0,000;0,000
siri_simulator;Kylie Jenner;0,250;0,250;0,000;0,250;0,250;0,000;0,000;0,000;0,000;0,000
siri_simulator;Beyoncé;0,009;0,000;0,336;0,000;0,000;0,000;0,000;0,009;0,647;0,000;0,000
siri_simulator;Taylor Swift;0,250;0,216;0,534;0,000;0,000;0,000;0,000;0,000;0,000;0,000;0,000
siri_simulator;Leo Messi;0,397;0,250;0,000;0,000;0,250;0,000;0,000;0,103;0,000;0,000
siri_simulator;Neymar jr;0,250;0,250;0,250;0,000;0,250;0,000;0,000;0,000;0,000;0,000;0,000
siri_simulator;Kendall Jenner;0,250;0,250;0,000;0,000;0,474;0,000;0,000;0,026;0,000;0,000
siri_simulator;Justin Bieber;0,250;0,000;0,259;0,000;0,250;0,241;0,000;0,000;0,000;0,000
```

Abbildung 14: Ergebnis CSV Score Berechnung

In Abbildung 14 ist ein Ausschnitt der CSV-Datei mit den Ergebnissen der Score-Berechnung dargestellt, aus der zur Weiterverarbeitung in Excel Anführungszeichen zu Beginn um am Ende jeder Zeile schon entfernt worden sind, die beim Schreiben in die CSV-Datei generiert worden sind. Zudem ist für die weitere Verarbeitung das standardmäßige Dezimaltrennzeichen in R „.“ durch das Dezimaltrennzeichen „,“ ersetzt worden.

Die Struktur der Datei ist so gewählt, dass ein Zeileneintrag den Suchbegriff mit den ermittelten Score-Werten der Kategorie 1-10 repräsentiert. Die Gruppierung erfolgt dabei nach der Plattform, also der Suchmaschine. Ergebnisse und Diskussion

In diesem Kapitel werden die Ergebnisse beschrieben, die im Zuge dieser Arbeit erzielt worden sind. Zunächst wird auf die Ergebnisse der Schnittmengen- und Frequenzanalyse eingegangen. Nachfolgend dann die Ergebnisse des Vergleichs mit der Methode des Rank-biased Overlap (RBO). Abschließend werden die Ergebnisse der Clusterbildung bzw. die daran angeschlossene Anwendung des Scores.

6 Ergebnisse

6.1 Schnittmengen – und Frequenzanalyse

Die Datenbasis für die Schnittmengen- und Frequenzanalyse enthält 23.427 Einträge, dies entspricht der Anzahl der insgesamt gesammelten Vorschläge zu allen Suchmaschinen im konzipierten Versuchsaufbau über den Erhebungszeitraum von 29 Tagen.

Unter Abwendung der Filterkonfiguration in Kapitel 4.1 beschrieben Pivot-Tabelle ergibt sich aus dieser Gesamtsumme der Vorschläge ein Vorkommen von 484 Unique Terms unter Einbeziehung aller Suchmaschinen.

Tabelle 4 zeigt die Top 10 der vorgeschlagenen Terme über alle Suchmaschinen hinweg. Zum einen ist in dieser Tabelle die gezählte Häufigkeit des jeweiligen Terms aufgeschlüsselt nach Suchmaschine dargestellt. Zum anderen in der letzten Spalte die gesamte Anzahl an Vorkommen. Diese zehn häufigsten Terme werden in der Summe schon 5653 Mal vorgeschlagen. Das macht in Bezug auf die die gesamte Anzahl von Vorschlägen in etwa ein Viertel aus.

Vorschlag	**Bing**	**DuckDuckGo**	**Google**	**Siri_ Simulator**	**Siri_ Tab**	**Yahoo**	**Gesamtergebnis**
instagram	399	338	361	377	416	144	2035
wikipedia	218	319	19			132	688
filme	58	87	109	58	70	36	418
alter			374		28		402
bilder		290	1			108	399
größe		29	121	140	93	12	395
steckbrief	356		8				364
songs		116	133	29	33	24	335
twitter	100	47	73	46	29	36	331
vermögen	58		108	58	62		286

Tabelle 4: Darstellung der Top 10 Suchvorschläge mit ihrer Gesamtanzahl und Aufsummierung über die verwendeten Suchmaschinen

Tabelle 5 weist die Anzahl der Unique Terms aufgeteilt nach Suchmaschinen aus. Durch diese Einzelbetrachtung ergibt sich eine Summe, die höher ist als die vorher genannte Gesamtsumme der Unique Terms.

Suchmaschine	Unique Terms
DuckDuckGo	118
Google	237
Bing	226
Siri_Simulator	85
Siri_Tab	90

Tabelle 5: Darstellung der jeweiligen Unique pro Suchmaschine

In dieser Betrachtung zeigt sich, dass die Anzahl der Unique Terms bei den Siri_Simulator und Siri_Tab mit 85 bzw. 90 verzeichneten Unique Terms fast identisch ist. Google und Bing weisen im Vergleich zu allen weiteren Suchmaschinen in der Darstellung eine weit höhere Anzahl an Unique Terms auf. Dies ist wahrscheinlich darauf zurückzuführen, dass es bei diesen beiden Plattformen mehr unterschiedliche Vorschläge zu einem Suchbegriff gegeben hat.

	DuckDuckGo	Google	Bing	Siri_Simulator	Siri_Tab
DuckDuckGo	-	62	40	32	31
Google	62	-	74	49	56
Bing	40	74	-	33	34
Siri_Simulator	32	49	33	-	69
Siri_Tab	31	56	34	69	-

Tabelle 6: Absolute Anzahl von gemeinsamen Unique Terms jeweils in der Betrachtung zweier Suchmaschinen

Eine Gegenüberstellung zweier Suchmaschinen ist in Tabelle 6 dargestellt. Diese dient dazu, die Schnittmenge an Unique Terms aufzuzeigen. Dieser Darstellung kann beispielsweise entnommen werden, dass die gemeinsame Anzahl von Unique Terms zwischen Siri_Simulator und Siri_Tablet 69 beträgt. In Bezug auf die reine Anzahl liegt die Überschneidung zwischen Google und Bing mit 74 nochmals höher. Eine geringe Anzahl an gemeinsamen Unique Terms weist DuckDuckGo zu beiden Siri-Suchen mit 31 bzw. 32 auf.

Tabelle 7 bildet eine Aufstellung der relativen Häufigkeiten von Überschneidungen der Unique Terms zwischen zwei Suchmaschinen ab.

Diese Tabelle ist wie folgt zu lesen: 81,18% der Unique Terms bei Siri_Simulator finden sich auch unter den Unique Terms von Siri_Tab, umgekehrt sind 76,67% der Unique Terms von Siri_Tab auch unter den Unique Terms bei Siri_Simulator zu finden. Dieser Unterschied liegt in der unterschiedlichen absoluten Anzahl der Unique Terms der jeweiligen Suchmaschinen (Tabelle 6) begründet.

Die relativen Häufigkeiten beider Siri-Suchen im Vergleich zu DuckDuckGo und Bing liegen bei 34,44% und 38,82%. Im deutlichen Gegensatz dazu sind 57,65% bzw. 62,22% der Unique Terms der Siri-Suche auch unter den Unique Terms bei Google zu finden.

	DuckDuckGo	Google	Bing	Siri_Simulator	Siri_Tab
DuckDuckGo	-	26,16%	17,70%	37,65%	34,44%
Google	52,54%	-	32,74%	57,65%	62,22%
Bing	33,90%	31,22%	-	38,82%	37,78%
Siri_Simulator	27,12%	20,68%	38,82%	-	76,67%
Siri_Tab	26,27%	23,63%	37,78%	81,18%	-

Tabelle 7: Prozentuale Darstellung gemeinsamer auftretenden Unique Terms im Vergleich von jeweils zwei Suchmaschinen

Das zeigt also abschließend betrachtet, dass es deutliche Gemeinsamkeiten zwischen den Vorschlägen der Suche in Siri_Simulator und Siri_Tab gibt. Zudem ist die Schnittmenge der Unique Terms von Google und beiden Suchumgebungen von Siri prozentual am höchsten.

6.2 Ergebnisse Rank-biased Overlap (RBO)

Nach erfolgt eine Beschreibung der Ergebnisse der berechneten RBO-Scores zu den 20 in Kapitel 3.3 vorgestellten Suchbegriffen.

Die Anwendung des RBO erfolgte für die zwei gewählten p-Werte 1.0 und 0.75 in den nachstehend aufgelisteten sieben Vergleichen:

- Siri_Tab / Siri_Simulator
- Siri_Tab / Google
- Siri_Tab / Bing
- Siri_Tab / DuckDuckGo
- Siri_Simulator / Google
- Siri_Simulator / Bing
- Siri_Simulator / DuckDuckGo

Diese Kombinationen von zwei verschiedenen festgelegten p-Werten, sieben Kombinationen aus den im Versuch betrachteten Suchmaschinen und den zugrunde liegenden 20 Suchbegriffen, bringt dabei insgesamt 280 Ergebnis-Plots hervor. Die in diesem Kapitel vorgestellten Ergebnisse werden aus Gründen der besseren Lesbarkeit auf zwei Nachkommastellen gerundet.

Hierbei zeigt sich das die Verläufe der RBO-Werte zu den verschiedenen Suchbegriffen über den Erhebungszeitraum sich ganz unterschiedlich verhalten. Diese Beobachtung gilt über alle sieben genannten Vergleiche hinweg.

Um diese Beobachtung zu spezifizieren, werden im Folgenden einige Beispiele aus der großen Gesamtheit an Daten herausgegriffen. Anhand dieser werden Gemeinsamkeiten und auffällige Unterschiede verdeutlicht. Dabei werden einzelne grafische Darstellungen

aus den Ergebnissen herausgegriffen. Die vollständige Sammlung der erstellten Plots der RBOs über den Zeitverlauf finden sich im Anhang 1.

6.2.1 Siri Tablet / Siri Simulation

Bei der Betrachtung der RBO-Verläufe von Siri-Tablet zu Siri-Simulation zeigt sich, dass die Kurvenverläufe über alle 20 Suchbegriffe hinweg sehr unterschiedlich ausfallen. Unabhängig von den gewählten p=0.75 und p=1, sind in diesem Vergleich nur zwei Suchbegriffe aus der Gesamtmenge, bei denen über den gesamten Zeitraum der Erhebung der RBO-Wert durchgängig das Maximum von 1 aufweist.

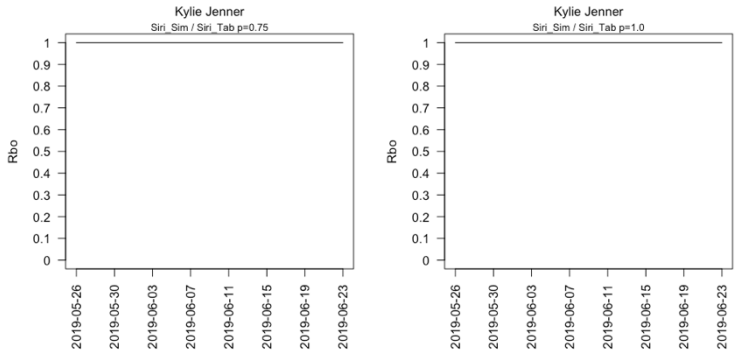


Abbildung 15: Darstellung des Verlaufs des RBO über den Erhebungszeitraum bei unterschiedlicher Gewichtung p=0.75 / p=1

Abbildung 15 zeigt den zeitlichen konstanten Verlauf des RBO bei „Kylie Jenner“. Der zweite Suchbegriff aus der Gesamtmenge, der oben genanntes Verhalten aufweist ist „Cristiano Ronaldo“.

Außerdem zeigt sich beim Blick auf die Kurvenverläufe, dass sich die Kurvenverläufe zwischen p=0.75 und p=1 kaum unterscheiden. Eher ist hier ein glättender Effekt bei einem p-Wert von 1 erkennbar.

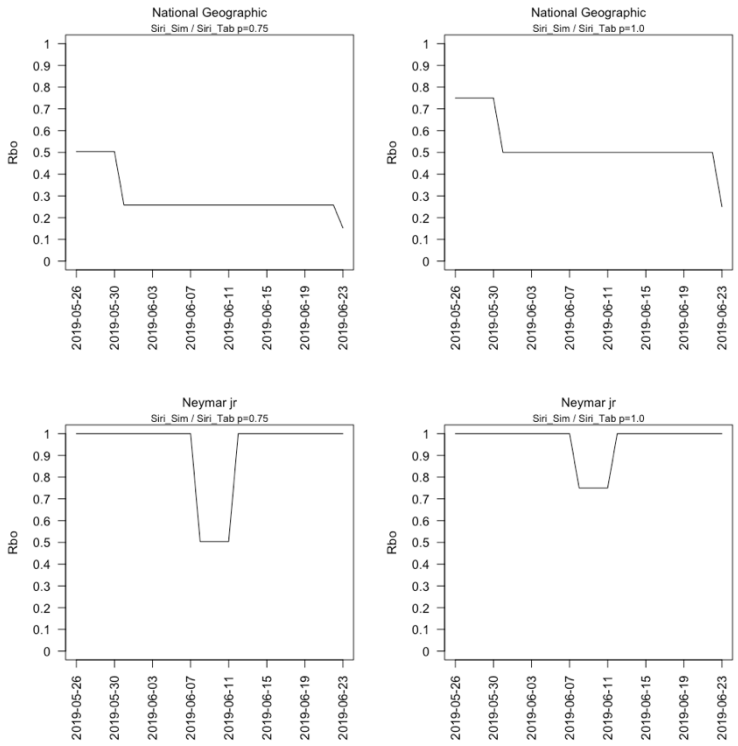


Abbildung 16: Darstellung des Verlaufs des RBO über den Erhebungszeitraum bei unterschiedlicher Gewichtung p=0.75 / p=1 für die Suchbegriffe National Geographic und Neymar jr.

Zudem zeigt sich, dass der RBO im Verlauf entweder höher ansetzt oder aber sich Schwankungen im Verlauf weniger deutlich abzeichnen (Abbildung 16).

Im Vergleich von Siri_Tablet zur Siri_Simulation ist erkennbar, dass es insbesondere dann zu Bewegung in vorher konstante RBO-Werten kommt, als ein Ortswechsel stattgefunden hat und die Suchanfragen nicht von Köln aus gesendet wurden. Dies erfolgte im Zeitraum vom 6. Juni bis 11. Juni. Innerhalb dieser festzustellenden vermehrten Bewegung zu diesem Zeitpunkt, zeigen sich dennoch Unterschiede. Ein vorher bei p=0.75 konstant bei 1.0 liegender RBO zum Suchterm „Neymar jr“ sinkt am Datum 8.6.2019 auf einen RBO von 0.5 ab und verbleibt bis zum 11.6.2019 auf diesem Wert, bis dieser ab dem 12.6.2019 erneut auf den vorherigen Ausgangswert 1.0 zurückgeht. Den gleichen Verlauf nimmt die Kurve bei einem p=1. In diesem Fall sinkt der RBO im gleichen Zeitraum auf 0.75 ab und steigt dann wieder auf den Ausgangswert 1.0 an (siehe Abbildung 16).

Beim Suchbegriff „Jennifer Lopez“ erfolgt im genannten Zeitraum wiederum ein kurzzeitiger Anstieg des RBO. Am 7.6.2019 springt der Wert, bezogen auf p=1.0, von einem Ausgangswert von 0.5 auf 1, um sofort am Folgetag wieder auf den vorherigen RBO zurückzufallen.

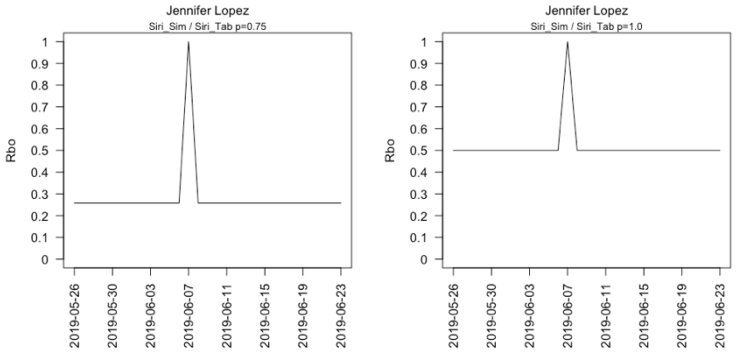


Abbildung 17: Darstellung des Verlaufs des RBO über den Erhebungszeitraum bei unterschiedlicher Gewichtung p=0.75 / p=1 für den Suchbegriff Jennifer Lopez

Eine parallele Veränderung ist auch beim Vergleich mit p=0.75 dokumentiert. Bei diesem p liegt der RBO für „Jennifer Lopez“ bei 0.26 und steigt einmalig sprunghaft auf einen RBO von 1 (Abbildung 17).

Bei wiederum weiteren Queryterms in diesem Vergleich ist eine Bewegung im RBO-Verlauf nicht ausschließlich auf den Zeitraum des genannten Ortswechsels beschränkt. So ergeben sich beispielweise für den Begriff „Beyoncé“ RBO-Werte von maximal 1 und 0.33 im Mininmum, bei Betrachtung von p=1. Ein ganz ähnlicher Verlauf zeigt sich hier auch mit mehrfachen Schwankungen von 0.75 maximal und geringen Werten, wie 0.35 und sogar minimalem RBO-Wert von 0.28 in den letzen drei Tagen der Erhebung (Abbildung 18).

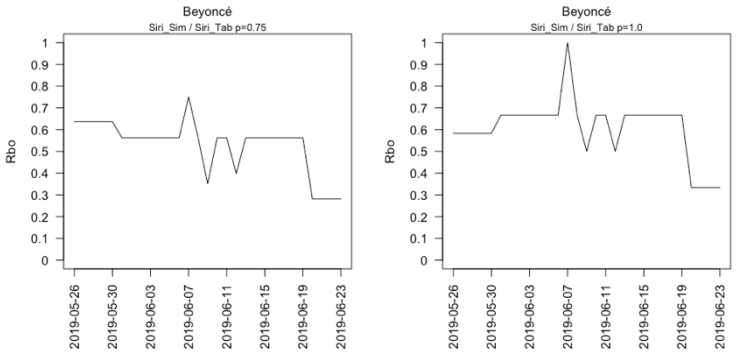


Abbildung 18: Darstellung des Verlaufs des RBO über den Erhebungszeitraum bei unterschiedlicher Gewichtung p=0.75 / p=1 für den Suchbegriff Beyoncé

Zusammenfassend lässt sich für die Betrachtung des RBOs bei einem Vergleich von Siri Suche am Tablet zur Siri Suche in der Simulation, dass die RBOs in aller Regel hoch sind, auch im Zeitverlauf. Dies zeigt sich durch einen durchschnittlichen RBO über alle Suchbegriffe und alle Tage im Erhebungszeitraum von 0.73 (bei p=0.75) und 0.78 (bei p=1). Das deutet auf meist nur kleine Unterschiede in den zurückgelieferten Vorschlägen hin.

6.2.2 Siri / DuckDuckGo

Der Vergleich von Siri Simulation bzw. Siri Suche am Tablet mit DuckDuckGo bestätigt das zuvor gewonnene Bild darüber, dass der Verlauf der RBO-Werte sich über die verschiedenen Suchbegriffe ganz unterschiedlich verhalten.

Dabei zeigen diese allerdings, egal ob nun in Kombination DuckDuckGo/Siri_Tab oder DuckDuckGo/Siri_Simulator sehr ähnliche Muster und auch die RBOs befinden sich bei den meisten der Suchbegriffe auf einem ähnlichen Niveau, wie sich unter anderem durch die in Abbildung 19 dargestellten Plots zeigt.

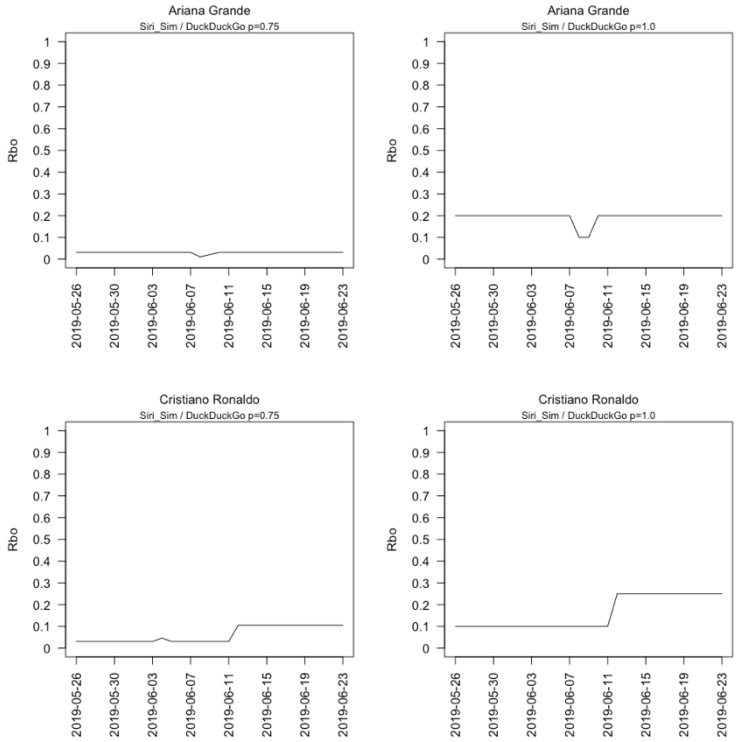


Abbildung 19: Darstellung des Verlaufs des RBO über den Erhebungszeitraum bei unterschiedlicher Gewichtung p=0.75 / p=1 für der Suchbegriffe Ariana Grande und Cristiano Ronaldo

Im Vergleich zu den Siri Suchen mit DuckDuckGo ist außerdem zu erkennen, dass die RBO-Werte fast durchgängig, über die meisten Suchbegriffe betrachtet, niedrig ausfallen. Lediglich in Ausnahmefällen werden Spitzen von mehr als 0.5 erreicht. Sowohl bei dem Vergleich von Siri_Simulation zu DuckDuckGo und Siri_Tablet Suche zu DuckDuckGo mit p=0.75 überschreiten die Suchbegriffe „Kendall Jenner“ mit 0.54 und „Khloe Kardashian“ mit 0.55 (Simulation) bzw. 0.58 (Tablet-Suche) diese Marke knapp. Bei einem gesetzten p=1 im Vergleich Siri_Tab/DuckDuckGo ist der RBO bei den Suchbegriffen „Nike“, „Khloe Kardashian“ und „Neymar jr“ zeitweise über dem genannten Wert. Im Vergleichspaar Siri_Simulator/DuckDuckGo bei einem p=1 liegt sowohl „Neymar Jr“ als auch „Nike“ mit 0.6 fast durchgängig über dieser Grenze. Bei „Nike“ ist hier innerhalb des Zeitraums des thematisierten Ortswechsels ein einmaliger Abfall auf einen RBO von 0.25 verzeichnet, der aber anschließend sofort wieder auf den vorherigen Wert zurückgeht.

Im Vergleichspaar Siri_Tab/DuckDuckGo bleibt der RBO sowohl bei p=0.75 als auch p=1 für den Suchbegriff „Kim Kardashian West“ über den gesamten Erhebungszeitraum bei 0.

```
siri_tab,Kim Kardashian West,2019-06-09,instagram,1
siri_tab,Kim Kardashian West,2019-06-09,twitter,2
siri_tab,Kim Kardashian West,2019-06-09,merch,3
siri_tab,Kim Kardashian West,2019-06-09,beauty,4

DuckDuckGo,Kim Kardashian West,2019-05-26,wiki,1
DuckDuckGo,Kim Kardashian West,2019-05-26,instagram kim kardashian west,2
DuckDuckGo,Kim Kardashian West,2019-05-26,kim kardashian kanye west,3
DuckDuckGo,Kim Kardashian West,2019-05-26,kim kardashian and kanye west,4
DuckDuckGo,Kim Kardashian West,2019-05-26,kim kardashian und kanye west,5
DuckDuckGo,Kim Kardashian West,2019-05-26,kim kardashian north west,6
DuckDuckGo,Kim Kardashian West,2019-05-26,kim kardashian saint west,7
```

Abbildung 20: Erhebungsausschnitt zur Verdeutlichung, dass es keine Überschneidungen der Suchmaschinen zum hier gezeigten Suchbegriff gibt

Abbildung 20 zeigt einen Erhebungsausschnitt für den Suchbegriff „Kim Kardashian West“. Es ist gut zu erkennen, dass es keinerlei Gemeinsamkeiten bei den Suchvorschlägen gibt.

Die Beobachtung von nur wenigen Suchbegriffen mit RBO über 0.5 zeichnet sich auch in den Durchschnittswerten der RBOs ab. Diese liegen bei den Vergleichen von Siri-Tablet mit DuckDuckGo bei 0.2 für ein p=0.75 bzw. 0.28 für p=1. Für den Vergleich der Siri Simulation mit DuckDuckGo liegen die durchschnittlichen Werte des RBO für die p-Werte nur ungleich höher. Nämlich bei 0.21 (p=0.75) und 0.31 (p=1).

6.2.3 Siri / Bing

In diesem Vergleich zeigt sich, dass sich bei über der Hälfte (12 von 20) der betrachteten Suchbegriffe paarweise ähnliche RBO-Werte abzeichnen. Dies bezieht sich auf eine Betrachtung der Kombination Siri-Tab/Bing und Siri-Simulator/Bing, jeweils für die Gewichtung p=0.75 und p=1, wie in Tabelle 8 grün umrandet dargestellt. Diese Beobachtung erfolgt auf Basis des durchschnittlichen RBO-Scores zum jeweiligen Suchbegriff über den Zeitverlauf.

	Bing/Siri_Tab p=0.75	**Bing/Siri_Tab p=1**	**Bing/Siri_Simulator p=0.75**	**Bing/Siri_Simulator p=1**
Instagram	0.05	0.09	0.05	0.09
Cristiano Ronaldo	0.40	0.59	0.40	0.59
Ariana Grande	0.21	0.23	0.24	0.29
Selena Gomez	0.50	0.25	0.50	0.25
The Rock	0.10	0.29	0.09	0.21
Kim Kardashian West	0.00	0.00	0.00	0.00
Kylie Jenner	0.35	0.45	0.35	0.45
Beyoncé	0.24	0.25	0.29	0.39
Taylor Swift	0.50	0.25	0.50	0.25
Leo Messi	0.30	0.28	0.33	0.31

Neymar jr	0.08	0.25	0.08	0.25
Kendall Jenner	0.55	0.42	0.55	0.41
Justin Bieber	0.20	0.33	0.20	0.34
National Geographic	0.29	0.27	0.40	0.50
Barbie	0.15	0.27	0.15	0.27
Khloe Kardashian	0.03	0.08	0.06	0.15
Jennifer Lopez	0.38	0.25	0.15	0.25
Miley Cyrus	0.49	0.33	0.52	0.34
Nike	0.37	0.54	0.37	0.54
Katy Perry	0.05	0.10	0.00	0.00

Tabelle 8: Durchschnitt der RBO-Scores zu ausgewählten Suchbegriffen über den Erhebungszeitraum. Aufgeteilt nach Vergleichspaarungen Bing/Siri_Tab und Bing/Siri_Simulator unter Berücksichtigung von p=0.75 und p=1

Entgegen dieses beschriebenen erkennbaren Musters, treten auch Fälle auf, die sich von dieser Beobachtung allerdings unterscheiden. So zeigt sich bei „Jennifer Lopez“ mit 0.38 ein deutlich höherer durchschnittlicher RBO beim Blick auf Siri_Tab und einem p von 0.75, als in den drei weiteren aufgezeigten Kombinationen.

Diese Auffälligkeit beim Suchbegriff „Jennifer Lopez“ zeigt sich im Übrigen auch in der Verlaufskurve der RBOs über genannte Vergleiche.

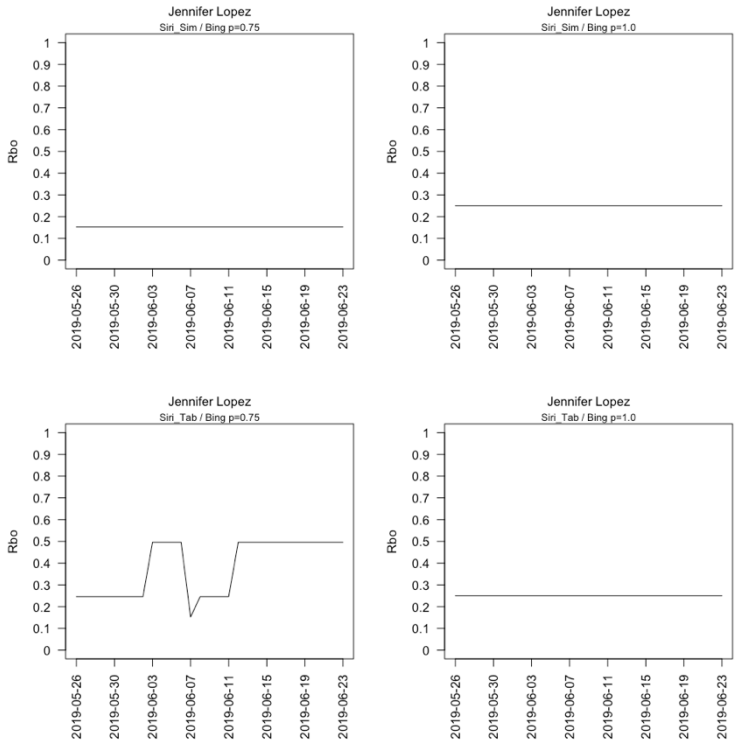


Abbildung 21: Darstellung des Verlaufs des RBO über den Erhebungszeitraum bei unterschiedlicher Gewichtung p=0.75 / p=1 für den Suchbegriff Jennifer Lopez unter der Betrachtung Bing / Siri_Simulator – Siri_Tablet

Während sich bei Bing/Siri_Simulator mit p=1 und gleicher Kombination mit p=0.75, sowie Bing/Siri_Tab mit p=1 jeweils einen geraden Verlauf abzeichnet, so unterliegt der Kurvenverlauf bei p=0.75 vergleichsweise starken Schwankungen (Abbildung 21). Diese sind durch Veränderungen in Inhalt und Reihenfolge der Vorschläge durch Bing begründet.

Dass es bei Bing bzw. den Vorschlägen der Siri Suche teilweise kaum zu Veränderungen innerhalb der gelieferten Query Suggestions über den Erhebungszeitraum kommt, zeigt sich durch mehrfach konstant auftretende auf einem Level verbleibende RBOs. Beim Vergleich von Tablet zu Bing mit p=0.75 weisen sieben Suchbegriffe aus der Gesamtmenge einen solchen Verlauf auf. Bei einem p=1 sind es acht aus 20 Suchbegriffen. Schaut man hier auf die Paarung Simulation zu Bing, sind es sogar mehr als die Hälfte (elf von 20), bei dem p-Wert 0.75 bzw. die Hälfte der in die Betrachtung eingeflossenen Suchbegriffe bei einem gesetzten p-Wert von 1.

Das schon bei der RBO-Betrachtung von DuckDuckGo zu Siri thematisierte Phänomen von einem durchgängigen RBO=0 bei Begriff „Kim Kardashian West“, zeigt sich genauso auch hier. Das auch unabhängig von gewählter Kombination und p.

Ein RBO von 0 ist hier auch beim Queryterm „Katy Perry“ sichtbar. Ausgenommen ist hiervon allerdings der Vergleich mit der Suche am Tablet. Dort steigt das RBO-Niveau zwischenzeitlich auf maximal 0.11 (p=0.75) bzw. 0.25, bei einem p=1 an.

Bezieht man alle Suchbegriffe, Suchumgebungsoptionen und beide p-Werte mit ein, zeigt sich, dass sich das RBO-Level auch hier auf einem meist geringen Stand bewegt. Die Grenze von 0.5 wird nur in wenigen Fällen und allermeist auch nur punktuell überschritten. Dies verdeutlicht sich auch durch die gebildeten Mittelwerte der ermittelten RBO-Scores über alle 20 Suchbegriffe und den Datenerhebungszeitraum. Die Konstellation Siri_Tab/Bing weist hier einen Wert von 0.26 (bezogen auf p=0.75) und 0.28 (p=1) auf. Genauso ist dieser durchschnittliche Wert bei der Zusammenstellung von Siri_Simulator und Bing bei 0.26, wenn das p auf 0.75 festgesetzt ist. Nur unwesentlich höher liegt der Schnitt bei p=1 mit einem Wert von 0.29.

6.2.4 Siri / Google

Bei der Betrachtung der ermittelten RBOs im Vergleich der Siri Suchen und Google zeigt sich, dass hier, anders als bei den vorher erläuterten Kombinationen, kaum durchgängig konstante RBO-Werte dokumentiert worden sind. Eine Ausnahme bilden die Suchbegriffe „Katy Perry“ in der Vergleichskonfiguration Siri_Simulator/Google mit p=1 und Leo Messi beim Vergleich Siri_Tab/Google mit p=1.

Abgesehen von dieser Ausnahme zeigen sich die Kurvenverläufe deutlich schwankend. Diese intensiven Bewegungen im Verlauf fallen je nach Suchbegriff aber wiederum trotzdem unterschiedlich aus. Außerdem fallen sie nicht nur um den bereits thematisierten Ortswechsel herum auf, sondern es zeichnen sich über den gesamten Erhebungszeitraum von 29 Tagen Veränderungen ab.

Insbesondere ist aber auch erkennbar, dass das RBO-Niveau über die unterschiedlichen Suchbegriffe ganz verschieden gelagert ist. Allerdings zeigt sich auch über alle Vergleichskombinationen hinweg, also Siri_Tab oder Siri_Simulator und p=0.75 oder p=1, dass augenscheinlich häufig hohe Spitzen in den RBO-Werten erreicht werden, die die Marke von 0.5 überschreiten.

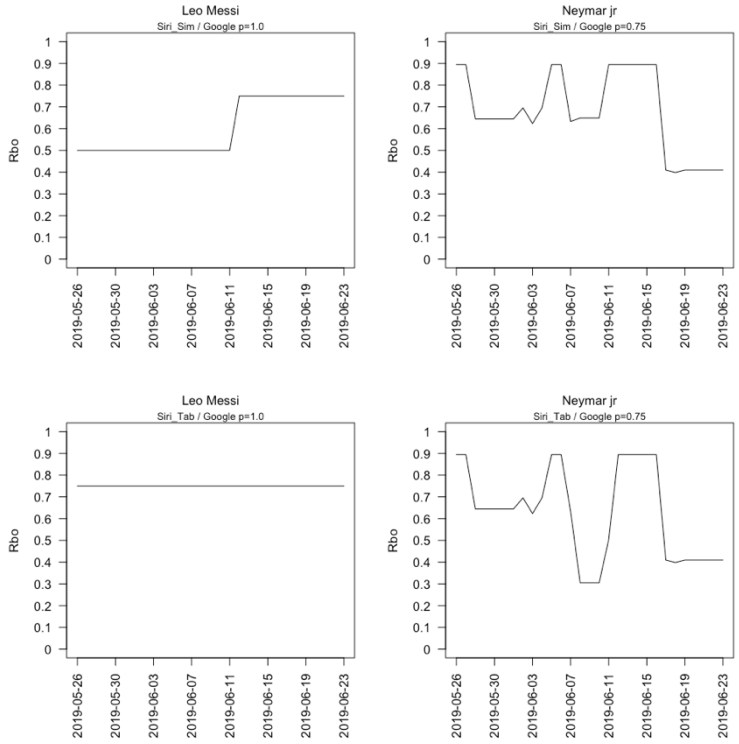


Abbildung 22: Darstellung des Verlaufs des RBO über den Erhebungszeitraum bei unterschiedlicher Gewichtung p=0.75 / p=1 für die Suchbegriffe Leo Messi und Neymar jr unter der Betrachtung Google / Siri_Simulator – Siri_Tablet

Diese visuell auffälligen Ausschläge (Abbildung 22) spiegeln sich darüber hinaus auch bei Betrachtung der durchschnittlichen RBOs zu einem Suchbegriff wider.

	Google/Siri_Tab p=0.75	**Google/Siri_Tab p=1.0**	**Google/Siri_Sim p=0.75**	**Google/Siri_Sim p=1.0**
Leo Messi	0.74	0.75	0.73	0.60
Neymar jr	0.63	0.68	0.68	0.68
Barbie	0.47	0.63	0.60	0.63
Nike	0.76	0.90	0.80	0.89

Tabelle 9: Durchschnitt der RBO-Scores zu ausgewählten Suchbegriffen über den Erhebungszeitraum. Aufgeteilt nach Vergleichspaarungen Google/Siri_Tab und Google/Siri_Simulator unter Berücksichtigung von p=0.75 und p=1

So zeigt sich beispielsweise beim Suchbegriff „Nike“, dass in allen in Tabelle 9 abgebildeten Vergleiche durchschnittlich hohe RBO-Werte von minimal 0.76 und 0.90 im

Maximum erreicht werden. Auch für den Begriff „Leo Messi“ sind in dieser Gegenüberstellung hohe Durchschnittswerte verzeichnet von minimal 0.6 bis 0.75 im Maximum.

Dem gegenüberstehen aber wiederum Begriffe, zu denen die errechneten RBOs nur gering ausfallen.

	Google/Siri_Tab p=0.75	Google/Siri_Tab p=1.0	Google/Siri_Sim p=0.75	Google/Siri_Sim p=1.0
Instagram	0.10	0.16	0.10	0.16
Kim Kardashian West	0.13	0.26	0.13	0.25
Khloe Kardashian	0.07	0.15	0.02	0.06
Katy Perry	0.19	0.26	0.33	0.25

Tabelle 10: Durchschnitt der RBO-Scores zu ausgewählten Suchbegriffen über den Erhebungszeitraum. Aufgeteilt nach Vergleichspaarungen Google/Siri_Tab und Google/Siri_Simulator unter Berücksichtigung von p=0.75 und p=1

Bei „Khloe Kardashian“ sind die Mittelwerte der RBOs im gleichen Zeitraum mit nur 0.02 bis maximal 0.15 errechnet. Auch bei Instagram ergeben sich mit 0.10 und 0.16 geringe Werte, wobei der Durchschnitt hier bei identischen p-Werten paarweise gleich ist.

Die Gegenüberstellung der Tabellen 9 und 10 zeigt nochmals auf, wie gegensätzlich sich die Verläufe der RBO-Scores bei verschiedenen Suchbegriffen gestalten.

Betrachtet man hier allerdings die durchschnittlichen RBOs über alle Suchbegriffe von Google/Siri_Tab bei p=0.75 und p=1, die bei 0.31 und 0.38 liegen, dann liegen diese höher als die der übrigen Vergleiche. Die Kombination Google/Siri_Simulator mit den genannten p-Werten, verzeichnet durchschnittliche RBO-Werte von 0.30 bzw. 0.40, was den höchsten Durchschnittswert über alle Vergleiche darstellt.

6.2.5 Zusammenfassende Betrachtung zum RBO

Zusammenfassend hat sich gezeigt, dass bei Anwendung des RBO keine der im Vergleich betrachteten Suchmaschinen sich mit einem konstant hohen RBO-Score im Vergleich mit den durchgeführten Siri-Suchen über alle angewendeten Suchbegriffe hinweg deutlich zu anderen abgrenzt.

Bei Betrachtung der durchschnittlichen RBO-Werte über alle Suchbegriffe hinweg, weist Google im Vergleich die höchsten Scores auf (Tabelle 11). Dies ist allerdings bedingt durch die zeitweisen auftretenden hohen Spitzen und den vereinzelt überdurchschnittlich hohen RBOs.

	p=0.75	p=1
Siri_Simulator/Siri_Tab	0.73	0.78
Siri_Tab/DuckDuckGo	0.20	0.28
Siri_Simulator/ DuckDuckGo	0.28	0.31
Siri_Tab/Bing	0.26	0.28
Siri_Simulator/Bing	0.26	0.29
Siri_Tab/Google	0.31	0.38
Siri_Simulator/Google	0.30	0.37

Tabelle 11: Durchschnitts RBO über alle Suchbegriffe zum jeweiligen Suchmaschinenvergleich mit den Gewichtungen p=0.75 und p=1

Bei den Suchmaschinen DuckDuckGo und Bing zeigte sich beim Blick auf diesen durchschnittlichen Score ein geringerer Wert. Dafür waren bei den beiden genannten Suchmaschinen die Verläufe der RBOs über den Erhebungszeitraum erkennbar konstanter, als die Kurvenverläufe bei Google. Diese Konstanz ist ein deutliches Zeichen für wenige Veränderungen der gelieferten Query Suggestions, sowohl in Bezug auf inhaltliche Veränderungen als auch in der Rangfolge. Genau entgegengesetzt ist dabei das Verhalten der RBO-Kurven im Vergleich Siri/Google zu interpretieren.

Eine Ausnahme von der beobachteten Konstanz zeigte sich im Zeitraum des mehrfach thematisierten Ortswechsels (Suchbegriff Barbie als Beispiel). Auch wenn sich das Verhalten über die Suchmaschinen und Suchbegriffe hinweg different zeigt, so zeigte sich häufig der Fall, dass Schwankungen von einem Ausgangswert aus, entweder hinauf oder hinunter, nach der Rückkehr nach Köln wieder auf diesen Ausgangwert zurückgingen. Diese Beobachtung der vermehrten Bewegung explizit in diesem Zeitraum spricht stark dafür, dass die im iOS Security Report von Apple aus dem Mai 2019 aufgeführte Verwendung von Standortdaten tatsächlich zutrifft. Wie diese Standortdaten genau verwendet werden, ist wie folgt beschrieben: „The approximate location of their device, if they have Location Services for Location-Based Suggestions turned on“ [42]
Ein weiterführendes Indiz genau für diese Beobachtung ist auch, dass die RBO-Scores im Vergleich von Siri-Suche am Tablet und Siri-Suche in der Simulation über alle Suchbegriffe hinweg häufig erkennbare Unterschiede aufweisen, was sich auch im durchschnittlichen RBO von 0.73 bzw. 0.78 (Tabelle 11) zeigt. Dass diese Werte nicht konstant höher liegen, ist ebenfalls genau auf Vorschläge mit standortspezifischem Bezug zurückzuführen, die bei Vorschlägen der Siri-Suche am Tablet auftauchen, in der Simulation jedoch nicht.

[42] *Apple Inc.* (Mai 2019): iOS Security: iOS 12.3, May 2019, S. 70.

6.3 Ergebnisse Score Berechnung zur Darstellung der durchschnittlichen Precision zu einer Kategorie

Um in Bezug auf den berechneten Score Ergebnisse vergleichen und bewerten zu können, wird die gelieferte Ausgabe des R-Skript zur Score-Berechnung, welches in Kapitel 5.5 erläutert wurde, verwendet.

Mit diesem Score ist die Möglichkeit geschaffen, folgende Fragestellung zu beantworten:

Wie hoch ist die Wahrscheinlichkeit, dass bei Suchbegriff X aus Population Y bei Suchmaschine S ein Vorschlag aus der Kategorie K kommt?

Es sollen also Gruppen bzw. Populationen betrachtet werden und ermittelt werden, aus welcher Kategorie des intellektuell konzipierten Clusters die Vorschläge der Suchmaschinen stammen, wenn der Suchbegriff aus einer bestimmten betrachteten Gruppe kommt.

Diese betrachteten Gruppen ergeben sich dabei aus den in Kapitel 3.3 vorgestellten Steckbriefen zu den Suchbegriffen. Dies lässt im ersten Schritt eine Aufteilung nach Geschlecht, hier Männer und Frauen, sinnvoll erscheinen. Darüber hinaus ist es zudem sinnhaft eine Betrachtung nach Berufsgruppen anzustreben. Diese Personen, die Schauspieler und Sänger sind, werden in der Gruppe der Künstler zusammengefasst. Eindeutig ergibt sich auch die Gruppe der Sportler. Des Weiteren wird aus den vier Familienmitgliedern des bekannten Kardashian-Clans eine eigene Gruppe gebildet, welche im weitesten Sinne ebenfalls die Profession der Personen widerspiegelt.

Je nach Fragestellung wird die Ergebnistabelle nach den zur Gruppe zugehörigen Suchbegriffen gefiltert. Alle Scores einer Kategorie werden dann aufsummiert. Somit wird eine Summe zu jeder Kategorie ermittelt, die einen Überblick darüber bietet, welche Kategorie, bezogen auf die zugehörigen Suchbegriffe der jeweils betrachteten Population, am häufigsten vorkommt. Allerdings ist zu beachten, dass die variierenden Größen der betrachteten Populationen ausgeglichen werden muss. Um dies zu erreichen werden die ermittelten Summen jeweils durch die Anzahl der jeweiligen Populationsgröße dividiert. Hiermit ist die Ermittlung des prozentualen Anteils einer Kategorie zur Gesamtheit der Kategorien gewährleistet.

	Siri_Simulator	**Siri_Tab**	**Google**	**DuckDuckGo**	**Bing**
Social Media/Information	0.23	0.24	0.13	0.22	0.38
Körpermerkmale	0.20	0.14	0.16	0.08	0.01
Profession	0.15	0.17	0.12	0.14	0.13
Statussymbole	0.05	0.07	0.14	0.00	0.09
Beziehung/Familie	0.20	0.20	0.21	0.18	0.10
Nacktheit/Tod	0.05	0.04	0.03	0.00	0.04
Orte/Sprache	0.00	0.00	0.00	0.04	0.06
Sonstiges	0.12	0.14	0.20	0.34	0.20

Tabelle 12: Vorschlagshäufigkeiten zur Population der Männer nach Kategorien über Suchmaschinen hinweg

	Siri_Simulator	Siri_Tab	Google	DuckDuckGo	Bing
Social Media/Information	0.23	0.26	0.12	0.33	0.44
Körpermerkmale	0.09	0.11	0.22	0.14	0.06
Profession	0.29	0.18	0.29	0.26	0.24
Statussymbole	0.02	0.02	0.05	0.00	0.03
Beziehung/Familie	0.24	0.25	0.19	0.16	0.10
Nacktheit/Tod	0.02	0.04	0.00	0.05	0.02
Orte/Sprache	0.02	0.05	0.01	0.00	0.00
Sonstiges	0.08	0.09	0.12	0.06	0.10

Tabelle 13: Vorschlagshäufigkeiten zur Population der Frauen nach Kategorien über Suchmaschinen hinweg

Bei Bing ist bei jeder betrachteten Population (Tabelle 12-16) mit deutlichem Abstand die Kategorie „Social Media/Information“ vorne.

Google ist bei jeder Population, außer die der Sportler, ist die Kategorie „Körpermerkmale“ mindestens unter den Top 3 Kategorien. Bei der Gruppe der Frauen ist diese Kategorie sogar relativ deutlich die zweithäufigste Vorschlagskategorie. Betrachtet man hier die Gruppe „The Kardashians“ liegt diese Kategorie bei mit großem Abstand sogar auf dem ersten Platz. Die genannten Top 3 Kategorien sind in den Tabellen farblich abgehoben. Dabei repräsentiert das dunkle grün die stärkste Kategorie, nachfolgende Kategorie ist mit einem hellen grün markiert und die drittstärkste Kategorie ist gelb hervorgehoben. Tauchen gleiche Werte für verschiedenen Kategorien auf, sind diese je nach Platzierung in diesem Ranking ebenfalls entsprechend markiert, sodass Ränge in den Top 3 mehrfach belegt und farblich unterlegt sein können.

Siri_Simulator und Siri_Tab bedienen häufig die gleichen Top-Kategorien. Bei Fokus auf die Gruppe der Männer ist die vorrangig vorgeschlagene Kategorie bei beiden genannten „Social Media/Information“. Hier zeigt sich sogar, dass sich beide Suchumgebungen in den nachfolgend stärksten Kategorien „Beziehung/Familie“ an zweiter Stelle und Kategorie „Profession“ an dritter Position gleich sind. Hier mit der Einschränkung, dass bei Siri_Simulator die Kategorien „Beziehung/Familie“ und „Körpermerkmale“ mit einem Wert von 0.20 genau gleich häufig vorgeschlagen werden.

In Bezug auf die Gruppe der Frauen setzt sich dieses Bild fort. Hier sind die Top 3-Kategorien über nahezu alle verglichenen Suchmaschinen hinweg „Social Media/Information“, „Profession“ und „Beziehung/Familie“. Es unterscheiden sich zwischen den einzelnen Suchmaschinen lediglich jeweils die Rangfolgen. Einzig Google variiert hier im übergreifenden Vergleich und weist als zweitstärkste Kategorie mit 0.22 „Körpermerkmale“ auf.

Vergleicht man allerdings die Gruppen der Frauen und Männer in Bezug auf die Kategorie „Körpermerkmale“, so wird diese Kategorie von Siri_Simulator und Siri_Tab für Männer sogar deutlich häufiger mit Vorschlägen bedient (0.20 bzw. 0.14), als diese Suchmaschinen das für Frauen tun. Hier liegen die ermittelten Werte bei 0.09 bzw. 0.11.

Diese Beobachtung gilt aber für Google wiederum genau umgekehrt. Hierfür liegt der Wert für Frauen mit 0.22 erkennbar höher als dieser für die Gruppe der Männer mit 0.16.

Bei Betrachtung der Gruppe der Künstler über alle Suchmaschinen hinweg, zeigt sich ein ähnliches Muster, wie zuvor schon bei der Gruppe der Frauen beobachtet. Hier zeigt sich, dass über fast alle betrachteten Suchmaschinen die Kategorie „Profession“ die stärkste Kategorie ist. Nur Bing bedient auch für diese Gruppe die Kategorie „Social Media/Information“ am häufigsten. Zudem wird hier erneut deutlich, dass Siri_Simulator und Siri_Tab die gleiche Reihenfolge der Top 3-Kategorien „Profession“, „Social Media/Information“ und „Beziehung/Familie“ aufweisen. Lediglich der Abstand zwischen Kategorie „Profession“ an erster Stelle und „Social Media/Information“ an zweiter Stelle ist bei Siri_Simulation größer.

In Bezug auf die Gruppe „The Kardashians“ zeigt sich die Kategorie „Beziehung/Familie“ als am stärksten ausgeprägten Kategorie bei den Suchmaschinen DuckDuckGo, Siri_Simulator und Siri_Tab. Bing und Google weisen dabei an erster Stelle, wie bereits erwähnt, „Social Media/Information“ bzw. „Körpermerkmale“ als stärkste Kategorie auf. Aus der Kategorie „Social Media/Information“ stammen allerdings bei DuckDuckGo und Siri_Simulator am zweithäufigsten die Vorschläge bei einer Suche zu einem Gruppenmitglied von „The Kardashians“. Für Siri_Tab gilt sogar der Fall, dass diese Kategorie sich mit einem Wert von 0.33 den erwähnten ersten Rang mit „Beziehung/Familie“ teilt.

In der Betrachtung der Berufsgruppe der Sportler zeigt sich, dass über die Suchmaschinen Siri_Simulator, Siri_Tab, Google und DuckDuckGo die Kategorie „Beziehung/Familie“ die häufigste bzw. zweithäufigste Kategorie darstellen. Sie weisen aber unabhängig von der suchmaschineninternen Platzierung der Kategorien hier mit 0.25 (Siri_Simulator), 0.25 (Siri_Tab), 0.27 (Google) und 0.27 (DuckDuckGo) ähnliche Werte auf. Hier verhält sich Bing abermals entgegengesetzt und weist die Kategorie „Social Media/Information“ als häufigste Kategorie auf. Auch bei allen weiteren Suchmaschinen ist diese Kategorie mindestens in den Top 3 platziert.

Auffallend ist, dass die Kategorie „Statussymbole“ für die Gruppen Frauen und Künstler quasi keine Rolle spielt. Für die Gruppe der Männer und The Kardashians ist der Wert bei Suchmaschine Google aber auffällig höher. Bei beiden Gruppen liegt er bei 0.14. Wirklich hervor sticht die Kategorie allerdings für die Gruppe der Sportler, hier bildet sie für die Suchmaschine Google mit 0.20 mit erkennbarem Abstand die zweithäufigste Vorschlagskategorie. Dies trifft mit einem Wert von 0.15 auch auf Bing zu.

	Siri_Simulator	Siri_Tab	Google	DuckDuckGo	Bing
Social Media/Information	0.17	0.22	0.10	0.32	0.42
Körpermerkmale	0.11	0.09	0.17	0.07	0.01
Profession	0.35	0.23	0.35	0.32	0.29
Statussymbole	0.00	0.00	0.02	0.00	0.01
Beziehung/Familie	0.16	0.19	0.15	0.06	0.06
Nacktheit/Tod	0.05	0.07	0.02	0.05	0.04
Orte/Sprache	0.03	0.07	0.01	0.02	0.03
Sonstiges	0.13	0.13	0.18	0.16	0.15

Tabelle 14: Vorschlagshäufigkeiten zur Population der Künstler nach Kategorien über Suchmaschinen hinweg

	Siri_Simulator	Siri_Tab	Google	DuckDuckGo	Bing
Social Media/Information	0.31	0.33	0.14	0.27	0.43
Körpermerkmale	0.13	0.16	0.34	0.27	0.16
Profession	0.06	0.06	0.06	0.09	0.08
Statussymbole	0.06	0.06	0.14	0.00	0.05
Beziehung/Familie	0.37	0.33	0.24	0.33	0.19
Nacktheit/Tod	0.00	0.00	0.00	0.03	0.01
Orte/Sprache	0.00	0.00	0.00	0.00	0.01
Sonstiges	0.07	0.06	0.09	0.03	0.07

Tabelle 15: Vorschlagshäufigkeiten zur Population der „The Kardashians“ nach Kategorien über Suchmaschinen hinweg

	Siri_Simulator	Siri_Tab	Google	DuckDuckGo	Bing
Social Media/Information	0.30	0.25	0.16	0.26	0.42
Körpermerkmale	0.17	0.15	0.14	0.09	0.00
Profession	0.17	0.17	0.12	0.10	0.15
Statussymbole	0.08	0.10	0.20	0.00	0.15
Beziehung/Familie	0.25	0.25	0.27	0.27	0.13
Nacktheit/Tod	0.00	0.00	0.00	0.00	0.00
Orte/Sprache	0.00	0.00	0.00	0.00	0.00
Sonstiges	0.03	0.08	0.11	0.28	0.15

Tabelle 16: Vorschlagshäufigkeiten zur Population der Sportler nach Kategorien über Suchmaschinen hinweg

Abschließend lässt sich für diese vergleichende Betrachtung zusammenfassend festhalten, dass die beiden Siri-Suchen sich im Bezug auf ihre Top 3-Kategorien, über alle betrachteten Populationen hinweg, kaum unterscheiden. Dies gilt sowohl für die Rangfolgen und auch der Größenordnung der ermittelten Scores. Das bedeutet, dass bei einer Suchanfrage zu einem Term aus einer bestimmten Gruppierung, eine nahezu gleich hohe Wahrscheinlichkeit herrscht, dass der Vorschlag sowohl bei Siri_Tab, als auch bei Siri_Simulator aus der gleichen Kategorie erfolgt.

Festzuhalten ist außerdem, dass die am stärksten ausgeprägten Kategorien über alle Populationen gesehen „Social Media/Information“, „Profession“ und „Beziehung/Familie“ darstellen.

Einige Suchmaschinen tun sich bei Ausprägung von Kategorien besonders hervor. So ist Bing, wie bereits erwähnt, durchgängig „Social Media/Information“ die Kategorie, aus der diese Suchmaschine für alle Populationen die meisten Vervollständigungen liefert.

Auch Google ist als einzige Suchmaschine besonders auffällig in Bezug auf die Kategorie „Körpermerkmale“. Teilweise bestätigen sich in diesem Vergleich sogar gängige Klischees. So ist für die Gruppe der Frauen die Wahrscheinlichkeit mit einem Wert von 0.22 erkennbar hoch, dass ein Vorschlag aus dieser Kategorie stammen. Für die Gruppe „The Kardashians“ liegt diese Wahrscheinlichkeit mit sogar nochmals höher mit einem Score von 0.34.

Im Übrigen ist dabei die Kategorie „Statussymbole“ ganz ähnlich auffällig. Vorschläge aus dieser Kategorie werden mit einer fast gleich hohen Wahrscheinlichkeit für die Gruppe der Sportler (0.20) bzw. die Gruppe der Männer (0.14) gemacht.

Teilweise ist die Kategorie „Sonstige“ die am stärksten belegte. Dies deutet darauf hin, dass die Zuteilung zu dieser Kategorie ein Sammelbecken für ganz unterschiedliche Vervollständigungen darstellt. Um diesen Effekt umzukehren, erscheint es rückblickend sinnvoll, diese Kategorien zu entschlacken und feiner zu gruppieren.

Es lassen sich also über alles Suchmaschinen hinweg Tendenzen erkennen, dass Vorschläge thematisch aus den gleichen Top-Kategorien gemacht werden. Trotzdem sind einige Ausreißer, wie oben beschrieben, bemerkenswert.

Die Betrachtung von Suchbegriffen zu Marken bzw. Firmen wurde in dieser Auswertung ausgeklammert, da sowohl die Vorschläge weniger breit geclustert wurden als auch eine sinnvolle Einteilung in zu betrachtenden Gruppen sich hier als wenig sinnvoll erwiesen hat.

7 Fazit

Im Laufe dieser Bachelorarbeit wurden zu 20 ausgewählten Suchbegriffen insgesamt 23.427 Datensätze erhoben, verarbeitet und ausgewertet.

Durch Anwendung der drei zuvor beschriebenen voneinander unabhängigen Methoden, Frequenz- und Schnittmengenanalyse, RBO und Clusterbildung mit anschließender Score Berechnung konnte nicht eindeutig bewiesen werden, ob eine Plattform derer, die während des Vergleichs beleuchtet worden sind, der „qualified partner“[43] ist, an den Apple nach eigener Aussagen im iOS Security Report aus dem Mai 2019 für einzelne Anfragen zu gängigen Worten und Phrasen weitergibt.[44]

Im Rahmen der Frequenzanalyse zeigten sich deutliche Gemeinsamkeiten zwischen den beiden in unterschiedlichen Setups durchgeführten Siri-Suchen. Die Unterschiede hier beziehen sich in aller Regel auf Vorschläge mit lokalem Bezug, die während einer Suche am Tablet vorgeschlagen werden, im Simulator so jedoch nicht auftauchen. Feststellbar war auch, dass die bei Siri, egal ob in der Simulation oder am Tablet, vorgeschlagenen Unique Terms, zu einem hohen Prozentsatz auch bei Google geliefert werden. Der Anteil liegt im Vergleich wesentlich höher, als bei den beiden anderen einbezogenen Suchmaschinen DuckDuckGo und Bing.

Bei Anwendung der Methode RBO zeigten sich hohe RBO-Scores beim Vergleich von Siri_Simulator/Siri_Tab (auch im Zeitverlauf), die oftmals auch konstant verliefen. Das deutet auf nur kleine Unterschiede in Inhalt und Reihenfolge der jeweils gegeben Vorschläge hin. Viele der Unterschiede sind auch hier auf Vorschläge mit geografischem Bezug zurückzuführen, die Siri_Tab liefert.

DuckDuckGo und Bing haben sich in Bezug auf den RBO zu den verschiedenen Suchbegriffen unterschiedlich verhalten. Beide wiesen aber meist einen geringen Score im Vergleich mit den Suchumgebungen von Siri auf. Wenngleich sich bei diesen beiden Websuchmaschinen ebenfalls recht konstante Kurvenverläufe nachweisen ließen, die auf nur kleinere Veränderungen der übermittelten Vervollständigungen im Verlauf der Erhebung schließen lassen. Google zeigt bei der Methode RBO von allen Suchmaschinen die meiste Bewegung im Zeitverlauf, weist aber durchschnittlich den höchsten RBO über alle Suchbegriffe im Erhebungszeitraum auf. Dieser hohe Durchschnitt ergibt sich aber weniger durch konstant hohe RBO, sondern ist bedingt durch vereinzelt sehr hohe Werte bei einzelnen Suchbegriffen und punktuell hohe Spitzen. Google bietet hier somit sogar noch die höchsten Scores, aber der RBO zeigt über alle durchgeführten Vergleiche keine klaren Muster. Daher grenzt sich auch keine der betrachteten Suchmaschinen ganz deutlich ab.

In Bezug auf die Clusterbildung und die daran angeschlossene Berechnung und Auswertung des daran anschließenden Scores konnten keine klaren Vergleiche zwischen

[43] Apple Inc. (Mai 2019): iOS Security: iOS 12.3, May 2019, S. 71.

[44] vgl. ebd.

den Suchmaschinen gezogen werden. Es zeigt sich lediglich, dass sowohl Siri_Tab als auch Siri_Simulator bei Suchanfragen zu einer betrachteten Population auch meist sehr ähnlich, aus der gleiche Kategorie, vorschlagen. Zudem waren die, am häufigsten vorkommenden Kategorien, über alle Suchmaschinen hinweg (ausgewiesen im Vergleich als Top 3) auffällig gleich. Trotz dieser gemachten Beobachtung zeigen sich aber spannende Tendenzen für einzelne Suchmaschinen. So werden doch schon in einem vergleichsweisen kleinen Rahmen, wie es der Versuchsaufbau im Rahmen dieser Arbeit ist, „Klischees“ bedient, zumindest für einzelne Suchmaschinen. Dies zeigte sich zum Beispiel dadurch, dass Vorschläge von Google deutlich häufiger aus der Kategorie „Statussymbole“ stammen, wenn der Suchgebegriff der Gruppe der Sportler angehört.

Über alle angewendeten Methoden betrachtet kristallisiert sich Google als die Suchmaschinen mit der tendenziell größten Überschneidung zu den Suchen in Siri heraus. Allerdings sind die ermittelten Werte nicht so hoch und konstant, dass nach der Analyse im Rahmen dieser Bachelorarbeit eine eindeutige Aussage darüber getroffen werden sollte, ob Google tatsächlich ein Kooperationspartner und somit die mögliche Quelle der Siri Suggestions ist.

Folglich konnte die zu Beginn gestellte Forschungsfrage durch Anwendung der gewählten Methoden für einen Vergleich nicht bestimmt beantwortet werden.

Nachfolgend ein kurzes Resümee zur Umsetzung der Arbeit mit einem Ausblick.

Rückblickend lässt sich sagen, dass das entwickelte JS-Skript zum Absenden der Suchanfragen an die jeweiligen APIs in dieser Arbeit rein auf Funktionalität ausgerichtet war. Anpassungen dort würden die zeitlich sehr aufwändige, aber erforderliche Bereinigung im Nachgang reduzieren können. Dies hätte eine deutliche Arbeitserleichterung bedeutet. Insbesondere, da dass das händische Erfassen der Autocompletions in Siri, durch das 40-fache Eintragen in ein Suchfeld und das manuelle Dokumentieren die Datenerhebung sehr zeitaufwändig gemacht hat.

Zudem ist festzuhalten, dass sich gegen eine zunächst beabsichtige tiefergehende Analyse entschieden wurde, bei der die jeweiligen Google-Trends des Vortags als Anfrageterme gedient haben.[45] Diese Daten zu diesen Trendthemen wurden über den Zeitraum von 29 Tagen zwar ebenfalls erhoben, allerdings zeigten sich durch die, in der Themenauswahl begründet liegenden, starken Vorprägung in Richtung Google teils große Lücken bei den zurückgelieferten Vorschlägen der übrigen Suchmaschinen. Da dadurch eine rein stichprobenartige und, bedingt durch sich täglich verändernde Suchbegriffe, nur punktuelle Vergleichbarkeit gegeben gewesen wäre, wurde diese Idee unter Anbetracht des zeitlichen Aufwands einer Analyse dieser, verworfen.

Das in dieser Arbeit angewendete Clustering hat, gemessen am abgesteckten Rahmen, schon interessante Erkenntnisse hervorgebracht. Denkbar wäre hier anstelle einer

[45] Google (o.D.): Suchtrends des Tages.

intellektuellen Herangehensweise ein automatisches Vorgehen. Möglich wäre zum Beispiel eine Umsetzung mit Hilfe des k-means-Algorithmus, wie sie Anastasiia Samokhina in ihrer Masterthesis “Analysing the systematics of search engine autocompletion functions by means of data mining methods” beschreibt. In dieser Arbeit vergleicht sie ebenfalls Vorschläge der Websuchmaschinen Google, Bing und DuckDuckGo zu Mitgliedern des deutschen Bundestags und verwendet dort k-Means Algorithmus zur Clusterbildung. [46]

Einen Vergleich von Autocompletions von Websuchmaschinen unter Anwendung der Methode RBO vollziehen auch Malte Bonart und Philipp Schaer unter anderem im Rahmen der Publikation „Intertemporal Connections Between Query Suggestions and Search Engine Results for Politics Related Queries“. In dieser wird allerdings kein übergreifender Vergleich vorgenommen, sondern die Stabilität der Ergebnisse gemessen. Auch diese Arbeit nutzt Suchterme aus dem politischen Kontext [47]. Eine ähnliche Betrachtung unter Anwendung des RBO als Stabilitätsmaß von Autocompletions zu Suchbergriffen aus dem Bereich von Social Media, wie sie in dieser Arbeit genutzt wurden, könnte eine ebenfalls interessante Blickrichtung auf diese Suchbegriffe aus einem Themenbereich des öffentlichen Interesses abseits der Politik sein.

[46] vgl. Samokhina, A. (2017): Analysing the systematics of search engine autocompletion functions by means of data mining methods.

[47] vgl. Bonart, M. /Schaer, P. (2018): Intertemporal Connections Between Query Suggestions and Search Engine Results for Politics Related Queries.

Material

Das zu dieser Arbeit erstellte Material, in Form von Quellcode, erhobene Daten und Dateien zur Analyse, finden sich in folgender Ordnerstruktrur abgelegt auf der dieser Arbeit beigelegten CD-ROM. Zudem sind diese Informationen auch in einem öffentlichen GitHub-Repository [48] abrufbar. Die erhobenen Daten der „dauerthemen_gesamt.csv“ sind zudem auf der Plattform Zenodo [49] hinterlegt.

Ordner	**Dateiname**	**Erklärung**
Skript_Datenerhebung	suchmaschinenAbfrage.js	Skript um Suchanfragen an Google/Bing/DuckDuckGo/Yahoo zu versenden
	dauer.txt	Suchanfragen, über die im Skript iteriert werden
	dauerthemen_gesamt.csv	Sammlung der Datenerhebung (Gesamtoutput aus Datenerhebung)
Frequenz_Schnittmengenanalyse	FrequenzSchnittmengenAnalyse.xlsx	Auswertung der Daten der Erhebung mittels einer Frequenz-/Schnittmengenanalyse (PivotTabelle & Ermittlung UniqueTerms)
RBO	dauerthemen_gesamt.csv	Sammlung der Datenerhebung / Basis zur Berechnung der RBO-Werte (Input für RBO-Berechnung)
	rbo.R	R-Skript um RBO-Werte zu bestimmen
	rbo_plot.R	R-Skript um ermittlelte RBO-Werte graphisch darzustellen
	rbo_Siri_Tab-Bing-0.75.csv rbo_Siri_Tab-Bing-1.0.csv rbo_Siri_Tab-DuckDuckGo-0.75.csv rbo_Siri_Tab-DuckDuckGo-1.0.csv rbo_Siri_Tab-Google-0.75.csv rbo_Siri_Tab-Google-1.0.csv rbo_Siri_Tab-siri_simulator-0.75.csv rbo_Siri_Tab-siri_simulator-1.0.csv rbo_Siri_Simulator-Bing-0.75.csv rbo_Siri_Simulator-Bing-1.0.csv rbo_Siri_Simulator-DuckDuckGo-0.75.csv rbo_Siri_Simulator-DuckDuckGo-1.0.csv rbo_Siri_Simulator-Google-0.75.csv rbo_Siri_Simulator-Google-1.0.csv	Sammlung der ermittelten RBO-Werte getrennt nach gewählten Parameter p
Scoring	scoring.R	R-Skript zur Score-Berechnung der Kategorie 1-10
	bereinigteGesamtScoring.csv	Sammlung der um die Kategorie erweiterte Datenerhebung
	ScoringErgebnis.csv	Ergebnis der Score-Berechnung für die Suchbegriffe über alle Suchmaschinen

Tabelle 17: Ablagestruktur, der der Bachelorarbeit beigelegten Daten bzw. Ordnerstruktur im GitHub Repository

[48] https://github.com/blangkam/QuerySuggestionMitSiri

[49] https://doi.org/10.5281/zenodo.3361813

Abbildungsverzeichnis

Tabellenverzeichnis

Literaturverzeichnis

Anderson, Bruce R. / Galvez, Kassandra (2018): „Apple“. In: Schintler Laurie A. / McNeely, Connie L. (Hrsg), Encyclopedia of Big Data, Springer International Publishing, Cham, DOI: 10.1007/978-3-319-32001-4_11-1.

Apple Inc (2019): iPhone Benutzerhandbuch: Für iOS 6 Software, <https://manuals.info.apple.com/MANUALS/1000/MA1658/de_DE/iphone_ios6_benutzerhandbuch.pdf>, aufgerufen am 05.08.2019.

Apple Inc. (Mai 2019): iOS Security: iOS 12.3, May 2019, <https://www.apple.com/business/site/docs/iOS_Security_Guide.pdf>, aufgerufen am 05.08.2019

Apple Inc., 20. Mai 2019: Suchfunktion auf dem iPhone, iPad oder iPod touch verwenden, 20.05.2019 <https://support.apple.com/de-de/HT201285>, aufgerufen am 05.08.2019.

Beza- Yates, Ricardo / Ribero-Neto, Berthier: Modern Information Retrieval. the concepts an technology behind search, 2. Auflage, Harlow 2011

Bonart, Malte (2019): rbo in r, <https://gist.github.com/bonartm/c9d07132f62519c957bb7e30f302e57c>, aufgerufen am 02.08.2019.

Bonart, Malte / Schaer, Philipp (2018): Intertemporal Connections Between Query Suggestions and Search Engine Results for Politics Related Queries, arXiv:1812.08585v2 [cs.IR].

Brunsmann, Jörg / Hauser, Dominik / Rodewig, Klaus M: Apps programmieren mit Swift, 1. Auflage, Bonn 2017.

DuckDuckGo (o.D.): We don’t collect or share personal information: That’s our privacy policy in a nutshell., <https://duckduckgo.com/privacy>, aufgerufen am 06.08.2019.

DuckDuckGo (2019): Result Sources <https://help.duckduckgo.com/duckduckgo-help-pages/results/sources/>, aufgerufen am 06.08.2019.

Ecma International (2017): Standaed ECMA-404. The JSON Data Interchange Syntax <http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf>, aufgerufen am 06.08.2019.

Gavin, Ryan (2018): Delivering Personalized Search Experiences in Windows 10 through Cortana, <https://blogs.windows.com/windowsexperience/2016/04/28/delivering-personalized-search-experiences-in-windows-10-through-cortana/>, aufgerufen am 06.08.2019.

Geißler, Otto / Ostler, Ulrike (2018): „Was ist ein Application-Programming-Interface (API)?“, <https://www.datacenter-insider.de/was-ist-ein-application-programming-interface-api-a-735797/>, aufgerufen am 06.08.2019.

Google (o.D): Suchtrends des Tages, https://trends.google.de/trends/trendingsearches/daily?geo=DE, aufgerufen am 06.08.2019.

Langkammerer, Birte (2019): QuerySuggestionMitSiri, <https://github.com/blangkam/QuerySuggestionMitSiri>, aufgerufen am: 06.08.2019.

Langkammerer, Birte (2019): query suggestion with siri - query suggestions for instagram accounts, <https://doi.org/10.5281/zenodo.3361813>, aufgerufen am: 06.08.2019.

Kerkmann, Christof / Scheuer, Stephan (2018): Elektronikmesse IFA: Siri, Alexa und Google Home – wie Sprachassistenten die Technikwelt verändern, <https://www.handelsblatt.com/unternehmen/it-medien/elektronikmesse-ifa-siri-alexa-und-google-home-wie-sprachassistenten-die-technikwelt-vera-endern/22971046.html>, aufgerufen am 05.08.2019.

Manning, Christopher D. / Raghavan, Prabhakar / Schütze, Hinrich: Introduction to Information Retriaval, Cambridge 2008.

Mattel (2019): Barbie – Lustige Spiele, Videos und Aktivitäten für Mädchen<https://play.barbie.com/de-de/>, aufgerufen am 05.08.2019.

Microsoft (o.D.): Erstellen einer PivotTable zum Analysieren von Arbeitsblattdaten, <https://support.office.com/de-de/article/erstellen-einer-pivottable-zum-analysieren-von-arbeitsblattdaten-a9a84538-bfe9-40a9-a8e9-f99134456576#OfficeVersion=Windows>, aufgerufen am 05.08.2019.

Minaj, Nicki (o.D): Barbie (@nickiminaj), <https://www.instagram.com/nickiminaj/>, aufgerufen am 05.08.2019.

NetMarketShare (2019): Marktanteile der Suchmaschinen - Mobil und stationär 2019, <https://de.statista.com/statistik/daten/studie/222849/umfr. age/marktanteile-der-suchmaschinen-weltweit/>, aufgerufen am 06.08.2019.

Ohne Autor (o.D): Einführung in JSON< https://www.json.org/json-de.html>, aufgerufen am 06.08.2019.

Panzarino, Matthew (2017) Apple switches from Bing to Google for Siri web search results on iOS and Spotlight on Mac, <https://techcrunch.com/2017/09/25/apple-switches-from-bing-to-google-for-siri-web-search-results-on-ios-and-spotlight-on-mac/ />, aufgerufen am 05.08.2019

Samokhina, Anastasiia (2017): Analysing the systematics of search engine autocompletion functions by means of data mining methods, <https://publiscologne.th-koeln.de/frontdoor/index/index/searchtype/simple/query/%2A%3A%2A/browsing/true/doctypefq/masterthesis/start/0/rows/10/facetNumber_author_facet/all/author_facetfq/Samokhina%2C+Anastasiia/docId/1042>, aufgerufen am 05.08.2019.

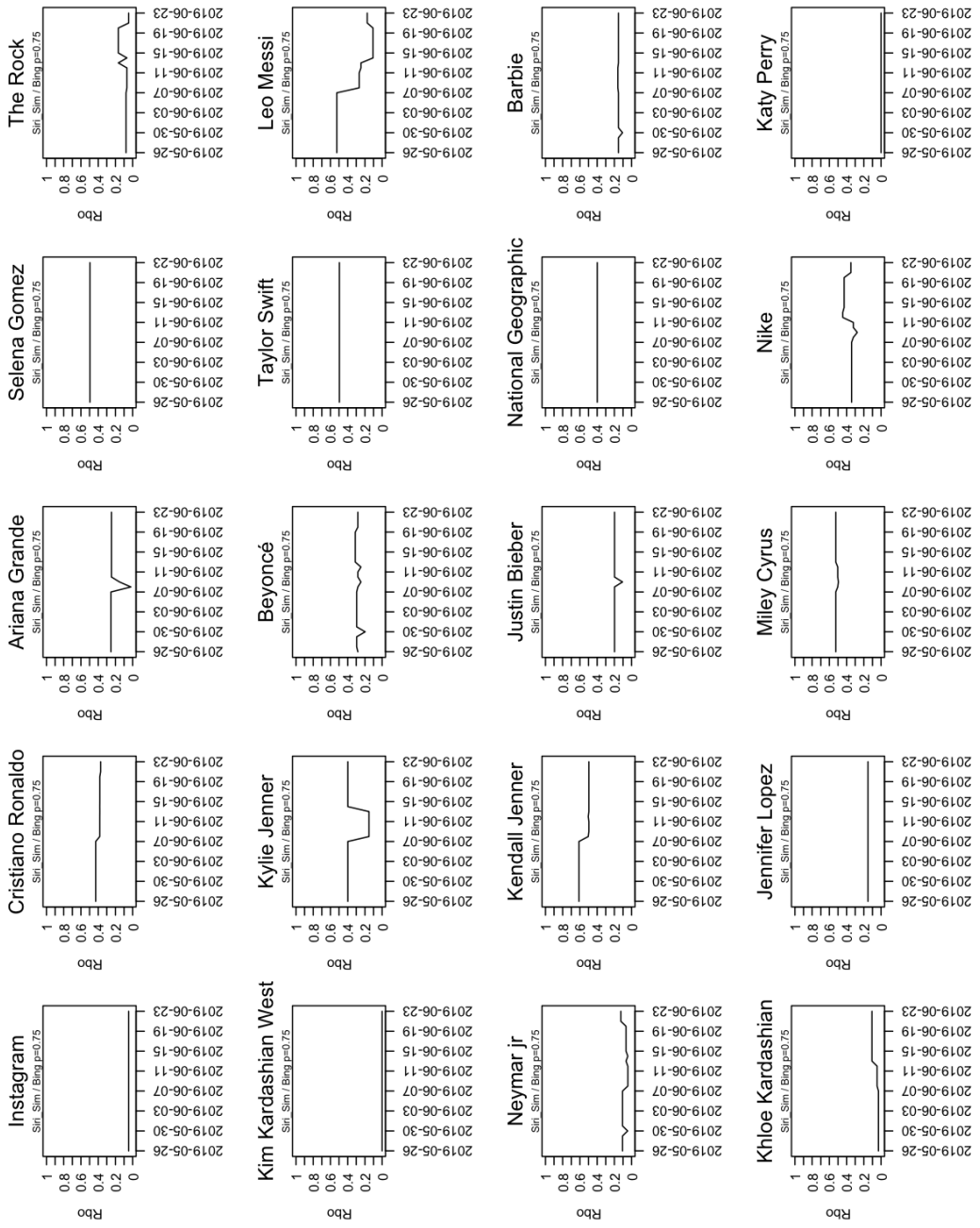
The R Foundation (o.D): What is R? <https://www.r-project.org/about.html>, aufgerufen am 05.08.2019.

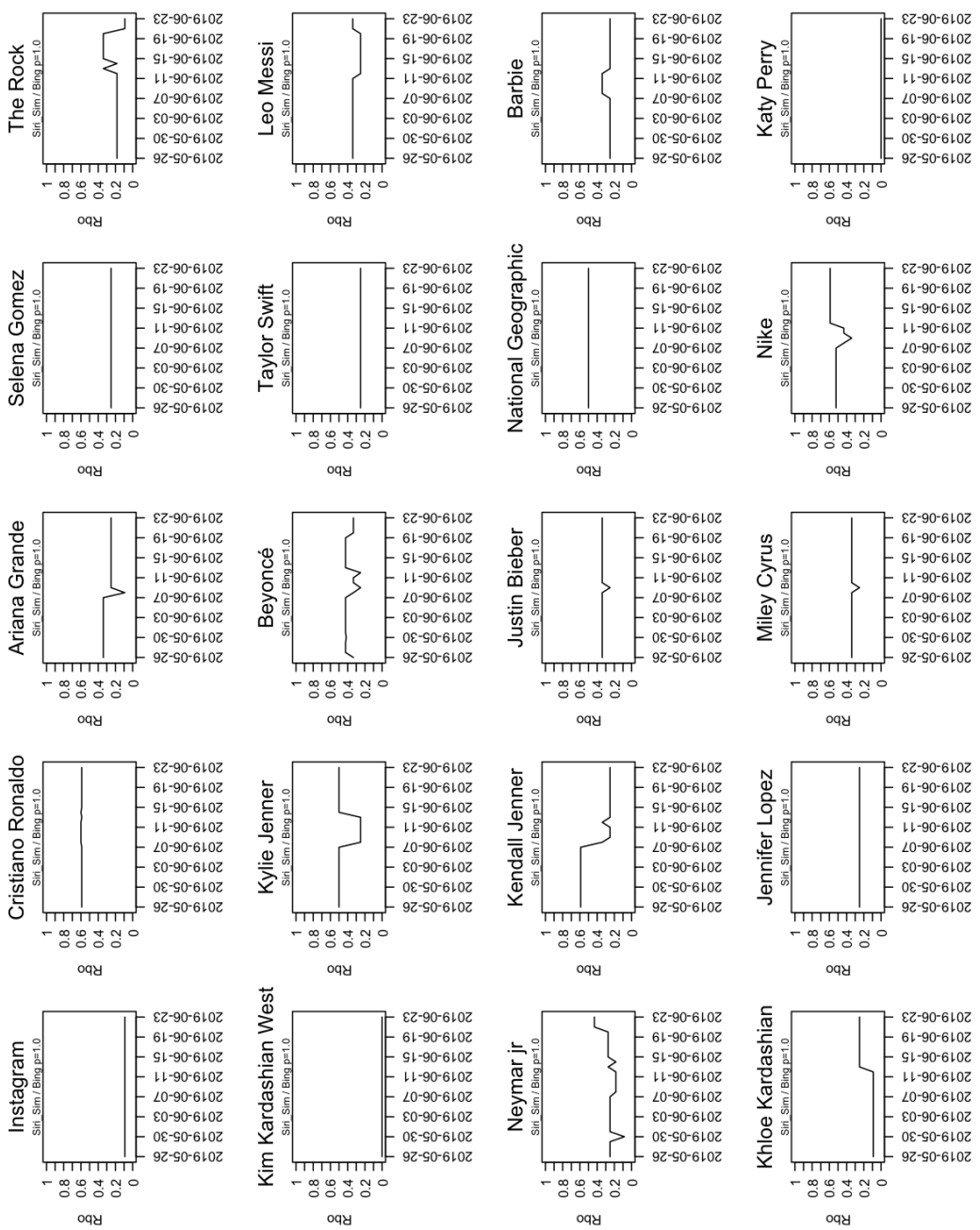
Trackalytics.com (2019): The Most Followed Instagram Profiles, 05.08.2019 <https://www.trackalytics.com/the-most-followed-instagram-profiles/page/1/>, aufgerufen am 05.08.2019.

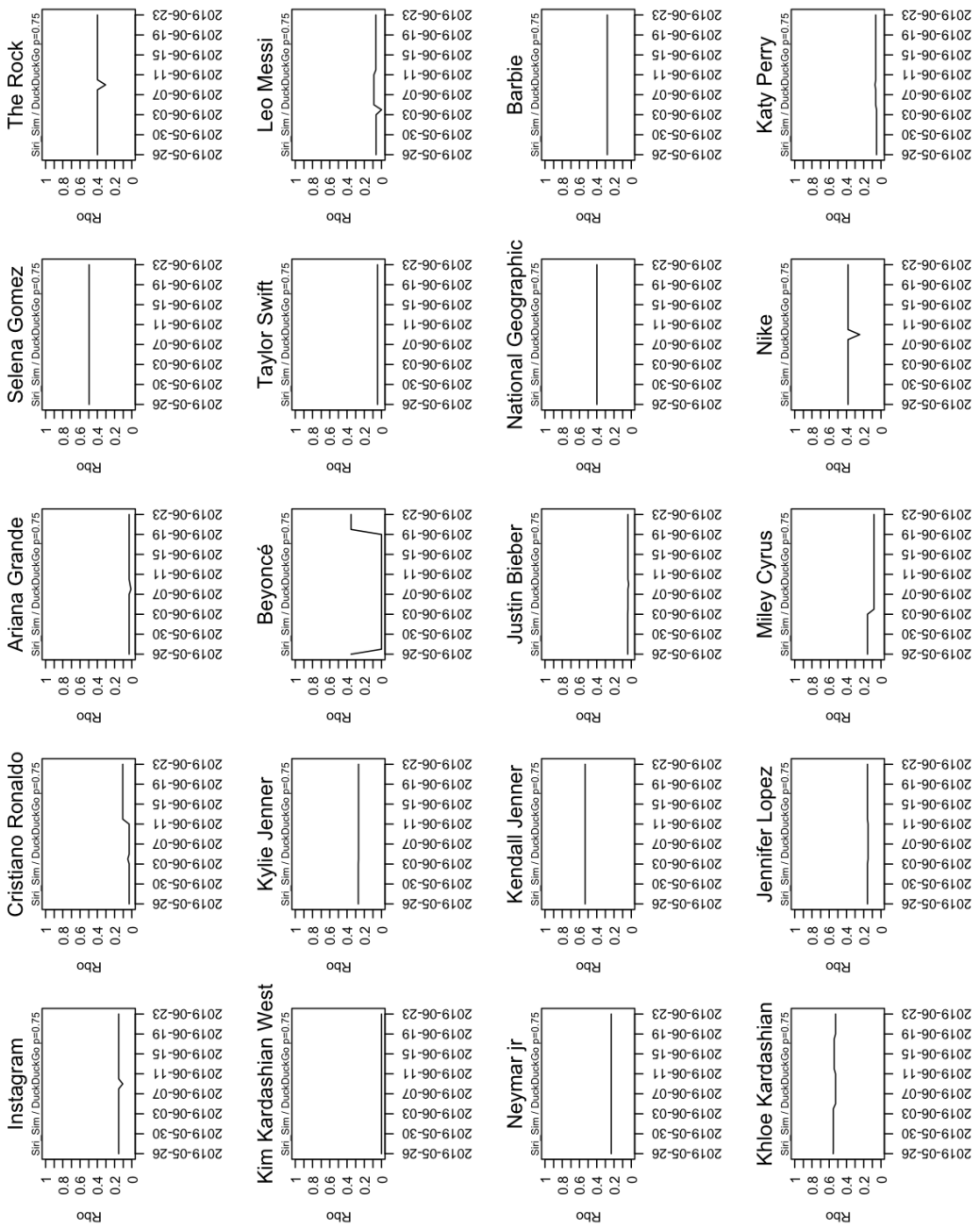
Webber, William / Moffat, Alistair / Zobel, Justin (2010): A similarity measure for indefinite rankings, ACM Transactions on Information Systems , 1–38, 10.1145/1852102.1852106.

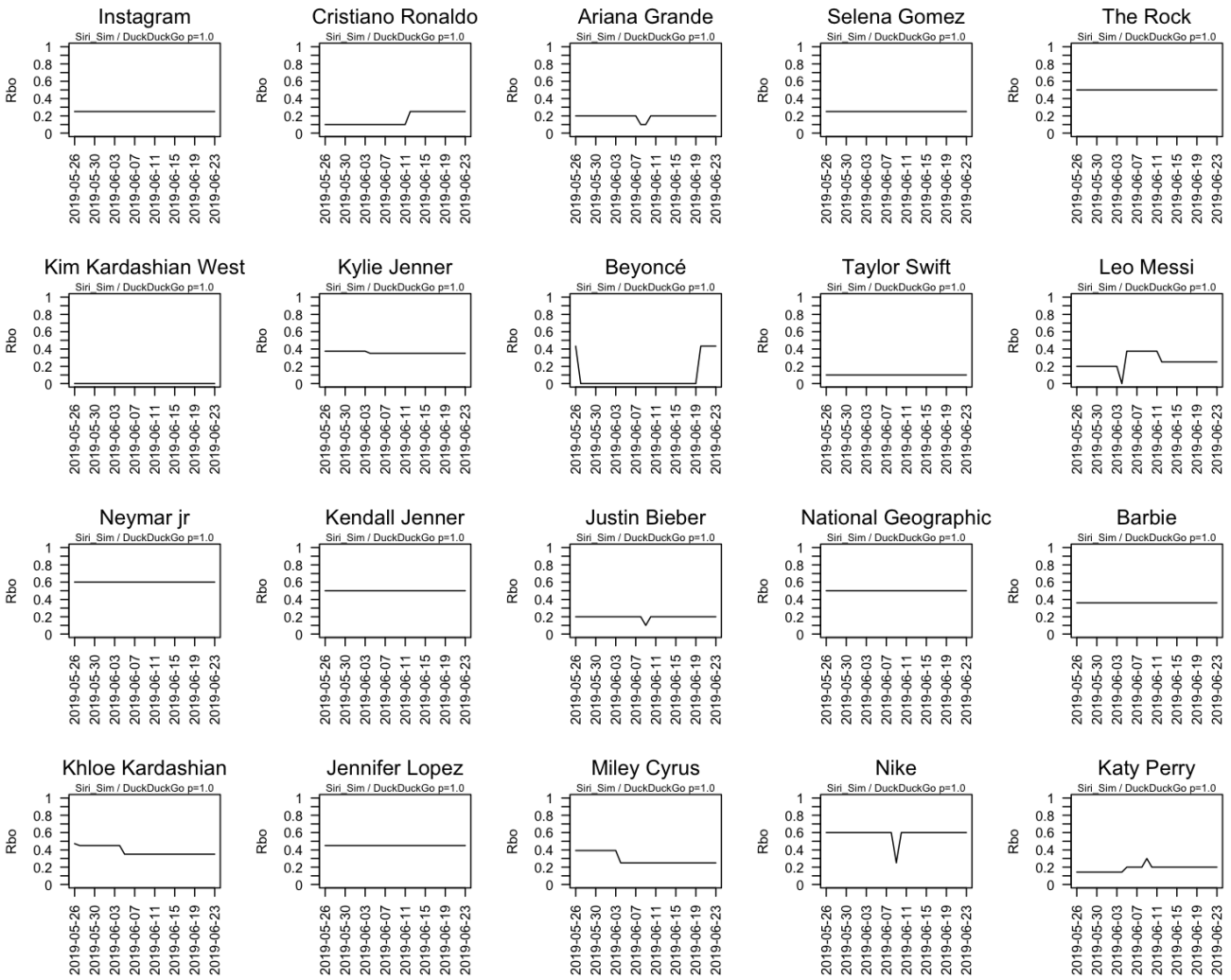
Anhang

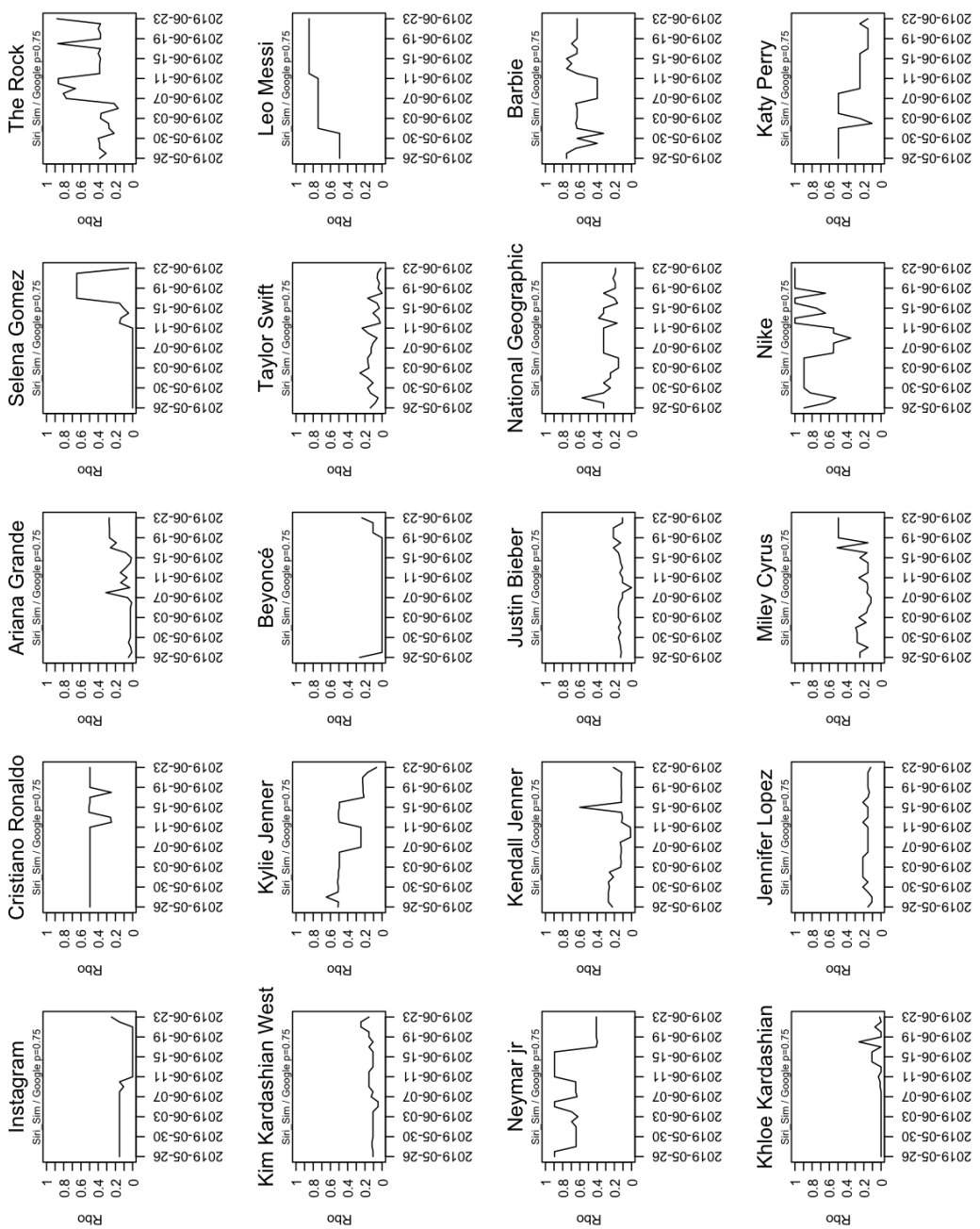
Anhang 1: RBO-Plots mit p = 0.75 und p = 1

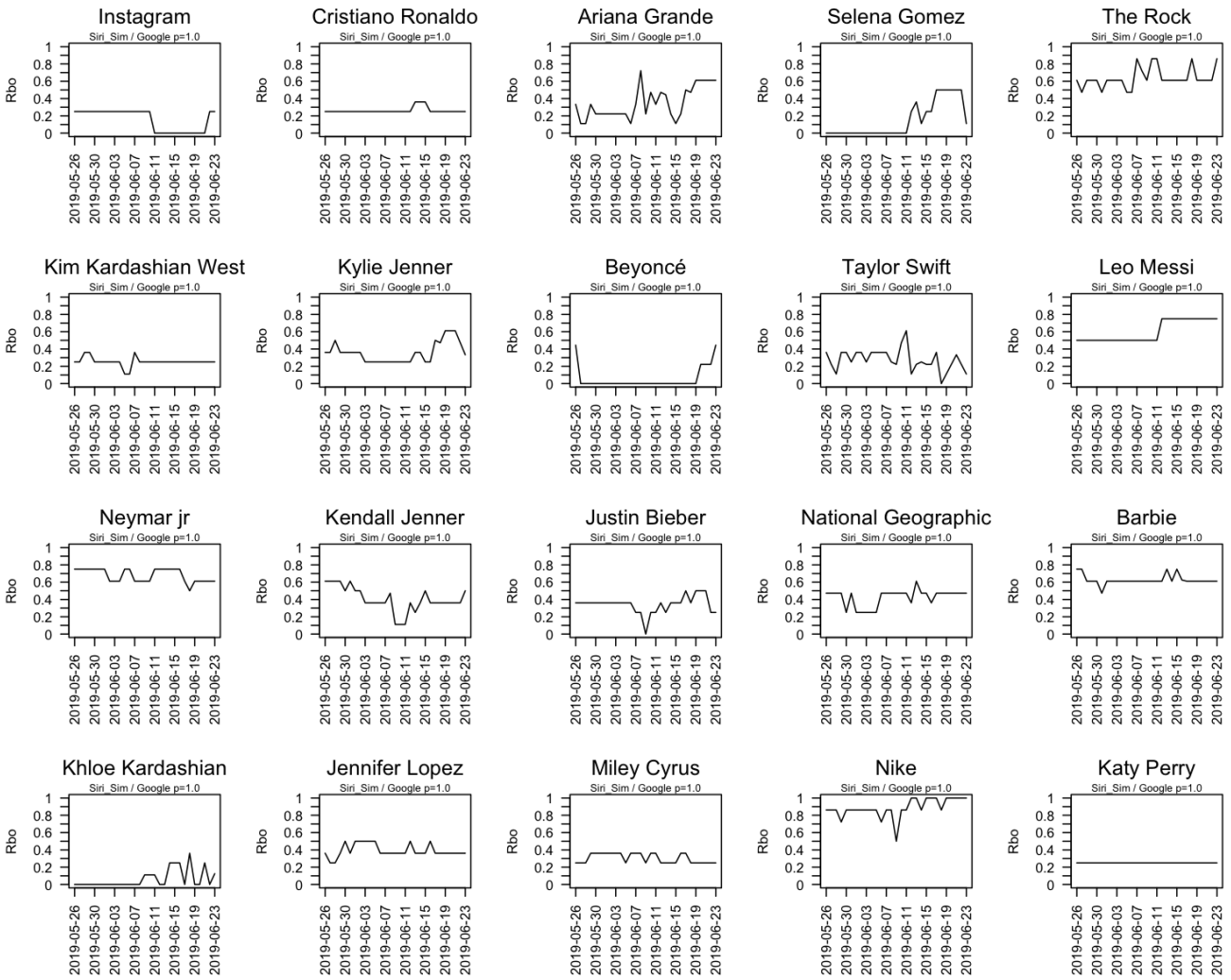


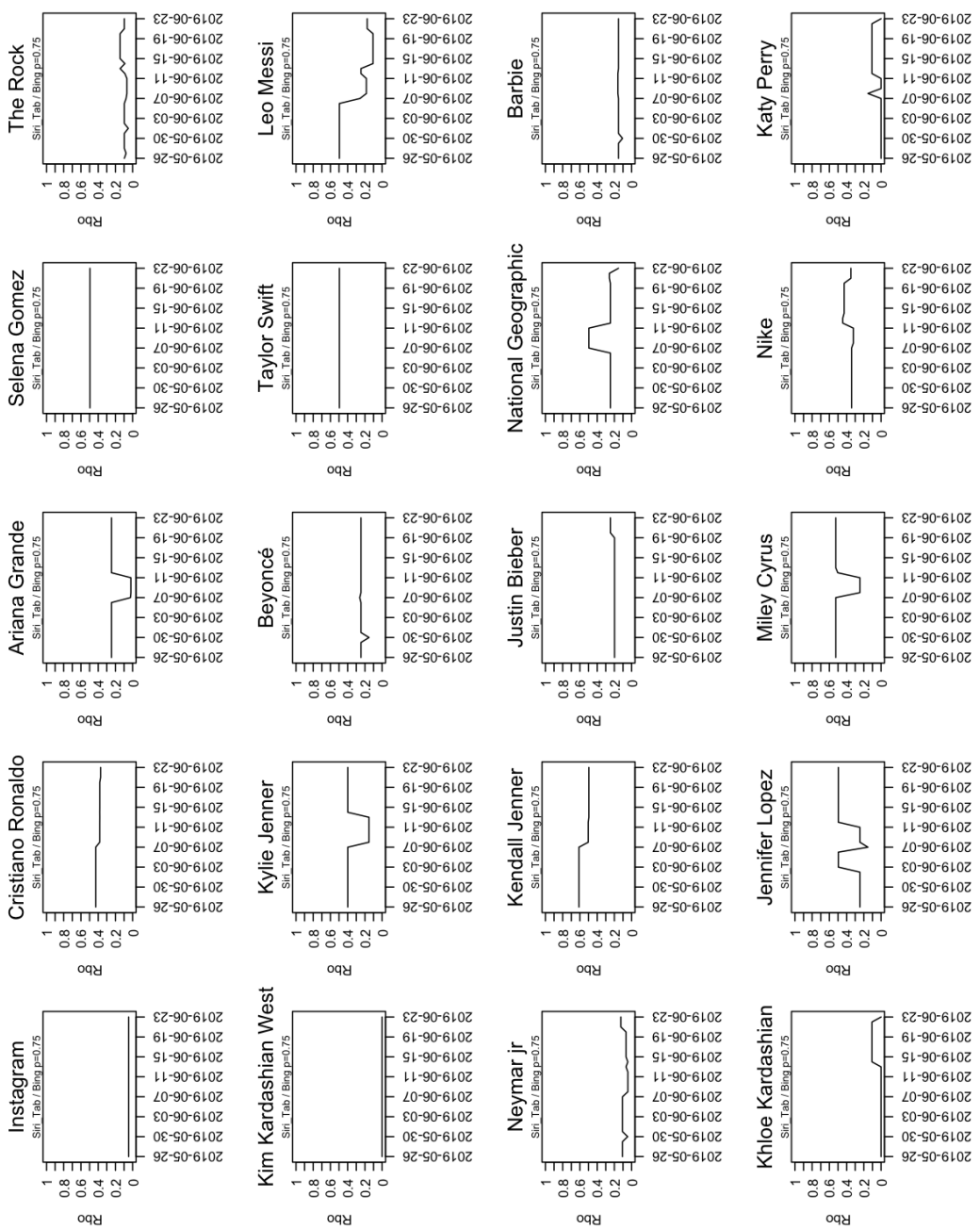

Instagram
Cristiano Ronaldo
Ariana Grande
Selena Gomez
The Rock
Kim Kardashian West
Kylie Jenner
Beyoncé
Taylor Swift
Leo Messi
Neymar jr
Kendall Jenner
Justin Bieber
National Geographic
Barbie
Khloe Kardashian
Jennifer Lopez
Miley Cyrus
Nike
Katy Perry
Siri_Sim / Bing p=1.0
Rbo
2019-05-26
2019-05-30
2019-06-03
2019-06-07
2019-06-11
2019-06-15
2019-06-19
2019-06-23

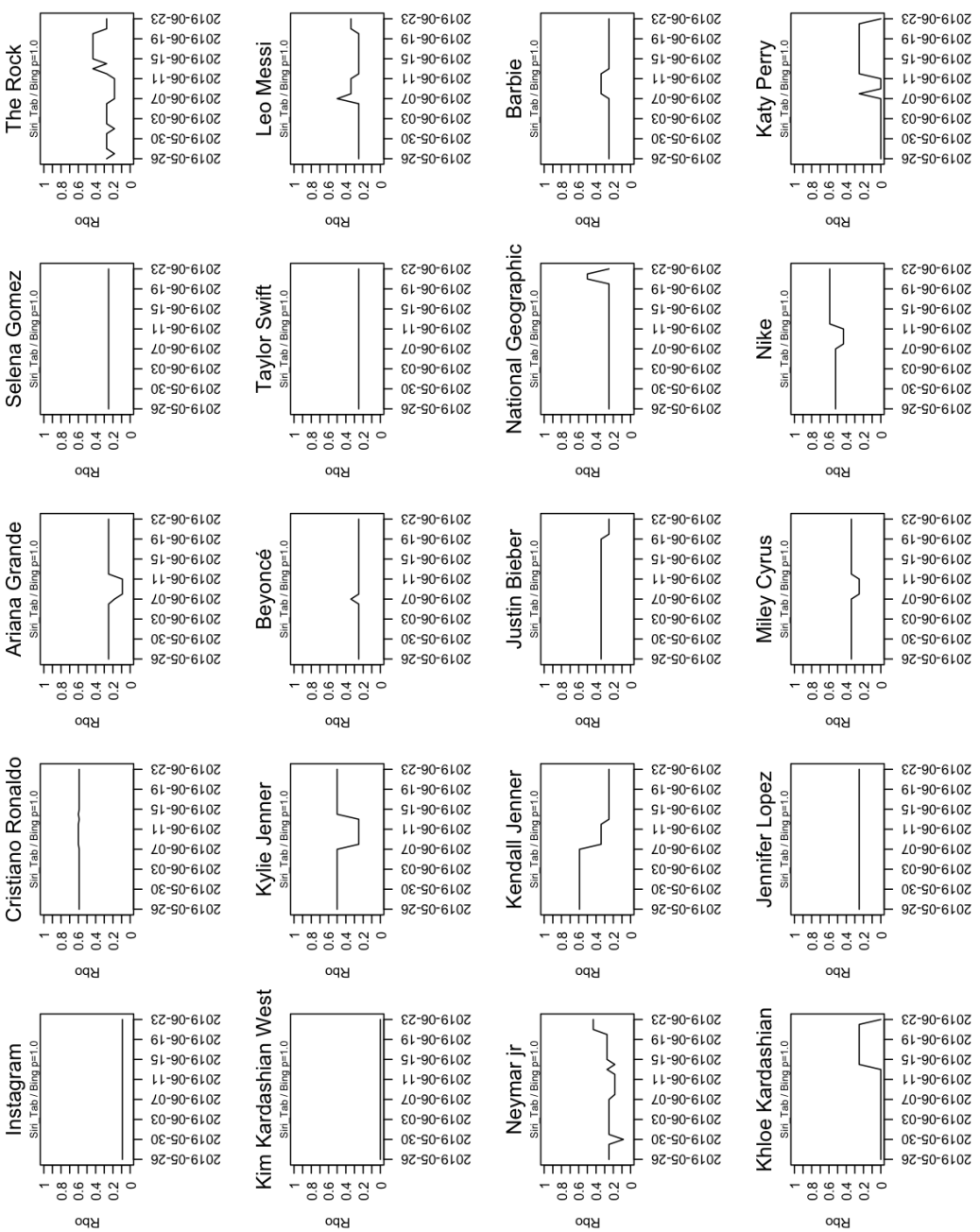

Instagram
Cristiano Ronaldo
Ariana Grande
Selena Gomez
The Rock
Kim Kardashian West
Kylie Jenner
Beyoncé
Taylor Swift
Leo Messi
Neymar jr
Kendall Jenner
Justin Bieber
National Geographic
Barbie
Khloe Kardashian
Jennifer Lopez
Miley Cyrus
Nike
Katy Perry
Siri Sim / DuckDuckGo p=0.75
Rbo
2019-05-26
2019-06-23

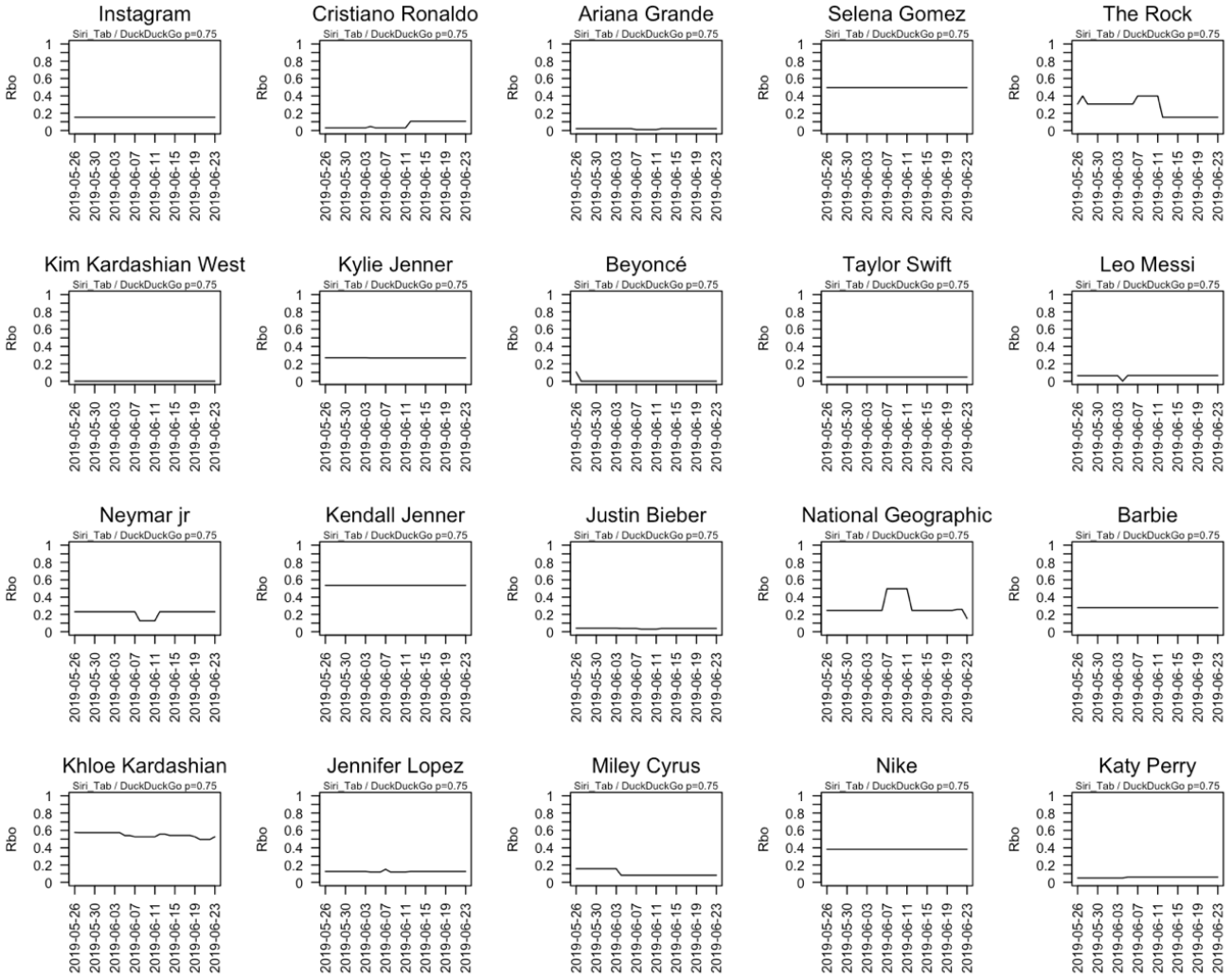

Instagram
Cristiano Ronaldo
Ariana Grande
Selena Gomez
The Rock
Kim Kardashian West
Kylie Jenner
Beyoncé
Taylor Swift
Leo Messi
Neymar jr
Kendall Jenner
Justin Bieber
National Geographic
Barbie
Khloe Kardashian
Jennifer Lopez
Miley Cyrus
Nike
Katy Perry
Siri_Sim / DuckDuckGo p=1.0
Rbo
0
0.2
0.4
0.6
0.8
1
2019-05-26
2019-05-30
2019-06-03
2019-06-07
2019-06-11
2019-06-15
2019-06-19
2019-06-23

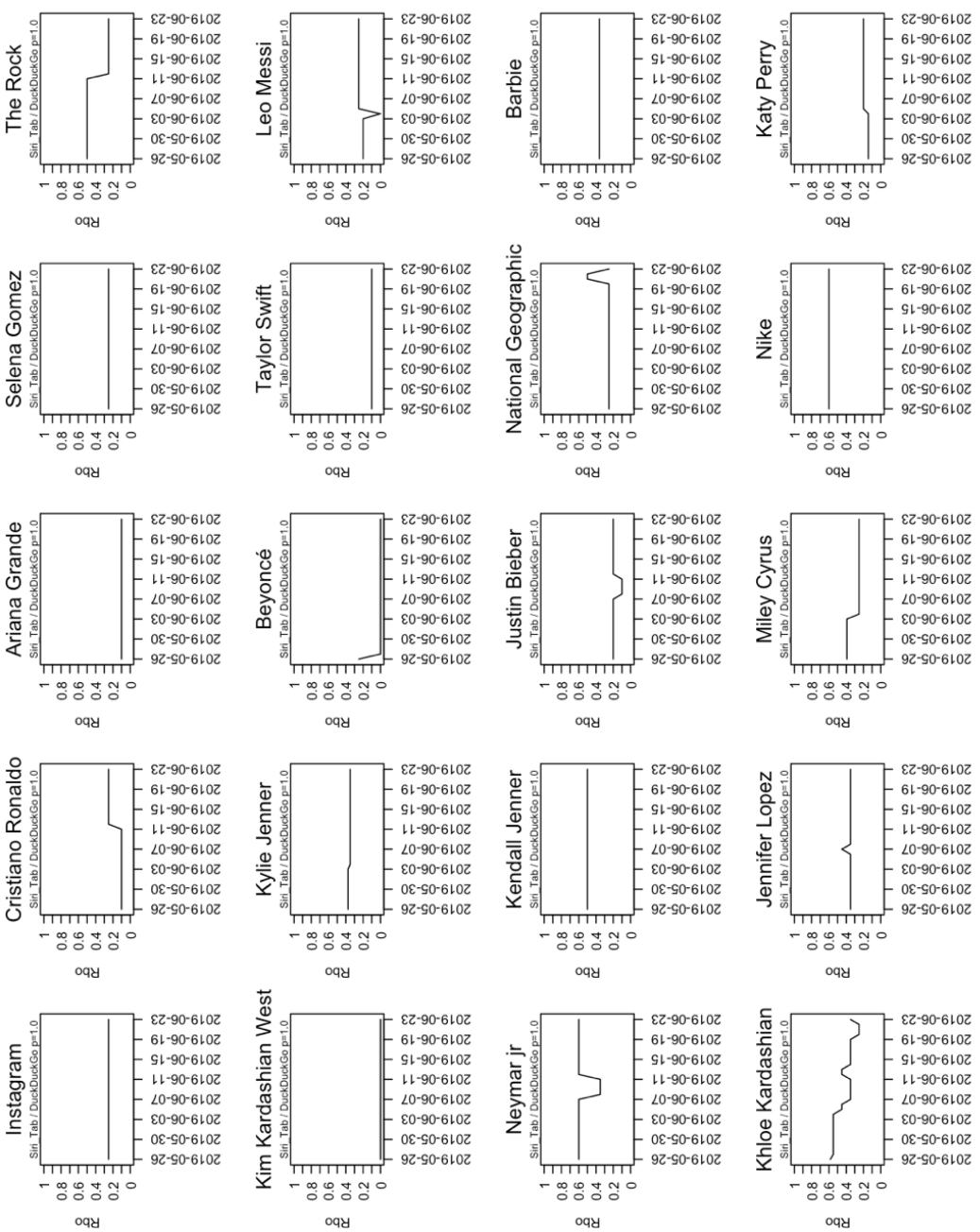

Instagram
Cristiano Ronaldo
Ariana Grande
Selena Gomez
The Rock
Kim Kardashian West
Kylie Jenner
Beyoncé
Taylor Swift
Leo Messi
Neymar jr
Kendall Jenner
Justin Bieber
National Geographic
Barbie
Khloe Kardashian
Jennifer Lopez
Miley Cyrus
Nike
Katy Perry
Siri_Sim / Google p=0.75
Rbo


Instagram
Cristiano Ronaldo
Ariana Grande
Selena Gomez
The Rock
Kim Kardashian West
Kylie Jenner
Beyoncé
Taylor Swift
Leo Messi
Neymar jr
Kendall Jenner
Justin Bieber
National Geographic
Barbie
Khloe Kardashian
Jennifer Lopez
Miley Cyrus
Nike
Katy Perry
Siri_Sim / Google p=1.0
Rbo
0
0.2
0.4
0.6
0.8
1
2019-05-26
2019-05-30
2019-06-03
2019-06-07
2019-06-11
2019-06-15
2019-06-19
2019-06-23

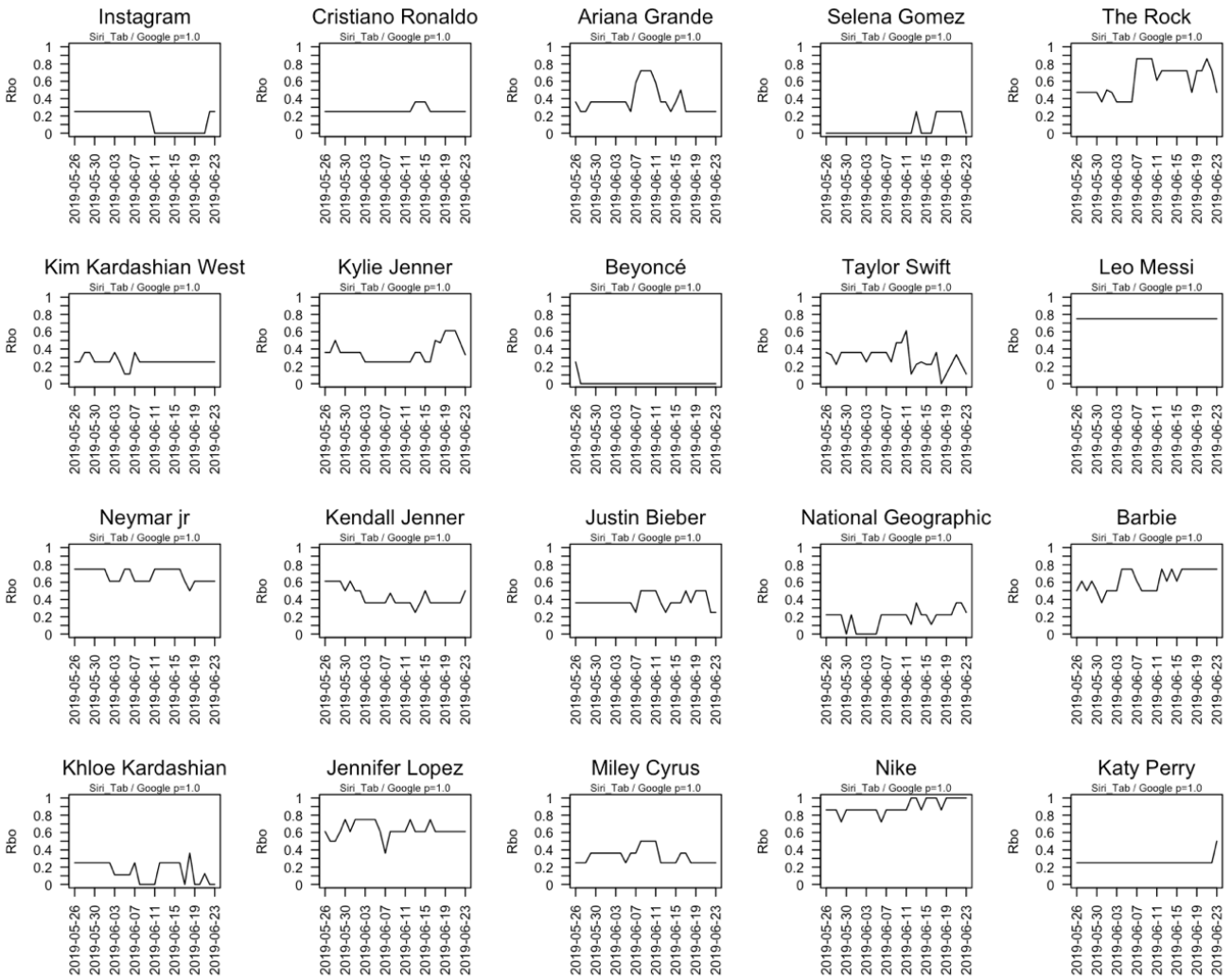

The Rock
Leo Messi
Barbie
Katy Perry
Selena Gomez
Taylor Swift
National Geographic
Nike
Ariana Grande
Beyoncé
Justin Bieber
Miley Cyrus
Cristiano Ronaldo
Kylie Jenner
Kendall Jenner
Jennifer Lopez
Instagram
Kim Kardashian West
Neymar jr
Khloe Kardashian
Siri_Tab / Bing p=0.75
Rbo
2019-05-26
2019-06-23

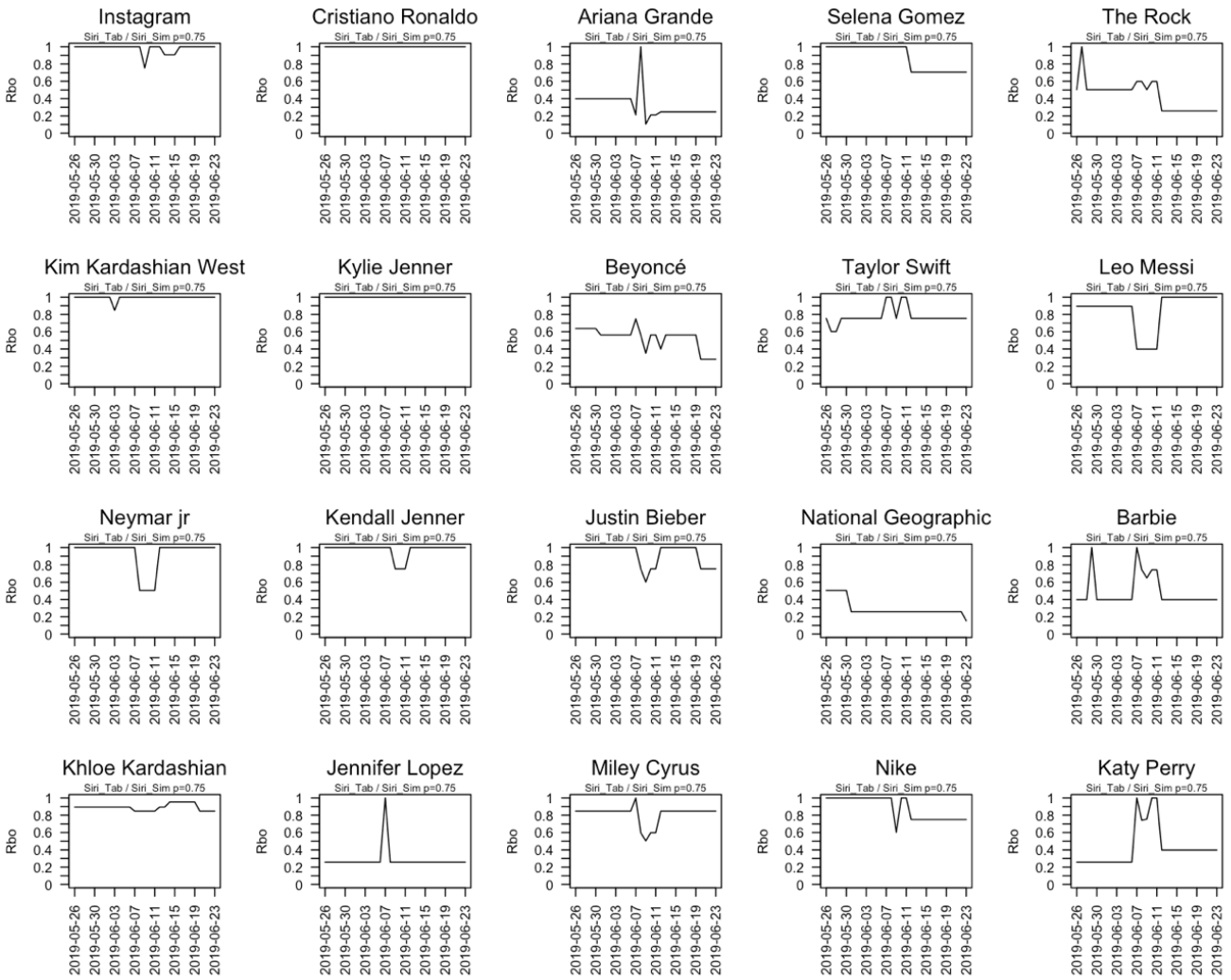

Instagram
Cristiano Ronaldo
Ariana Grande
Selena Gomez
The Rock
Kim Kardashian West
Kylie Jenner
Beyoncé
Taylor Swift
Leo Messi
Neymar jr
Kendall Jenner
Justin Bieber
National Geographic
Barbie
Khloe Kardashian
Jennifer Lopez
Miley Cyrus
Nike
Katy Perry
Sin_Tab / Bing p=1.0
Rbo
2019-05-26
2019-06-23

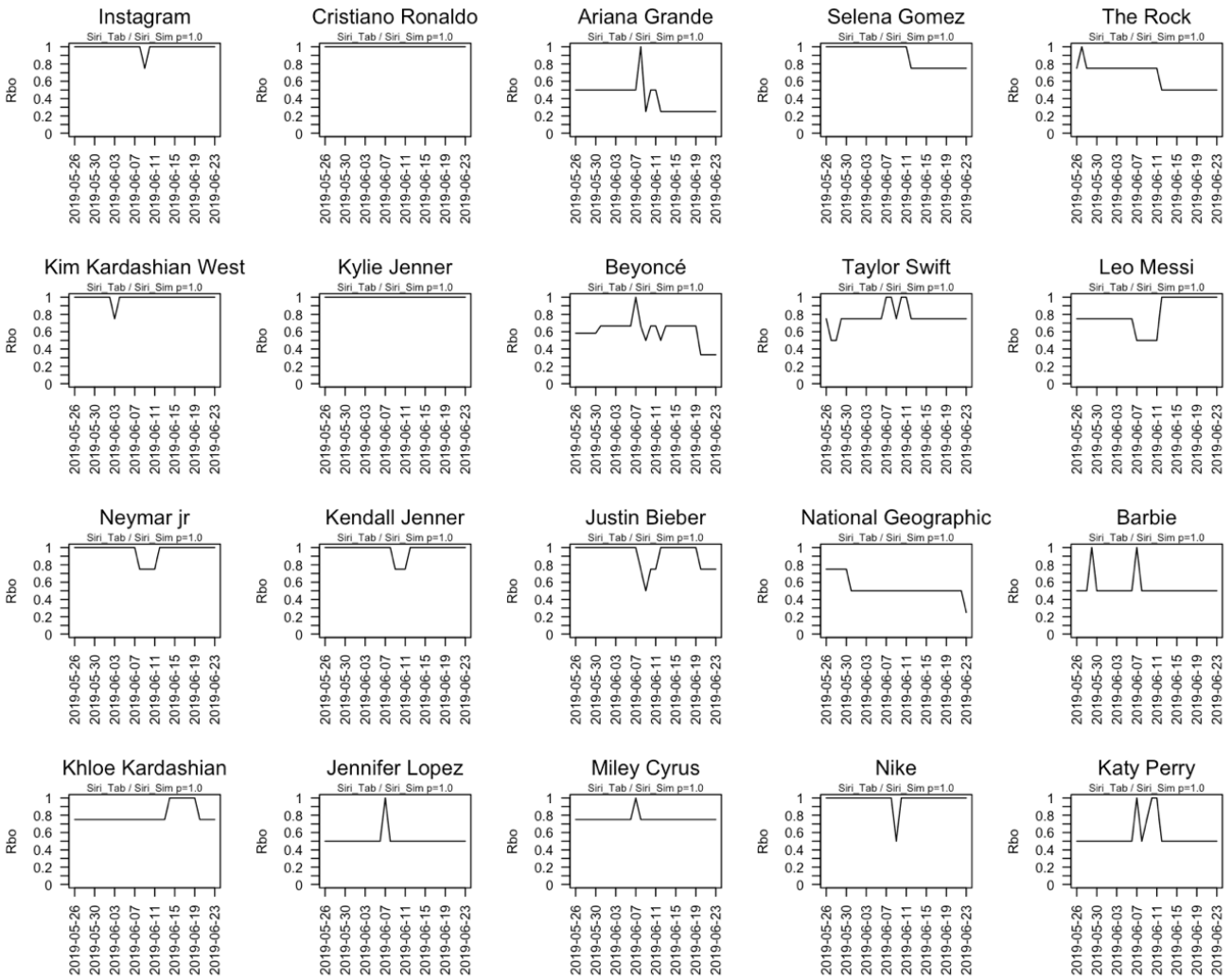

Instagram
Cristiano Ronaldo
Ariana Grande
Selena Gomez
The Rock
Kim Kardashian West
Kylie Jenner
Beyoncé
Taylor Swift
Leo Messi
Neymar jr
Kendall Jenner
Justin Bieber
National Geographic
Barbie
Khloe Kardashian
Jennifer Lopez
Miley Cyrus
Nike
Katy Perry
Siri_Tab / DuckDuckGo p=0.75
Rbo
0
0.2
0.4
0.6
0.8
1
2019-05-26
2019-05-30
2019-06-03
2019-06-07
2019-06-11
2019-06-15
2019-06-19
2019-06-23


Instagram
Siri_Tab / DuckDuckGo p=1.0
Cristiano Ronaldo
Ariana Grande
Selena Gomez
The Rock
Kim Kardashian West
Kylie Jenner
Beyoncé
Taylor Swift
Leo Messi
Neymar jr
Kendall Jenner
Justin Bieber
National Geographic
Barbie
Khloe Kardashian
Jennifer Lopez
Miley Cyrus
Nike
Katy Perry
Rbo
1
0.8
0.6
0.4
0.2
0
2019-05-26
2019-05-30
2019-06-03
2019-06-07
2019-06-11
2019-06-15
2019-06-19
2019-06-23

Instagram
Cristiano Ronaldo
Ariana Grande
Selena Gomez
The Rock
Kim Kardashian West
Kylie Jenner
Beyoncé
Taylor Swift
Leo Messi
Neymar jr
Kendall Jenner
Justin Bieber
National Geographic
Barbie
Khloe Kardashian
Jennifer Lopez
Miley Cyrus
Nike
Katy Perry
Siri_Tab / Google p=0.75
Rbo
2019-05-26
2019-05-30
2019-06-03
2019-06-07
2019-06-11
2019-06-15
2019-06-19
2019-06-23


Instagram
Cristiano Ronaldo
Ariana Grande
Selena Gomez
The Rock
Kim Kardashian West
Kylie Jenner
Beyoncé
Taylor Swift
Leo Messi
Neymar jr
Kendall Jenner
Justin Bieber
National Geographic
Barbie
Khloe Kardashian
Jennifer Lopez
Miley Cyrus
Nike
Katy Perry
Siri_Tab / Google p=1.0
Rbo
0
0.2
0.4
0.6
0.8
1
2019-05-26
2019-05-30
2019-06-03
2019-06-07
2019-06-11
2019-06-15
2019-06-19
2019-06-23


Instagram
Cristiano Ronaldo
Ariana Grande
Selena Gomez
The Rock
Kim Kardashian West
Kylie Jenner
Beyoncé
Taylor Swift
Leo Messi
Neymar jr
Kendall Jenner
Justin Bieber
National Geographic
Barbie
Khloe Kardashian
Jennifer Lopez
Miley Cyrus
Nike
Katy Perry
Siri_Tab / Siri_Sim p=0.75
Rbo
0
0.2
0.4
0.6
0.8
1
2019-05-26
2019-05-30
2019-06-03
2019-06-07
2019-06-11
2019-06-15
2019-06-19
2019-06-23


Instagram
Cristiano Ronaldo
Ariana Grande
Selena Gomez
The Rock
Kim Kardashian West
Kylie Jenner
Beyoncé
Taylor Swift
Leo Messi
Neymar jr
Kendall Jenner
Justin Bieber
National Geographic
Barbie
Khloe Kardashian
Jennifer Lopez
Miley Cyrus
Nike
Katy Perry
Siri_Tab / Siri_Sim p=1.0
Rbo
0
0.2
0.4
0.6
0.8
1
2019-05-26
2019-05-30
2019-06-03
2019-06-07
2019-06-11
2019-06-15
2019-06-19
2019-06-23

Anhang 2: Auflistung der für das Clustering bereinigten Vorschläge

unbereinigt;bereinigt
wiki; wikipedia
insta;instagram
beyonce halo;halo
selena gómez songs;songs
instagram kim kardashian west;instagram
und justin bieber;justin bieber
beyonce instagram;instagram
kim kardashian kanye west;kanye west
beyonce crazy in love;crazy in love
kim kardashian north west;north west
kim kardashian and kanye west;kanye west
beyonce steckbrief;steckbrief
beyonce konzert deutschland 2019;konzert deutschland 2019
selena gómez bilder;bilder
kim kardashian saint west;saint west
beyonce köln;köln
kim kardashian und kanye west;kanye west
beyonce tour 2019;tour 2019
beyonce tour;tour
beyonce jay z köln;jay z köln
beyonce homecoming;homecoming
beyonce feminismus;feminismus
beyonce schwester;schwester
beyonce luxusleben;luxusleben
beyonce größe;größe
beyonce lemonade;lemonade
beyonce solange;solange
kim kardashian west instagram account;instagram account
beyoncé b’day;b'day

Anhang 3: Gebildetes Cluster auf Basis der ermittelten UniqueTerms der Kategorie 1-10

1	2	3	4	5
Social Media/Information	**Körpermerkmale**	**Profession**	**Statussymbole**	**Beziehung/Familie**
instagram	alter	songs	vermögen	freund
wikipedia	größe	filme	gehalt	kinder
steckbrief	früher	trikot	net worth	frau
twitter	tattoo	tour	auto	sohn
bilder	frisur	parfum	haus	schwanger
youtube	ungeschminkt	schuhe	house	baby
news	age	roar	milliardärin	boyfriend
fotos	hair	friends	bugatti	liam hemsworth
snapchat	height	halo	im luxus	hochzeit
facebook	lippen	ain't your mama	autos	orlando bloom
instagram account	vorher nachher	wrecking ball	luxusleben	girlfriend
biography	before after	7 rings	dekadent	tristan thompson
promiflash	haare	shake it off		tristan
crew twitter	acne	firework		jordyn woods
charts twitter	abgenommen	me		vater
snapchat instagram	diet	beauty		freundin
news english	größe und gewicht	365		jr
superiorpics	ops	transfermarkt		familie
website	fat	alben		hailey baldiwn
instagram stories	op	back to you		harry styles
nachrichten	feets	me lyrics		mac miller
updates instagram	konfektionsgröße	cosmetics		justin bieber
twitter account	weight	dark horse		kanye west
fansite	frisieren	on the floor		north west
javi twitter	masse	tour 2019		baby name
app	before	bad blood		liebe
biografie	fitness	love yourself		saint west
celebmafia	operationen	style		kind
pics	tattoos	blank space		tristan thompson beziehung
fakten	ernährung figur	mode		pete davidson
steckbrief englisch	diät	lip kit		mann
leben	früher heute	sorry		wedding
tumblr	old	problem		boyfriends list
video	lips	psg		ben simmons
lebenslauf		focus		schwester
		listen		family pic
		wrestler		alex rodriguez
		juventus		father
		single ladies		junior

ed sheeran song

song

zayn

wechsel

chained

rise

she is coming

aint your mama

skills

crazy in love

make up

makeup

konzert

if i were a boy

nothing breaks like a heart

lemonade

i am… sasha fierce

nothing breaks like a heart text

365 mp3 download

part of me stream

lipstick

reputation tour

jay z köln

musikkarriere

never really over

amor

konzert 2019 deutschland

chained to the rhythm

black mirror

medicine

me songtext

good for you

365 lyrics

rise andreid remix mp3 download

filme die besten 10

shake it off analyse

coachella

marktwert

unconditionally

roar deutsche übersetzung

konzert 2019

new album

erfolge

tristan jordyn

tochter

solange

orlando

hailey baldwin

travis scott

zedd

daughter

partner

hailey

freundin 2019

lamar

mother's daughter

tickets

teenage dream

alle songs

lover

bon appetit

deutschland tour 2019

real madrid

lieder

konzert deutschland 2019

thank you next lyrics

konzert köln

lippenstift

new song

never really over lyrics

black mirror song

schuhe astrea silber

neues album

you need calm down

s reputation

swish swish

tickets köln

reputation

album

nothing breaks like a heart chords

lyrics

beyoncetour 2019

lovestory

verletzung

calm down

you need to calm down

i can't get enough

imagines he m

louvre song

dream lyrics

friends

transfer

7 rings lyrics

homecoming

6	7	8	9	10
Nackheit/Tod	**Orte/Sprache**	**Sonstiges**	**Produkte**	**Services**
tot	karlsruhe	2016	zeitschrift	suche
po	zanzibar	2019	your shot	story viewer
nackt	malibu	2015	world	störung
unten ohne	traunreut	2017	wild	store kerpen
hot	cannes	shop	vapormax	store
nacktvideo	bremen	merch	und die 12 tanzenden prinzessin stream	snkrs
hintern bilder	helgoland	y horror picture show	und das geheimnis von oceana stream	shop
schwanzbild	antalya	film	und das diamantschloss stream deutsch	programm heute
skandalfotos	portugal	beyonce	traveler	programm
po implantate	köln	2018	traumvilla abenteuer	profilbild zoom
krankheit	deutschland	knowles	teleskop	profilbild vergrößern
krank	deutsch	hund	story	profilbild größe
vergewaltigung anzeige		johnson	spiele 1001	outlet kerpen
		wallpaper	spiele	outlet
		beyonce knowles	society	online shop
		2014	shoes	mediathek
		fels der entscheidung	schwanensee	tv
		shakira	schuhe damen	löschen
		y horror show	schuhe	login
		law school	rucksack	konto löschen
		di gaspare instagram	puzzle	id
		trainingsplan	puppen	hashtags
		autounfall vor haus	prinzessinnen akademie stream deutsch	follower kaufen
		diet plan	presto	follower free
		geständnis	pferd	follower
		training	meerjungfrau	filter
		selena gomez	magazine	download
		ng	koffer	down
		pepsi	kids deutsch	deutschland live stream
		statue	kids	channel
		krass schule	kalender 2019	bilder
		carolina lemke	katie's face	anmelden
		puma	katie's	abo
		stream	kalender	registrieren
		kopfhörer	internationalist	
		7 live stream	huarache	
		spiele	haus	
		rekorde	grabeskirche	
		di gaspare	games	

bettwäsche	free
taylor swift	filme deutsch
complex 2015	filme
katy perry	fairytopia stream deutsch
headphones	doku
katzen	camper
ed sheeran	bücher
chat	auto
h&m	ausmalbilder
hillsong	air max 97
sonnenbrille	air max 270
meme	air max
gntm	air force 97
met 2018	air force 1
wallpaper hd	air force
y horror show neunkirchen	air
eteer	account löschen
teste dich	account
eminem	videos
bill murray	720
tom cruise	270
met gala 2019	reisen
jeans	
nicki minaj	
fanartikel	
inglot	
tochter singt	
handmaid's tale	
beyonce&ie	
knwoles	
beyoncé b’day	
1996	
petition	
präsentation	
feminismus	
dwayne johnson	
et man	
tn	
stories heimlic an-schauen	
sprüche	
sdp	
schrift	
sb	

s

ro

re el figaro

re da roberto

r von sevilla lugano

r von sevilla

r trier

r simmern

r köln

r koblenz

r eric hamburg

r emad

r bonn

r

praktikum

nickelodeon

logo

lena meyer landrut

kostüm

kelly

heidi klum

girl

friesennerz

erkencikus

droge

dolunay

dieter bohlen

daniela katzenberger

bibisbeautypalace

bella kraus

aktie

/jayalvarrez

.de

.com

barbie

national geographic

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Dies gilt auch für Quellen aus eigenen Arbeiten.

Ich versichere, dass ich diese Arbeit oder nicht zitierte Teile daraus vorher nicht in einem anderen Prüfungsverfahren eingereicht habe.

Mir ist bekannt, dass meine Arbeit zum Zwecke eines Plagiatsabgleichs mittels einer Plagiatserkennungssoftware auf ungekennzeichnete Übernahme von fremdem geistigen Eigentum überprüft werden kann.

Ort, Datum	Rechtsverbindliche Unterschrift