

Generische Webwissenschaften in der Praxis – Prüfung und Reflexion von Forschungs- methoden und -instrumenten

Masterarbeit

Markt- und Medienforschung 13

Fakultät für Informations- und Kommunikationswissenschaften

Institut für Informationswissenschaft

Technische Hochschule Köln

vorgelegt von:

Stephan Jülich

am 20.11.2015 bei: Prof. Dr. disc. pol., M.A. Helmut Volpers
M.Sc. Miriam Schmitz

Abstract

The following paper is based on the field of web science. More precisely, it builds upon a web-native research method created by Richard Rogers, called “search as research”. This paper explains how to gather data “born” in the web, via web search engines. It aims to show, that culture is reflected in the use of the web, particularly employing *Google*, to develop various methods, in regard to gather the relevant data. With these different approaches of the method, data representative for 20 European countries are generated, explored and evaluated. Therefore representative studies of European Institutions (European Monitoring Centre for Drugs and Drug Addiction – “European Drug Report 2015”, European Values Project – “European Values Study”, World Health Organisation – “Global status report on alcohol and health”) are used to compare the outcomes and evaluate the reliability and validity of the data generated via *Google*. Further it aims to show, that based on this data, it is possible to conduct a cultural comparison of countries. While the comparison of the representative studies with the *Google* data shows, that *Google* does not necessarily reflect a culture in line with the above mentioned studies, it yet clearly can be shown, that with the use of different methods, cultural differences between countries can be gathered and evaluated. Therefore some of the methods provide a possibility to uplift a first glimpse when comparing different countries on a cultural basis and may for instance be helpful to enhance the quality when constructing questionnaires. Nevertheless it is also shown, that further research is needed to improve the developed methods, especially in regards to improve the data quality on basis of reliability and validity.

Inhaltsverzeichnis

Abbildungsverzeichnis	V
Tabellenverzeichnis	VII
Abkürzungsverzeichnis	VIII
1 Einleitung	1
2 Theorie.....	7
2.1 Status Quo – Die Forschung im Web	7
2.1.1 Netnographie – Methoden zur Analyse der Online Gemeinschaft.....	7
2.1.2 Digital Methods – Das Web als Datenquelle.....	8
2.1.3 Die Webwissenschaften – Das Web verstehen	12
2.2 Analysemethode & Operationalisierung.....	15
2.2.1 Online-Inhaltsanalyse	15
2.2.2 Erhebung des Datenmaterials bei dynamischen Web-Inhalten	17
2.2.3 Suchmaschinen – Indexierung von Web-Inhalten.....	18
2.2.4 Virtual Private Network und Tunnelung	23
2.3 Soziologie: Validierung der Daten.....	26
2.3.1 Kulturvergleichende Forschung	27
2.3.2 Vergleichsdaten	28
3 Praxis.....	30
3.1 Generierung des Datenmaterials	30
3.2 Stichprobenziehung und Analyseeinheiten.....	33
3.3 Kodier-Prozess & Bildung der Indices auf Basis von Google.....	35
4 Auswertung der Google Indices.....	41
4.1 Index: „Homophobie“ nach Google	41
4.1.1 Vergleichsdaten: Homophobie in der EU nach EVS-Index	41
4.1.2 Homophobie auf Basis der Google Suchvervollständigung.....	44
4.1.3 Homophobie auf Basis der Google Textsuche (TOP 100).....	49
4.1.4 Homophobie auf Basis der Indexierung von Google Textsuche und Suchvervollständigung (TOP 100 & DMI).....	52
4.1.5 Relevante Themenkreise zur Homophobie nach Google Textsuche.....	54
4.2 Index: „Ausländerfeindlichkeit“ nach Google.....	57
4.2.1 Vergleichsdaten: Ausländerfeindlichkeit in der EU nach EVS-Index	57
4.2.2 Ausländerfeindlichkeit auf Basis der Suchvervollständigung mit Google Autocomplete	59
4.2.3 Ausländerfeindlichkeit auf Basis der Google Textsuche	61

4.2.4	Relevante Themenkreise zur Ausländerfeindlichkeit nach Google Textsuche	64
4.3	Index: „Toleranz von Rauschmitteln“ nach Google	67
4.3.1	Vergleichsdaten: Toleranz von Rauschmitteln (WHO / EMCDDA)	67
4.3.2	Toleranz von Rauschmitteln auf Basis der Google Suchervollständigung	69
4.3.3	Toleranz von Rauschmitteln auf Basis der Google Textsuche (TOP 100)	70
5	Exkurs: Analyse der Google Bildersuche	73
5.1	Homophobie auf Basis der Google Bildersuche	74
5.2	Ausländerfeindlichkeit auf Basis der Google Bildersuche	76
6	Ergebnisse: (K)Eine Abbildung der Gesellschaft mit Hilfe von Google	79
7	Fazit	83
	Literaturverzeichnis	IX
	Anhang I	XIII
	Anhang II - Codebuch	XXVI
	Eidesstattliche Erklärung	XXXIX

Abbildungsverzeichnis

Abbildung 1-1 – (links) Anzahl der Suchanfragen bei Google weltweit in den Jahren 2007 bis 2014 (in Milliarden), (rechts) Suchmaschinen Marktanteile	4
Abbildung 2-1 - Die Struktur des Webs	20
Abbildung 2-2 - Werbeanzeigen vs. Web-Suche google.de, Suchbegriff "Schuhe"	22
Abbildung 2-3 - Suchvervollständigung google.de, Suchbegriff "Köln"	22
Abbildung 2-4 - Internetverbindung ohne VPN Server	24
Abbildung 2-5 – Internetverbindung mit VPN Server	25
Abbildung 3-1- Beispiel 4 Analyseeinheiten Textsuche bei google.de zum Thema Homosexualität	33
Abbildung 3-2 - Beispiel 4 Analyseeinheiten Suchvervollständigung bei google.de zum Thema Homosexualität	34
Abbildung 3-3 - Beispiel 10 Analyseeinheiten Suchvervollständigung via Google Autocomplete (DMI) zum Thema Homosexualität	34
Abbildung 3-4 - Beispiel 1 Kodierung der Textsuche zum Thema Homosexualität	37
Abbildung 3-5 - Beispiel 2 Kodierung der Textsuche zum Thema Homosexualität	38
Abbildung 3-6 - Beispiel 3 Kodierung der Textsuche zum Thema Homosexualität	38
Abbildung 4-1 - Gesamtindex für Homophobie je Land nach EVS	42
Abbildung 4-2 - Gesamtindex Homophobie je Land: Google Suchvervollständigung, manuelle Suche	44
Abbildung 4-3 - EVS Index Homophobie vs. Homophobie Index manuelle Suchvervollständigung	45
Abbildung 4-4 - Gesamtindex Homophobie je Land: Google Suchvervollständigung, DMI Tool	47
Abbildung 4-5 - EVS Index Homophobie vs. Homophobie Index der Google Suchvervollständigung mittels Google Autocomplete	48
Abbildung 4-6 - Themenebene Homophobie Suchvervollständigung	49
Abbildung 4-7 - Gesamtindex für Homophobie je Land nach Google Textsuche	50
Abbildung 4-8 - EVS Index Homophobie vs. Homophobie Index der Google Textsuche	51
Abbildung 4-9 - Gesamtindex für Homophobie je Land nach Google Index aus Textsuche & Suchvervollständigung	52
Abbildung 4-10 - EVS Index Homophobie vs. Homophobie Index der Google Textsuche und Suchvervollständigung	53
Abbildung 4-11 – links: Themenebene Homophobie Textsuche, rechts: relative Anzahl relevanter Analyseeinheiten je Land	54
Abbildung 4-12 - Homophobie auf Basis der Google Textsuche im Themenbereich Religion	55
Abbildung 4-13 - Homophobie auf Basis der Google Textsuche im Themenbereich Gesellschaft	55
Abbildung 4-14 - Homophobie auf Basis der Google Textsuche im Themenbereich Krankheit	56
Abbildung 4-15 - Homophobie auf Basis der Google Textsuche im Themenbereich Gewalt	57
Abbildung 4-16 - Gesamtindex für Ausländerfeindlichkeit je Land nach EVS	58
Abbildung 4-17 - Gesamtindex für Ausländerfeindlichkeit je Land nach Google Suchvervollständigung mittels Google Autocomplete	60
Abbildung 4-18 - EVS Index Ausländerfeindlichkeit vs. Index Ausländerfeindlichkeit Google Suchvervollständigung mittels Google Autocomplete	61
Abbildung 4-19 - Gesamtindex für Ausländerfeindlichkeit je Land nach Google Textsuche	62
Abbildung 4-20 - Index Ausländerfeindlichkeit EVS vs. Index Ausländerfeindlichkeit Google Textsuche	62
Abbildung 4-21 - links: Themenebene Ausländerfeindlichkeit Textsuche, rechts: relative Anzahl relevanter Analyseeinheiten nach Standortbezug je Land	64
Abbildung 4-22 - Ausländerfeindlichkeit auf Basis der Google Textsuche im Themenbereich Gesellschaft ..	65
Abbildung 4-23 - Ausländerfeindlichkeit auf Basis der Google Textsuche im Themenbereich Politik	65
Abbildung 4-24 - Ausländerfeindlichkeit auf Basis der Google Textsuche im Themenbereich Gewalt	66
Abbildung 4-25 - Toleranz von Alkohol auf Basis der getrunkenen Menge	67
Abbildung 4-26 - Toleranz von Marihuana auf Basis Anzahl der Konsumenten je Land	68
Abbildung 4-27 - WHO Index Toleranz von Alkohol vs. Toleranz von Alkohol auf Basis der Google Textsuche	70
Abbildung 4-28 - EMCDDA Index Toleranz von Marihuana vs. Toleranz von Marihuana auf Basis der Google Textsuche	72
Abbildung 5-1 - Gesamtindex für Homophobie je Land nach Google Bildersuche	75
Abbildung 5-2- Gesamtindex für Ausländerfeindlichkeit je Land nach Google Bildersuche	77
Abbildung A 1 - European Values Study 1981 - 2008: Participating countries	XIII
Abbildung A 2 - Landessprachen in den Gemeinden, 2000 (Schweiz)	XIV
Abbildung A 3 - Setup Google Autocomplete (DMI), Google Landesversion Deutschland	XV

Abbildung A 4- Homophobie Index EVS vs. Google Suchvervollständigung (manuell).....XVIII
Abbildung A 5 - Homophobie Index EVS vs. Google Suchvervollständigung (Google Autocomplete).....XVIII
Abbildung A 6 - Homophobie Index EVS vs. Google Textsuche..... XIX

Tabellenverzeichnis

<i>Tabelle 3-1 - Übersicht der genutzten Serverstandorte, ccTLDs und Such-Sprachen.....</i>	<i>31</i>
<i>Tabelle 4-1 - Korrelation der Google Indices zum EVS Index für Homophobie (SPSS).....</i>	<i>43</i>
<i>Tabelle 4-2 - Zuordnungsübersicht Cluster 1-3 EVS vs. Google manuelle Suchvollständigkeit - Homophobie</i>	<i>46</i>
<i>Tabelle 4-3 - Zuordnungsübersicht Cluster 1-3 EVS vs. Google Suchvollständigkeit mit Google Autocomplete - Homophobie</i>	<i>48</i>
<i>Tabelle 4-4 - Zuordnungsübersicht Cluster 1-3 EVS vs. Google Textsuche - Homophobie.....</i>	<i>51</i>
<i>Tabelle 4-5 - Zuordnungsübersicht Cluster 1-3 EVS vs. Google Textsuche & Suchvollständigkeit – Homophobie</i>	<i>53</i>
<i>Tabelle 4-6 - Korrelation der Google Indices zum EVS Index für Ausländerfeindlichkeit.....</i>	<i>59</i>
<i>Tabelle 4-7 - Zuordnungsübersicht Cluster 1-3 EVS vs. Google Suchvollständigkeit - Ausländerfeindlichkeit</i>	<i>61</i>
<i>Tabelle 4-8 - Zuordnungsübersicht Cluster 1-3 EVS vs. Google Textsuche - Ausländerfeindlichkeit</i>	<i>63</i>
<i>Tabelle 4-9 - Korrelation der Google Indices zum WHO Index für Alkohol.....</i>	<i>69</i>
<i>Tabelle 4-10 - Korrelation der Google Indices zum EMDCCA Index für Marihuana</i>	<i>69</i>
<i>Tabelle 4-11 - Zuordnungsübersicht Cluster 1-3 EVS vs. Google Textsuche – Toleranz von Alkohol.....</i>	<i>71</i>
<i>Tabelle 4-12 - Zuordnungsübersicht Cluster 1-3 EVS vs. Google Textsuche – Toleranz von Marihuana</i>	<i>72</i>
<i>Tabelle 5-1 - Korrelation der Google Indices der Bildersuche zum EVS Index Homophobie</i>	<i>75</i>
<i>Tabelle 5-2 - Zuordnungsübersicht Cluster 1-3 EVS vs. Google Bildersuche - Homophobie.....</i>	<i>76</i>
<i>Tabelle 5-3 - Korrelation der Google Indices der Bildersuche zum EVS Index Ausländerfeindlichkeit.....</i>	<i>77</i>
<i>Tabelle 5-4 - Zuordnungsübersicht Cluster 1-3 EVS vs. Google Bildersuche - Ausländerfeindlichkeit</i>	<i>78</i>
<i>Tabelle A 1 - Suchbegriffe Google für Indices</i>	<i>XVII</i>
<i>Tabelle A 2 - Abgleichstabelle Kappa-Wert zu Datenreliabilität.....</i>	<i>XIX</i>
<i>Tabelle A 3 - Kreuztabelle & Cohen's Kappa EVS vs. Index Google Homophobie: Suchvollständigkeit manuell.....</i>	<i>XX</i>
<i>Tabelle A 4 - Kreuztabelle & Cohen's Kappa EVS vs. Index Google Homophobie: Suchvollständigkeit DMI Tool</i>	<i>XXI</i>
<i>Tabelle A 5 - Kreuztabelle & Cohen's Kappa EVS vs. Index Google Textsuche Homophobie.....</i>	<i>XXII</i>
<i>Tabelle A 6 - Kreuztabelle & Cohen's Kappa EVS vs. Index aus Google Textsuche & Suchvollständigkeit Homophobie</i>	<i>XXII</i>
<i>Tabelle A 7 – Kreuztabelle & Cohen's Kappa EVS vs. Index Ausländerfeindlichkeit Google Textsuche ...</i>	<i>XXIII</i>
<i>Tabelle A 8 - Kreuztabelle & Cohen's Kappa EVS vs. Index Toleranz von Alkohol Google Textsuche.....</i>	<i>XXIV</i>
<i>Tabelle A 9 – Kreuztabelle & Cohen's Kappa EVS vs. Index Toleranz von Marihuana Google Textsuche</i>	<i>XXIV</i>
<i>Tabelle A 10 - Kreuztabelle & Cohen's Kappa EVS vs. Index Google Bildersuche Homophobie</i>	<i>XXV</i>
<i>Tabelle A 11 - Kreuztabelle & Cohen's Kappa EVS vs. Index Google Bildersuche Ausländerfeindlichkeit</i>	<i>XXV</i>

Abkürzungsverzeichnis

ccTLD	Country Code Top Level Domain
CTR	Click Through Rate
DMI	Digital Methods Initiative
EMCDDA	European Monitoring Centre for Drugs and Drug Addiction
EVS	European Values Studie
GA_DMI	Google Autocomplete (Digital Methods Initiative)
GUI	General User Interface
IE	Internet Explorer
IP	Internetprotokoll
ISP	Internet Service Provider
ÖVP	Österreichische Volkspartei
PC	Personal Computer
SEA	Search Engine Advertising
SEO	Search Engine Optimization
SVP	Schweizerische Volkspartei
TS	(Google) Textsuche
VPN	Virtual Private Network
VPNC	Virtual Private Network Client
WHO	World Health Organisation

1 Einleitung

Forschungsgrundlage

Im Web hat sich in den letzten Jahren ein soziales Gefüge entwickelt, das heute vereinfacht durch den Begriff *Social Media* beschrieben wird: Der Nutzer hat sich vom reinen Konsumenten zum Inhalte konsumierenden Produzenten entwickelt (sog. *Prosumer*). (Mühlichen, 2014, S. 95) Informationen und Daten, die die Prosumer *generieren*¹ (z.B. in sozialen Netzwerken wie *Facebook, Twitter, YouTube, etc.*), entstehen losgelöst von den Daten die mittels Erhebungsmethode (z.B. Online-Befragung) im Web *produziert*² werden. Im digitalen Zeitalter jedoch wird das Web von vielen Wissenschaften (z.B. den Geisteswissenschaften, der Soziologie, den Kommunikationswissenschaften) nach wie vor hauptsächlich als Mittel zur künstlichen Datenproduktion und -erhebung verstanden. Es steht „einerseits die Forschung über Online-Medien (z.B. Usability³) und andererseits die Forschung mittels Online-Medien (z.B. Online-Befragung) im Vordergrund.“ (Döring, 2008, S. 357) Grundsätzlich unterscheiden sich Online-Befragungen dabei lediglich durch den Erhebungsprozess mittels Web (Wagner & Hering, 2014) von den Offline Erhebungsmethoden. (vgl. Reinecke, 2014, S. 602, f.) Eine Passantenbefragung beispielsweise wird nicht länger in der Fußgängerzone durchgeführt, sondern ins Web „verlagert“. Vor allem haben Online Befragungen im Vergleich zu den Offline-Verfahren den Vorteil einer schnelleren und vereinfachten Datenerhebung. (Reinecke, 2014, S. 662-663) Viele Unternehmen haben mittlerweile allerdings erkannt, dass sich solche Daten die im Web *generiert* werden, von den Daten, die im Web *produziert* werden, unterscheiden. Generische Daten aus sozialen Netzwerken zum Beispiel lassen sich für Unternehmen „werbewirksam und repräsentativ instrumentalisieren“. (Krause & Golkowsky, 2014, S. 151) Mittels Social Media Analysen beispielsweise werden die im Web generierten Daten erhoben und ausgewertet. Zugrunde liegt dabei die Annahme, dass Nutzer im (sozialen) Web ein Abbild Ihrer realen Person widerspiegeln. Das Web und das reale Leben sind in ihrer Sozialisierung, der Kommunikation, ihren Ausdrücken und der Verständigung durch die heutige Mediennutzung inzwischen untrennbar miteinander verbunden. (Kozinets, 2013)

¹ Daten die im Web durch die individuelle Nutzung des Webs entstehen

² Daten die mit Hilfe des Webs künstlich erstellt werden

³ Direkte Übersetzung durch den Autor: „die Benutzerfreundlichkeit“; passender nach Definition durch Rampl: „Die wohl treffendste Übersetzung für den Begriff Usability ist „Gebrauchstauglichkeit“, die auch in Normen und Standards verwendet wird. Dieser Begriff impliziert Notwendigkeit, nicht freiwillige Serviceorientierung.“ (2014)

Die im Web generierten Daten (Meinungen, Wünsche, Interessen, etc.) eines Nutzers bieten daher die Basis, aus den digitalen Informationen des sozialen Webs auf die Wünsche, das Verhalten und die Interessen des Nutzers in der realen Welt rückschließen zu können. In Zeiten von Big Data hinterlassen Nutzer / -innen⁴ ihre Spuren im Web indes nicht lediglich beim Austausch in sozialen Netzwerken. Mittels IP (Internetprotokoll) Adresse kann z.B. die Verhaltensweise von Individuen im Web nachvollzogen werden. Sobald sich Nutzer auf Webseiten namentlich registrieren, ist auch die Verknüpfung zu personenbezogenen Daten möglich. Website-Betreiber können personalisierte Werbebotschaften ausspielen, Bewegungsprofile erstellen und Verhaltensmuster lernen oder Produktinteressen ableiten. (Weiss, 2010, S. 134, ff.) Auch Suchmaschinen (*Google, Bing, Yahoo, etc.*) werden von den Web-Nutzern mit (privaten) Informationen (wie z.B. Fragen, Standorten, Meinungen) „gefüttert“ und lernen dadurch den, bzw. die Nutzer Tag für Tag besser kennen. Auf Basis verschiedener Algorithmen werden die Ergebnisse einer Suchanfrage nach unwichtig und relevant getrennt und den Nutzern in den Suchergebnissen in einem Ranking nach Bedeutsamkeit sortiert präsentiert. Es lässt sich argumentieren, dass diese Daten, die auf dem Interesse und dem Nutzungsverhalten der Nutzer basieren, authentischer sind als solche Daten, die künstlich produziert werden (z.B. durch Umfragen). Zudem werden nicht länger nur Social Media Kanäle als relevante Datenquellen angesehen, sondern das Web als Ganzes. Diese im Web generierten Daten ermöglichen der Forschung somit einen neuen Blickwinkel auf das Web.

Gegenstand der Masterarbeit

Neben der Möglichkeit, das Web als Mittel zur Datenerhebung zu begreifen, kann es ferner also auch als Untersuchungsobjekt, das seine Daten selbstständig generiert, betrachtet werden. Während klassische Massenmedien (TV, Radio, Print) primär unidirektional senden, funktioniert die Kommunikation im Web (meist) bidirektional. (Mühlichen, 2014, S. 95) Das Web löst die klassische Sender-Empfänger-Beziehung auf und ermöglicht es, relativ einfach Rückmeldungen „mit beliebig großen und kleinen, professionellen wie unprofessionellen Mitteln zu produzieren“. (Mühlichen, 2014, S. 95) War der Nutzer früher somit reiner Empfänger einer Botschaft, ist er heute Rezipient und Sender zugleich. Er kann auf direktem Weg, über das Web, auf die gesendeten Inhalte reagieren – seine Meinung ist für den Sender einer Botschaft und auch für andere Rezipienten nicht länger „stumm“. Wann

⁴ Im Folgenden wird aufgrund der besseren Lesbarkeit in der Regel die männliche Schreibweise verwendet. Dennoch weise ich hiermit ausdrücklich darauf hin, dass nachfolgend sowohl die männliche als auch die weibliche Schreibweise gemeint ist

immer Daten von Nutzern generiert werden, bestehen – neben der Social Media Analyse – somit viele weitere Möglichkeiten das Web als Forschungsgegenstand zu analysieren. Hierzu dienen webwissenschaftliche Untersuchungsmethoden. Die Idee von webwissenschaftlichen Methoden besteht darin, die im Web generierten Daten als Basis für die Forschung in verschiedenen wissenschaftlichen Bereichen, zum Beispiel in der Medizin (*Google Flu Trends*), der Klimaforschung (*Hyperlink- und Resonanzanalyse*), oder den Medienwissenschaften (*Screen Cast Documentary*), zu nutzen. (vgl. Schmitz, 2014)

Gegenstand der Masterarbeit ist es, einen sozialwissenschaftlichen Bereich mit einer webwissenschaftlichen Methode zu erforschen. Als Untersuchungsobjekt im Web dient die Suchmaschine Google. Mit Hilfe der Suchmaschine Google soll ein Abbild der verschiedenen kulturellen Ausprägungen von europäischen Gesellschaften erstellt werden. Die Arbeit beschäftigt sich mit der Frage, ob es möglich ist, mittels Suchmaschine (Google) valide und reliable Rückschlüsse auf die Einstellung zu verschiedenen Themen, der realen Gesellschaft zu ziehen. Die Forschungsfrage lautet somit: Ist es möglich, mit Hilfe der Suchmaschine Google ein Abbild der Gesellschaft widerzuspiegeln? Ziel der Arbeit ist es, auf Grundlage der generierten Daten eine innereuropäische kulturvergleichende Forschung zu erstellen. Hierdurch sollen qualitative Aussagen zu den Einstellungen und Meinungen zu den Themengebieten *Homophobie*, *Ausländerfeindlichkeit* und *Toleranz von Rauschmitteln* in 20 verschiedenen europäischen Kulturen (Ländern) getroffen werden. In diesem Rahmen werden keine Analysen und Aussagen zu quantitativen Daten getätigt, da diese Aussagen (zurzeit) durch die Google Suche nicht abbildbar sind. Eine generelle Repräsentativität der Daten für alle Einwohner eines Landes soll durch den Abgleich mit Daten aus einer Studie zur europäischen Wertegemeinschaft, bzw. aus Studien über den Genuss von Rauschmitteln in der EU, validiert werden. Daher dient zum einen die *European Values Studie (EVS)*, eine Studie der Europäischen Union zum Thema Werteeinstellungen in der EU, als Abgleich zur Validierung der mittels Google generierten Daten zum den Themen *Homophobie* und *Ausländerfeindlichkeit*. Zum anderen dienen die Studien der WHO (World Health Organisation), sowie der EMCDDA (European Monitoring Centre for Drugs and Drug Addiction), als Abgleich zum Thema *Toleranz von Rauschmitteln*. Sowohl die Studie der WHO, als auch die der EMCDDA bieten sich zum Datenabgleich im Bereich Toleranz von Rauschmitteln in der EU an, da sie die Themen Alkohol- und Cannabiskonsum sowohl seriös, als auch sehr umfangreich abdecken.

Durch den Abgleich der Studiendaten mit den per Google generierten Daten, soll nachgewiesen werden, dass die mittels webwissenschaftlicher Methode generierten

qualitativen Daten Validität und Reliabilität besitzen. Ein solcher Abgleich zwischen Google Daten und repräsentativen Studiendaten wurde in dieser Art wissenschaftlich bislang nicht realisiert. Eine hohe Güte der Google Daten (im Abgleich mit den Studiendaten) würde den Rückschluss zulassen, dass Webinstrumente wie Google repräsentative Befragungen ab- bzw. nachbilden können. Dies würde im wissenschaftlichen Kontext bedeuten, dass eine Untersuchung von Google Daten eine repräsentative Befragung ersetzen kann und somit Zeit und Kosten Effizienz gesteigert werden.

Untersuchungsobjekt und Vorgehensweise

Im Jahr 2014 hatte Google im Suchmaschinenmarkt in Europa einen Marktanteil von 92,8%. Weltweit gingen im selben Jahr ca. 2.095,1 Milliarden Suchanfragen bei Google ein. (Abbildung 1-1)

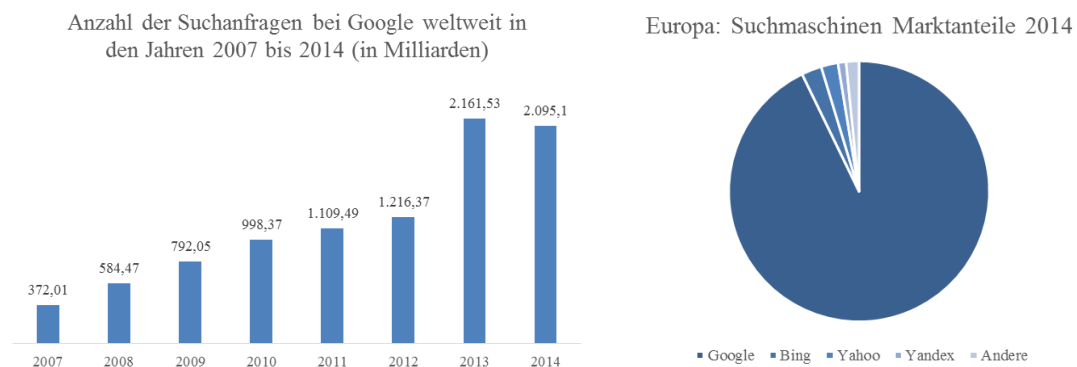


Abbildung 1-1 – (links) Anzahl der Suchanfragen bei Google weltweit in den Jahren 2007 bis 2014 (in Milliarden), (Internet Live Stats (statisticbrain.com/), 2015, Zugriff am 15.07.2015 von: <http://de.statista.com/statistik/daten/studie/71769/umfrage/anzahl-der-google-suchanfragen-pro-jahr/>), (rechts) Suchmaschinen Marktanteile, (luna-park GmbH, 2014)

Demzufolge dominiert Google den Suchmaschinen-Markt in Europa und kann aufgrund seiner hohen Nutzungszahlen als repräsentativ für die Nutzung und damit die Suchanfragen im Web angesehen werden. Diese Suchanfragen bei Google enthalten neben Suchen in Themenbereichen wie Arbeit, Nachrichten oder Information auch Fragen, Probleme, persönliches Befinden, private Meinungen, Wünsche, Interessen, usw. der (u.a. europäischen) Google Nutzer. Durch diese Suchanfragen entstehen somit Daten und Informationen über die Suchmaschinen-Nutzer selbst. Für jede dieser Suchanfrage erstellt Google u.a. anhand der Sprache, des Geo-Targeting⁵ und der besuchten länderspezifischen Top Level Domain (ccTLD⁶) ein länderspezifisches, bzw. ein anhand der Standorterkennung

⁵ Geo-Targeting = „grounded Web“ / lokalisiertes Web = Möglichkeit das Land, bzw. den Standort eines Nutzers anhand seiner IP Adresse zu bestimmen. Im nachfolgenden Verlauf der Arbeit impliziert der Begriff Internetprotokoll (IP) das Geo-Targeting

⁶ country code top-level domain = länderspezifische Top Level Domain; z.B. google.de - .de = ccTLD für Deutschland

lokalisiertes Suchergebnis für den Google Nutzer. (Google Inc., 2015e) Das heißt, dass ein Nutzer, der die Website *google.de* aufruft (ccTLD = *.de*, für [Google] Deutschland), einen Suchbegriff auf Deutsch eingibt (z.B. „Fußball“) und eine deutsche IP Adresse besitzt, ein anderes Suchergebnis erhält, als ein Nutzer, der die Website *google.co.uk* aufruft (ccTLD = *.co.uk*, für [Google] England), den gleichen Suchbegriff auf Englisch eingibt (infolge dessen „soccer“) und eine englische IP Adresse besitzt. Die Suchmaschine Google gibt u.a. anhand der Sprache, der IP (siehe Fußnote 5) und der angewählten ccTLD des Google Nutzers entsprechend zur gleichen Suchanfrage (also demselben Wort in der jeweiligen Landessprache) andere, länderspezifische Suchergebnisse aus. Zudem wird in der jeweiligen „Google-Landesversion“ (ccTLD) aufgrund der unterschiedlichen Suchergebnisse nachfolgend durch den Google Algorithmus auch eine andere Rangreihenfolge der Suchergebnisse erstellt. Die Rangreihenfolge stellt ein Ranking der besten, bzw. der relevantesten Treffer der Suchergebnisse von Google dar. Ergebnisse, die an erster Stelle in den Suchergebnissen erschienen (ausgenommen von Werbeanzeigen), werden von Google als relevanter bzw. wichtiger eingestuft, als die Suchergebnisse, die nachfolgen. Auf Basis der verschiedenen Google-Versionen (ccTLD) wird somit für den Nutzer (u.a. auf Basis Sprache, IP, ccTLD und Google Algorithmus) eine individuelle Rangreihenfolge ermittelt. Es ist daher anzunehmen, dass die Suchergebnisse bei Google mit der jeweils zugehörigen Landessprache, der IP Adresse eines Nutzers und dem Aufrufen der zugehörigen ccTLD die Fragen, Meinungen, Interessen, Gedanken, etc., der Einwohner eines Landes reflektieren könnten. Im praktischen Teil der Arbeit werden mit Hilfe eines VPN Hosters (Virtual Private Network⁷) die länderspezifischen Suchergebnisse der Autovervollständigung⁸ und der Textsuche⁹ von Google qualitativ anhand einer Online-Inhaltsanalyse untersucht. Über das Untersuchungsobjekt Google werden mittels Suchanfragen Daten zu den Themenbereichen *Homophobie, Ausländerfeindlichkeit* und der *Toleranz von Rauschmittel* generiert, mit dem Ziel, die Einstellung und Meinungsbilder verschiedener europäischer Gesellschaften auf Basis der kulturellen Unterschiede widerzuspiegeln. Dazu werden authentische Suchanfragen bei verschiedenen Google Landesversionen zu den genannten Themenbereichen durch die Kombination von Landessprache, ccTLD und IP Adresse (via

⁷ VPN = Virtual Private Network. Keine einheitliche Definition gegeben; im Kontext der Arbeit zu verstehen als Dienst zur Umgehung von Geo-Targeting (ausführlicher Kapitel 2.2.4)

⁸ Dienst von Google, der „Vorschläge zur automatischen Vervollständigung“ zur Anfrage im Suchfeld angezeigt. (Google Inc., 2015b) Bezieht sich darauf, wonach im Zusammenhang mit dem gesuchten Begriff „sonst noch oft gesucht wird.“ (ebd) Im nachfolgenden Text auch Suchvervollständigung, oder Auto-Fill.

⁹ Beschränkt auf die Top 100 Suchergebnisse

VPNC¹⁰) simuliert. Mit einem generischen Suchbegriff zu den Themen werden Textsuche und Suchvervollständigung von Google genutzt, um soziokulturell-vergleichende Aussagen zum Thema europäische Wertegemeinschaft treffen zu können. Die erhobenen Daten werden indexiert und mit den Daten der anderen Studien verglichen, um das Vorgehen der webwissenschaftlichen Methode zu validieren.

¹⁰ VPNC = Virtual Private Network Client; Software, die das Endgerät mit dem VPN verbindet.

2 Theorie

2.1 Status Quo – Die Forschung im Web

2.1.1 Netnographie – Methoden zur Analyse der Online Gemeinschaft

Die Netnographie¹¹ nach Kozinets ist ein methodologischer Leitfaden, der die sozialen Interaktionen durch das Web und verwandte Informations- und Kommunikations-Technologien aus einer ethnologischen Sichtweise studiert. (Kozinets, 2013, S. 3) Mit der fortschreitenden Digitalisierung unserer Welt haben sich nach Ansicht des Autors teilweise zentrale soziale Aktivitäten und Begegnungen in den virtuellen Raum verschoben. Das heutige Verhalten und der Alltag sind durch das Web geprägt – unser Arbeits-, und Sozialleben funktionieren zum Teil nur noch im und durch das Web. (Kozinets, 2013, S. 3) Dabei bezieht sich der Autor vor allem auf das Studium der Kommunikation in „Foren, Chats und Nachrichtengruppen, [...] Blogs, audiovisuellen Gemeinschaften, Foto- und Podcast-Gemeinschaften, virtuellen Welten, Online-Spielen, mobilen Gemeinschaften und sozialen Netzwerken.“ (Kozinets, 2013, S. 3) Um die Gesellschaft als Ganzes zu verstehen, ist es nach Kozinets für die Forschung zwingend erforderlich, auch die Online Kultur und die Gemeinschaften¹² im Web zu studieren. (Kozinets, 2013, S. 1) Der Autor sieht das Problem darin, dass die Wissenschaften der „Kultur- und Sozialforschung, der Volkswirtschaftslehre, die Rechts-, und Informationswissenschaften, Betriebswirtschafts- und Managementlehren, Kommunikationswissenschaften, Humangeographie, Medizin und Anthropologie¹³“ (Kozinets, 2013, S. 3) ihren jeweils eigenen Ansatz im Studium der sozialen Interaktionen im Web entwickelt haben. Hierdurch entstehen nach Kozinets unterschiedliche unbestimmte Methoden. (2013) Unter dem Begriff Netnographie möchte der Autor die Methoden, verschiedene Ansätze sozialer Interaktionen im Web zu studieren, systematisieren und definieren. (Kozinets, 2013, S. 3) Netnographie sieht der Autor somit als eine „Teilnehmerbeobachtung basierend auf einer Online Feldforschung.“ (Kozinets, 2013, S. 60) Das methodische Vorgehen zur Datenerhebung bei einer Online-Beobachtung wird von Kozinets daher folglich an das Vorgehen zur Datenerhebung in der Ethnographie angelehnt. (vgl. Kozinets, 2013, S. 58, ff.) Darüber hinaus sieht Kozinets den Bedarf, den Term Netnographie von der Ethnographie abzugrenzen. Beide Methoden versuchen

¹¹ Übersetzung aus dem Englischen durch den Autor; Original nach Kozinets: „Netnography“

¹² Übersetzung aus dem Englischen durch den Autor; Original nach Kozinets: „communities“

¹³ Übersetzung aus dem Englischen durch den Autor; Original nach Kozinets: „...cultural studies, sociology, economics, law, information science, business and management fields, communication studies, human geography, nursing and healthcare, and anthropology.“

Verhaltensweisen, sowie das kulturelle, soziale und politische Zusammenleben einer Gesellschaft zu verstehen. Die Netnographie unterscheidet sich dabei nach Kozinets allerdings vor allem durch die Art der Datenentstehung von der Ethnologie. Während die Ethnologie Daten erhebt, die auf zwischenmenschlicher Ebene entstehen (Verhaltensbeobachtungen, Mimik, Gestik, etc.), erhebt die Netnographie Daten, die durch reine Online-Interaktionen entstehen, d.h. es kann weder Mimik, noch Gestik, etc. untersucht werden. (Kozinets, 2013, S. 60) Zudem ändert sich der Prozess der Schriftführung von „Beobachtungen“ und die erhobenen Datenmengen variieren. Auch die Möglichkeit spezielle Analysetools und Methoden anzuwenden, ändert sich, wenn die Daten bereits in digitaler Form vorliegen. (Kozinets, 2013, S. 3)

Zusammenfassend kann gesagt werden, dass die Netnographie eine Vereinheitlichung verschiedener Methoden zur Analyse der sozialen Interaktionen der „Gemeinschaften und Kultur“ im Web darstellt. Hierbei sind die Datenquellen auf soziale Knotenpunkte (z.B. Foren, Gemeinschaften, soziale Netzwerke) im Web begrenzt. Die Forschungsmethoden werden dabei aus verschiedenen Wissenschaften entliehen, abgewandelt und auf das Web angewandt.

2.1.2 Digital Methods – Das Web als Datenquelle

Einführung in die digitalen Methoden

Der Begriff „Digital Methods“ ist vor allem durch Richard Rogers geprägt, der 2013 unter selbigem Begriff verschiedene methodische Ansätze zur Forschung im Web in seinem gleichnamigen Hauptwerk veröffentlichte. Rogers geht davon aus, dass es neuer Forschungsmethoden bedarf, um das Web aufgrund seiner neuartigen Beschaffenheit und seiner Struktur zu verstehen. Er bezieht sich daher mit den „digitalen Methoden“ als Grundlage auf die Untersuchung von Daten, die im Web *geborene*¹⁴ sind. Grundsätzlich unterscheidet Rogers zwischen den im Web *angeborenen digitalen*¹⁵ und den *digitalisierten*¹⁶ Daten, bzw. *digitalisierten Methoden*. (Rogers, 2013, S. 1-10) Während die einen Daten im Web generiert werden – nach Rogers *geboren* –, werden die anderen lediglich in das Web migriert (*digitalisiert*) mit dem Hintergrund, Daten zu produzieren. (Rogers, 2009, S. 1) Nach Rogers sind z.B. die Methoden der Human- und Sozialforschung (z.B. Online Befragungen) lediglich ins Web migrierte Methoden. Sie produzieren Daten im Web, die ohne die Online Befragung nicht entstünden. Diese Daten und Methoden sind somit

¹⁴ Übersetzung aus dem Englischen durch den Autor; Original nach Rogers: „born in the web“

¹⁵ Übersetzung aus dem Englischen durch den Autor; Original nach Rogers: „natively digital“

¹⁶ Übersetzung aus dem Englischen durch den Autor; Original nach Rogers: „digitized“

künstlich digitalisiert. Folglich nutzen derartige Methoden das Web als Medium, repräsentieren es aber nicht, da sie nicht im Web selbst entstanden sind. Digitalisierte Methoden versuchen im Web zu forschen, sind in ihrer Art selbst allerdings nur bedingt dafür geeignet, da sie digitalisiert wurden, um Daten zu produzieren. Sie sind jedoch nicht in der Lage, im digitalen Raum generierte Daten zu erfassen. (Rogers, 2009, S. 1) Nach Rogers gilt es, dem Medium zu folgen und die Forschungsmethoden an das Medium (hier das Web) anzupassen, und im Umkehrschluss nicht das Medium „Web“ an die Forschungsmethoden (z.B. Offline Erhebungsmethoden). (Rogers, 2013, S. 1) Rogers geht daher davon aus, dass das Web als Datenquelle selbst durch verschiedene Organisationen, Inhalte, Geräte und Umfelder und deren Kombinationsmöglichkeiten eine ausreichend Datenmenge generiert, um eigene, neue Forschungsansätze und –Methoden zu entwickeln. Diese neuen Methoden nennt Rogers „*digital Methods*“ – digitale Methoden. Generell gilt laut Rogers demzufolge, dass im Web geborene Daten mit digitalen Forschungsmethoden erhoben werden müssen, um valide Aussagen über die reale Kultur und Gesellschaft treffen zu können. Bei den digitalen Methoden stehen das Web und die im Web geborenen Daten als Untersuchungsobjekt somit an erster Stelle. (Rogers, 2009, S. 1) Für Rogers ist das Web nicht nur ein reines Untersuchungsobjekt, sondern vor allem eine (Daten-)Quelle selbst. (Rogers, 2009, S. 3) Nach Rogers sind im Web generierte Daten solche, die Rückschlüsse auf das reale Leben und die Kultur der Nutzer zulassen. Zwei Umstände ermöglichen es nach Rogers, das kulturelle Verhalten verschiedener Nationen im Web zu verfolgen: Zum einen wurde durch Daniel Miller und Don Slater (2000) nachgewiesen, dass verschiedene Kulturkreise das Web an ihre unterschiedlichen kulturellen Bedürfnisse anpassen. Das Verhalten im und die Anpassung des Webs nach den jeweiligen Bedürfnissen der Nutzer bestimmter Kulturen, spiegelt somit die Kultur selbiger im Web wider. (Rogers, 2009, S. 2) Zum anderen entsteht zeitgleich im Jahr 2000 mit dem von Rogers als „grounded-web“¹⁷ bezeichneten Geo-Targeting die Möglichkeit, Nutzer im Web anhand ihrer IP Adresse zu geo-lokalisieren und somit eindeutig dem Land zuzuordnen, aus dem sie sich ins Web eingewählt haben. Seitdem werden Nutzer beim Surfen im Web automatisch anhand ihrer IP dem jeweiligen Land zugeteilt, in dem sie sich mit ihrem Endgerät ins Web einwählen. (Rogers, 2009, S. 4) Die Kombination der Theorie von unterschiedlichen Kulturkreisen im Web mit der technischen Entwicklung der Geo-IP ermöglicht nach Rogers den Rückschluss aus dem Web auf die Realität. Die im Web geborenen Daten dienen als Spiegel für die

¹⁷ Übersetzung durch den Autor: *lokalisiertes Web* = Möglichkeit das Land, bzw. den Standort eines Nutzers anhand seiner IP Adresse zu bestimmen

unterschiedlichen Kulturen und deren Verhaltensmuster und Einstellungen in der Realität. Das lokalisierte Web erlaubt dazu eine „Trennung“ der Daten auf länderspezifischer Ebene. Daher sind laut Rogers soziale und kulturelle Veränderungen in der Realität im Web nachzuvollziehen. (Rogers, 2009, S. 3) Als Untersuchungsobjekte im Web stellt Rogers verschiedene Analyse-Objekte, bzw. Methoden vor. Unter anderem nennt er Hyperlinks („*The Link and the Politics of Web Space*“), Webseiten („*The Website as Archived Object*“), Suchmaschinen („*Search as Research*“), getrennte Netze („*National Web Studies*“), soziale Netzwerke („*Social Media & Postdemographics*“) und Gemeinschaftsarbeiten („*Wikipedia as Cultural Reference*“) als Datenquellen. (Rogers, 2013, S. 39-165) Der Autor zeigt auf, wie sich beispielsweise Internetarchive für soziale Forschung nutzen lassen, warum die Veränderung von Webseiten festgehalten werden muss, oder wie IP Adressen genutzt werden können, um Länder und deren kulturelles Profil zu definieren. (Rogers, 2013, S. 61 ff., 83-124)

Search as Research – Datenerhebung mittels Suchmaschine

Als Ansatzpunkt für das methodische Vorgehen zur Datenerhebung im praktischen Teil der Arbeit ist vor allem der Punkt *Suchmaschinen* („*Search as Research*“, nach Rogers, 2013, S.95-124) relevant. Nachfolgend wird daher insbesondere auf die Theorie zum Bereich Suchmaschinen („*Search as Research*“) eingegangen. Bei der Forschung mittels Suchmaschinen kategorisiert Rogers das Web in drei Sphären¹⁸: die Web-Sphäre¹⁹, die Blog-Sphäre²⁰ und die Nachrichten-Sphäre²¹. Durch die Abgrenzung per Suchmaschine besitzt nach Rogers jeder dieser Bereiche eine eigenes Kriterium zur Betrachtung der Suchergebnisse (z.B. Suche per Google News nach Nachrichteninhalten). Für die Nachrichten-Sphäre zum Beispiel kann es nach Rogers sinnvoll sein, absteigend von aktuellen Inhalten hin zu alten Inhalten zu sortieren. (Rogers, 2013, S. 97) Im nachfolgenden praktischen Teil der Arbeit bezieht sich die Analyse der Suchergebnisse auf die Web-Sphäre, die Gesamtsichtbarkeit im Web. In der Web-Sphäre stufen Suchmaschinen per Indexierung / Algorithmus (siehe auch Kapitel 2.2.3) alle relevanten sichtbaren Web Inhalte²² auf Basis

¹⁸ A sphere: Eine durch das Gerät (z.B. die Suchmaschine) vordefinierte Quellen Sammlung. (nach Rogers, 2013, S32-33)

¹⁹ Zentrale Ebene des Webs, alle Bereiche die für Suchmaschinen sichtbar (z.B. alles was durch Google, Yahoo, etc. gefunden werden kann) (nach Rogers, 2013, S32-33)

²⁰ Der „Volks-“ oder „Nutzer-“ Teil des Webs, alle Bereiche die für „Blog“ Suchmaschinen sichtbar (z.B. alles was durch Technorati, etc. gefunden werden kann) (nach Rogers, 2013, S32-33)

²¹ Der Nachrichten-Teil des Webs, alle Bereiche die für „News“ Suchmaschinen sichtbar (z.B. alles was durch Google News, etc. gefunden werden kann) (nach Rogers, 2013, S32-33)

²² Nicht sichtbare Web-Inhalte sind Inhalte die nicht von Suchmaschinen wie Google gefunden werden können (z.B. Inhalte im Deep Web)

der Suchanfrage anhand des Suchbegriffs / der Suchbegriffe ein. Der Algorithmus bezieht dabei unter anderem sowohl die Anzahl der Verlinkung von anderen Websites auf eine Seite (*Inlinks*), als auch die Relevanz von Inhalten, z.B. die Anzahl der Klicks von Nutzern auf die jeweils angezeigten Suchergebnisse, ein. (Rogers, 2013, S. 97) Hinzu kommen weitere Faktoren wie „Aktualität, Sprache, Kontext und geographische Analysen“. (Rogers, 2013, S. 3) Die Rangreihenfolge bei der Sortierung von Suchmaschinen in der Web-Sphäre basiert somit auch auf dem Interesse der Nutzer an den verfügbaren Inhalten im Web. Suchergebnisse von Suchmaschinen sind nach Rogers eine Auflistung von aktuellen, relevanten Inhalten (Datenquellen), die das gesellschaftliche Interesse wiedergeben, da sie akut von den Nutzern konsumiert (hohe CTR) werden. Hoher Konsum bezeugt insoweit, dass Fragestellung und Information gesellschaftlich relevant sind. Der Autor folgert, dass Suchmaschinen Quellen darstellen, die Informationen über ihre Nutzer preisgeben, weil sie automatisiert gesellschaftlich relevante Themen in eine Rangreihenfolge übersetzen. (Rogers, 2013, S. 100) Daher gilt nach Rogers: „Wir schauen uns Google Suchergebnisse an und sehen die Gesellschaft, anstatt Google²³.“ In Kombination mit der Möglichkeit Nutzer per ccTLD und mittels IP Adresse separiert (als Land) zu betrachten, entsteht somit auch die Möglichkeit, Gesellschaften getrennt voneinander via Google zu betrachten und zu vergleichen. Einer Verzerrung durch z.B. den Algorithmus, sprachliche Überschneidungen, Suchterme, etc. kann dabei durch verschiedene Methoden entgegengewirkt, gänzlich ausgeschlossen werden kann diese jedoch nicht. (Rogers, Digital methods, S. 95, ff., 2013)

Zusammenfassung

Rogers plädiert dafür, wie Kozinets auch, das Reale und das Virtuelle nicht länger getrennt zu betrachten. Durch die heutigen technischen Möglichkeiten und die Nutzung des Webs verschmelzen das reale und das digitale „Leben“. Laut Rogers muss das Digitale als eine Verlängerung des realen Raumes angesehen werden. Im Web wird nicht eine „Online-Kultur“ als solche repräsentiert, sondern Spiegelbilder der verschiedenen Gesellschaften gelebt. Aufgrund der veränderten Kommunikationsstruktur mit neuen Technologien muss die Forschung im Web mit digitalen Methoden erfolgen, um die Validität und Reliabilität der Ergebnisse zu gewährleisten. Während Kozinets die „Online Forschung“ auf Basis gegebener Forschungsmethoden verschiedener wissenschaftlicher Disziplinen forciert, plädiert Rogers dafür, eigene, der Quelle (dem Web) angepasste Methoden zu entwickeln

²³ Übersetzung durch den Autor: „We look at Google results and see society, instead of Google“

und zu nutzen. Die digitalen Methoden zur Datenerhebung von Rogers sind daher den angepassten digitalisierten Methoden von Kozinets zur Forschung im Web vorzuziehen.

2.1.3 Die Webwissenschaften – Das Web verstehen

Als Basiswerke zur Begründung der Disziplin der Webwissenschaften dienen die Arbeiten „*Creating a Science of the Web*“ und „*A Framework for Web Science*“ von Tim Berners-Lee. (Scherfer & Volpers, 2013, S. 1) In Anlehnung an diese Ideen erfolgte im Jahr 2008 auch im deutschsprachigen Raum erstmals die Auseinandersetzung mit der Thematik: „*Warum und zu welchem Zweck benötigen wir eine Webwissenschaft?*“ (Volpers, 2008, S. 31, ff.) Nach Helmut Volpers liegt die Notwendigkeit der Begründung der Webwissenschaften zum einen in der „Entstehung völlig neuer Gegenstandsbereiche in der sich entwickelnden Gesellschaft.“ (2008, S. 32) Damit meint der Autor die sich stetig entwickelnde Struktur des Webs und die immense Bedeutung dieser Technologie. Entwickelt als „plattformübergreifendes Kommunikations- und Informationsnetzwerk für Forscher“ (Andelfinger & Hänisch, 2015, S. 2) durchdringt das Web heute auf PCs, Tablets und Smartphones alle Lebensbereiche, ist in der Arbeitswelt und der privaten Nutzung, in der Wirtschaft und der Freizeit nicht mehr entbehrlich. Das Web beeinflusst Ökonomie, Politik, Gesellschaft und Kultur. (Scherfer & Volpers, 2013, S. 8) Es ist zur festen Instanz im alltäglichen Leben geworden und entwickelt sich stetig weiter. Zudem erfordert die „zunehmende Komplexität wissenschaftlicher Erkenntnis“ das Herausbilden eines eigenständigen Wissenschaftszweiges. (Volpers, 2008, S. 32) Das Web als solches sprängt den Rahmen der gegebenen Forschungsmethoden und fordert einen „Wandel der anerkannten Probleme und Methoden eines Forschungsgebietes“ (nach Kuhn, 1976, S. 25). (Volpers, 2008, S. 32) Seine Betrachtung durch die gegebenen Wissenschaften (z.B. Kommunikations-, Medienwissenschaft, Soziologie, Informatik, etc.) unterliegt den jeweiligen fachlichen Perspektiven und verzerrt somit die Erkenntnisse aufgrund des jeweiligen „fachlichen Paradigmas“. (Volpers, 2008, S. 45, f.) Die Kommunikations- und Medienwissenschaften beispielsweise begrenzen das Web auf den Gegenstandsbereich als Medium. Fachfremde Wissenschaften studieren das Web, die es mit ihren fachlich nicht geeigneten Methoden zu ergründen versuchen. (Scherfer & Volpers, 2013, S. 9) Dabei fehlt es den Wissenschaften „an einer einheitlichen Terminologie und Fachkultur“. (Scherfer & Volpers, 2013, S. 33) Daher erfolgt oftmals interdisziplinär die Verwendung falscher, bzw. nicht trennscharfer Termini. Der Begriff „Web“ wird mit dem Begriff „Internet“ gleichgestellt. Auch Termini wie „Website“ und „Webseite“ werden falsch genutzt und

deuten auf ein mangelndes Fachverständnis hin. (Scherfer & Volpers, 2013, S. 9) Zudem rückt die „Notwendigkeit der immanenten Weiterentwicklung des Webs auf der Grundlage anwendungsbezogener Forschung“ hinzu. (Volpers, 2008, S. 33) Dies steht im Bezug zur Grundidee von Tim Berners-Lee, dass die Webwissenschaften eine Analyse, Regulierung und Entwicklung des Webs leisten und hierbei neben den technologischen Aspekten die rechtlichen und sozialen Bedingungen reflektieren sollen. Die Webwissenschaften betrachten das Web somit auf zwei Ebenen.

1. Die Mikro Ebene

Auf der Mikro-Ebene wird „zur Analyse, Regulierung und Entwicklung“ die technische Seite des Webs, also seine Infrastruktur beleuchtet. (Scherfer & Volpers, 2013, S. 9) Wir haben verstanden das Web zu nutzen, wissen aber in der Regel nicht warum es wie funktioniert. (O’Hara & Hall, 2008, S. 1) Ziel der Mikro Ebene ist es daher zum einen, eine einheitliche Fachterminologie für das Web zu erstellen. Zum anderen soll der Aufbau – die Infrastruktur des Webs mit ihren Protokolle und Sprachen – verständlich gemacht, bzw. überwacht werden. Dies geschieht, um sicher zu stellen, dass das Web eine für die Gesellschaft nutzbringende Technologie ist und bleibt. (O’Hara & Hall, 2008, S. 1)

2. Die Makro-Ebene

Die Makro-Ebene untersucht Verhaltensmuster und Ausdrucksformen, also die soziale Interaktion zwischen den Menschen im Web. (Scherfer & Volpers, 2013, S. 9) Dies bedeutet, das Web bietet neben der Möglichkeit, es auf seine technische Infrastruktur hin zu untersuchen auch die Möglichkeit, es auf seine Nutzer, auf Kultur und Gesellschaft hin zu untersuchen. Während bei Richard Rogers die digitalen Methoden zur Untersuchung der Gesellschaft und der Kultur im Vordergrund stehen, stehen bei den Webwissenschaften die „technologischen und mathematischen Aspekte der Webentwicklung“ (Scherfer & Volpers, 2013, S. 9) im Zentrum. Dennoch wollen auch die Webwissenschaften das Web dahin gehend ergründen, „soziale Bedingungen zu reflektieren“ (Volpers, 2008, S. 38) und die „Interaktion zwischen Menschen, die Informationen kreieren, verlinken und konsumieren“ (Scherfer & Volpers, 2013, S. 9) zu verstehen. Somit werden neben den genannten technologischen und mathematischen Aspekten auch die „geistes- und sozialwissenschaftlichen Perspektiven in den Blick gefasst.“ (Scherfer & Volpers, 2013, S. 9)

Die vorab definierten digitalen Methoden fallen somit in den Bereich der Makro-Ebene der Webwissenschaften und stellen ein Methoden-Set zur Erhebung und zur Analyse der im

Web generierten Daten dar. Im Verlauf der Arbeit liegt die Theorie der Webwissenschaften als Basis für das weitere Vorgehen zu Grunde. Zum einen bieten die Webwissenschaften, wie vorab definiert, ein Portfolio diskriminierender Termini, die ein trennscharfes wissenschaftliches Arbeiten ermöglichen. Zum anderen finden im praktischen Teil der Arbeit, in Anlehnung an die Untersuchung der Makro Ebene des Webs, digitale Verfahren zur Daten-Generierung im Web nach Richard Rogers Anwendung. Zur Analyse und Interpretation der Daten orientiert sich der methodische Schwerpunkt vor allem an der Online-Inhaltsanalyse.

2.2 Analysemethode & Operationalisierung

2.2.1 Online-Inhaltsanalyse

Im praktischen Teil der Arbeit wird die Online-Inhaltsanalyse zur Erhebung der im Web generierten Daten verwendet. Auch wenn sich der Begriff der Online-Inhaltsanalyse sowohl auf das „Forschungsinstrument, als auch auf den Forschungsgegenstand beziehen kann“, bietet sich der Term nach Volpers an. (2013, S. 15) Immanent ist, dass der Vorgang der Online-Inhaltsanalyse als eigenes, standardisiertes Verfahren verstanden wird. Das Ziel ist es, im Web generierte, quantifizierbare Daten (Text, Bild, Audio-Visuell) zu erheben und auszuwerten. Die Gütekriterien der Online-Inhaltsanalyse orientieren sich im „Hinblick auf Validität und Reliabilität“ an denen der in der Kommunikationswissenschaften entwickelten Inhaltsanalyse. (Volpers, 2013, S. 15) Im Web generierte Daten werden mittels Online-Inhaltsanalyse in ihrer Komplexität reduziert, das heißt kategorisch und eindeutig durch eine Kodierung operationalisiert. Diese Transformation ermöglicht neben der „Komplexitätsreduktion“ mannigfaltiger Daten auch die Möglichkeit, Aussagen über die soziale Wirklichkeit zu treffen (Inferenzschlüsse). (Volpers, 2013, S. 15&16) Die Datenerhebung kann sowohl manuell als auch automatisch erfolgen. Aufgrund der Komplexität der Daten erfolgt die Erhebung im Bereich der Online-Inhaltsanalyse in der Regel meist manuell, „da bisher kein Verfahren für deren automatische Kodierung zur Verfügung steht“. (Volpers, 2013, S. 16) Aufgrund der Beschaffenheit des Webs stehen der Methodik vor allem vier Problemfelder gegenüber:

1. Die „Unendlichkeit und Unübersichtlichkeit des Webs“ beschreibt das Problem, eine vollständige Grundgesamtheit verlässlich zu ermitteln. Die Autoren führen hier das Beispiel an, bei einer Hyperlink-Analyse eine Liste *aller* Hyperlinks (Abbildung 2-1 - Die Struktur des Webs. (Broder, et al. 2000, S. 318)) im Web zu erstellen (Volpers, 2013, S. 22)
2. Die „Nonlinearität und Hypertextualität“ des Webs verhindert eine abgeschlossene thematisch zusammengehörige Aussage zu einem Thema, dies wiederum führt zu „Problemen bei der Bestimmung und Abgrenzung von [...] Analyseeinheiten“. (Volpers, 2013, S. 22)
3. Der „flüchtige, bzw. dynamische Aggregatzustand des Contents“ befasst sich mit dem Problem der Wandelbarkeit von Inhalten im Web. D.h., dass die Inhalte von Webseiten jederzeit ohne viel Aufwand verändert, bzw. Nutzer-spezifisch angepasst werden können (Volpers, 2013, S. 22)

4. Die „verschiedenen Angebotsformen der Inhalte“ im Web verhindern eine einheitliche Kodierung. Daten (Texte, Audio-Files, Bilder, Videos), die im Web miteinander verknüpft sind, unterscheiden sich oft dadurch, dass sie eine ungleiche Kodier Logik erfordern – man spricht auch von Multicodität (Volpers, 2013, S. 22)

Datenabgrenzung mit Hilfe von Suchmaschinen

Die Abgrenzung der Untersuchungseinheit und der Analyseeinheit stellt nach Seibold vor allem ein Problem aufgrund der Größe (Anzahl der Webseiten) des Webs und der zusammenhängenden Inhalte (Verlinkungen der Webseiten durch Hyperlinks) im Web dar. (Seibold, 2002, S. 46-47) Um der Problematik zur Reduktion einer Analyseeinheit entgegen zu wirken (siehe Punkt 1, Seite 15), muss das Web im Gegensatz zu den Offline Medien daher anhand seiner individuellen Möglichkeiten erschlossen werden. Zur Erfassung einer repräsentativen Grundgesamtheit dient nicht (wie z.B. Offline) die Marktstruktur (Anbieter gegen Anbieter) des Webs, sondern der angebotene Inhalt (Content gegen Content) im Web. (Volpers, 2013, S. 26) Daher ist sowohl die „Findability“²⁴ von Beiträgen, als auch die Zugriffszahlen auf die Beiträge ein entscheidendes Kriterium für die Erfassung einer repräsentativen Grundgesamtheit im Web. Die Sichtbarkeit und die Zugriffszahlen von Web-Inhalten werden vor allem von Suchmaschinen bestimmt. (Volpers, 2013, S. 16) Suchmaschinen, insbesondere Google, bieten sich daher als relevante Hilfsmittel an. Sie ermöglichen einen unkomplizierten Zugang zu den Daten im Web und dienen zeitlich zur Abgrenzung der Analyseeinheiten. Um repräsentative Daten im Web zu generieren, bieten zum Beispiel die Suchergebnisse von Suchmaschinen eine Möglichkeit, die Eingrenzung der Analyseeinheiten vorzunehmen. Durch die Rangreihenfolge der Suchergebnisse, bzw. die Sortierungsmöglichkeit von Suchergebnissen, kann die Stichprobe anhand der „zeitlichen, räumlichen, oder mengenmäßigen Selektion von Auswahleinheiten“ bestimmt werden. (Volpers, 2013, S. 26) D.h., es kann je nach Bedarf eine beliebige Anzahl z.B. aktueller, geographischer, oder relevanter Treffer aus den Suchergebnissen als Analyseeinheit untersucht werden.

Datenbasis mit Hilfe von Suchmaschinen

Auch die „Nonlinearität und Hypertextualität“ stellt bei kulturvergleichender Forschung im Web nur bedingt ein Problem für die Erarbeitung eines Themenfelds mittels Suchmaschine dar. Davon ausgehend, dass das Web die Gesellschaft und die Kultur abbildet (siehe auch

²⁴ Übersetzung durch den Autor: die Auffindbarkeit. Hier im Zusammenhang mit Sichtbarkeit von Web-Inhalten. Kann ein Web Inhalt z.B. nicht von der Google Suche gefunden werden, besitzt er für Google keine *Findability*.

Kapitel 2.1.2), so sollte sich folglich auch die Meinung der Gesellschaft zu relevanten Themen im Web abbilden – und daher auch mit Suchmaschinen wiederzufinden sein. Bei kulturvergleichenden Fragestellungen ergibt sich die Abgrenzung der Analyseeinheiten des Themas durch die Forschungsfrage selbst. Die thematische Eingrenzung erfolgt, wie die Eingrenzung der Untersuchungs- und Analyseeinheit auch, automatisch durch die Trefferliste von Suchmaschinen, bzw. den für die Suchanfrage verwendeten Suchbegriff. Zu analysieren sind somit als kleinste Einheit die Ergebnisse der Trefferliste. Dabei besteht zum Beispiel die Möglichkeit, die Treffer rein anhand ihres Textes zu analysieren oder aber deren Verlinkung, bzw. Volltexte zu untersuchen. Die Trefferliste der Suchmaschine engt das Themenfeld ein, so dass z.B. ein *kultureller Vergleich* von verschiedenen Google Versionen (ccTLDs) ermöglicht wird. Im praktischen Teil der Arbeit dienen, wie vorab definiert, die ersten 100 Treffer der Suchmaschine Google allein auf Text Basis (Verlinkungen werden nicht untersucht) als Analyseeinheiten.

2.2.2 Erhebung des Datenmaterials bei dynamischen Web-Inhalten

Die Wandelbarkeit von dynamischen Inhalten im Web stellt insofern ein Problem dar, als Inhalte im Web einfach gelöscht werden können, bzw. nicht länger auffindbar sind, da Inhalte dynamisch ihre Position verändern können. D.h., dass der Inhalt der Startseite einer Website jederzeit durch wenige Klicks auf eine beliebige andere Webseite verschoben, oder gänzlich gelöscht werden kann und somit unter dem „alten“ Hyperlink nicht mehr, oder nur noch mit anderem Inhalt abrufbar ist. (Seibold, 2002) Auch die Trefferlisten von Suchmaschinen ändern sich bei verschiedenen Such-Durchläufen, weil ihr Inhalt dynamisch durch den Algorithmus der Suchmaschine bestimmt wird. Dem flüchtigen „Aggregatzustands“ des Webs kann nach Seibold jedoch mittels Datenspeicherung entgegen gewirkt werden. Zum einen besteht die Möglichkeit, die generierten Daten (z.B. das Suchergebnis) manuell aus dem Web-Browser zu speichern²⁵ und somit zu archivieren. Zum anderen können die Daten durch den Serverbesitzer direkt in einer Datenbank auf dem Server gespeichert werden. Dies erfordert jedoch den direkt Kontakt zum Server-Eigner, als auch evtl. Zugriffsrechte auf den Server selbst. (Seibold, 2002, S. 51-52) Gerade bei kulturvergleichenden Forschungen im Web kann die Dynamik und Wandelbarkeit von Web-Inhalten allerdings sogar vor Vorteil sein – auf lange Sicht gesehen ist sie gar zwingend erforderlich, um einen Wandel in der Gesellschaft abbilden zu können. Eine kulturvergleichende Studie im Web könnte zum Beispiel untersuchen wie schnell

²⁵ Beispiel Browser „Mozilla Firefox“ (V 38.0.1): Open Menu -> Save Page -> Data Type: Web Page, complete

Tageszeitung (in der News-Sphäre) auf Nachrichten aus unterschiedlichen Themenfeldern wie Politik, Wirtschaft, Kultur, Sport, Gesellschaft, etc. reagieren. Über die Dynamik in den jeweiligen Bereichen ließe sich somit eventuell eine Aussage zur Medien-Nutzung, bzw. zum Informationsinteresse unterschiedlicher Gesellschaften treffen. Eine schnelle Wandelbarkeit von Inhalten in einem Bereichen wäre dann z.B. ein Indikator für hohes Interesse (Grundgedanke es werden viele Inhalte benötigt und daher schnell ausgetauscht) und in dieser Form nur durch die Schnelllebigkeit des Webs nachzuvollziehen.

Die Erhebung der von Seibold diskutierten Klick- und Nutzungspfade, sowie Probleme durch Multicodität sind für die praktische Umsetzung der durchgeführten Online-Inhaltsanalyse (vgl. Seibold, 2002, S. 48, f.) vor dem Hintergrund der Datenerhebung mittels Google irrelevant. Wie vorab definiert, filtert die Suchmaschine die Analyseeinheiten auf Basis ihrer gesellschaftlichen Relevanz. Eine tiefergehende Analyse der Nutzungs- bzw. Klickpfade²⁶, wie die Kodierung von Bildern (Exkurs mit eigenem Ansatz) ist auf Basis der vorliegenden Forschungsfrage nicht gegeben.

2.2.3 Suchmaschinen – Indexierung von Web-Inhalten

Wie bereits in Kapitel 2.1.2 angerissen, geht es im nachfolgenden Kapitel um die Funktionsweise von Suchmaschinen, insbesondere um die von Google. Hierbei steht nicht der Aspekt der Spiegelung der Gesellschaft, sondern vielmehr die technische Funktionsweise der Suchmaschine im Vordergrund. Es wird vor allem untersucht, mit welchen technischen Mitteln Suchmaschinen (Google) arbeiten, um nachvollziehen zu können, wie diese in der Forschung („*Search as Research*“) zu nutzen sind. Dazu wird vorab eine Beschreibung zur Indexierung des Webs mittels Suchmaschinen vermittelt, nachfolgend werden zudem die Google Text und die Google Suchvervollständigung genauer betrachtet.

Bei Suchmaschinen handelt es sich um computer-, bzw. serverbasierte Suchsystem. Diese Systeme bestehen aus einer Vielzahl zusammengeschlossener Computer und / oder Webserver, „die unterschiedliche Funktionen erfüllen und gemeinsam die Suchmaschine bilden.“ (Lewandowski, 2015, S. 29) Wie vorab bereits durch Rogers definiert, beschränkt sich die Möglichkeit der Suche dabei auf das „sichtbare Web“. Ein zentrales Web als Datenquelle gibt es nicht – alle Daten sind zunächst unabhängig voneinander auf den verschiedenen Computern und Webservern gespeichert. Erst durch die Verbindung (Entstehung von „Hyperlinks“) zwischen verschiedenen Dokumenten entsteht „ein Netz“ –

²⁶ Die CTR beispielsweise ist bereits in der Rangreihenfolge durch den Google-Algorithmus abgedeckt

das Web. Suchmaschinen verfolgen die verschiedenen Verlinkungen zwischen den Dokumenten im Web und erfassen diese Inhalte (sog. *Crawling*²⁷). Durch ein ständiges Wiederholen des Crawlings werden alte Inhalte aktualisiert, neue Inhalte gefunden (sofern diese sichtbar verlinkt werden) und in den Kreislauf eingebunden. Die Idee des Crawlings ist es, eine Abbildung des kompletten sichtbaren Webs zu erstellen (*Indexierung*). Die Verlinkung der Inhalte ermöglicht dem Nutzer und den Suchmaschinen daher die Navigation im Web. (Lewandowski, 2015, S. 29-30) Google beispielsweise ist als Suchmaschine eine Benutzerschnittstelle, die es dem Nutzer ermöglicht „Suchanfragen einzugeben und Ergebnisse ansehen zu können.“ (Lewandowski, 2015, S. 30) Die Anordnung der Suchergebnisse erfolgt nach Lewandowski nach „systemseitig angenommener Relevanz“. (2015) Das heißt zum einen, dass die Suchergebnisse auf Basis eines Algorithmus entstehen, zum anderen, dass es keine „richtige“ Sortierung der Ergebnisse geben kann. Für unterschiedliche Nutzer können (bei der gleichen Suchanfrage) unterschiedliche Dokumente oder Quellen relevant sein. Bei den Ergebnislisten von Suchmaschinen handelt es sich somit oftmals nicht um „einfache Listen“, sondern um „komplexe Formen der Ergebnisdarstellung“. (Lewandowski, 2015, S. 30)

Google Textsuche

Google funktioniert grundsätzlich auf die vorab definierte Art und Weise. Mittels Webcrawler²⁸ (bei Google als *Googlebot* bezeichnet) werden regelmäßig die Verlinkungen im Web geprüft und im Google Index²⁹ abgespeichert. „Dabei wird besonders auf neue Websites, Änderungen an bestehenden Websites und "tote" Links geachtet.“ (Google Inc., 2015c) Der Google Index ist somit ein Abbild des sichtbaren Webs. Die Google Suche erfolgt daher nicht in Echtzeit im Web, sondern im Index, der auf das Web verweist. Sehr aktuelle Inhalte können – sofern sie noch nicht gecrawlt wurden – somit ggf. auch nicht von Google gefunden werden. Neben der Aktualität der Inhalte (*Echtzeit vs. Index Suche*) wird oftmals kritisiert, dass Google das Web nicht als Ganzes widerspiegelt. Verlinkungen von Webseiten bestehen nicht in alle Bereiche des Webs.

²⁷ Durchsuchen des Webs nach In- und Outlinks auf Webseiten, zur Indexierung von Webseiten. D.h. zur Speicherung der Daten im Index. Für ausführlichere Informationen zum Thema Crawling siehe: Baeza-Yates, R., & Ribeiro-Neto, B. (2011). *Modern information retrieval: The concepts and technology behind search*, S.515, ff.

²⁸ Software die den Vorgang des Crawlings voll automatisiert ausführt. (automatisiert das Web auf Hyperlinks durchsucht und diese im Google Index abspeichert)

²⁹ Datenbasis, die vom Crawler gefüllt wird und die Daten so aufbereitet, dass sie effizient durchsuchbar werden

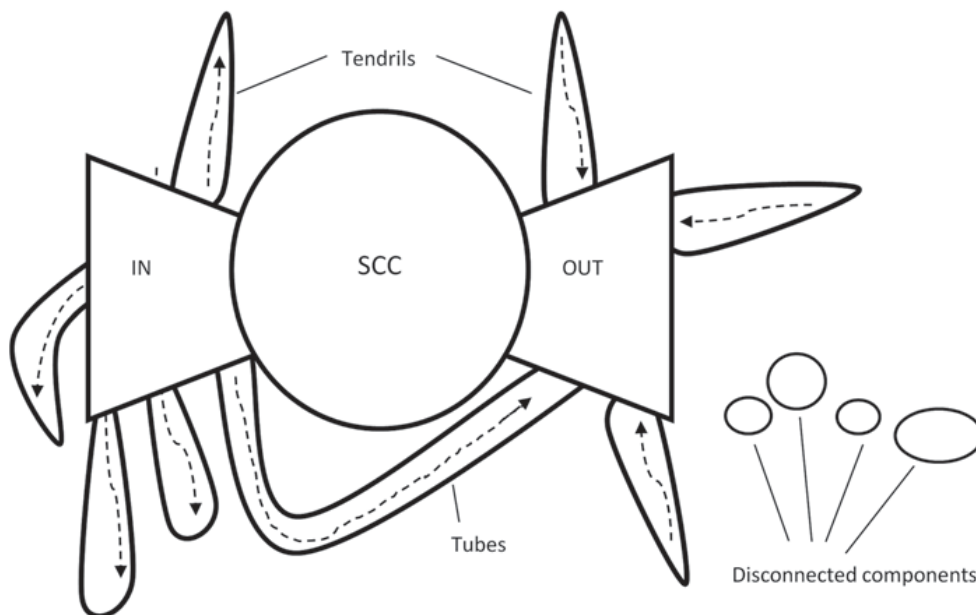


Abbildung 2-1 - Die Struktur des Webs. (Broder, et al. 2000, S. 318)

Zur Erfassung des Webs dient ein „Kernbereich“ (bei Google der *Google Index*) in dem alle Dokumente durch Hyperlinks miteinander verbunden sind. D.h., dass alle Dokumente im Google Index durch Hyperlinks aufeinander verweisen (In- und Outlinks³⁰). Zusätzlich zum Kernbereich gibt es den „Out-Bereich“. Dies sind alle Verlinkungen, die aus dem Kernbereich auf Inhalte außerhalb des Kerns verweisen. Diese Inhalte können problemlos vom Crawler gefunden werden. Hyperlinks im „In-Bereich“ wiederum sind Hyperlinks, die aus dem Out-Bereich auf den Kern verweisen, auf die aber nicht aus dem Kern zurück verwiesen wird. Diese Hyperlinks sind daher in der Regel nicht von Crawlern zu erfassen. Die betreffenden Webseiten werden somit nicht von Google indexiert. Demzufolge sind meist auch Verbindungen zwischen dem In- und dem Outbereich („Tubes“) nicht für die Crawler sichtbar. Zudem bestehen sogenannte „Tendrils“, die mit einem der 3 großen Bereiche (Kern-, In-, Out-Bereich) verbunden, „jedoch insgesamt relativ isoliert sind.“ (Lewandowski, 2015, S. 34-35) Die Indexierung durch Webcrawler dient somit der Organisation von Informationen im Web. Durch die Indexierung „lernt“ die Suchmaschine Google, wo im Web welche Begriffe und Themen zu finden und abzurufen sind. (Google Inc., 2015c)

Ziel einer Web-Suche ist es nach Google Inc. in der Regel allerdings nicht, die Seite zu finden, die einen Suchbegriff am häufigsten verwendet. Bei einer Suche mit dem Begriff

³⁰ Inlink: Ein Hyperlink von einer anderen Webseite auf die eigene Webseite; Outlink: Ein Hyperlink von der eigenen Webseite auf eine andere Webseite

„Katze“ ist zum Beispiel in der Regel nicht das Ziel, die Webseite zu finden, die den Begriff „Katze“ am häufigsten verwendet. Zweck der Suche ist es wahrscheinlich viel mehr, „Fotos, Videos oder eine Liste“ mit Katzenrassen zu erhalten. (Google Inc., 2015c) Daher werden zusätzlich zur Indexierung von Webseiten Algorithmen von Google eingesetzt, die passende Antworten zu dem / den Suchbegriff(en) liefern. (Google Inc., 2015c) Derzeit besteht der „Algorithmus von Google aus über 200 Signalen, die darauf hinweisen, was das beste Ergebnis für eine Suchanfrage sein könnte.“ (Google Inc., 2015a) Unter anderem beinhalten diese Signale auch, wie oft ein Inhalt von anderen Nutzern angeklickt wird (*CTR*), die „Begriffe auf Websites, die Aktualität des Inhalts, den eigenen Standort sowie den Page Rank.“ (Google Inc., 2015a) Das heißt, dass die Suchergebnis-Liste auch an den physikalischen Aufenthaltsort und an die Sprache, die ein Nutzer verwendet, geknüpft ist. Wie bereits vorab diskutiert, gibt Google u.a. anhand der Sprache, der IP und der angewählten ccTLD des Google Nutzers entsprechend zur gleichen Suchanfrage (also demselben Wort in der jeweiligen Landessprache) andere, länderspezifische Suchergebnisse aus. Zudem wird in der jeweiligen „Google-Landesversion“ (ccTLD) aufgrund der unterschiedlichen Suchergebnisse nachfolgend durch den Google Algorithmus auch eine unterschiedliche Rangreihenfolge der Suchergebnisse erstellt. Eine bestimmte Anzahl der unterschiedlichen Suchergebnisse von Google je ccTLD kann daher miteinander verglichen werden, um kulturelle Unterschiede zwischen den Google Versionen herauszuarbeiten. Die Analyse der Textsuche bietet somit theoretisch ein Mittel, eine kulturvergleichende Forschung zu erstellen.

Neben der Auflistung von gesuchten Inhalten auf Basis des Algorithmus ist es zudem möglich, Werbung (Anzeigen) bei Google zu platzieren. Werbeplatzierungen werden in der Regel über den ersten generischen Suchergebnissen angezeigt und sind als solche in der Regel durch ein gelbes Feld mit der Aufschrift „Anzeige“ kenntlich gemacht. (Abbildung 2-2 - Werbeanzeigen vs. Web-Suche google.de, Suchbegriff "Schuhe"; Screenshot Autor 23.07.2015) Diese „Anzeigenplätze“ werden von Unternehmen z.B. auf Klick-Basis gekauft und sind in dem Sinne nicht wie generische Suchergebnisse durch den Page Rank, bzw. den Google Algorithmus beeinflusst. Werbeanzeigen werden daher nicht im praktischen Teil der Arbeit betrachtet, da diese nicht, wie durch Rogers definiert generisch, im Web entstanden sind, sondern künstlich als relevanter Inhalt zu bestimmten Suchbegriffen durch SEA³¹

³¹ SEA = Search Engine Advertising, bedeutet: Platzierung von Werbeinhalten bei Suchmaschinen in Zusammenhang mit bestimmten Suchbegriffen

platziert wurden. Suchergebnisse die durch SEO³² beeinflusst sind können in der nachfolgenden Studie nicht ausgeschlossen werden, da SEO Maßnahmen via Google nicht direkt erkennbar sind.

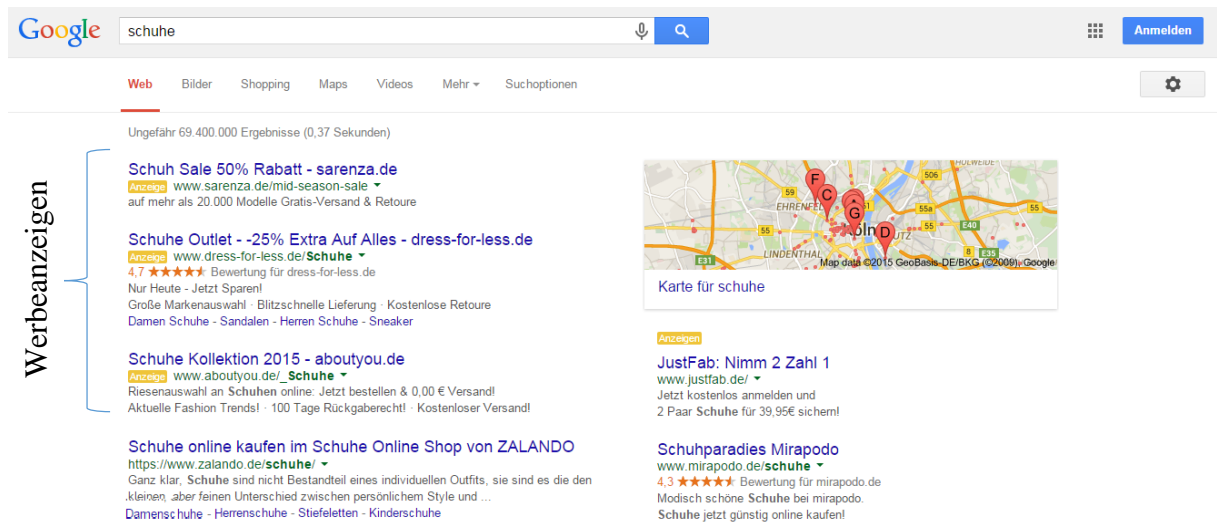


Abbildung 2-2 - Werbeanzeigen vs. Web-Suche google.de, Suchbegriff "Schuhe"; Screenshot Autor 23.07.2015

Google Suchvervollständigung

Bei der Eingabe einer Suchanfrage in die Suchmaske von Google, liefert die Suchmaschine adhoc eine automatische Vervollständigung, die mit dem eingegebenen Suchbegriff in direktem Zusammenhang steht. Dabei werden dem Nutzer von Google diverse Vorschläge unterbreitet (Abbildung 2-3), die er / sie im Zusammenhang mit Suchbegriff wahrscheinlich ausführen möchte und die in Zusammenhang mit seinen Nutzerdaten stehen (Formulardaten, Cookies, Google Account Daten, Suchverlauf etc.).

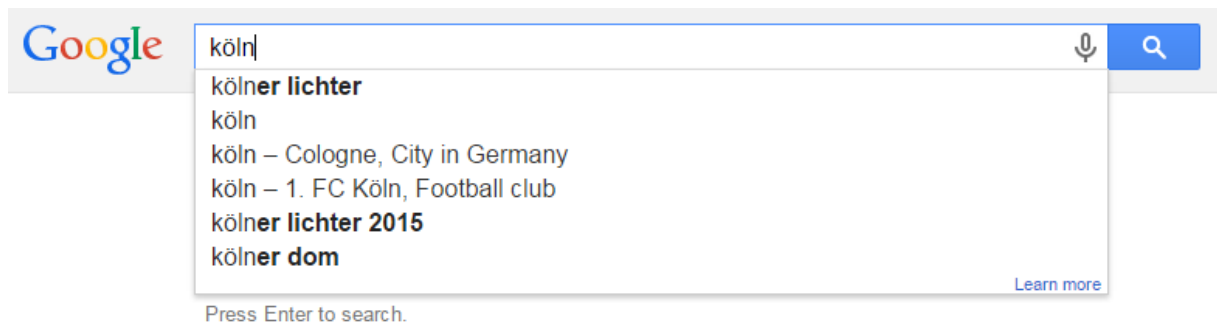


Abbildung 2-3 - Suchvervollständigung google.de, Suchbegriff "Köln"; Screenshot Autor 23.07.2015

Die automatische Vervollständigung des Suchbegriffs³³ wird von Google auf Basis „häufiger Suchanfragen in Bezug auf“ den Suchbegriff selbst eingeblendet. (Google Inc., 2015b) „Die über die automatische Vervollständigung angezeigten Suchanfragen stammen

³² SEO = Search Engine Optimization bedeutet: Optimierung von Web Inhalte, so dass die Findability steigt

³³ Siehe auch Fußnote 8: Autovervollständigung, Suchvervollständigung, oder Auto-Fill

aus den Suchaktivitäten anderer Nutzer und aus dem Inhalt von Webseiten.“ (Google Inc., 2015b) Dabei basiert die automatische Vervollständigung der Suchbegriffe, wie die Suche selbst auch, auf einem Algorithmus (Geo-IP, ccTLD, Sprache, Page Rank, CTR, Nutzerdaten, etc.) und erfolgt daher nach Google auf der „Grundlage objektiver Faktoren“. (Google Inc., 2015b) Unter anderem wird vom Algorithmus beachtet, wie viele „Nutzer in der Vergangenheit nach einem Begriff“ in Zusammenhang mit einem anderen Begriff gesucht haben, oder welche Suchergebnisse nachfolgend angeklickt wurden. (Google Inc., 2015b) Für das oben aufgeführte Beispiel gilt bei der hier aufgeführten Suche nach „Köln“ somit, dass um den 23.07.2015 viele Nutzer und oder viele Webseiten im Raum Deutschland (Köln) zuletzt verstärkt den (Such-)Begriff „Kölner Lichter“, „Kölner Lichter 2015“, oder „Kölner Dom“ in Zusammenhang mit dem Suchbegriff „Köln“ gesucht, bzw. verwendet haben. Begriffe, die in Zusammenhang mit dem Suchbegriff durch die automatische Vervollständigung von Google generiert werden, sind somit Begriffe die Nutzer häufig in Zusammenhang mit dem Suchbegriff setzen. Die Suchvervollständigung beachtet zudem auch „Signale“ wie die Sprache, den lokalen Standort (Geo-Lokalisierung) eines Nutzers und die angewählte ccTLD. Vereinfacht ausgedrückt bedeutet dies: die Auto-Vervollständigung von Google repräsentiert das Suchverhalten, bzw. das Interesse einer Vielzahl der Nutzer je Land (wie vorab definiert durch: Geo-Lokalisierung / ccTLD / Sprache) in der Art, dass Sie die am häufigsten gesuchten und geklickten Inhalte zum Suchbegriff miteinander kombiniert. Die Auto-Vervollständigung kann daher als eine Art grobe Zusammenfassung des Such- und Nutzungsinteresses von den Google-Nutzern je Land angesehen werden. Mit Hilfe der Geo-Lokalisierung und der Anwahl verschiedener ccTLDs ist daher davon auszugehen, dass auch die Analyse der Auto-Vervollständigung von Google ein Mittel zur kulturvergleichenden Forschung bereitstellen könnte. Unter anderem ist es auch denkbar, die Analyse der Textsuche mit der Analyse der Auto-Vervollständigung zu kombinieren, um die Trennschärfe und die Reliabilität der Gesamtaussage zur kulturvergleichenden Forschung zu verbessern.

2.2.4 Virtual Private Network und Tunnelung

Wie bereits vorab diskutiert, werden die Suchergebnisse von Google auch durch den physikalischen Aufenthaltsort eines Nutzers (Geo-IP) beeinflusst. Ein kultureller Vergleich mittels Google muss daher die Standortlokalisierung einbeziehen, um Verzerrungen in den Suchergebnissen aufgrund der Standortlokalisierung bei der Google Suche zu vermeiden. Bei der Anwahl verschiedener ccTLDs muss daher eine zur ccTLD passende lokale IP

Adresse je Land vorliegen. Wird z.B. *google.co.uk* aufgerufen, muss dies mit einer englischen IP Adresse erfolgen; *google.tk* muss mit einer türkischen, *google.se* hingegen mit einer schwedischer IP Adresse, usw. angewählt werden. VPN Provider ermöglichen es, IP Adressen zu wechseln und die Standortlokalisierung zu beeinflussen. Da es allerdings mehrere Arten von „virtual private networks“³⁴ (VPN) gibt, liegt keine einheitliche Definition des Begriffs VPN vor. VPNs können z.B. genutzt werden, um Mitarbeitern einen sicheren Fernzugriff auf ein Firmennetzwerk zu gewährleisten, oder aber z.B., um die eigene Identität beim Surfen im Web zu verschleiern. Vereinfacht ausgedrückt handelt es sich bei einem VPN um eine Netzwerk Virtualisierung, „die ein öffentliches Netzwerk benutzt“ um (private) Daten zu transportieren. (Lipp, 2006, S. 17) Im Rahmen der Masterarbeit wird ein VPN Tunnel genutzt, um unterschiedliche Standorte mit Hilfe verschiedener VPN Server zu simulieren.

Üblicherweise verbindet sich ein Nutzer mit dem Endgerät, mit dem er sich ins Web einwählt, über eine verschlüsselte Verbindung direkt mit dem Internet Service Provider (ISP). Der ISP vergibt für den Nutzer eine IP Adresse, mit der der Nutzer im Web sichtbar ist und an die die Inhalte der aufgerufenen(n) Webseite(n) zurückgesandt werden. Der Nutzer kann daher über seine IP / den ISP identifiziert und lokalisiert werden. (Abbildung 2-4)

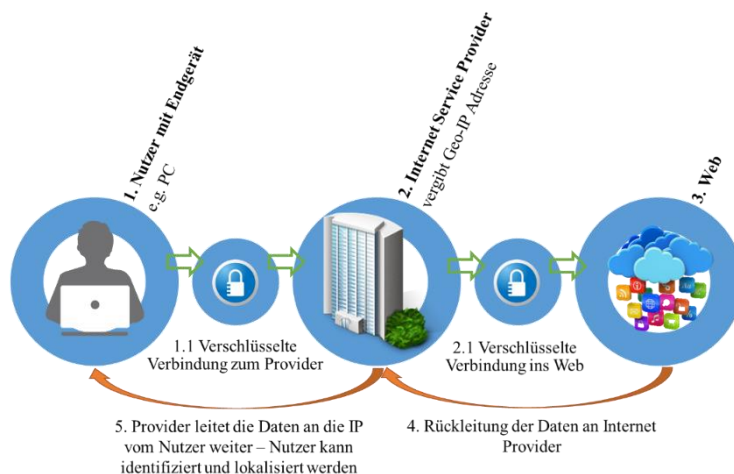


Abbildung 2-4 - Internetverbindung ohne VPN Server (Schaubild von Autor)

Verbindet sich der Nutzer hingegen mit einem VPN Server, werden die Daten nach der Anmeldung beim ISP zunächst an den VPN Server gesendet. Der VPN Server vergibt nun wiederum je nach Server-Standort eine neue IP Adresse für den Nutzer. Die ursprünglich durch den ISP vergebene IP wird durch die IP des VPN Servers ersetzt. Wird bei einem VPN

³⁴ Übersetzung nach Baun: virtuelle private Netze; Definition: „logische Teilnetze innerhalb öffentlicher Netze (zum Beispiel das Internet)“ (2012, S. 185)

Provider beispielweise ein Server in Istanbul angewählt, verbindet sich der Rechner nach der Anmeldung beim ISP automatisch mit dem VPN Server in Istanbul. Nachfolgend werden alle Anfrage, die über das Endgerät des Nutzer laufen (z.B. das Aufrufen einer Webseite) über den VPN Server, also mit einer türkische IP Adresse, getätigt. Webseiten z.B. gehen somit davon aus, die Anfrage stamme aus der Türkei, da nun der VPN Server die Anfrage bei der Webseite tätigt, nicht länger der ISP. Die Daten werden von der Webseite also an den VPN Server zurückgeleitet. Der VPN Server wiederum schickt das Datenpaket direkt an das Endgerät, bzw. den Nutzer – der ISP wird bei diesem Vorgang somit komplett umgangen. Es entsteht ein VPN Tunnel („Tunnelung“), die es ermöglicht, von einem bestimmten physikalischen Standort aus (z.B. Deutschland) verschiedene Geo-Lokalisierungen über einen VPN Server zu simulieren (z.B. England, Türkei, Schweden). Der ISP kann die Verbindung des Nutzers nach der Einwahl beim VPN Server nicht mehr verfolgen. Alle Daten werden fortan über den VPN Tunnel mit einer „neuen IP“ direkt an den Nutzer gesendet. (Abbildung 2-5) Die Möglichkeit, ein anderes Land mittels VPN zu simulieren (bzw. die IP Adresse eines anderen Landes zu simulieren), hängt demzufolge von der Anzahl der Server in verschiedenen Ländern ab, die ein VPN Provider bereitstellt.

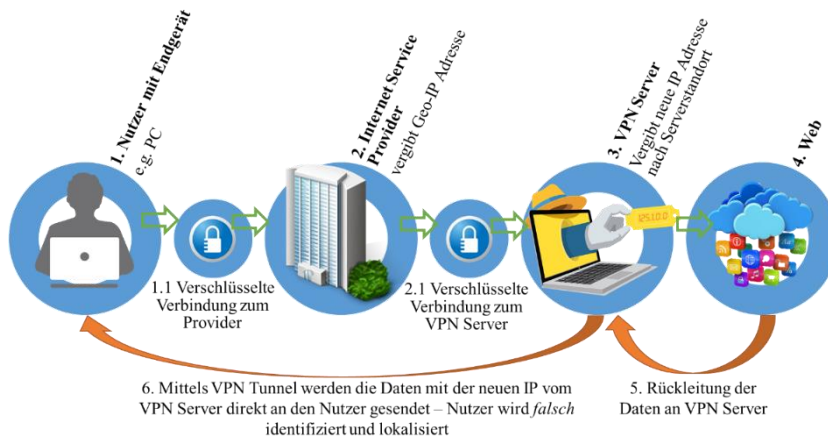


Abbildung 2-5 – Internetverbindung mit VPN Server (Schaubild von Autor)

Die Einwahl bei einem VPN Server kann zum Beispiel durch eine vom VPN Provider bereitgestellte Software erfolgen, oder z.B. im Netzwerk- und Freigabecenter bei Windows eingerichtet werden. Software, die vom VPN Provider bereitgestellt wird, ermöglicht oftmals eine intuitive Bedienung mittels graphischer Benutzeroberfläche (engl. GUI³⁵). Zudem erleichtert die Bedienung des VPN mittels GUI den Serverwechsel.

³⁵ General User Interface; Definition: “Als Graphical User Interface (GUI) bezeichnet man die grafische Benutzeroberfläche, die das Bedienen der Computer erleichtert. Die grafische Benutzeroberfläche gehört zum

2.3 Soziologie: Validierung der Daten

Da die in der Praxis angewandte webwissenschaftliche Methode (*siehe auch Kapitel 2.1.2 – Search as Research*) auf einen Kulturvergleich zwischen mehreren Mitgliedsstaaten (Gesellschaften) aus Europa abzielt, wird nachfolgend eine kurze Einführung in den Begriff der Soziologie vermittelt. Dabei wird bewusst auf ein tiefergehendes Studium der soziologischen Grundlagen verzichtet, um dem Rahmen der Masterarbeit gerecht zu werden.

Der Begriff Soziologie ist auf den Wissenschaftler August Comte (1798 – 1857) zurückzuführen und ist die „Wissenschaft vom Sozialen“. (Schäfers, 2013, S. 19) Soziologie wird nach Schäfers definiert als,

„eine empirische Sozialwissenschaft. Sie untersucht die Strukturen des sozialen Handelns und die Formen der Vergemeinschaftung und Vergesellschaftung, unter Berücksichtigung der Normen und Werte, sozialen Prozesse und Institutionen, die die Integration der Gesellschaft und den sozialen Wandel bewirken.“ (2013, S. 19)

Die Soziologie untersucht nach Schäfers somit auf wissenschaftlicher Basis das Handeln – die Aktionen und Reaktionen – von Menschen in einem gemeinsamen sozialen Kontext und dessen Entwicklung. Auch nach Meulemann spricht man von der Soziologie als „Wissenschaft vom Zusammenleben der Menschen“. (2013, S. 29) Das Zusammenleben der Menschen begründet Meulemann, neben der Befriedigung der Elementar-Bedürfnisse, durch ein bestimmtes Verhalten, bzw. bestimmte Grundsätze, wie „Grußformeln, Gesetze, Ehen, Nationalstaaten“, etc. (2013, S. 32) Diese gemeinsamen Zusammenhänge sind nach Meulemann die Regeln des Zusammenlebens und auch der Grund dafür, dass der Mensch sich eine „Sozialnatur“ aneignen muss. Das heißt, dass die gemeinsamen Kontexte, die auf „Technik, Zivilisation, Wirtschaft, Kultur und Tradition“ beruhen, durch ihr Bestehen und ihre Adaption durch den Menschen, Gesellschaften erschaffen (Meulemann, 2013, S. 33) Wenn die Soziologie nach Meulemann also die *Wissenschaft vom Zusammenleben der Menschen* ist, dann untersucht sie Technik, Zivilisation, Wirtschaft, Kultur und Tradition einer Gesellschaft im Rahmen des sozialen Gefüges (der Gesellschaft). Auch Strutz und Wagner (1980, S. 8) definieren Soziologie in ähnlicher Weise. Nach Strutz und Wagner bilden sich mit der Zeit durch die Weitergabe („Lehren und Lernen“) verschiedener „Normen, Verhaltensregeln, Sitten und Gebräuche“ unterschiedliche Gesellschaften. (Strutz & Wagner, 1980, S. 13)

Betriebssystem und erleichtert dem Benutzer den Aufruf von Anwendungsprogrammen, Diensten und Funktionen. Grafische Benutzeroberflächen sind übersichtlich aufgebaut und umfassen je nach Anwenderprogramm mehrere Funktionsbereiche, wie die Menüleiste, Symbolleiste, Funktionsleiste, Statusleiste, Taskleiste und Bildlaufleisten.“ (Lipinski, 2015)

[sic!] So vermittelt zum Beispiel die Muttersprache (Sprache ist wesentliches Element einer Kultur!) gemäß ihrer jeweiligen Form allen ihren Trägern eine gewisse gemeinsame Denkweise, die sich in vielem von anderen Sprachen unterscheidet.

Somit wird ein Individuum nach Strutz und Wagner durch sein Umfeld sozialisiert und passt sich an die Gesellschaft an. Die Soziologie untersucht diesen Vorgang und das Verhalten der Menschen in diesem sozialen Gefüge. Darüber hinaus sehen sie die Funktionen der Soziologie in 3 Bereiche gegliedert: die Informationsfunktion, die Stabilisierungsfunktion und die kritische Funktion. (1980, S. 8) Die Informationsfunktion dient dazu, „systematische Informationen über die Gesellschaft, ihre Strukturen und Wandlungen“ bereit zu stellen. Je nach Funktion und Einsatz kann die Soziologie den Autoren zufolge daher „politisch-gesellschaftlich stabilisierend wirken.“ Die kritische Funktion fordert darüber hinaus, dass die Soziologie unabhängig ist, so dass sie „[sic!] ein kritisches Bewusstsein in der Öffentlichkeit erzeugen und wachhalten, die Aufmerksamkeit auf bestimmte gesellschaftliche Missstände hinlenken“ kann. (Strutz & Wagner, 1980, S. 8)

2.3.1 Kulturvergleichende Forschung

Die Theorie zur kulturvergleichenden Forschung dient zum einen der Auswertung und Interpretation der mittels webwissenschaftlichen Methode erhobenen Daten. Zum anderen werden die *European Values Studie* (EVS), der *European Drug Report 2015* (EMCDDA) und der *Global status report on alcohol and health* (WHO) als Studien der kulturvergleichenden Forschung, zur Validierung der Daten herangezogen. Gemäß der abgefassten Definition von Soziologie, ist Kultur als ein Teil-Aspekt der Soziologie zu verstehen. Sie prägt das Zusammenleben der Menschen und bietet u.a. eine Möglichkeit der Abgrenzung verschiedener Gesellschaften auf Basis verschiedener kultureller Grundgedanken. Eine einheitliche Definition des Begriffs Kultur ist nach Rippel und Seipel jedoch nicht gegeben; folglich ist auch der Begriff *kulturvergleichende Forschung* bisweilen nicht einheitlich zu definieren. (Rippel & Reipel, 2015, S. 15) Bei beiden Begriffen handelt es sich den Autoren zufolge eher um „weit gefasste Sammelbegriffe“. (2015, S. 16) Im Folgenden wird Kultur als ein „System von geteilten Werten“ definiert. (Rippel & Reipel, 2015, S. 17) Werte sind in diesem Sinne zu verstehen als „Kriterien, die Menschen zur Selektion, Evaluation und Rechtfertigung von Handlungen dienen“. (ebd.) Sie repräsentierten somit auch die unterschiedlichen Auffassungen und Einstellungen verschiedener Kulturen zu bestimmten Themen. Werte stehen in diesem Kontext somit in einem sozialen Zusammenhang, da sie von den Individuen einer Kultur geteilt werden („geteilte Wertvorstellungen“). (ebd.) Die kulturvergleichende Forschung ermöglicht es im

Sinne dieser Definition somit, verschiedene Kulturen auf Basis ihrer Werte – also ihrer Auffassungen und Einstellungen – zu untersuchen und zu vergleichen. Dazu können verschiedene Kulturen anhand ihrer jeweiligen Gesellschaften (z.B. Land, Sprache) erfasst werden. Die vergleichende Methode bietet ein systematisches Verfahren zur empirischen Generalisierbarkeit, der Überprüfung von Thesen, bzw. zur Formulierung von Hypothesen im Hinblick auf die Gesellschaft(en) / die Kultur(en) an. Die kulturvergleichende Forschung kann dementsprechend deskriptiver, definierender, prognostizierender oder interpretativer Natur sein. Es können Besonderheiten und Abweichungen, oder Übereinstimmungen zwischen den Untersuchungsobjekten hervorgehoben werden. Die Typologisierung und Katalogisierung der verschiedenen Merkmale von Gesellschaften ermöglicht es, Untersuchungsobjekte mit gleichen Merkmalen zusammenzufassen, bzw. die Untersuchungsobjekte voneinander abzugrenzen. Auf dieser Basis können Gesellschaften mit möglichst homogenen Merkmalen zu „Clustern“ zusammengefasst werden und sich wiederum zu den anderen heterogenen Gesellschaften / Clustern abgrenzen. In Anlehnung an Meulemann kann demzufolge dann von vergleichender Sozialforschung gesprochen werden, wenn die Wahrnehmung von Technik, Zivilisation, Wirtschaft, Kultur und Tradition in verschiedenen Gesellschaften empirisch erhoben und statistisch ausgewertet wird, so dass die Ergebnisse miteinander verglichen werden können.

2.3.2 Vergleichsdaten

Als Hauptquelle für den Datenabgleich der Themen Homophobie und Ausländerfeindlichkeit wird die European Values Study (EVS) genutzt. Diese Themen bieten sich für die Nutzung der EVS zum Abgleich mit den Google Daten besonders an, da sie durch die Studie mit mehreren Items erhoben werden. Die EVS ist eine „transnationale empirische Langzeitstudie“, die in den 1970er Jahren von Sozialforschern initiiert wurde. (GESIS – Leibniz-Institut für Sozialwissenschaften e.V., 2015) Ziel der EVS ist es herauszufinden, ob innerhalb der EU Staaten eine gemeinschaftliche Wertebasis besteht oder nicht und wie sich die Werteeinstellung der Europäer verändert. Die Studie befasst sich mit „den Ideen, dem Glaube, den Vorlieben, Einstellungen, Werten und Meinungen der europäischen Bürger“³⁶ und versucht die „Vorstellungen der Europäer zu Familie, Arbeit, Religion, Politik und Gesellschaft“ zu ermitteln. (ebd) Eine der Grundsatzfragen des Projekts

³⁶ Übersetzung durch den Autor: „...the ideas, beliefs, preferences, attitudes, values and opinions of citizens all over Europa“ (European Values Study, 2014a)

lautet daher: „Wohin entwickelt sich Europa?“³⁷. Die Stichprobe zur Studie wird aus der Grundgesamtheit aller volljährigen Einwohner (18 Jahre und älter) eines Land gezogen. Der Prozess der Stichprobenziehung basiert auf einer *mehrstufigen Stichprobenauswahl*, oder auf einer *geschichteten Zufallsstichprobe*. (European Values Study, 2014b) Die Datenerhebung erfolgt auf Basis standardisierter face-to-face Interviews. (ebd) Seit 1981 wurden 4 Befragungswellen, die auf „die Analyse von Werteunterschieden und – Gemeinsamkeiten sowie Wertewandel unter den Bürgern der EVS Mitgliedsländer gerichtet“ sind, durchgeführt. (GESIS – Leibniz-Institut für Sozialwissenschaften e.V., 2015) Die Liste der teilnehmenden Staaten wurde dabei stetig erweitert. (Abbildung A 1) Die aktuellsten Daten der Studie beinhalten Angaben aus 47 europäischen Staaten von ~70.000 Interview-Partnern und beziehen sich auf das Jahr 2008. (European Values Study, 2014a) Heute ist die EVS Foundation Träger der Studie. Alle Daten und die Dokumentation zum Projekt stehen frei zugänglich über das Web im GESIS Datenarchiv kostenlos zur Verfügung.

Der Abgleich der Daten zum Thema *Akzeptanz von Rauschmitteln* hingegen erfolgt zum einen auf Basis des „*Global status report on alcohol and health*“ der WHO. Zum anderen auf den Daten des *European Monitoring Centre for Drugs and Drug Addiction – “European Drug Report 2015”* (EMCDDA). Da die EVS die Items zum Thema Alkoholkonsum, bzw. Konsum von leichten Drogen (Marihuana) seit den 1980er Jahren in einem Großteil der Länder nicht weiter erhebt, wird auf die aktuelleren Daten der WHO, bzw. der EMCDDA zurückgegriffen. Die Daten der WHO beziehen sich dabei auf den pro-Kopf Konsum (15 Jahre und älter) reinen Alkohols je Land pro Jahr. Die Daten des EMCDDA bilden den Marihuana Konsums nach europäischen Ländern, auf Basis der Menge an Konsumenten je Land, im Jahr 2014 ab. Beide Studien bieten sowohl in ihrer Seriosität, als auch in ihrem Umfang eine valide Datenbasis zum Abgleich mit den erhobenen Google Daten.

³⁷ Übersetzung durch den Autor: „Where is Europe heading?“ (GESIS – Leibniz-Institut für Sozialwissenschaften e.V., 2015)

3 Praxis

3.1 Generierung des Datenmaterials

Für den Zugriff auf Google wurde der Webbrowser *Internet Explorer* (IE; Version 11.0.9600.18015) verwendet. Im Browser wurde keine Startseite festgelegt (*Internetoptionen – Startseite: about:blank*), so dass beim Öffnen des Browsers keine Webseite(n) angewählt wurde. Der Standortzugriff für Webseiten war über die Browser Einstellungen freigegeben, so dass der Standort von besuchten Webseiten über die IP Adresse im Browser erfasst werden konnte. Der Browserverlauf wurde mit dem Schließen des Internet Explorers gelöscht, auch temporäre Daten wie „Cookies, gespeicherte Kennwörter und Webformulardaten“ (Microsoft Corporation, 2015) wurden mit dem Beenden des Browsers gelöscht. Nach dem Öffnen des Browsers wurden keine weiteren Webseiten als die jeweils relevante Google Landesversion (IP / Sprache / ccTLD) geöffnet. Nach dem Abschluss der Suche zu einem Themenbereich durch Eingabe eines Suchbegriffs wurden die Daten (Suchvervollständigung und Textsuche) gesichert und der Browser geschlossen. Damit wurde sichergestellt, dass in den nachfolgenden Suchanfragen zu anderen Themen, bzw. auf anderen ccTLDs keine Beeinflussung durch den Verlauf, Formulardaten oder die Speicherung von Cookies (aus den vorigen Suchanfragen) gegeben war. Um eine valide und reliable Suche je Google Landesversion (ccTLD) durchführen zu können, wurde zudem der VPN Hoster *Hide My Ass (HMA)* verwendet. Die Soft-, bzw. Hardware des Anbieters erlaubt es dem Nutzer, wie in Kapitel 2.2.4 beschrieben, unterschiedliche Standorte mit Hilfe eines VPN Servers zu simulieren. HMA wurde aufgrund der großen Anzahl an Servern gewählt. Der Provider bietet mit „890 VPN Servern an 340 Standorten in 190 Länder“ mehr Zugangsoptionen als alle anderen VPN Hoster. (Privax LTD (GB), 2015) Um Verzerrungen aufgrund geographischer Lokationen durch verschiedenen Serverstandorte je Land (Großstadt / Kleinstadt / ländliche Gebiete) zu umgehen, wurden nach Möglichkeit Server in den Landeshauptstädten, bzw. in Ballungsgebieten angewählt. Wenn in der Landeshauptstadt kein Server via HMA verfügbar war, wurde ein Server in der nächstgrößten Stadt (Bevölkerungsanzahl) gewählt. Teilweise war je Land nur ein bestimmter Serverstandort gegeben, weshalb dieser angewählt werden musste. Weitere Ausnahmen zur Wahl des Serverstandorts werden zudem in der nachfolgenden Tabelle (3-1) erörtert. Die Tabelle zeigt die jeweils angewählten Serverstandorte via HMA, die genutzte ccTLD, sowie die zugehörige genutzte Sprache für die Suche auf der jeweiligen Google Landesversion.

Staat	Serverstandort (VPN)	Besonderheit	ccTLD (Google Landesversion)	Such-Sprache
Dänemark*	Kopenhagen	Landeshauptstadt	google.dk	Dänisch
Deutschland	Berlin	Landeshauptstadt	google.de	Deutsch
England	London	Landeshauptstadt	google.co.uk	Englisch
Finnland*	Helsinki	Landeshauptstadt	google.fi	Finnisch
Frankreich*	Paris	Landeshauptstadt	google.fr	Französisch
Irland	Dublin	Landeshauptstadt	google.ie	Englisch
Italien*	Mailand	Rom nicht gegeben, Mailand größter wählbarer Standort	google.it	Italienisch
Kroatien*	Zagreb	Landeshauptstadt	google.hr	Kroatisch
Niederlande*	Amsterdam	Landeshauptstadt	google.nl	Niederländisch
Norwegen*	Oslo	Landeshauptstadt	google.no	Norwegisch
Österreich	Vienna	Landeshauptstadt	google.at	Deutsch
Polen*	Warschau	Landeshauptstadt	google.pl	Polnisch
Rumänien*	Bukarest	Landeshauptstadt	google.ro	Rumänisch
Schweden*	Stockholm	Landeshauptstadt	google.se	Schwedisch
Schweiz	Zürich	Deutschsprachige Schweiz, Kodierung auf Deutsch möglich. Zudem repräsentativ für die Schweiz, da flächenmäßig ein Großteil der Schweiz Deutschschweiz (Abbildung A 2)	google.ch	Deutsch
Spanien*	Madrid	Landeshauptstadt	google.es	Spanisch
Türkei*	Istanbul	Landeshauptstadt	google.com.tr	Türkisch
Ukraine*	Odessa	Kein anderer Serverstandort via HMA in der Ukraine gegeben	google.com.ua	Russisch
Ungarn*	Budapest	Landeshauptstadt	google.hu	Ungarisch
Weißrussland*	Minsk	Landeshauptstadt	google.by	Russisch

Tabelle 3-1 - Übersicht der genutzten Serverstandorte, ccTLDs und Such-Sprachen (eigene Darstellung)

Die Google-Suche wurde somit kohärent zum Server auf der zugehörigen ccTLD in der Regel in der jeweiligen Landessprache, bzw. der territorial am häufigsten genutzten Sprache durchgeführt. (Tabelle A 1) Aufgrund der Vielzahl der untersuchten Länderversionen von Google und der vielen verschiedenen (Such-)Sprachen, wurden die Suchbegriffe, bzw. die Suchergebnisse mit Hilfe von Übersetzern, zumeist Muttersprachlern oder Menschen, die längere Zeit vor Ort in dem jeweiligen Land gelebt haben, übersetzt. Die Google Suche in einem mit ‚Sternchen‘ (*) gekennzeichneten Land (Tabelle 3-1) erfolgte mit Hilfe eines Übersetzers. Das heißt, es wurden sowohl die Suchbegriffe mit dem jeweiligen Übersetzer abgestimmt (korrekte Rechtschreibung, richtige Bedeutung des Wortes, richtige Übersetzung, etc.), als auch nachfolgend die Suchergebnisse übersetzt. Der Autor der Masterarbeit war bei der Übersetzung anwesend und konnte so simultan zur Übersetzung die Kodierung der Analyseeinheiten vornehmen. Zum einen wurde somit eine korrekte Übersetzung der Suchbegriffe und der Analyseeinheiten forciert, zum anderen konnten die Übersetzer die jeweilige Kultur und konnten somit ggf. auf Besonderheiten (z.B. Differenz in sprachlicher Nutzung des Webs) bei der Suche hinweisen. So wurde beispielsweise auf Empfehlung der Übersetzer (beide muttersprachlich in den jeweiligen Ländern groß geworden), Russisch als Such-Sprache in den Ländern Ukraine und Weißrussland genutzt, da laut Übersetzer Russisch im Alltag im Web eher Gebrauch findet als die jeweilige Landessprache (Ukrainisch, bzw. Weißrussisch) selbst. Alle nicht gekennzeichneten Länder wurden ohne Übersetzungshilfe direkt vom Autor untersucht und verkodet. Die komplette Datenerhebung erfolgte im Zeitraum vom 01.09.2015 bis zum 30.09.2015. Als Suchbegriff zur Datengenerierung zu den Themenkreisen Homophobie, Ausländerfeindlichkeit und Toleranz von Rauschmitteln dienten die Wörter „Homosexualität“, „Ausländer“ und „illegale Drogen“ in der jeweiligen Landessprache. Bei der Suche der Begriffe „Homosexualität“ und „Ausländer“ handelt es sich um ein explizites Vorgehen zur Generierung des Datenmaterials, d.h. Analyseeinheiten können sowohl negative, neutrale, als auch positive Aussagen zu den Themen beinhalten und müssen daher explizit auf deren Inhalte untersucht werden. Da der Suchbegriff „illegale Drogen“ bereits einen negativen, bzw. implizit wertenden Suchbegriff enthält (*illegal*), wurde das Vorgehen für den Themenkreis Toleranz von Rauschmitteln methodisch leicht abgewandelt. Die 100 Analyseeinheiten zum Thema Toleranz von Rauschmitteln wurden nicht konkret auf den Inhalt, sondern lediglich auf die Anzahl der Analyseeinheiten, die die Wörter Marihuana, bzw. Alkohol enthielten, untersucht. Da der Suchbegriff „illegale Drogen“ den negativen Zusammenhang zum Thema bereits impliziert, wird das Erscheinen der Wörter Alkohol oder

Marihuana als intolerant gegenüber dem jeweiligen Rauschmittel gewertet. Im Folgenden wird daher zwischen einem expliziten und impliziten Vorgehen, bzw. einer expliziten und impliziten Suche unterschieden.

3.2 Stichprobenziehung und Analyseeinheiten

Der Stichprobenumfang wurde vorab je Land auf die ersten 100 Treffer (Top 100), sowie auf die Ausgabe der Suchvervollständigung (4, bzw. 10 Analyseeinheiten je Land) der Google Suchergebnisse begrenzt. Die gesamte Stichprobe umfasst somit 6.840 Analyseeinheiten (20 Länder zu 3 Themen (Suchbegriffe) à 114 Analyseeinheiten). Für die Datenerhebung im Exkurs wurden zudem zusätzlich weiter 4.000 Analyseeinheiten kodiert. Es wurde angenommen, dass die ersten 100 Treffer, bzw. die Suchvervollständigung der Google Suchergebnisse, gemäß der Funktionsweise des Google Algorithmus die relevantesten und wichtigsten Themen und Inhalte zur jeweiligen Suchanfrage der Google Länderversionen repräsentieren. Auch aufgrund des Umfangs der Arbeit wurden die Anzahl der Analyseeinheiten (Anzahl an Übersetzungen, Kodierung und Auswertung von insgesamt 10.840 Analyseeinheiten) auf die Top 100 Treffer der Suchergebnisse in der Trefferliste der Textsuche von Google je Land begrenzt. Hyperlinks in der Trefferliste von Google wurden nicht beachtet, d.h., es wurden keine Hyperlinks von Google auf andere Webseiten geöffnet. Lediglich die ersten 100 Suchergebnisse wurden als eigenständige, rein textuelle Analyseeinheiten betrachtet. Die Analyse und Kodierung der Suchergebnisse basiert daher rein auf dem via Google sichtbaren Text. Im nachfolgenden Screenshot (Abbildung 3-1) sind exemplarisch 2 Analyseeinheiten dargestellt. Je Analyseeinheit wurde sowohl die Überschrift (1, blau), als auch der Hyperlink (als reiner Text; 2, grün) und der Textkörper (3, grau) unter dem Hyperlink betrachtet und analysiert.



Abbildung 3-1- Beispiel 4 Analyseeinheiten Textsuche bei google.de zum Thema Homosexualität (Screenshot, google.de)

Zudem wurde jeweils die Suchervollständigung von Google je Länderversion untersucht. Die Suchervollständigung beschränkt sich in der vorliegenden Masterarbeit auf 4, bzw. 10 Analyseeinheiten je Land. Da Google zu den Suchanfragen mit den vorab definierten Browser Einstellungen (keine Cookies, keine User Informationen per Browser) nicht mehr als 4 Suchervollständigungen je Suchanfrage ausgibt, wurde zusätzlich das Tool *Google Autocomplete* der Digital Methods Initiative³⁸ (DMI) verwendet, um weitere Suchervollständigungen zu erhalten. Via *Google Autocomplete* werden 10 Suchervollständigungen je Land ausgegeben. Die DMI bietet über ihre Website freien Zugriff auf das Tool.³⁹ Die nachfolgende Abbildung 3-2) zeigt 4 Analyseeinheiten der manuellen Suchervollständigung der deutschen Google Landesversion. Diese wurde manuelle über die Google Textsuche (IE, ccTLD, Sprache, Geo-IP) erstellt.

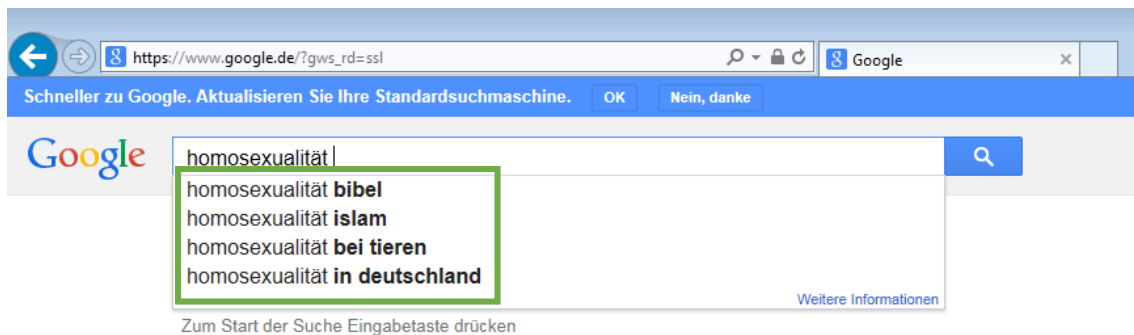


Abbildung 3-2 - Beispiel 4 Analyseeinheiten Suchervollständigung bei google.de zum Thema Homosexualität (Screenshot von: google.de)

Abbildung 3-3 zeigt einen Screenshot des *Google Autocomplete* Tools, ebenfalls mit den gewählten Einstellungen für die Google Deutschland Version. (Abbildung A 3)

word	country	language	suggestion
Homosexualität	Germany	German	homosexualität
Homosexualität	Germany	German	homosexualität im islam
Homosexualität	Germany	German	homosexualität bei tieren
Homosexualität	Germany	German	homosexualitäten
Homosexualität	Germany	German	homosexualität ursachen
Homosexualität	Germany	German	homosexualität in der bibel
Homosexualität	Germany	German	homosexualität krankheit
Homosexualität	Germany	German	homosexualität definition
Homosexualität	Germany	German	homosexualität in russland
Homosexualität	Germany	German	homosexualität deutschland

Abbildung 3-3 - Beispiel 10 Analyseeinheiten Suchervollständigung via Google Autocomplete (DMI) zum Thema Homosexualität (Screenshot von: <https://tools.digitalmethods.net/beta/scrapeGoogle/autocomplete.php>)

³⁸ Digital Methods Initiative, Mediastudies, University of Amsterdam

³⁹ Aufzurufen über: <https://tools.digitalmethods.net/beta/scrapeGoogle/autocomplete.php>

Die Ergebnisse für Google Deutschland zeigen, dass sich die ersten 4 Suchbegriffe bei der manuellen Methode im Vergleich mit dem DMI Tool lediglich zu 50% überschneiden. Wie das Tool genau arbeitet, ob es beispielsweise eine Anfrage mit kohärenter IP Adressen zur Auswahl „Land“⁴⁰ für die Suchanfrage an Google sendet um die Daten zu erhalten, oder ob die Daten auf einem anderen Weg gewonnen werden, ist nicht bekannt. Insgesamt finden sich die 4 Begriffe, die im manuellen Verfahren zur Suchvervollständigung ermittelt wurden, auch in der Liste der 10 Suchvervollständigungen des Tools *Google Autocomplete* – wenn auch in anderer Reihenfolge – wieder. Die 19 restlichen in dieser Arbeit untersuchten Länder wurden ebenfalls mit der jeweils relevanten Einstellung über das *Google Autocomplete* Tool der DMI ausgewertet und mit dem manuellen Verfahren abgeglichen. Das heißt, das „Input Feld für die Ausgabe“⁴¹ wurde mit dem Suchbegriff in der jeweiligen Landessprache befüllt. Die Felder „Sprache“⁴² und „Land“ mit der jeweils zugehörigen relevanten Such-Sprache, bzw. dem jeweils relevanten Land (ccTLD), das betrachtet werden sollte. (Tabelle A 1) Der Abgleich der manuellen Suchvervollständigung mit dem Abgleich derjenigen des Tools *Google Autocomplete* zeigt, dass sich alle Suchvervollständigungen, wenn auch nicht in den ersten 4 Suchergebnissen, insgesamt gleichen. Es kann bei beiden Methoden vorkommen, dass Google keine Suchvervollständigung zu einem Suchbegriff anbietet. Suchvervollständigungen werden von Google nicht ausgegeben wenn:

„Der Suchbegriff nicht beliebt genug ist, der Suchbegriff zu neu ist, oder der Suchbegriff (irrtümlich) als Richtlinienverstoß angesehen wurde. Dabei kann es vorkommen, dass eine Suchanfrage in einer bestimmten Sprache nicht ausgeführt wird, die in einer anderen Sprache vollkommen akzeptabel wäre. So kann beispielsweise versehentlich eine Wortgruppe nicht angezeigt werden, weil darin die Übersetzung eines anstößigen Worts aus einer anderen Sprache enthalten ist.“
(Google Inc., 2015b)

Somit konnten für einige Länder keine Daten via Suchvervollständigung erhoben werden. Das Datenmaterial zur Suchvervollständigung wurde nach der jeweiligen Suche per Screenshot gespeichert, bzw. in der Textsuche über das Menü *Extras – Datei – Speichern unter...* aus dem Internet Explorer als komplette Webseite im HTML-Format exportiert.

3.3 Kodier-Prozess & Bildung der Indices auf Basis von Google

Die Kodierung der Analyseeinheiten erfolgte anhand eines Codebuchs, analog zur Übersetzung der Analyseeinheiten durch den Autor der Masterarbeit selbst. Die deutsch- und englischsprachigen Google Landesversionen wurden ohne Übersetzungshilfe durch den

⁴⁰ Im Tool: „Select one or more countries”

⁴¹ Im Tool: „Input words to retrieve autocomplete for, one per line”

⁴² Im Tool: „Select one or more languages:”

Autor analysiert und entsprechend des Codebuchs verkodet. (siehe Anhang II - Codebuch) Alle anderen Google Landesversionen wurden simultan zur Übersetzung durch den Autor analysiert und kodiert. Das heißt, die Übersetzung wurde in der Regel in einem face-2-face Aufbau durchgeführt. Hierzu wurde dem jeweiligen Übersetzer ein PC bereitgestellt, an dem der relevante Datensatz (die Suchvervollständigung und Suchergebnisse der jeweiligen ccTLD / Landessprache) in den Browser geladen wurde. Der Autor / Coder⁴³ hatte einen zweiten PC neben dem Übersetzer, an dem die Analyseeinheiten simultan zur Übersetzung verkodet wurden. Somit konnten Rückfragen des Autors, bzw. des Übersetzers direkt je Analyseeinheiten besprochen werden. Die Übersetzung für Niederländisch, Französisch und Norwegisch erfolgte in Einzelgesprächen via Skype, da die Übersetzer nicht im Raum Köln zugegen waren. Die Übersetzer erhielten die vom Autor erstellten Datensätze per Mail und mussten diese selbstständig an ihrem Rechner im Web-Browser öffnen. Nach dem Abgleich der Daten (der Google Suchergebnisse je relevanter Landesversion) die dem Übersetzer / dem Autor vorlagen, konnte auch per Skype die simultane Übersetzung und Kodierung der Analyseeinheiten erfolgen. Aufgrund der Anwesenheit des Autors bei der Übersetzung, waren die Übersetzer selbst nicht direkt in den Kodier-Prozess eingebunden. Eine Kodier Schulung für die Übersetzer war daher nicht notwendig, eine schlechte Intercoder-Reliabilität kann daher ausgeschlossen werden. Die Messung der Intracoder-Reliabilität erfolgte anhand einer Test Kodierung der Länder Deutschland und Österreich zum Thema Homophobie. Die erste Kodierung fand am 10.08.2015 statt, die zweite Messung erfolgte im Abstand einer Woche. Die Intracoder-Reliabilität bei der Test Kodierung für Deutschland lag bei 85,71%, (1. Kodierung: Homophobie-Index 10%, 2. Kodierung: Homophobie-Index 12%) und bei der Test Kodierung für Österreich bei 100% (1. Kodierung: Homophobie-Index 20%, 2. Kodierung: Homophobie-Index 20%). Insgesamt betrug die Intracoder-Reliabilität somit 92,85%.

Im Kodier-Prozess wurde zunächst die Suchvervollständigung betrachtet und anhand des Codebuchs bewertet, nachfolgend wurde die Google Suche geöffnet, so dass auch die Textsuche betrachtet, analysiert und kodiert werden konnte. Im Kodier-Prozess wurde zwischen themenrelevanten (1 = relevant) und nicht themenrelevanten (0 = irrelevant) Inhalten unterschieden – ein nicht themenrelevanter Inhalt hatte keinen weiteren Einfluss auf die gesammelten Daten. Bei den relevanten Themeninhalten wurde zwischen den positiven, bzw. neutralen Aussagen und den negativen Aussagen unterschieden. Negative

⁴³ Im Folgenden aufgrund besserer Lesbarkeit lediglich als Autor bezeichnet

Aussagen waren für die Kodierung als solche Aussagen relevant, als das sie zur Bildung der Indices (Homophobie, Ausländerfeindlichkeit, Toleranz von Drogen) verwendet wurden. Negative Aussagen sind laut Codeplan definiert als solche Aussagen, die das Thema als schlecht darstellen, Aussagen zur Abneigung (der Gesellschaft) gegen das Thema, als Verunglimpfungen / Beleidigungen von Individuen oder Gruppen, als verletzenden und beleidigenden Kommentare, sowie als Aussagen der Missachtung. Alle negativen Aussagen wurden nachfolgend zudem auf ihre Inhaltsebene untersucht und verkodet. Die erarbeiteten relevanten Inhaltsebenen z.B. im Bereich der Homophobie waren: religiöse homophobe Äußerungen (3), Darstellung von Homosexualität als Krankheit (4), gesellschaftlich homophobe Äußerungen (5) und körperliche oder psychische Gewalt gegen Homosexuelle (6). Zudem gab es die Möglichkeit, dass die Analyseeinheit einen Bezug zu einem anderen Land aufweist. In diesem Fall wurde auf der nominalen Skala eine 7 vergeben. Um einen Einblick in den Kodier-Prozess zu geben, werden nachfolgend exemplarisch 3 Beispiele der Textsuche zum Thema Homophobie erläutert:



Abbildung 3-4 - Beispiel 1 Kodierung der Textsuche zum Thema Homosexualität (Screenshot, google.de)

Abbildung 3-4 weist eine neutrale Überschrift auf (1, blau) – es wird objektiv berichtet, dass ein Mann sich als homosexuell geoutet hat. Die Analyseeinheit hat somit einen klaren Bezug zum Thema Homosexualität (V1=1). Ein Bezug muss allerdings nicht zwingend in der Überschrift gegeben sein, sondern kann sich auch in Hyperlink oder Textkörper befinden. Auch der Hyperlink (2, grün) beinhaltet das Outing als neutrale Aussage. Im Textkörper selbst (3, grau) findet eine negative Aussage zum Thema Homosexualität statt – der Mann hat seine Homosexualität aufgrund gesellschaftlichen Drucks verheimlicht. Insgesamt wäre dies somit als Homophob (V2 = 2) auf gesellschaftlicher Ebene (V2.5 = 5) zu bewerten. Da es sich bei der Überschrift allerdings deutlich erkennbar um einen britischen Rugbyspieler handelt und die Suchanfrage über die deutsche Google Version (google.de) gestellt wurde, wird die Analyseeinheit mit inhaltlich relevant (V1=1), allerdings mit Bezug zu einem anderen Land, gewertet (V2=90 / V2.5=7). Der gesellschaftliche Druck in Großbritannien, sich als homosexueller Rugby Spieler nicht zu outen, sagt nicht auch direkt etwas über den gesellschaftlichen Druck gegen das Outing Homosexueller in Deutschland aus. Somit sollte

sich auch die Analyseeinheit nicht direkt positiv oder negativ auf den Homophobie Index für (Google) Deutschland auswirken.

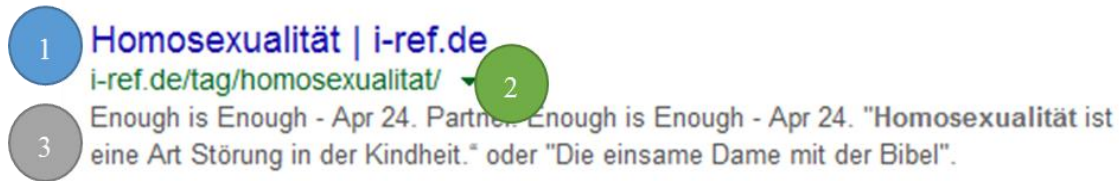


Abbildung 3-5 - Beispiel 2 Kodierung der Textsuche zum Thema Homosexualität (Screenshot, google.de)

Abbildung 3-5 hat sowohl eine neutral Überschrift (1, blau), als auch eine neutrale Aussage im Hyperlink (2, grün). Beide Merkmale der Analyseeinheit lassen auf einen Bezug zum Thema Homosexualität schließen – der Inhalt ist somit relevant für die Datenerhebung ($V1=1$). Im Textkörper (3, grau) wird eine negative Aussage zum Thema Homosexualität getroffen. Dort heißt es: „Homosexualität ist eine Art Störung in der Kindheit.“ Dies ist eine eindeutig homophobe Aussage ($V2 = 2$), da sie explizit davon ausgeht, dass Homosexuelle traumatisiert sind und erst durch ein kindliches Trauma überhaupt homosexuell wurden. Diese Aussage stellt Homosexualität somit als eine Art psychische Krankheit / Störung dar ($V2.5 = 4$).



Abbildung 3-6 - Beispiel 3 Kodierung der Textsuche zum Thema Homosexualität (Screenshot, google.de)

Abbildung 3-6 hat in allen untersuchten Merkmalen (Überschrift, Hyperlink und Textkörper) einen klaren Bezug zum Thema ($V1 = 1$) und beinhaltet in allen Merkmalen neutrale Aussagen. Die Analyseeinheit ist somit als neutral zu bewerten ($V2 = 1$).

Die Datenerhebung zu den Themen Homophobie und Ausländerfeindlichkeit wurde analog des beschriebenen expliziten Verfahrens durchgeführt. Das Thema Toleranz von Drogen wurde wie in Kapitel 3-1 beschrieben anhand einer impliziten Suche verkodet. D.h., dass beim impliziten Verfahren die ersten 100 Analyseeinheiten zum Thema Toleranz von Rauschmitteln nicht konkret auf den Inhalt, sondern lediglich auf die Anzahl der Analyseeinheiten, die die Wörter Marihuana bzw. Alkohol enthielten, untersucht wurden. Da der Suchbegriff „illegale Drogen“ den negativen Zusammenhang zum Thema bereits impliziert, wird das Erscheinen der Wörter Alkohol oder Marihuana als intolerant gegenüber dem jeweiligen Rauschmittel gewertet. (Kapitel 3.1) Wie beschrieben, wurden zur

Datengenerierung selbst die Suchbegriffe „Homosexualität“, „Ausländer“ und „illegale Drogen“ in der jeweils zur ccTLD und IP zugehörigen Landessprache (Tabelle A 1) genutzt. Für die Indices der Themenbereiche Homophobie und Ausländerfeindlichkeit basiert der Google Index jeweils auf der Anzahl der negativen Analyseeinheiten in Relation zur Anzahl der relevanten Analyseeinheiten. Der Index auf Basis der Suchvervollständigung ergibt sich aus der Anzahl aller negativen Analyseeinheiten in Relation zur Anzahl aller relevanten (positiven / neutralen / negativen) Analyseeinheiten zum Thema. Enthält also die Suchvervollständigung eine negative und drei neutrale Aussagen zum Thema, liegt der Index bei 25% (Relation: 1 negative Aussagen zu 4 relevanten Aussagen). Auch der Index für die Textsuche wird durch die Relation der Menge an negativen Analyseeinheiten zu den insgesamt relevanten (positiven / neutralen / negativen) Analyseeinheiten berechnet. Analyseeinheiten, die keine relevante Aussage zum Thema beinhalten ($V1 = 0$, $V2 = 99$), oder die sich auf ein anderes Land beziehen ($V1 = 0$, $V2 = 90$), haben somit bezogen auf die unterschiedlich betrachteten Länder keinen positiven oder negativen Einfluss auf den jeweiligen Index. Zudem ermöglicht die Abgrenzung nicht relevanter Inhalte die Option zu vergleichen, welche Länder im Zusammenhang mit dem Thema sehr oft auf andere Länder verweisen. Liegen bei der Textsuche also zum Beispiel 5 Treffer mit Bezug zu einem anderen Land, 18 negative und 77 neutrale oder positive Aussagen zum Thema vor, liegt der Index auf Basis der Google Textsuche bei 19% (Relation: 18 negative Analyseeinheiten / 95 relevanten Analyseeinheiten, 5 Analyseeinheiten nicht relevant, da mit Bezug zu anderen Ländern). Zusätzlich zum Index selbst sind auf Basis der Kodierung Aussagen zu Themenfeldern möglich, die die verschiedenen Ebenen darstellen, auf denen negative Gedanken zu den Themen geäußert werden. Es gilt die Annahme: je höher die Anzahl der negativen Äußerungen bei Google zu einem Thema ist, desto homophober, bzw. ausländerfeindlicher ist das jeweilige Land.

Im Zusammenhang mit dem Suchbegriff „illegale Drogen“ soll ein Index zur Toleranz von Alkohol, bzw. Marihuana der untersuchten Länder erstellt werden. Die Autovervollständigung, sowie die Textsuche wurden daher in Zusammenhang mit dem Suchbegriff „illegale Drogen“ auf die Anzahl der Analyseeinheiten, die das Wort „Alkohol“, bzw. „Marihuana“ enthielten, geprüft (implizites Vorgehen). Dabei galt die Annahme: je höher die Anzahl der Analyseeinheiten, die das Wort Alkohol, bzw. Marihuana in Kombination mit der Suchanfrage „illegale Drogen“ enthielten, desto geringer ist die Akzeptanz der jeweiligen Droge in einem Land. Enthalten in einem Land also zum Beispiel 14 der 100 Analyseeinheiten das Wort Marihuana und 6 der 100 Analyseeinheiten das Wort

Alkohol, berechnet sich der jeweilige Index aus der Anzahl der Analyseeinheiten mit einer Nennung der jeweiligen Substanz in Relation zur Anzahl aller Analyseeinheiten (100). Im genannten Beispiel liegt der Index für die Toleranz von Marihuana somit bei 14%, der Index für Toleranz von Alkohol bei 6%. Diese Indices können nachfolgend mit anderen Studien, sowie über die untersuchten Länder hinweg verglichen werden. Demzufolge ermöglicht sowohl das implizite, als auch das explizite Verfahren eine länderübergreifende, kulturelle Vergleichbarkeit der Google Landesversionen und somit ggf. der Länder selbst.

4 Auswertung der Google Indices

4.1 Index: „Homophobie“ nach Google

4.1.1 Vergleichsdaten: Homophobie in der EU nach EVS-Index

Um die mittels Google generierten Daten validieren zu können, werden Vergleichsdaten aus relevanten Studien benötigt, die sich mit den Themen *Homophobie*, *Ausländerfeindlichkeit* und *Toleranz von Drogen* im Zusammenhang mit kulturvergleichender Forschung befassen. Das Gesis Panel bietet u.a. den öffentlichen Zugang zu den kompletten Umfragedaten der European Value Study. Die Studie wird nachfolgend als Datenbasis zum Abgleich der mittels generischer Websuche (Google) erhobenen Daten verwendet. Die EVS beinhaltet 3 Fragen zum Thema Homosexualität bzw. Homophobie, die ausgewertet wurden, um einen (EVS-)Index zum Thema Homophobie zu erstellen. Die nachfolgenden Fragen wurden in allen in dieser Masterarbeit untersuchten Ländern (Italien ohne Frage *f118 - do you justify: homosexuality*) gestellt:

1. a124_09 – *don't like as neighbours: homosexuals*: Auf dieser Liste finden Sie verschiedene Gruppen von Menschen. Bitte nennen Sie solche, die Sie nicht als Nachbarn akzeptieren würden: Homosexuelle⁴⁴ (bivariat: 1 genannt, 0 nicht genannt) (EVS, 2015, S. 194)
2. d026_01 – *homosexual couples - adopt children*: Wie denken Sie über die nachfolgenden Aussagen? Stimmen Sie diesen zu, oder lehnen Sie diese ab? Homosexuelle Paare sollten Kinder adoptieren dürfen.⁴⁵ (Skala 1-5, wobei: 1 = stimme zu, bis 5 = lehne ab) (EVS, 2015, S. 322)
3. f118 – *do you justify: homosexuality*: Auf einer Skala von 1-10, sagen Sie Ihre Meinung zu den folgenden Aussagen: Homosexualität ist...⁴⁶ (Skala 1-10, wobei 10 = vertretbar / legitimierbar, 1 = niemals vertretbar / legitimierbar) (EVS, 2015, S. 767)

Der in dieser Masterarbeit erstellte Index mit den Daten der EVS zum Thema Homophobie je Land wurde folglich aus den Top Box Werten der Fragen d026_01 (Box 4+5 = „lehne ab, dass Homosexuelle Kinder adoptieren dürfen“) und den Bottom Box Werten der Fragen f118

⁴⁴ Original: „On this list are various groups of people. Could you please sort out any that you would not like to have as neighbours?“ Im Text Übersetzung durch den Autor.

⁴⁵ Original: „How would you feel about the following statements? Do you agree or disagree with them? Homosexual couples should be able to adopt children.“ Im Text Übersetzung durch den Autor.

⁴⁶ Original: „Please tell me for each of the following statements whether you think it can always be justified, never be justified, or something inbetween, using this card. For: Homosexuality“

(1+2 = Homosexualität ist nicht vertretbar / legitimierbar) gebildet. Zudem wurde die Aussage „als Nachbar nicht akzeptiert“ (a124_09 = 1) als homophob gewertet. Basierend auf den Daten der EVS ergibt sich daraus das nachfolgende Schaubild zum Thema Homophobie.

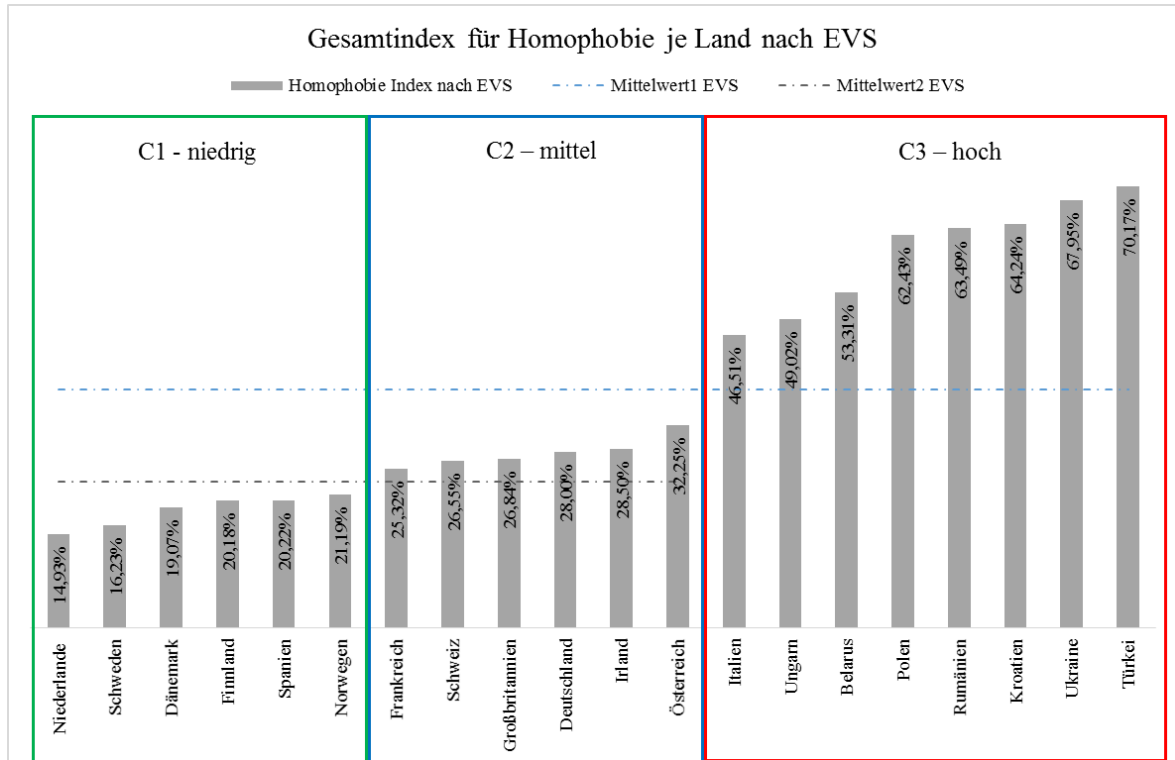


Abbildung 4-1 - Gesamtindex für Homophobie je Land nach EVS (eigene Darstellung)

Die Studie zeigt deutlich, dass das Thema Homophobie nach wie vor ein Problem im europäischen Raum darstellt. Um die Daten der EVS mit den Daten des Google Index vergleichbar zu machen, wurden die Länder in aufsteigender Reihenfolge sortiert und nachfolgend in 3 ca. gleich große Cluster (6 Länder, 6 Länder, 8 Länder) eingeteilt, die den Grad der Homophobie der Länder im direkten Vergleich widerspiegeln. Im dritten Cluster (C3) befinden sich die Länder, die über dem durchschnittlichen Homophobie-Index ($X > 37,82\%$, Mittelwert1 EVS) der 20 abgebildeten Länder liegen. Das mittlere Cluster (C2) bildet solche Länder ab, die unter dem Durchschnitt aller 20 Länder, aber über dem Durchschnitt der 12 Länder selbst, liegen ($23,27\% < X < 37,82\%$, Mittelwert2 EVS). Cluster 1 bildet solche Länder ab, die auch unter dem Mittelwert der Länder aus C1 und C2 liegen ($X < 23,27\%$). Nachfolgend ergeben sich auf Basis der EVS folgende Cluster:

C1 EVS: Niederlande, Schweden, Dänemark, Finnland, Spanien, Norwegen

C2 EVS: Frankreich, Schweiz, Großbritannien, Deutschland, Irland, Österreich

C3 EVS: Italien, Ungarn, Weißrussland, Polen, Rumänien, Kroatien, Ukraine

Die Auswertungen der Google Indices werden in ähnlicher Weise geclustert und sortiert, so dass die Indices auf Ebene der Cluster vergleichbar sind. Werden die Länder über den Google Index in die gleichen Clustern zugeordnet wie über den EVS Index, werden die Daten als valide angesehen.

Datenkorrelation

		Korrelationen					
		Homophobie Index_EVS_Gesamt	Google TEXT_SUCHE	Google AUTO_FILL_man	Google AUTO_FILL_DMI	GESAMTINDEXICH	GESAMTINDEXDMI
Homophobie_Index_EVS_Gesamt	Korrelation nach Pearson	1	,745**	,746**	,629**	,793**	,806**
	Signifikanz (2-seitig)		,000	,000	,005	,000	,000
	N	20	20	18	18	20	20
Google_TEXT_SUCHE	Korrelation nach Pearson	,745**	1	,639**	,425	,833**	,944**
	Signifikanz (2-seitig)	,000		,004	,079	,000	,000
	N	20	20	18	18	20	20
Google_AUTO_FILL_man	Korrelation nach Pearson	,746**	,639**	1	,533*	,958**	,857**
	Signifikanz (2-seitig)	,000	,004		,023	,000	,000
	N	18	18	18	18	18	18
Google_AUTO_FILL_DMI	Korrelation nach Pearson	,629**	,425	,533*	1	,542*	,514*
	Signifikanz (2-seitig)	,005	,079	,023		,020	,029
	N	18	18	18	18	18	18
GESAMTINDEXICH	Korrelation nach Pearson	,793**	,833**	,958**	,542*	1	,969**
	Signifikanz (2-seitig)	,000	,000	,000	,020		,000
	N	20	20	18	18	20	20
GESAMTINDEXDMI	Korrelation nach Pearson	,806**	,944**	,857**	,514*	,969**	1
	Signifikanz (2-seitig)	,000	,000	,000	,029	,000	
	N	20	20	18	18	20	20

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Tabelle 4-1 - Korrelation der Google Indices zum EVS Index für Homophobie (SPSS)

Eine hohe Güte bei der Korrelation der Google Daten zu den EVS Daten ist nach Hair im Themenbereich Homophobie insgesamt in allen Fällen gegeben. Nach Hair gilt, dass wenn $0 \leq r < 0,3$ eine schwache, $0,3 < r < 0,5$ eine mittlere und $0,5 < r \leq 1$ eine starke Korrelation vorliegt. (Hair, 2006) Lediglich der mittels *Google Autocomplete* (DMI) erstellte Index liegt relativ knapp über der Vorgabe zur starken Korrelation ($r=0.629$). Die höchste Güte besteht bei den Gesamtindices (Kombination von Textsuche und Suchvervollständigung, $r=0,806$). (Tabelle 4-1) Auch eine hohe Signifikanz ist im Abgleich zwischen den EVS Daten und den Google Daten gegeben ($0,000 < \alpha > 0,0025$, einseitiger Test). Dennoch bleibt festzuhalten, dass der EVS Index sich (aufgrund der Fragestellungen der Studie) vor allem auf Themen mit gesellschaftlichem Inhalt bezieht (Adoption, Homosexuelle im persönlichen Umfeld und Akzeptanz im Alltag). Zudem besteht eine zeitliche Diskrepanz von ca. 7 Jahren zwischen den Daten der EVS (Erhebung 2008/2009) und den mittels Google erhobenen Daten (2015).

4.1.2 Homophobie auf Basis der Google Suchvervollständigung

Homophobie auf Basis der manuellen Suchvervollständigung

Ausgehend von der Aussage, dass das Web die Gesellschaft widerspiegelt, gilt für die nachfolgenden Auswertung: je höher der Homophobie Index eines Landes in der EVS, desto mehr negative Aussagen, bezogen auf die relevante Anzahl an Analyseeinheiten je Land zum Thema Homosexualität, sind zu erwarten. Daher ist anzunehmen, dass ein Land mit einem hohen Homophobie Index in der EVS auch einen hohen Homophobie Index mittels Google (Suchvervollständigung / Textsuche) aufzeigt. Ein hoher Anteil an negativen Suchvervollständigungen via Google spiegelt daher einen hohen Grad an Homophobie im jeweiligen Land wider. Bei der Erstellung des Homophobie Indexes auf Basis der manuellen Suchvervollständigung bestand vor allem das Problem, einer sehr geringen Anzahl an Analyseeinheiten (4) je Land. Aufgrund der gleichen Intervalllängen ergibt sich in der Auswertung daher ein eher undifferenziertes Bild im Vergleich der Länder im jeweiligen Cluster.

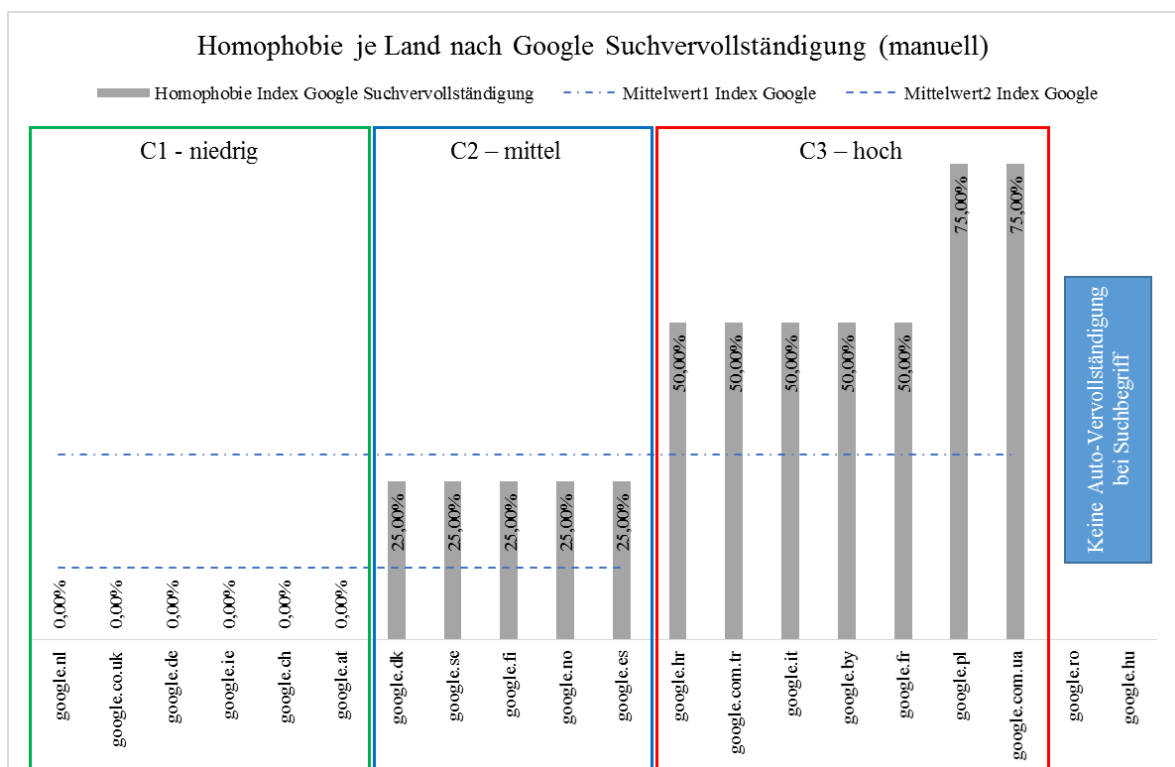


Abbildung 4-2 - Gesamtindex Homophobie je Land: Google Suchvervollständigung, manuelle Suche (eigene Darstellung)

Zudem wurde von den Google Versionen Rumänien und Ungarn zur jeweiligen Suche (google.ro = „homosexualitate“, google.hu = „homoszexualitás“) keine Suchvervollständigung ausgegeben. Die genauen Gründe hierfür sind nicht bekannt, dies könnte unter anderem auf eine insgesamt zu geringe Anzahl an Suchanfragen zum Thema

hindeuten, ist letztlich jedoch nicht eindeutig zu klären. (vgl. Kapitel 2.2.3) Die Länder Rumänien und Ungarn wurden daher bei der Erstellung des Google Index mittels manueller Suchvervollständigung, sowie beim Abgleich mit den EVS Daten nicht berücksichtigt. Trotzdem lassen sich analog der in Kapitel 4.1.1 erstellten Zuordnung der Länder in 3 Cluster auch die mittels manueller Suchvervollständigung erhobenen Daten in 3 Cluster einordnen. Die Sortierung der Länder erfolgt dazu, wie vorab auch, nach aufsteigendem Homophobie Index. Bei der manuellen Suchvervollständigung werden somit 5 von 20 Ländern an der gleichen Platzierung wie in der EVS zugeordnet. (Abbildung A 4) Dies ist allerdings vor allem dem Umstand der gleich bleibenden Intervalllänge geschuldet, die aufgrund der geringen Datenbasis (4 Analyseeinheiten) entsteht. Die Länder je Cluster können somit fast frei im Cluster verschoben werden. Aufgrund der gleichbleibenden Intervalllängen in den 3 Clustern liegt auf Basis der gleichen Platzierung im Index kein Indikator für eine valide Zuordnung der Länder vor. Zudem zeigt der Abgleich der 3 Cluster (Abbildung 4-3, Tabelle 4-2), dass die Methode insgesamt eher ungenau arbeitet.

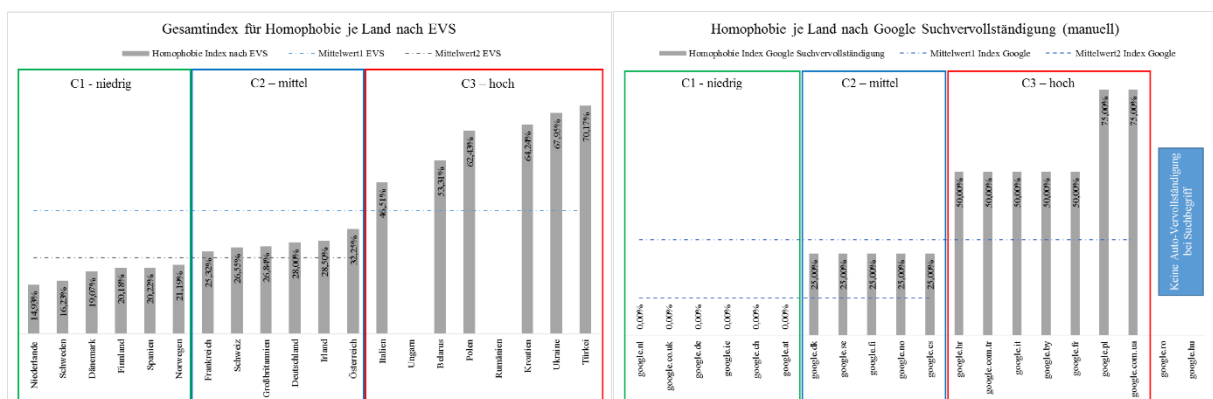


Abbildung 4-3 - EVS Index Homophobie vs. Homophobie Index manuelle Suchvervollständigung

- C1 EVS: Niederlande, Schweden, Dänemark, Finnland, Spanien, Norwegen
 C2 EVS: Frankreich, Schweiz, Großbritannien, Deutschland, Irland, Österreich
 C3 EVS: Italien, Weißrussland, Polen, Kroatien, Ukraine, Türkei
 C1 GSV⁴⁷ manuell: Niederlande, Großbritannien, Deutschland, Irland, Schweiz, Österreich,
 C2 GSV manuell: Dänemark, Schweden, Finnland, Norwegen, Spanien
 C3 GSV manuell: Kroatien, Türkei, Italien, Weißrussland, Frankreich, Polen, Ukraine

⁴⁷ Google Suchvervollständigung

Abgleich C1	1	Abgleich C2	0	Abgleich C3	1
	0		0		1
	0		0		1
	0		0		1
	0		0		1
	0		0		1
	0		0		1
	0		0		0
	16,67%		0,00%		83,33%
Anzahl Länder	18				
Falsch zugeordnet	11	61,11%			
Richtig zugeordnet	7	38,89%			

Tabelle 4-2 - Zuordnungsübersicht Cluster 1-3 EVS vs. Google manuelle Suchvervollständigung - Homophobie (eigene Darstellung)

Lediglich 38,89% (7 von 18) der Länder werden auf Basis der mittels manueller Suchvervollständigung generierten Daten in die gleichen Cluster wie durch die EVS Daten zugeordnet. Zudem liegt Cohen's Kappa ($k=0,083$, Tabelle A 3) mit einer Datenreliabilität von 0-4% eher in einem sehr geringen Bereich. (Tabelle A 2) Die mittels manueller Suche erhobenen Daten besitzen aufgrund der ungleichen Zuordnung zu den Clustern und der geringen Datenmenge (lediglich 4 Analyseeinheiten je Land) keine Validität und eine sehr geringe Aussagekraft. Eine Zuordnung der Themen, bzw. Ebenen durch die Homophobie sichtbar wird, wurde für die manuelle Suchvervollständigung aufgrund der geringen Datenmenge daher nicht vorgenommen.

Homophobie auf Basis der Suchvervollständigung mit Google Autocomplete

Im Vergleich zur manuellen Erhebung werden nachfolgend die mittels *Google Autocomplete* erhobenen Daten der Suchvervollständigung ausgewertet. Aufgrund des erweiterten Datenumfanges (10 Analyseeinheiten je Land) zeigt der Homophobie Index auf Basis der Suchvervollständigung via *Google Autocomplete*⁴⁸ bereits ein leicht differenziertes Bild im Vergleich der Länder im jeweiligen Cluster. Auch die Sortierung der Länder nach aufsteigender Reihenfolge des Homophobie Index zeigt eine Veränderung im Vergleich zur Darstellung der manuellen Suchvervollständigung. (Abbildung 4-4)

⁴⁸ In nachfolgenden auch bezeichnet als *GA_DMI* (*Google Autocomplete Digital Methods Initiative*)

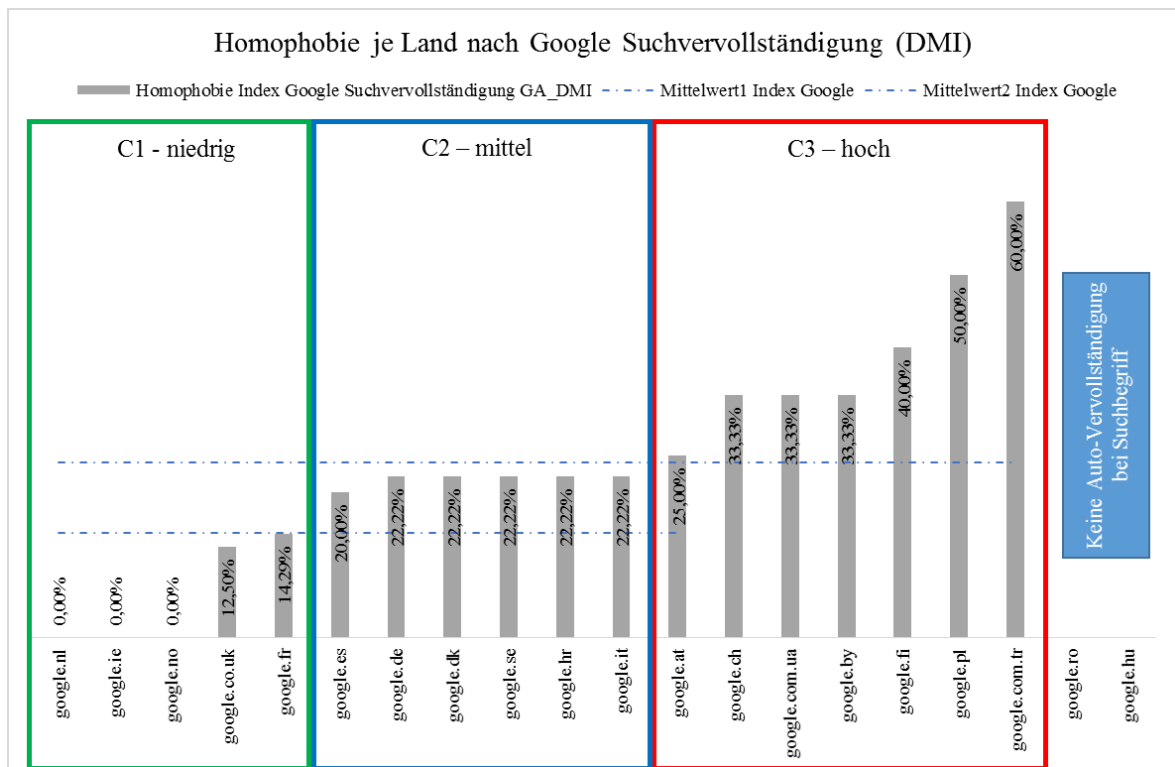


Abbildung 4-4 - Gesamtindex Homophobie je Land: Google Suchervollständigung, DMI Tool (eigene Darstellung)

Analog zur manuellen Suchervollständigung geben die Google Versionen Rumänien und Ungarn zur jeweiligen Suche (google.ro = „homosexualitate“, google.hu = „homoszexualitás“) auch mittels *Google Autocomplete* keine Suchervollständigung aus. Zudem lieferte *Google Autocomplete* für die Länder Ukraine und Weißrussland keine Suchervollständigung für die Eingaben mit kyrillischer Schreibweise (гомосексуализм⁴⁹). Daher wurden die Daten für Weißrussland und die Ukraine im vorliegenden Datensatz mit Hilfe der lateinischen Schreibweise (*gomoseksualizm*⁵⁰) ermittelt. Wie vorab, erfolgt die Sortierung der Länder in die jeweiligen Cluster anhand der Mittelwerte. Bei Sortierung der Länder nach aufsteigendem Homophobie Index zeigt sich, dass sich die Intervalllängen auch bei der *Google Autocomplete* Suche in den Clustern relativ ähnlich sind. Dennoch liegt im Vergleich zu den manuell erhobenen Daten eine verbesserte Darstellung durch leicht unterschiedliche Intervalllängen in den Clustern selbst vor. Somit werden auf Basis der *Google Autocomplete* Daten lediglich 16,67% (3 von 18) der Länder an der exakt gleichen Platzierung wie in den EVS Daten eingeordnet. (Abbildung A 5)

⁴⁹ Übersetzung = Homosexualität

⁵⁰ Übersetzung = Homosexualität

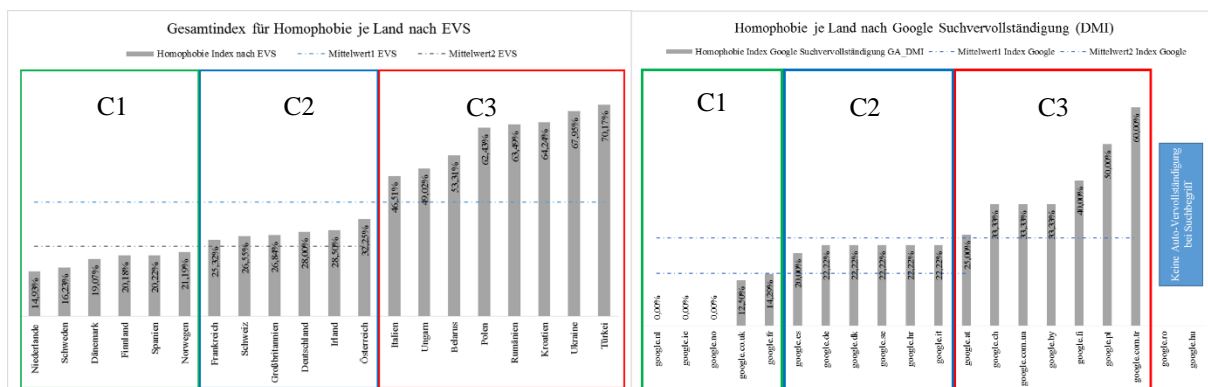


Abbildung 4-5 - EVS Index Homophobie vs. Homophobie Index der Google Suchvervollständigung mittels Google Autocomplete (eigene Darstellung)

- C1 EVS: Niederlande, Schweden, Dänemark, Finnland, Spanien, Norwegen
 C2 EVS: Frankreich, Schweiz, Großbritannien, Deutschland, Irland, Österreich
 C3 EVS: Türkei, Italien, Ungarn, Weißrussland, Polen, Rumänien, Kroatien, Ukraine
 C1 GA_DMI⁵¹: Niederlande, Irland, Norwegen, Großbritannien, Frankreich
 C2 GA_DMI: Spanien, Deutschland, Dänemark, Schweden, Kroatien, Italien
 C3 GA_DMI: Österreich, Schweiz, Ukraine, Weißrussland, Finnland, Polen, Türkei

Abgleich C1	1	Abgleich C2	0	Abgleich C3	1
	0		0		0
	0		0		1
	0		1		1
	0		0		0
	1		0		1
	33,33%		16,67%		66,67%

Anzahl Länder	18		
Falsch zugeordnet	11	61,11%	
Richtig zugeordnet	7	38,89%	

Tabelle 4-3 - Zuordnungsübersicht Cluster 1-3 EVS vs. Google Suchvervollständigung mit Google Autocomplete - Homophobie (eigene Darstellung)

Auch durch die mittels Google Autocomplete generierten Daten werden, wie bei der manuellen Suchvervollständigung, lediglich 38,89% (7 von 18) der Länder in die gleichen Cluster wie durch die EVS Daten zugeordnet. Somit liegt auch das Cohen's Kappa ($k=0,083$, Tabelle A 4) bei einer Datenreliabilität von 0-4% wieder in einem sehr geringen Bereich. (Tabelle A 2) Im Vergleich zum Homophobie Index nach manueller Suchvervollständigung zeigt sich somit keine Verbesserung. Dennoch werden nachfolgend solche Daten, die mit dem Tool *Google Autocomplete* generiert wurden, denen, die mittels manueller Suche generiert wurden, bevorzugt. Letztlich bieten die mittels *Google Autocomplete* erhobenen Daten mehr Analyseeinheiten. Die manuelle Suche wurde je Land ausgeführt und verkodet,

⁵¹ Google Autocomplete Digital Methods Initiative

wird aufgrund der geringen Anzahl von Analyseeinheiten jedoch für die anderen Themen als irrelevant betrachtet. Allerdings besitzen auch die mittels *Google Autocomplete (DMI)* erhobenen Daten im Themenindex Homophobie aufgrund der relativ zufälligen Zuordnung zu den jeweiligen Clustern, der geringen Datenmenge (lediglich 10 Analyseeinheiten je Land) und dem schwachen Kappa-Wert insgesamt keine Validität und eine eher geringe Aussagekraft.

Relevante Themenkreise zur Homophobie nach Google Autocomplete Suche

Dennoch wurde die Suchervollständigkeit, neben den erhobenen Daten zur Bildung des Homophobie Index auch auf Themen hin untersucht, die Homophobie auf unterschiedlichen Ebenen ausdrücken. Bei der Betrachtung des durchschnittlichen Themenspektrums in der Suchervollständigkeit zeigt sich die eindeutige Tendenz, dass Homosexualität in der Suchervollständigkeit mit einer Krankheit gleichgesetzt wird. (Abbildung 4-6)

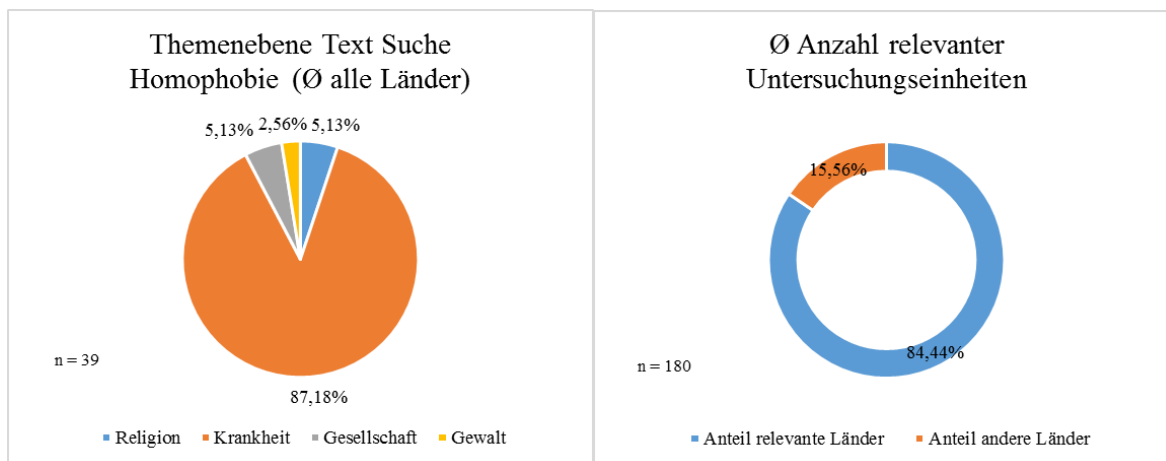


Abbildung 4-6 - Themenebene Homophobie Suchervollständigkeit (eigene Darstellung)

Vor allem die Suchbegriffe „Krankheit“, „Heilung“, oder „Ursachen“ werden häufig im Zusammenhang mit dem Suchbegriff Homosexualität in der Suchervollständigkeit angezeigt. Im Durchschnitt weisen zudem 15,56% aller Suchervollständigungen zum Thema Homosexualität einen Bezug zu einem anderen Land auf. Aufgrund der insgesamt geringen Validität der Daten, die mittels Suchervollständigkeit erhoben wurden, werden die Themenspektren im Zusammenhang mit der Suchervollständigkeit jedoch vorerst nicht weiter thematisiert.

4.1.3 Homophobie auf Basis der Google Textsuche (TOP 100)

Die Auswertung der mittels Textsuche generierten Daten zeigt ein, auch aufgrund der gesteigerten Anzahl an Analyseeinheiten (100 je Land), differenziertes Bild im Vergleich der einzelnen Länder. Die Sortierung der Länder in aufsteigender Reihenfolge anhand des Homophobie Index ermöglicht es wie zuvor auch, die Länder in 3 etwa gleich große Cluster

(6 Länder, 6 Länder, 8 Länder) einzuordnen. Ebenso gilt wie bei der Analyse der Suchvervollständigung, dass C1 ein eher niedriges, C2 ein mittleres und C3 ein hohes Niveau (alle Länder über dem Durchschnitt) in Bezug auf den Homophobie Index der Länder widerspiegelt. (Abbildung 4-7)

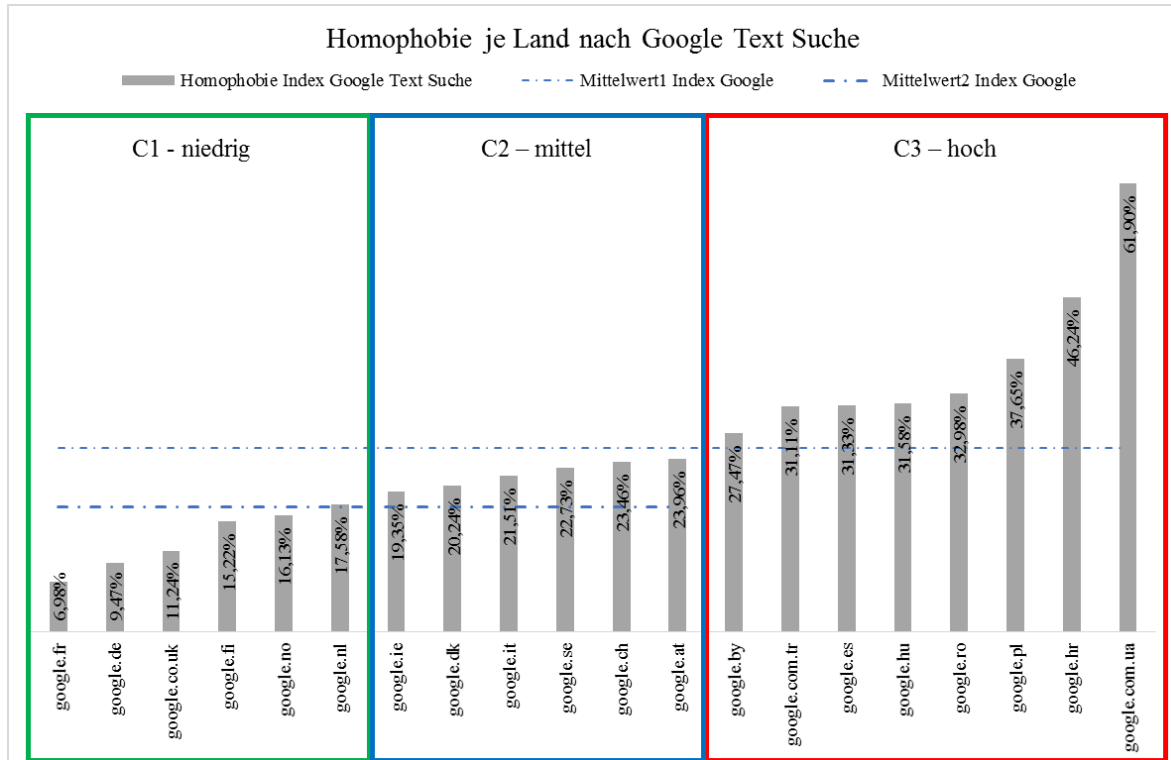


Abbildung 4-7 - Gesamtindex für Homophobie je Land nach Google Textsuche (eigene Darstellung)

Da für alle 20 relevanten Länder Daten generiert werden konnten, wurden keine Länder aus der Untersuchung ausgeschlossen. Aufgrund der sehr geringen Abweichung zum Mittelwert ($<0,005$), wurden die Niederlande C1 zugeteilt. Die Anzahl der Länder je Cluster entspricht im vorliegenden Datensatz somit der zahlenmäßigen Zuordnung der Länder je Cluster auf Basis der EVS Daten. Der direkte Vergleich der Platzierungen der Länder in aufsteigender Reihenfolge anhand des Homophobie Index zeigt, dass die Methode 20% (4 von 20) der Länder an der exakt gleichen Position einordnet, wie in der Zuordnung mittels EVS Daten. (Abbildung A 6) Die mittels Textsuche erhobenen Daten arbeiten daher im direkten Vergleich genauer, als die für die Suchvervollständigung bevorzugte Methode mittels *Google Autocomplete*. Auch der direkte Abgleich der Cluster zwischen der Google Textsuche und der EVS zeigt, dass die Daten der Textsuche insgesamt zuverlässiger sind, als die Daten der Suchvervollständigung. (Abbildung 4-8 & Tabelle 4-4) Darüber hinaus zeigt sich, dass auch Cohen's Kappa stark verbesserte Werte aufweist ($k=0,470$, Tabelle A

4.1.4 Homophobie auf Basis der Indexierung von Google Textsuche und Suchvervollständigung (TOP 100 & DMI)

Die Indexierung der Google Textsuche in Kombination mit der Suchvervollständigung soll eine Verbesserung in der Gesamtzuordnung der Länder Cluster bewirken. Die Indices der Textsuche, sowie die Indices der Suchvervollständigung je Land, werden dazu miteinander addiert und nachfolgend durch 2 dividiert. Somit ergibt sich je Land ein neuer Index, der die Einheiten Textsuche und Suchvervollständigung kombiniert abbildet. Für die Länder Ungarn und Rumänien wurden die Indices der Textsuche übernommen, da mittels Suchvervollständigung für beide Länder keine Daten erhoben werden konnten. Die nachfolgende Abbildung (Abbildung 4-9) zeigt die Sortierung der Länder nach aufsteigender Reihenfolge des kombinierten Index.

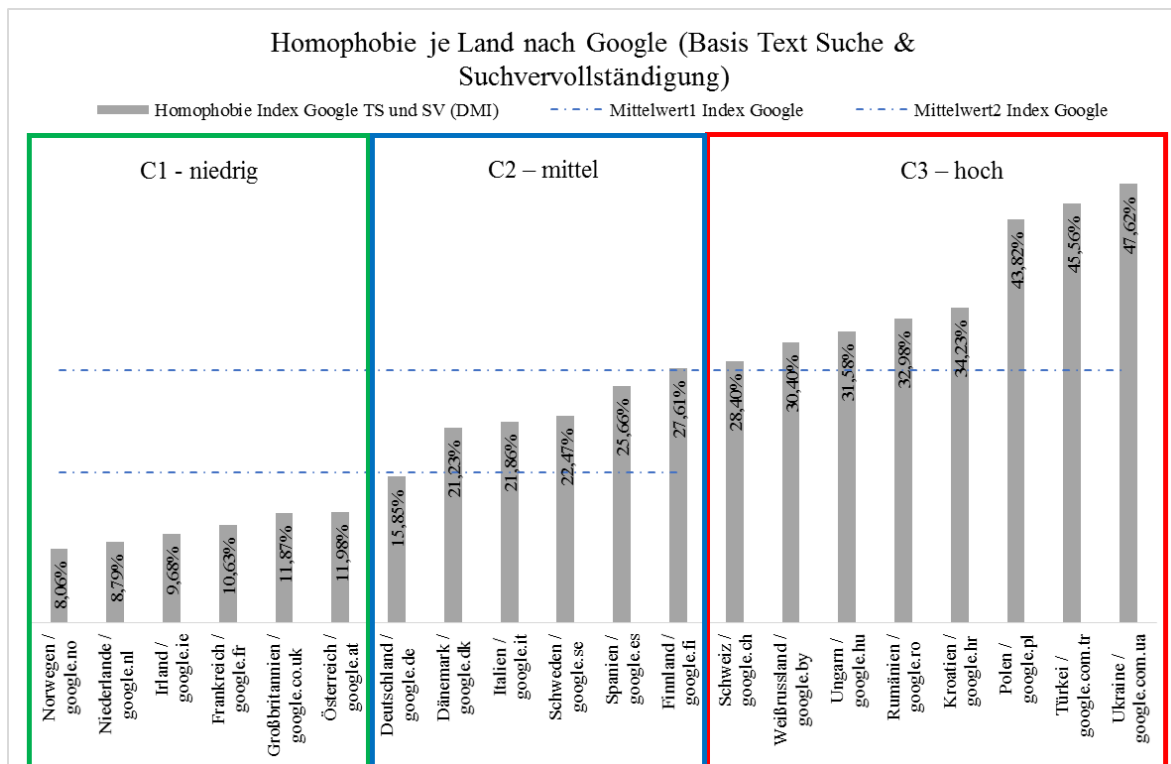


Abbildung 4-9 - Gesamtindex für Homophobie je Land nach Google Index aus Textsuche & Suchvervollständigung (eigene Darstellung)

Die Anzahl der Länder je Cluster im vorliegenden Datensatz entspricht der zahlenmäßigen Zuordnung der Länder mittels EVS Daten. Sowohl Deutschland, als auch Finnland werden aufgrund der geringen Unter- bzw. Überschreitung ($<0,005$) des Mittelwert2 in C2 eingeordnet. Der direkte Vergleich der Platzierungen der Länder in aufsteigender Reihenfolge anhand des Homophobie Index zeigt, dass die Methode 10% (2 von 20) der Länder an der exakt gleichen Position einordnet, wie in der Zuordnung mittels EVS Daten. Dies spricht für eine unzureichende Zuordnung des neu gebildeten Index. Auch der direkte

4.1.5 Relevante Themenkreise zur Homophobie nach Google Textsuche

Neben den erhobenen Daten zur Bildung des Homophobie Index selbst ermöglicht es die Analyse der via Textsuche generierten Google Daten zudem, Themenkreise zu ermitteln, die in Zusammenhang mit homophoben Äußerungen stehen. Zudem zeigt sich, auf Basis der mittels Textsuche erhobenen Daten, dass ca. 10% der Suchergebnisse je Land im Durchschnitt einen externen Bezug zu anderen Ländern haben und somit irrelevant für das jeweils verkodete Land sind. (Abbildung 4-11) Vor allem der Bezug zu den Vereinigten Staaten, aber auch zu Ländern, in denen Homosexualität verboten ist oder durch den Staat geahndet wird (Russland, Marokko, Mozambique), findet besonders häufig statt. In diesem Zusammenhang sind vor allem Berichterstattungen über die Kriminalisierung Homosexueller zu finden.

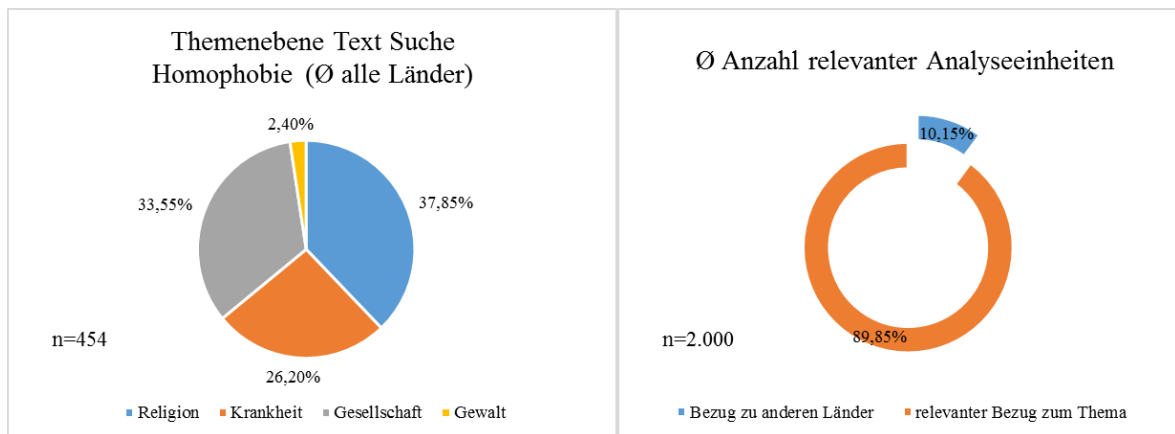


Abbildung 4-11 – links: Themenebene Homophobie Textsuche (eigene Darstellung), rechts: relative Anzahl relevanter Analyseeinheiten je Land

Der größte Bezug zum Thema Homophobie entsteht laut Google Textsuche im religiösen Kontext. Aussagen zur Homophobie auf Grundlage religiöser Ansichten werden in der EVS nicht erhoben. Dennoch zeigt sich in der Google Text Analyse, dass dieser Themenbereich in der Entstehung von Homophobie eine wichtige Rolle spielt – 35,68% aller negativen Äußerungen zum Thema Homosexualität haben einen religiösen Hintergrund. (Abbildung 4-11) Vor allem Aussagen wie „Homosexualität ist eine Sünde / ist eine Strafe von Gott / ist Sodomie“ sind häufige, länderübergreifende homophobe Aussagen.

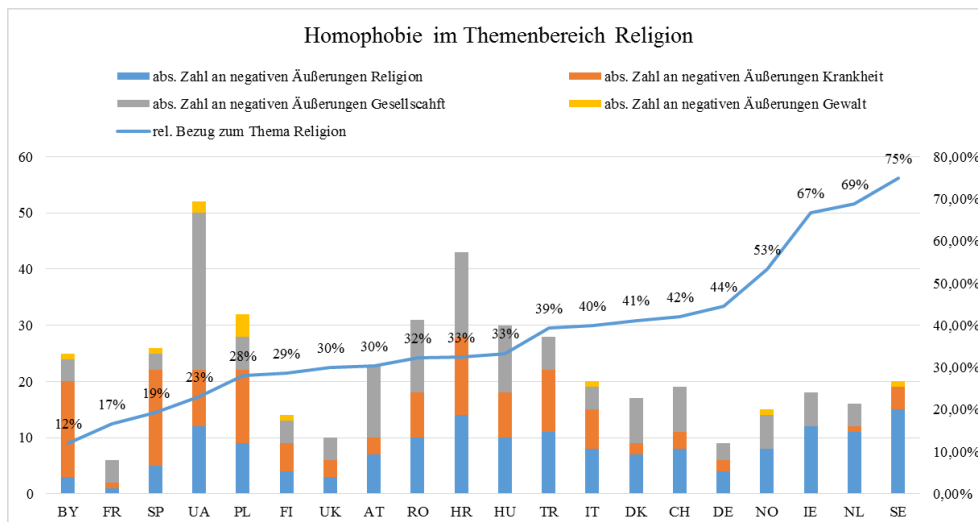


Abbildung 4-12 - Homophobie auf Basis der Google Textsuche im Themenbereich Religion (eigene Darstellung)

Relativ betrachtet sind zufolge der Google Textsuche homophobe Äußerungen mit religiösem Hintergrund insbesondere in Norwegen (53%), in Irland (67%), den Niederlanden (69%) und in Schweden (75%) der Hauptgrund für Homophobie. Die höchste Anzahl absoluter negativer Aussagen zum Thema Homosexualität auf einer religiösen Grundlage zeigen sich in Kroatien, der Türkei, in Ungarn, Irland und Schweden (mehr als 10 Analyseeinheiten je Land). Insgesamt zeigt sich, dass das Thema Religion in diesem Kontext auch in anderen Gesellschaften weit verbreitet ist und einen starken Bezug zum Thema Homophobie aufweist. (Abbildung 4-12)

Auch der Bezug zur generellen Homophobie in der Gesellschaft ist mit 33,26% laut Google Textsuche signifikant. Analog zur EVS sind in diesem Bereich die „generelle Akzeptanz von Homosexuellen“, das Thema „Adoption von Kindern durch Homosexuelle“ und „eingeschränkte Rechte von Homosexuellen“ ein Indikator für homophobe Äußerungen.

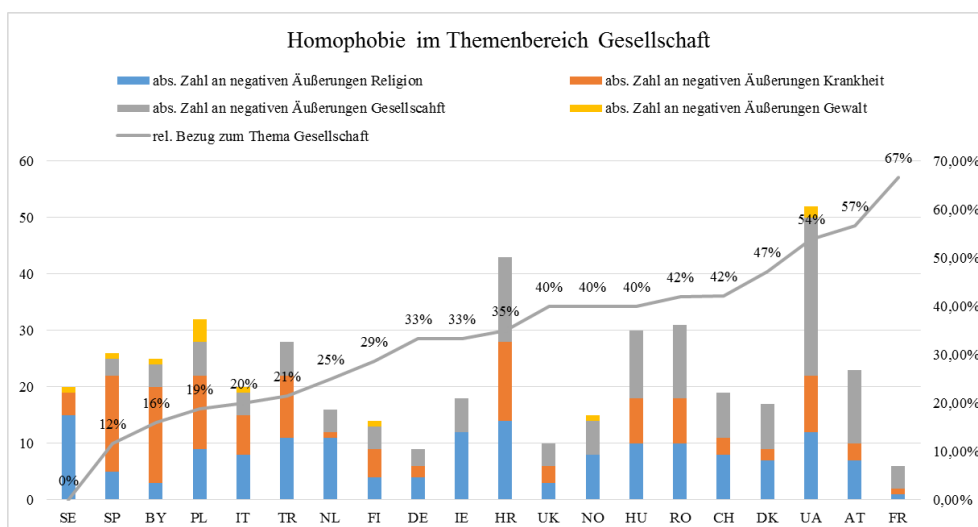


Abbildung 4-13 - Homophobie auf Basis der Google Textsuche im Themenbereich Gesellschaft (eigene Darstellung)

Vor allem in Dänemark (47%), der Ukraine (54%), Österreich (57%) und Frankreich (67%) ist die Entstehung von Homophobie auch auf eine Diskrepanz in der gesellschaftlichen Akzeptanz von Homosexuellen zurückzuführen. Die Anzahl absoluter Bezüge jedoch zeigt, dass auch in Kroatien, Ungarn und Rumänien, sowie in Österreich ein starker Bezug zu Homophobie in der Gesellschaft besteht. (Abbildung 4-13)

Als dritter großer Faktor ist zudem das Gleichsetzen von Homosexualität mit einer Krankheit als relevantes Themengebiet zu extrahieren. In etwa jede vierte (28,41%) als homophob verkodete Analyseeinheit weist einen Bezug zum Thema Homosexualität als Krankheit auf. Dabei sind vor allem die Themen „Heilung von Homosexuellen / Homosexualität“, „Homosexualität als psychische Erkrankung“ und der Kontext „Homosexualität als Krankheit / Virus“ wiederkehrende Phänomene.

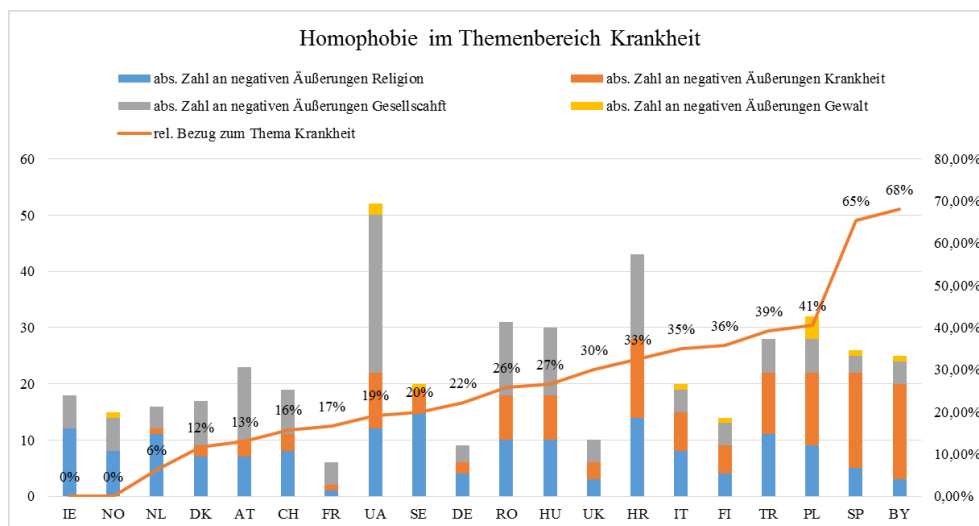


Abbildung 4-14 - Homophobie auf Basis der Google Textsuche im Themenbereich Krankheit (eigene Darstellung)

Besonders in Polen (41%), Spanien (65%) und Weißrussland (68%) ist die Wahrnehmung homosexueller Menschen stark mit dem homophoben Gedanken einer krankhaften Psyche oder einer körperlichen Erkrankung verknüpft. Jedoch sind die absoluten Nennungen der Thematik Homosexualität im Bezug zu Krankheit auch in Kroatien und der Türkei verstärkt aufgetreten. (Abbildung 4-14)

Zudem finden sich vereinzelt homophobe Äußerungen in Form von physischer oder psychischer Gewalt. Insbesondere „verunglimpfende Witze über Homosexuelle“, oder „Berichte über körperliche Gewalt gegen Homosexuelle“ fallen in diesen Themenkreis, der zufolge der Google Textsuche mit lediglich 2,64% einen eher geringen Einfluss auf das Thema Homophobie nimmt.

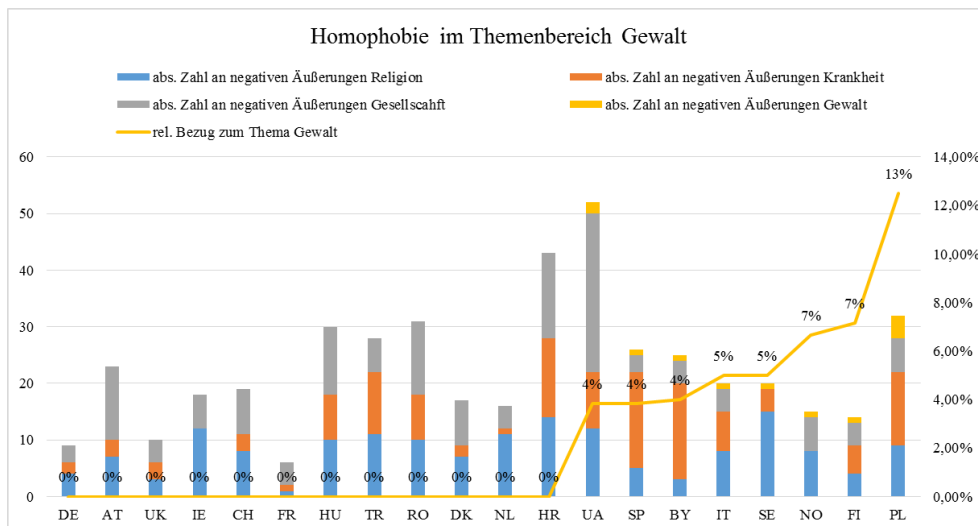


Abbildung 4-15 - Homophobie auf Basis der Google Textsuche im Themenbereich Gewalt (eigene Darstellung)

In einem Großteil der untersuchten Länder finden sich basierend auf dem Themenkreis Gewalt keine negativen Äußerungen zum Thema Homosexualität. Einzig in Polen fällt das Thema mit 13% der untersuchten Analyseeinheiten relativ hoch aus. (Abbildung 4-15)

4.2 Index: „Ausländerfeindlichkeit“ nach Google

4.2.1 Vergleichsdaten: Ausländerfeindlichkeit in der EU nach EVS-Index

Um auch die mittels Google erhobenen Daten zum Thema Ausländerfeindlichkeit validieren zu können, werden die Vergleichsdaten der EVS genutzt, auf die nachfolgend kurz eingegangen wird. Die EVS bietet mehrere Items zu den Themen Immigration, Ausländer, Religion, Glaube, usw. Der auf Basis der EVS Daten ermittelte Index bezieht sich auf 3 Fragen zum Thema Ausländerfeindlichkeit. Die nachfolgenden Items wurden in allen Ländern, die in dieser Masterarbeit untersucht wurden, in der EVS erhoben:

1. a124_02 – *don't like as neighbours: people of different race*: Auf dieser Liste finden Sie verschiedene Gruppen von Menschen. Bitte nennen Sie solche, die Sie nicht als Nachbarn akzeptieren würden: Menschen einer anderen „Rasse“ / Ethnie⁵² (bivariat: 1 = genannt, 0 = nicht genannt) (EVS, 2015, S. 187)
2. a124_06 – *don't like as neighbours: immigrants / foreign workers*: Auf dieser Liste finden Sie verschiedene Gruppen von Menschen. Bitte nennen Sie solche, die Sie nicht als Nachbarn akzeptieren würden: Einwanderer / ausländische Arbeitskräfte⁵³ (bivariat: 1 = genannt, 0 = nicht genannt;) (EVS, 2015, S. 191)

⁵² Original: „On this list are various groups of people. Could you please sort out any that you would not like to have as neighbours? Answer: people of different race“ - Im Text Übersetzung durch den Autor.

⁵³ Original: „On this list are various groups of people. Could you please sort out any that you would not like to have as neighbours? Answer: immigrants / foreign workers“ - Im Text Übersetzung durch den Autor.

3. e033 – *political view: left / right*: Im politischen Kontext spricht man von „der Linken“ und „der Rechten“. Wie würden Sie ihre Ansicht einordnen?⁵⁴ (Skala 1-10, wobei 1 = links, 10 = rechts) (EVS, 2015, S. 300)

Der für die Masterarbeit erstellte Index mit den Daten der EVS zum Thema Ausländerfeindlichkeit setzt sich folglich aus der Aussage „Menschen einer anderen „Rasse“ / Ethnie“ (a124_02 = 1), sowie „Einwanderer / ausländische Arbeitskräfte“ (a124_06 = 1) „als Nachbar ungewollt“ zusammen. Zudem werden die Top Box Werte des Items e033 (Top Box 9+10 = „politische Gesinnung Rechts“) einbezogen. Basierend auf den Daten der EVS ergibt sich daraus das nachfolgende Schaubild (Abbildung 4-16) zum Thema Ausländerfeindlichkeit.

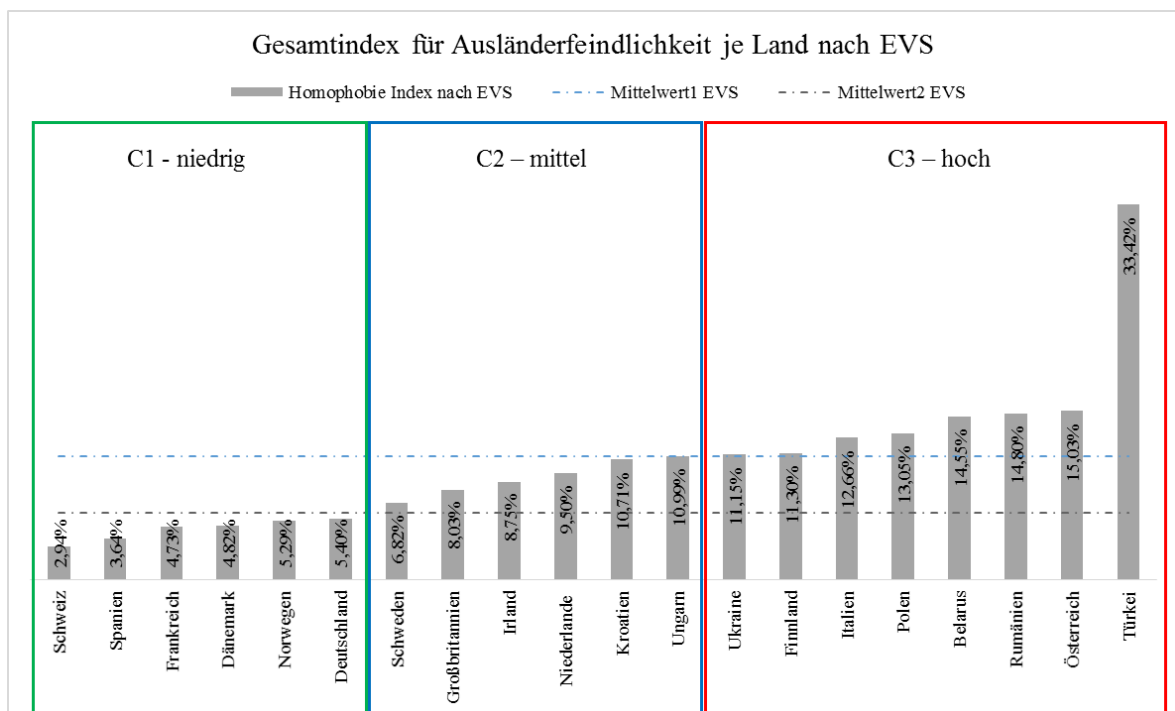


Abbildung 4-16 - Gesamtindex für Ausländerfeindlichkeit je Land nach EVS (eigene Darstellung)

Wie zuvor bereits mehrfach beschrieben, werden die Länder in aufsteigender Reihenfolge sortiert und anhand der Mittelwerte in 3 Cluster (6 Länder, 6 Länder, 8 Länder) eingeteilt, um die Daten der EVS generell mit den Daten des Google Index vergleichbar zu machen. Im ersten Cluster (C1) befinden sich wieder 6 Länder mit einem vergleichsweise geringen Index ($X < 5,99\%$). Das mittlere Cluster (C2) beinhaltet Länder mit mittelhohem Index ($5,99\% > X < 11\%$), während das letzte Cluster (C3: 8 Länder) alle Länder beinhaltet, die über dem durchschnittlichen Index ($X > 11\%$) der 20 abgebildeten Länder liegen:

⁵⁴ Original: „In political matters, people talk of 'the left' and 'the right'. How would you place your views on this scale generally speaking?“

C1 EVS: Schweiz, Spanien, Frankreich, Dänemark, Norwegen, Deutschland

C2 EVS: Schweden, Großbritannien, Irland, Niederlande, Kroatien, Ungarn

C3 EVS: Ukraine, Finnland, Italien, Polen, Weißrussland, Rumänien, Österreich, Türkei

Datenkorrelation

Korrelationen

		EVS_Index	Google TEXT_SUCHE	Google AUTO_FILL_man	Google AUTO_FILL_DMI	GESAMTINDEX_Text_DMI
EVS_Index	Korrelation nach Pearson	1	-,152	-,313	-,255	-,266
	Signifikanz (2-seitig)		,523	,178	,277	,256
	N	20	20	20	20	20
Google_TEXT_SUCHE	Korrelation nach Pearson	-,152	1	,637**	,257	,698**
	Signifikanz (2-seitig)	,523		,002	,274	,001
	N	20	20	20	20	20
Google_AUTO_FILL_man	Korrelation nach Pearson	-,313	,637**	1	,642**	,799**
	Signifikanz (2-seitig)	,178	,002		,002	,000
	N	20	20	20	20	20
Google_AUTO_FILL_DMI	Korrelation nach Pearson	-,255	,257	,642**	1	,871**
	Signifikanz (2-seitig)	,277	,274	,002		,000
	N	20	20	20	20	20
GESAMTINDEX_Text_DMI	Korrelation nach Pearson	-,266	,698**	,799**	,871**	1
	Signifikanz (2-seitig)	,256	,001	,000	,000	
	N	20	20	20	20	20

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Tabelle 4-6 - Korrelation der Google Indices zum EVS Index für Ausländerfeindlichkeit

Eine hohe Korrelation der Indices zum Thema Ausländerfeindlichkeit auf Basis der per Google erhobenen Daten besteht nach Pearson im Abgleich mit dem auf den EVS Daten basierendem Index nicht. Die Google Indices zur Text Suche ($r=-0,152$), zur Suchvervollständigung ($r=-0,255$), sowie der Gesamtindex ($r=-0,266$) liegen in einem schwachen Korrelationsbereich. Lediglich der Index der mittels manueller Suchvervollständigung erstellt wurde, liegt knapp über der Grenze der mittleren Korrelations-Güte ($r=-0,313$) Zudem zeigt sich, dass insgesamt eine relativ geringe Signifikanz der Google Daten vorliegt ($0,089 < \alpha > 0,262$, einseitiger Test). (Hair, 2006) Dennoch wird auch für die Google Indices zur Ausländerfeindlichkeit nachfolgend eine Clusterung und ein Abgleich zu den EVS Daten vorgenommen.

4.2.2 Ausländerfeindlichkeit auf Basis der Suchvervollständigung mit Google Autocomplete

Analog zum vorherigen Vorgehen gilt auch für die folgenden Google Auswertung (Suchvervollständigung und Textsuche) zum Thema Ausländerfeindlichkeit die Annahme: je höher der Index eines Landes in der EVS, desto mehr negative Aussagen zum Thema Ausländer sind bei den Analyseeinheiten von Google zu erwarten. Daher ist anzunehmen, dass ein Land mit einem hohem Index für Ausländerfeindlichkeit in der EVS, auch einen hohen Index für Ausländerfeindlichkeit mittels Google (Suchvervollständigung / Textsuche)

aufweist. Die automatische Suchvervollständigung mittels *Google Autocomplete* liefert wie zuvor bereits beschrieben ein sehr undifferenziertes Bild im Vergleich der Länder im jeweiligen Cluster. (Abbildung 4-17)

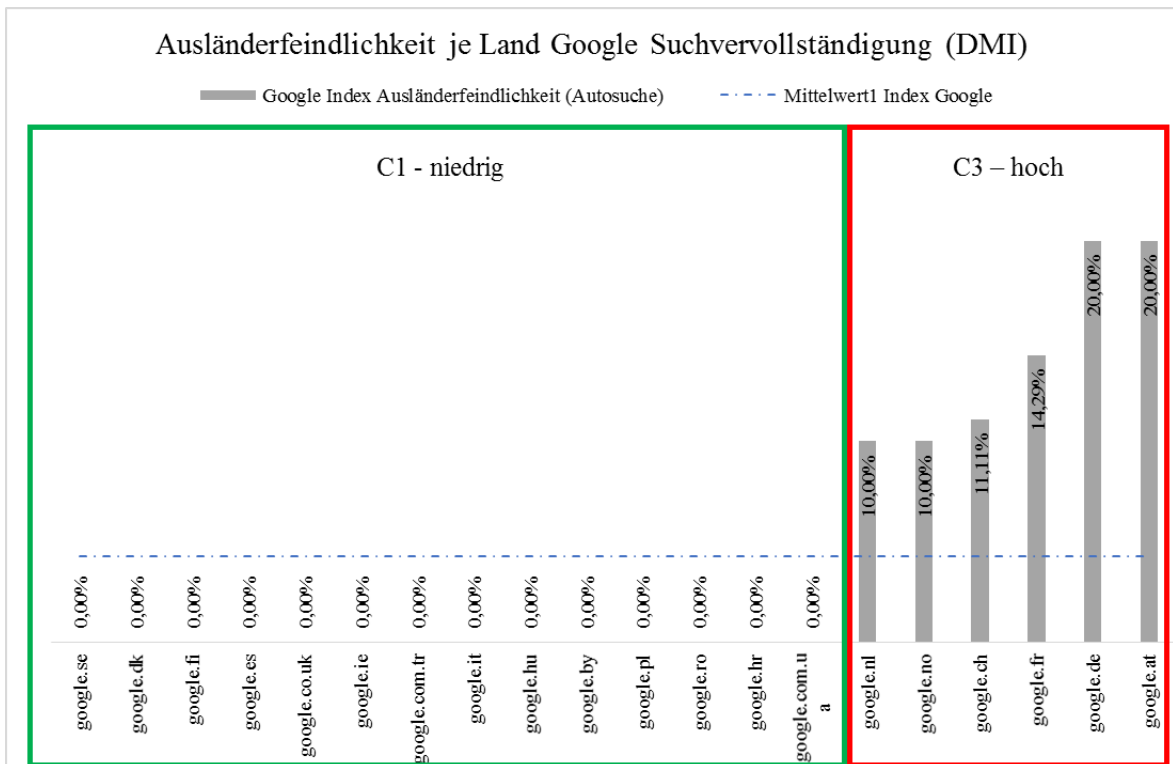


Abbildung 4-17 - Gesamtindex für Ausländerfeindlichkeit je Land nach Google Suchvervollständigung mittels Google Autocomplete (eigene Darstellung)

Dies ist zum einen auf die geringe Anzahl der Analyseeinheiten, zum anderen auf die sehr geringe Trennschärfe des Suchbegriffs selbst zurückzuführen. Zudem ist fraglich, inwiefern Google selbst in den Algorithmus eingreift, um z.B. rassistische Inhalte ggf. aus der Suchvervollständigung zu entfernen. Laut Google kann es vorkommen, dass eine Autovervollständigung nicht angezeigt wird, „weil darin die Übersetzung eines anstößigen Worts aus einer anderen Sprache enthalten ist.“ (Google Inc., 2015b) Dies könnte auch in Bezug auf rechtsradikale und rassistische Äußerungen gelten. Bei Betrachtung der Zuteilung der Länder auf die Cluster zeigt sich, dass die Methode in Kombination mit dem Suchbegriff „Ausländer“ nicht funktioniert. Eine valide Aussage zum Thema Ausländerfeindlichkeit mittels Google Autocomplete kann nicht vorgenommen werden. Zudem besteht unter anderem eine Problematik mit dem Suchbegriff selbst, die nachfolgend genauer in Kapitel 4.2.3 diskutiert wird.

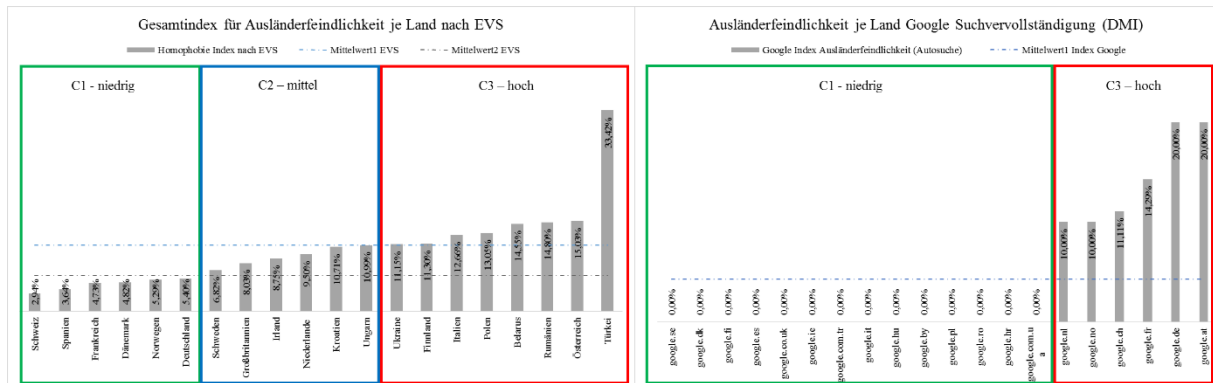


Abbildung 4-18 - EVS Index Ausländerfeindlichkeit vs. Index Ausländerfeindlichkeit Google Suchvervollständigung mittels Google Autocomplete (eigene Darstellung)

Auf Basis der Google Daten entstehen lediglich 2 Cluster. Diese bestehen aus Google Versionen, die ausländerfeindliche Analyseeinheiten in der Suchvervollständigung aufweisen und solchen, die es nicht tun. Ein valider Abgleich der Cluster ist daher nur bedingt möglich. (Abbildung 4-18)

Abgleich C1	0	Abgleich C2	0	Abgleich C3	0
	1		0		0
	0		0		0
	1		0		0
	0		0		0
	0		0		0
					1
					0
	33,33%		0,00%		12,50%
Anzahl Länder	20				
Falsch zugeordnet	17	85,00%			
Richtig zugeordnet	3	15,00%			

Tabelle 4-7 - Zuordnungsübersicht Cluster 1-3 EVS vs. Google Suchvervollständigung - Ausländerfeindlichkeit (eigene Darstellung)

Insgesamt werden lediglich 3 von 20 Ländern in das entsprechende Cluster zugeteilt. Die Quote ließe sich durch eine andere Anordnung der Länder im jeweiligen Cluster selbst verbessern (14 Länder mit 0%), wäre in ihrer Aussagekraft aufgrund der geringen Anzahl an Analyseeinheiten, sowie der gleichen Intervalllänge bei 14 Ländern (0%) aber nicht valide. (Tabelle 4-7) Daher wird der ermittelte Index zur Ausländerfeindlichkeit auf Basis der Suchvervollständigung mittels Google Autocomplete nicht weiter betrachtet.

4.2.3 Ausländerfeindlichkeit auf Basis der Google Textsuche

Die Google Textsuche zeigt ein weitaus differenzierteres Bild, als die Google Autocomplete Suche. Eine Zuordnung der Länder in 3 etwa gleichgroße Cluster ist möglich. (Abbildung 4-19)

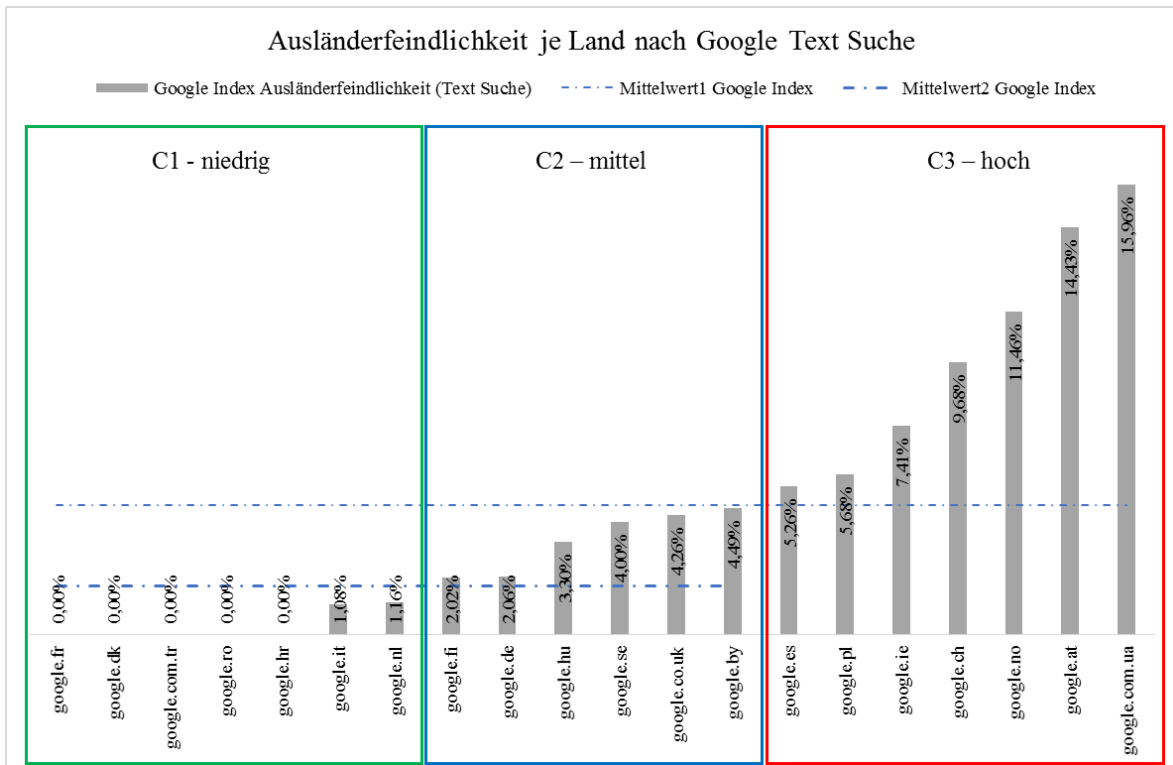


Abbildung 4-19 - Gesamtindex für Ausländerfeindlichkeit je Land nach Google Textsuche (eigene Darstellung)

Allerdings zeigt der Abgleich der Zuteilung zu den Clustern, dass die Methode im vorliegenden Themenkreis nicht analog zu den EVS Daten arbeitet. (Abbildung 4-20 & Tabelle 4-8) Auch Cohen's Kappa ($k=0,098$) und eine sehr geringe Signifikanz (0,268, da einseitig) deuten auf eine schwache Übereinstimmung der Werte hin. (Tabelle A 7)

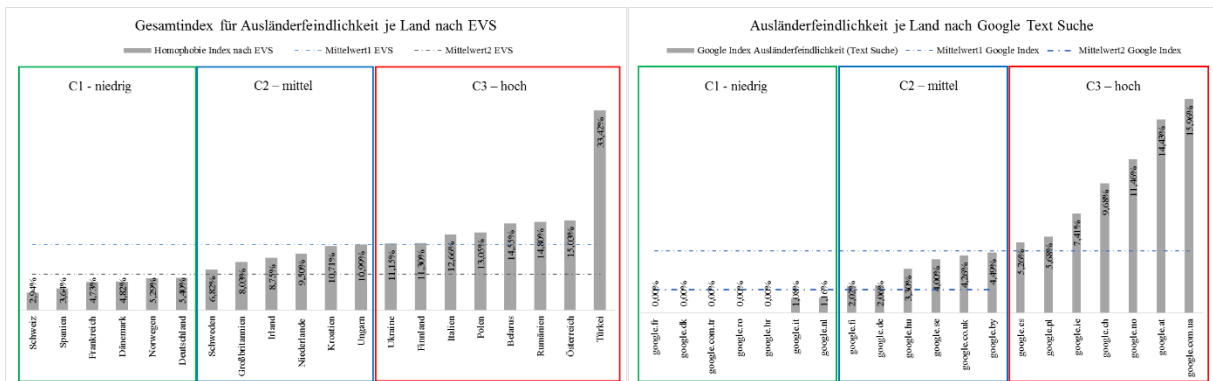


Abbildung 4-20 - Index Ausländerfeindlichkeit EVS vs. Index Ausländerfeindlichkeit Google Textsuche (eigene Darstellung)

Per Google Suche werden 3 Cluster gebildet, die zum einen zum Teil von der Größe der jeweiligen EVS Cluster abweichen: in C1 werden 7 Länder, in C2 6 Länder und in C3 wieder 7 Länder zugeordnet. Zum anderen differieren die jeweiligen Cluster hinsichtlich der Zuordnung der Länder selbst.

C1 EVS: Schweiz, Spanien, Frankreich, Dänemark, Norwegen, Deutschland

C2 EVS: Schweden, Großbritannien, Irland, Niederlande, Kroatien, Ungarn

C3 EVS: Ukraine, Finnland, Italien, Polen, Weißrussland, Rumänien, Österreich, Türkei

C1 TS : Frankreich, Dänemark, Türkei, Rumänien, Kroatien, Italien, Niederlande

C2 TS: Finnland, Deutschland, Ungarn, Schweden, Großbritannien, Weißrussland

C3 TS: Spanien, Polen, Irland, Schweiz, Norwegen, Österreich, Ukraine

Abgleich C1	0	Abgleich C2	0	Abgleich C3	1
	0		1		0
	1		0		0
	1		0		1
	0		0		0
	0		1		0
					1
					0
	33,33%		33,33%		37,50%
TOTAL	20				
Falsch zugeordnet	13	65,00%			
Richtig zugeordnet	7	35,00%			

Tabelle 4-8 - Zuordnungsübersicht Cluster 1-3 EVS vs. Google Textsuche - Ausländerfeindlichkeit (eigene Darstellung)

Insgesamt werden lediglich 35% der Länder über die Textsuche korrekt zugeordnet. (Tabelle 4-8) Dies kann zum Beispiel auf eine sehr schwache Trennschärfe des Suchbegriffs hinweisen. Das Wort „Ausländer“ ist in vielen Landessprachen nicht so eindeutig definiert, wie in der deutschen Sprache. Im türkischen steht das Wort „ecnebi“ zum Beispiel auch in Bezug zum Adjektiv *ausländisch*. In diesem Sinne entsteht ein häufiger Bezug zu den Themen *ausländische Serien*, *ausländische Filme* und *ausländische Musik*, *ausländische Waren*, *ausländische Sprachen*, etc. Das Wort Ausländer auf Türkisch ist Nomen und Adjektiv zugleich, was eine Auswertung im deutschen Wortsinn, bzw. bezogen auf den Index verhindert. Auch im kroatischen hat das Wort einen stärkeren Bezug zu *ausländischen* Filmen, Büchern und Serien, als zum gesuchten Thema „Ausländer“ selbst. Zudem besteht, in Zusammenhang mit dem Wort Ausländer, bei der Google Suche Kroatien ein Großteil der Bezüge zu Albert Camus Buch „Der Fremde“. Im englischsprachigen Raum hingegen (UK, Irland) zeigt das Wort eine Verzerrung auf Basis der inhaltlichen Bedeutungen. Das Wort „foreigner“ (engl. für Ausländer, oder auch Fremder), steht oftmals nicht in Zusammenhang mit dem Thema Ausländer selbst, sondern häufig mit der Band *Foreigner*. Fast die Hälfte aller Analyseeinheiten für google.co.uk (Großbritannien) bezieht sich auf Musikvideos, Lieder, Konzerte, etc. der Band. Diese Problematik zeigte sich auch bereits bei der Auswertung der Suchvervollständigung. Zudem ist auch in der Textsuche fraglich, ob und wenn ja, inwiefern der Google Algorithmus selbst Suchergebnisse mit rassistischen Inhalten zensiert. Eine länderübergreifende Auswertung zum Thema Ausländerfeindlichkeit anhand der Suchmaschine Google ist auf Basis des Suchbegriffs „Ausländer“ (sowie ausgeführt

nach Tabelle A 1) daher nur bedingt möglich. Eine Verbesserung der Daten auf Basis der Kombination von Textsuche und Suchvervollständigung ist nicht zu erwarten. Aufgrund der schwachen Aussagekraft von Suchvervollständigung und Textsuche wird daher auf die Berechnung eines Index, der die beiden Methoden kombiniert, verzichtet.

4.2.4 Relevante Themenkreise zur Ausländerfeindlichkeit nach Google Textsuche

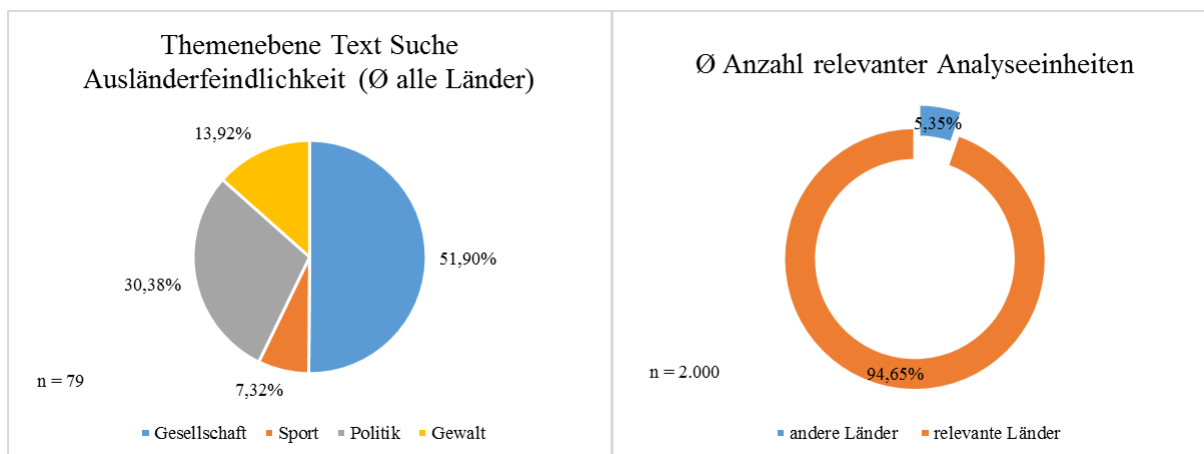


Abbildung 4-21 - links: Themenebene Ausländerfeindlichkeit Textsuche (eigene Darstellung), rechts: relative Anzahl relevanter Analyseeinheiten nach Standortbezug je Land

Trotz der insgesamt schwachen Zuteilung der Länder im Index selbst, konnten mittels Google Textsuche anhand der Analyseeinheiten Themenkreise gebildet werden, die als Basis für ausländerfeindliche Aussagen dienen. (Abbildung 4-21) Auch der Vergleich zu länderübergreifenden Bezügen ist gegeben. Lediglich 5,35% der Analyseeinheiten zum Thema Ausländer verweisen durchschnittlich auf andere Länder. Aussagen, die sich auf ein anderes Land beziehen, sind bei dem Suchbegriff Ausländer vor allem im Zusammenhang mit Nachrichten gegeben. Es werden vorwiegend Aussagen wie „Ausländer für Terroranschlag in Bangkok verantwortlich“ getätigt. Zudem hat der Begriff gerade im türkischen, wie vorab diskutiert einen starken Bezug zu ausländischen Gütern. Generell finden sich in den rechtsradikalen Analyseeinheiten zumeist radikale Parolen oder Aussagen die sich auf die Gesellschaft beziehen (51,9% aller Analyseeinheiten). (Abbildung 4-21)

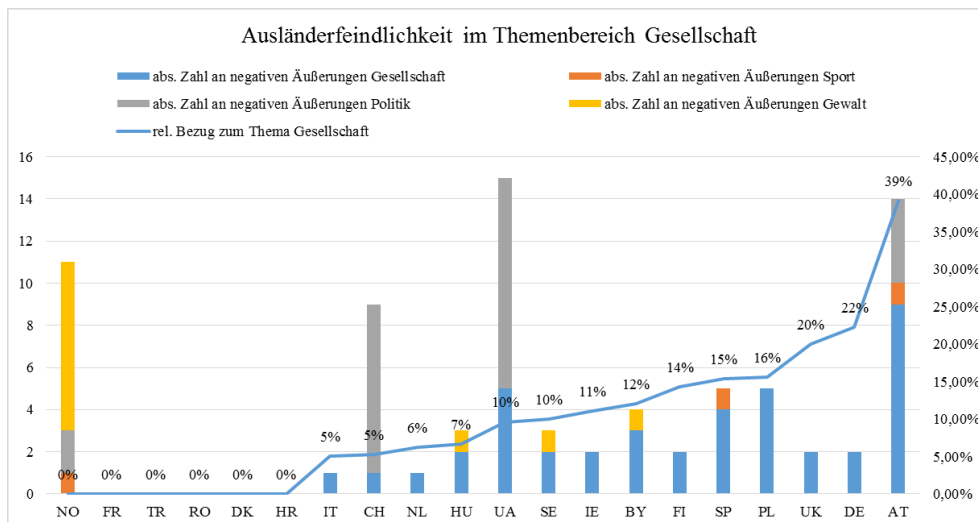


Abbildung 4-22 - Ausländerfeindlichkeit auf Basis der Google Textsuche im Themenbereich Gesellschaft (eigene Darstellung)

In den gesellschaftlichen Themenkreis sind vor allem sinngemäße Aussagen zum Thema, wie „Ausländer sind schlecht für unser Land, da sie unsere Kultur verwässern“, einbezogen. Dies geschieht z.B. in Ungarn und Polen durch die Aussage, dass Ausländer das Land und / oder die Immobilien aufkaufen, die Arbeitsplätze wegnehmen und so den Ungarn, bzw. den Polen selbst wertvollen Wohnraum und Wohlstand entziehen. In Weißrussland hingegen fallen vor allem Aussagen zur negativen Beeinflussung der Kultur („Sprache wird durch die Ausländer negativ beeinflusst“) auf. In Ländern wie Deutschland, Spanien und Österreich sind es hingegen eher Plattitüden wie „Ausländer raus“, bzw. „Zigeuner raus“, die die gesellschaftliche Intoleranz im Netz repräsentieren. Auch Facebook Gruppen mit ähnlichen rechtsradikalen Aussagen sind in diesen Ländern via Google zu finden. (Abbildung 4-22) Zudem finden sich politische Merkmale (30,38% der Analyseeinheiten) in den Analyseeinheiten in Zusammenhang mit ausländerfeindlichen Äußerungen.

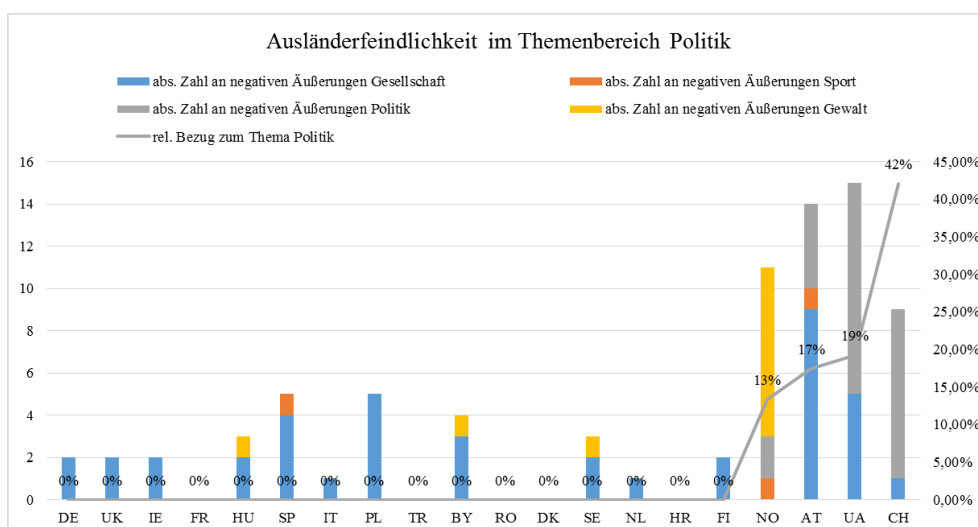


Abbildung 4-23 - Ausländerfeindlichkeit auf Basis der Google Textsuche im Themenbereich Politik (eigene Darstellung)

Die Länder Österreich und Schweiz weisen in den Suchergebnissen (starke) Bezüge zu rechtsradikalen Parteien auf (ÖVP & junge SVP). In der Ukraine wird dagegen eher von „Russen“ als „Ausländer“ gesprochen und die Ausländerfeindlichkeit auf der politischen Ebene sichtbar. Die „Ausländer“ haben einen negativen Einfluss auf die Regierung und sind daher „schlechte Ausländer“, bzw. schlecht für die Ukraine. (Abbildung 4-23)

Analyseeinheiten, die in Zusammenhang mit dem Thema Gewalt stehen, basieren meist auf Berichterstattungen von Zeugen oder Zeitungen, die von Gewalttaten mit rechtsradikalem Hintergrund berichten. Diese sind länderübergreifend allerdings nicht sehr weit verbreitet (13,92% der Analyseeinheiten).

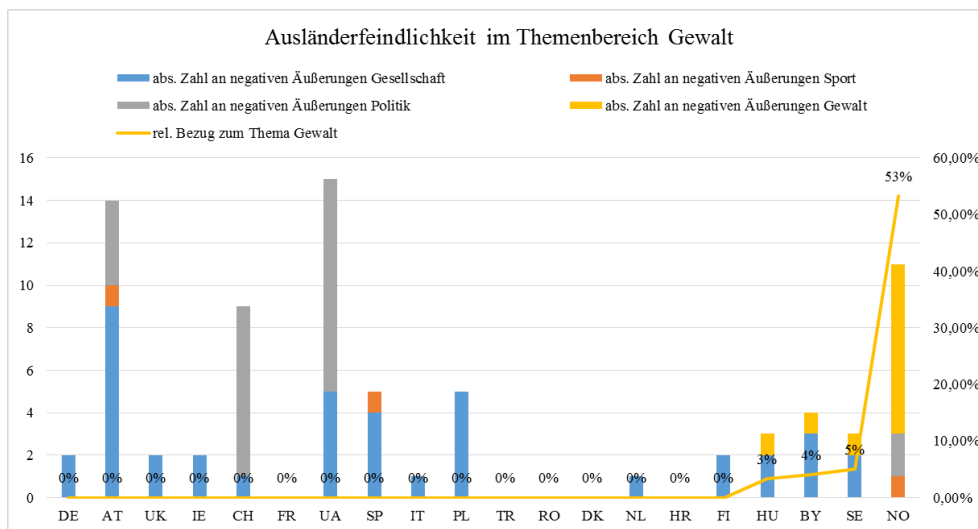


Abbildung 4-24 - Ausländerfeindlichkeit auf Basis der Google Textsuche im Themenbereich Gewalt (eigene Darstellung)

Vor allem in Norwegen wird vermehrt von einer Gewalttat gegen einen Ausländer berichtet, der aufgrund seiner dunklen Hautfarbe geschlagen wurde. Auch in Weißrussland, Ungarn und Schweden finden sich Berichte über Gewaltanwendung mit ausländerfeindlichem Hintergrund. (Abbildung 4-24)

4.3 Index: „Toleranz von Rauschmitteln“ nach Google

4.3.1 Vergleichsdaten: Toleranz von Rauschmitteln (WHO / EMCDDA)

Wie in Kapitel 2.3.2 bereits angemerkt, bietet die EVS keinen ausreichend großen Datensatz für den Abgleich zum Thema Toleranz von Rauschmitteln. Daher werden für den nachfolgenden Abgleich der Indices zum Thema Toleranz von Alkohol, bzw. Marihuana die Daten der WHO und der EMCDDA verwendet. Der Index zur Toleranz von Drogen nach WHO (Alkohol), bzw. EMCDDA (Marihuana) bezieht sich auf die in den zwanzig untersuchten Ländern konsumierte Menge der jeweiligen Substanz. Die konsumierten Mengen, bzw. die Menge an Konsumenten, unterscheidet sich in den jeweiligen Kulturkreisen innerhalb Europas. Die Ausgangsthese zum Abgleich der Daten mit den Google Daten lautet daher: je mehr von einer Substanz konsumiert wird, bzw. je häufiger Menschen es tun, desto weiter verbreitet ist die Droge in der jeweiligen Gesellschaft. Eine Droge, die in einer Gesellschaft weiter verbreitet ist als eine andere, erfährt innerhalb der Gesellschaft mehr Akzeptanz. Das heißt im Umkehrschluss, je weniger Liter reinen Alkohols pro Kopf in einem Land getrunken werden, bzw. je weniger Menschen in einer Gesellschaft Marihuana konsumieren, desto geringer ist die Toleranz für die jeweilige Substanz innerhalb der Gesellschaft. Der Index für Toleranz zum Thema Alkohol bezieht sich somit auf die Anzahl der pro-Kopf konsumierten Liter reinen Alkohols pro Jahr je Land, für Personen ab einem Alter von 15 Jahren. Folglich gilt: Je weniger Alkohol pro-Kopf in einem Land getrunken wird, desto geringer ist die Akzeptanz für Alkohol in diesem Land.

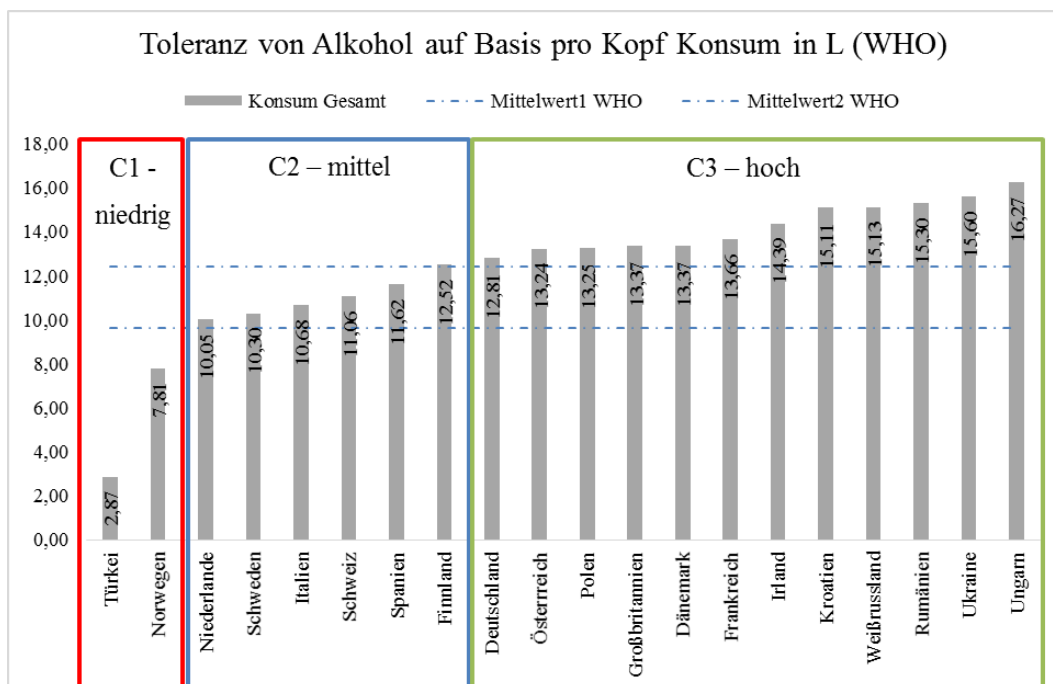


Abbildung 4-25 - Toleranz von Alkohol auf Basis der getrunkenen Menge (World Health Organization, 2011)

Eine Zuteilung der Länder in 3 Cluster nach WHO Daten ist möglich. Dabei gilt, dass Länder in C1 ($<9,6$ L) eine eher geringe Toleranz für Alkohol haben (geringer Konsum), Länder in C2 eine mittlere (Konsum von $9,6 < X < 12,4$ L reinen Alkohols pro anno) und Länder in C3 (mehr als 12,5 Liter pro-Kopf Konsum pro anno) eine hohe Toleranz aufweisen. (Abbildung 4-25)

Der Toleranz-Index für Marihuana hingegen bezieht sich auf den relativen Anteil der Konsumenten je Land. Auch hier gilt: Je weniger Menschen in einem Land Marihuana konsumieren, desto geringer ist die Toleranz für die Substanz.

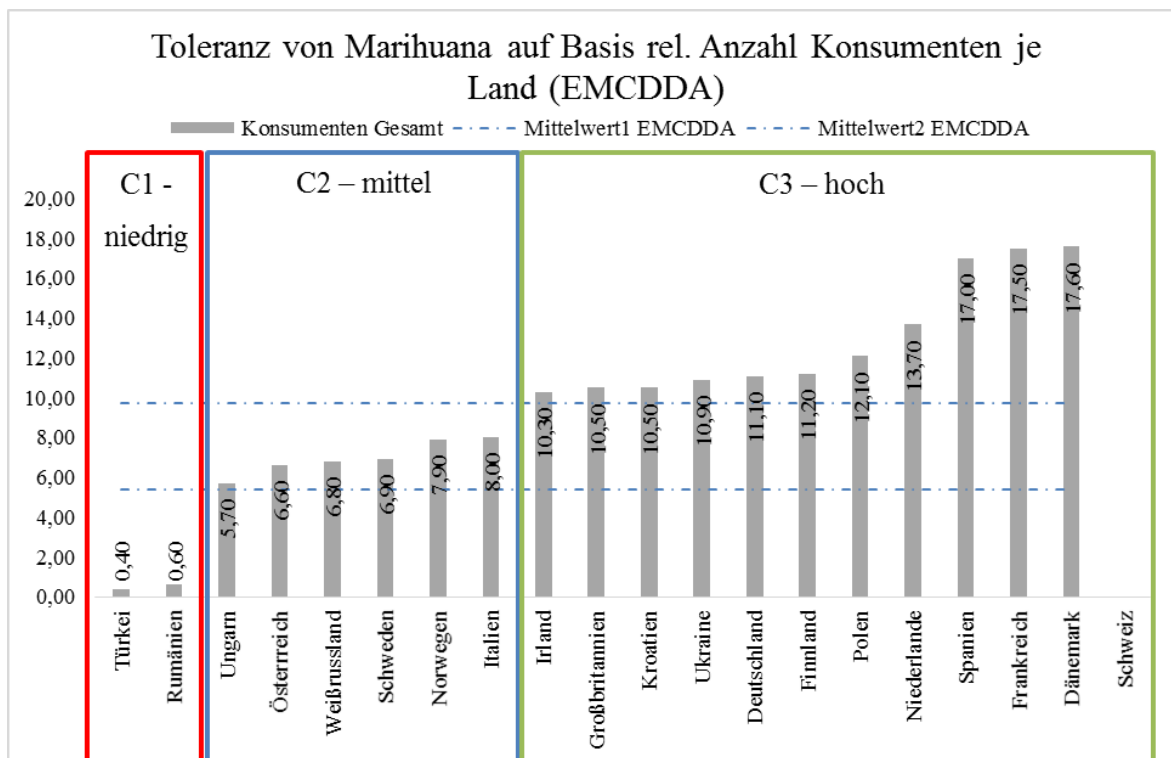


Abbildung 4-26 - Toleranz von Marihuana auf Basis Anzahl der Konsumenten je Land (EMCDDA, 2014)

Die Studie der EMCDDA bezieht sich dabei auf die Altersgruppe 15 - 34 Jahre je Land. In der Schweiz wurden für die Studie der EMCDDA keine Daten erhoben, daher wird die Schweiz nicht in den Abgleich einbezogen. Dennoch ist auch hier eine Zuteilung der Länder in 3 Cluster möglich. C1 beinhaltet Länder mit geringer Toleranz ($<5,7\%$ der Gesellschaft konsumiert Marihuana), C2 beinhaltet Länder, in denen vergleichsweise mäßige viele Menschen Marihuana konsumieren (5,7 - 10%). C3 bildet Länder ab, in denen eher viel Marihuana konsumiert wird (mehr als 10% der Einwohner), dies spricht für eine hohe Toleranz. (Abbildung 4-26)

Datenkorrelation

Nach Hair ist keine signifikante Korrelation zwischen dem Alkoholkonsum je Land im Vergleich mit dem Google Index der Textsuche zur Toleranz von Alkohol gegeben ($r=-$

0,253). Auch die Signifikanz (0,141, einseitige Korrelation) der Daten ist eher gering. (Tabelle 4-9) Die Korrelation zwischen der relativen Anzahl von Marihuana-Konsumenten je Land im Vergleich zu den Google Daten weist zudem auch einen geringen Pearson Wert ($r=-0,204$) aus. Die Signifikanz der Daten (0,101 einseitige Korrelationen) hingegen zeigt eine leichte Verbesserung der Werte im Vergleich zu den Google Indices zum Thema Alkohol. (Tabelle 4-10) Nachfolgend werden trotzdem sowohl die Google Daten zur Toleranz von Alkohol, als auch die Daten zur Toleranz von Marihuana, detailliert mit den jeweiligen Studiendaten abgeglichen.

Korrelationen

		WHO	TS_ALK
WHO	Korrelation nach Pearson	1	-,253
	Signifikanz (2-seitig)		,282
	N	20	20
TS_ALK	Korrelation nach Pearson	-,253	1
	Signifikanz (2-seitig)	,282	
	N	20	20

Tabelle 4-9 - Korrelation der Google Indices zum WHO Index für Alkohol (SPSS)

Korrelationen

		EMCDDA	TS_MAR
EMCDDA	Korrelation nach Pearson	1	-,204
	Signifikanz (2-seitig)		,403
	N	19	19
TS_MAR	Korrelation nach Pearson	-,204	1
	Signifikanz (2-seitig)	,403	
	N	19	19

Tabelle 4-10 - Korrelation der Google Indices zum EMCDDA Index für Marihuana (SPSS)

4.3.2 Toleranz von Rauschmitteln auf Basis der Google Suchervollständigkeit

Die Suchervollständigkeit mittels Google Autocomplete liefert in Zusammenhang mit dem Suchbegriff „illegale Drogen“ keinen Bezug zu den Themen Alkohol, bzw. Marihuana. Das heißt, es werden Suchervollständigungen ausgegeben, jedoch gibt keine der untersuchten Google Landesversionen in der Suchervollständigkeit in Zusammenhang mit dem Suchbegriff „illegale Drogen“ die Wörter „Alkohol“, „Marihuana“, o.ä. aus. Häufig wird in Zusammenhang mit dem Suchbegriff „illegale Drogen“ über die Suchervollständigkeit eine Auflistung der illegalen Substanzen oder der Handel im Web thematisiert. Eine Untersuchung der Suchervollständigkeit mit dem Suchbegriff „illegale Drogen“ mittels implizitem Verfahren ist in Bezug auf die Themen Alkohol und Marihuana daher nicht möglich. Auch die Suchervollständigkeit mit den Begriffen *Alkohol* oder *Marihuana* selbst

liefert bei Google Autocomplete keine zum Thema passenden Suchergebnisse. Eine weitere Untersuchung der Suchervollständigkeit wird daher nicht vorgenommen.

4.3.3 Toleranz von Rauschmitteln auf Basis der Google Textsuche (TOP 100)

Anders als die Suchervollständigkeit, liefert die Textsuche Treffer im Zusammenhang des Suchbegriffs „illegale Drogen“ mit den Begriffen „Alkohol“ und / oder „Marihuana“. Eine Zuteilung der Google Landesversionen auf 3 Cluster ist daher möglich. Für die nachfolgenden Analysen gilt, dass das Cluster C1 der jeweiligen Abgleichsstudie eine geringe Akzeptanz zur jeweiligen Substanz in der Gesellschaft abbildet. Dies gilt, da die Studien einen geringen Konsum der Droge in der jeweiligen Gesellschaft abbilden – dies deutet, wie bereits diskutiert, auch auf eine geringe Toleranz hin. C2 bildet in den Abgleichsstudien somit eine mittlere, C3 eine hohe Akzeptanz der jeweiligen Substanz ab. Für die jeweilige Google Landesversion gilt, dass eine hohe Anzahl Analyseeinheiten mit den Wörtern „Alkohol“, bzw. „Marihuana“, in Zusammenhang mit dem Suchbegriff „illegale Drogen“, auf eine geringe Toleranz hindeuten (Cluster 3). Daher gilt, dass je geringer die Menge an Analyseeinheiten, die bei einer Suche mit dem Suchbegriff „illegale Drogen“ die Wörter Alkohol / Marihuana enthalten, desto höher die Akzeptanz in der jeweiligen Gesellschaft. Cluster 1 wird für die Google Suche daher als Indikator für eine hohe Akzeptanz der Substanz in der jeweiligen Gesellschaft interpretiert, während C2 eine mittlere Akzeptanz (mittlere Anzahl an Analyseeinheiten, die die Wörter Alkohol oder Marihuana mit dem Suchbegriff „illegale Drogen“ verbindet) der Substanz in der Gesellschaft abbildet.

Alkohol

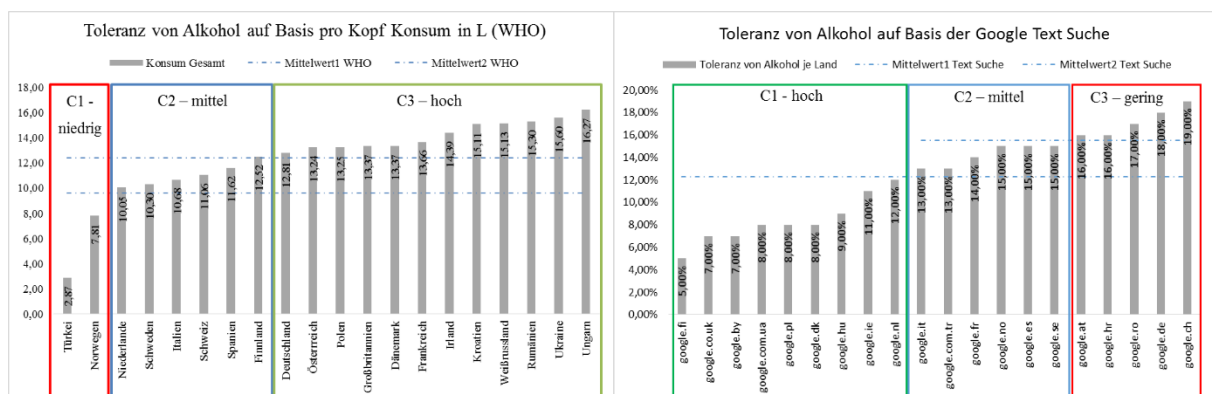


Abbildung 4-27 - WHO Index Toleranz von Alkohol vs. Toleranz von Alkohol auf Basis der Google Textsuche (eigene Darstellung)

Die per Google Suche erhobenen Daten werden zu 3 Clustern sortiert. Zum Teil weicht die Größe der Cluster, die mittels Google Daten erstellt wurden, von der Größe der jeweiligen

5 Exkurs: Analyse der Google Bildersuche

Aufgrund der großteils schwachen Übereinstimmung der generierten Google Daten mit den jeweiligen Studiendaten, wird anhand eines Exkurses eine weitere Möglichkeit aufgezeigt, Google als Quelle zur Datengenerierung zu nutzen. Neben der Suchvervollständigung und der Textsuche bietet Google unter anderem auch die Möglichkeit der Bildersuche. Die Generierung des Datenmaterials besteht für die Bildersuche aus dem in Kapitel 3.1 beschriebenen Vorgehen. Hierfür wurden, wie bereits diskutiert, die Geo-IP (mittels VPN) und die Landessprache passend zur jeweiligen Google Landes ccTLD genutzt, um mittels Bildersuche Daten zu einem bestimmten Thema zu generieren. Jedoch wurde die Kodierung der Bilder ohne die Hilfe von Übersetzern vorgenommen, da in diesem Exkurs angenommen wurde, dass Bilder universell verständlich sind und daher keiner Übersetzungshilfe bedürfen. Da es bislang kein einheitliches Verfahren zur qualitativen Bild Inhaltsanalyse gibt, wird nachfolgend kurz das Vorgehen zur Kategorisierung der Bilder erklärt. Zudem werden aufgrund des begrenzten Umfangs der Masterarbeit lediglich die Themen des expliziten Verfahrens (*Homophobie* und *Ausländerfeindlichkeit*) mit Hilfe der Google Bildersuche untersucht. Die Anzahl der Analyseeinheiten wird dabei analog zur Textsuche auf 100 Analyseeinheiten (100 Bilder pro Google Landesversion) begrenzt. Im Rahmen der Themen *Homophobie* und *Ausländerfeindlichkeit* begrenzt sich die inhaltliche Analyse der Bilder in diesem Exkurs zudem auf die Wirkung, bzw. die Aussage, die das Bild im Zusammenhang mit dem jeweiligen Thema hat. Dabei werden alle anderen quantitativen und qualitativen (zeitliche, räumliche, soziale, etc.) Bezüge im Bild und zum Bild ausgeblendet. Das heißt, die Darstellung auf dem Bild wird lediglich in Bezug auf das jeweilige Thema betrachtet. So gilt für die nachfolgende Analyse im Rahmen dieses Exkurses, dass ein Bild einen negativen, neutralen, oder positiven Bezug zu dem jeweiligen Thema (Homosexualität / Ausländer) haben kann. Negative Äußerungen können sich in Bildinhalten zum Beispiel in Form von Beschimpfungen (Text im Bild), der Abbildung von Gewaltanwendung oder von Verunglimpfungen von Menschen und / oder Symbolen einer Ideologie oder Bewegung, o.ä. ausdrücken.

Bei einer Verzerrung der inhaltlichen Interpretation von Bildern spielt vor allem die subjektive Betrachtung und die soziale, bzw. empathische Prägung der Coder eine große Rolle. Da im vorliegenden Versuchsaufbau allerdings lediglich der Autor kodiert, ist eine Verzerrung aufgrund sozialer und empathischer Unterschiede zwischen den Codern nicht möglich. Für einen möglichst exakten Kodier-Prozess wurde zudem der für die Textsuche

bzw. Suchvervollständigung konzipierte Codeplan genutzt, um zu generalisieren, wann ein Bild negativ, neutral oder positiv bewertet wird. Zudem wurde anhand der Intra-Coder Reliabilität eine gute interne Codier-Güte in Höhe von 85,2% gemessen. Somit kann via Google Bildersuche (anhand der Kodierung), über die Anzahl der als negativ eingestuften Analyseeinheiten in Relation zu den relevanten Analyseeinheiten, ein Index zur Homophobie, bzw. zur Ausländerfeindlichkeit ermittelt werden. Dabei ist aufgrund fehlender Sprachkenntnisse eine Verzerrung der Ergebnisse nicht zwingend auszuschließen. Hierbei sind vor allem Memes und sonstige Texte, die über die Bildsuche zu finden sind, gemeint. Als Vergleichsdaten zur Validierung der mittels Google generierten Daten werden wieder die bereits in Kapitel 4.1.1 und 4.2.1 verwendeten Daten der EVS herangezogen. Die Kodierung der Bilder selbst erfolgte analoge zur Kodierung der Text- und Suchvervollständigung. D.h., dass zwischen themenrelevanten (1 = relevant) und nicht themenrelevanten (0 = irrelevant) Inhalten unterschieden wurde. Zudem wurden die Bilder nachfolgend anhand auch auf ihre Inhaltsebene hin untersucht und verkodet. (siehe auch Kapitel 3.3)

5.1 Homophobie auf Basis der Google Bildersuche

Die Vergleichsdaten der EVS zum Thema Homophobie wurden in Kapitel 4.1.1 bereits ausgiebig diskutiert. Auch die per Bildersuche generierten Daten werden nachfolgend mit diesen EVS Daten abgeglichen. Dabei gilt auch für die Indexierung der Bildersuche, dass je höher ein Land im Homophobie Index der EVS abschneidet, desto mehr negative Analyseeinheiten auf die ersten 100 Analyseeinheiten (Bilder) zu erwarten sind. Somit repräsentiert die Anzahl der negativen Analyseeinheiten in Relation zu den 100 untersuchten Bildern pro Google Landesversion den Index für Homophobie mittels Bildersuche.

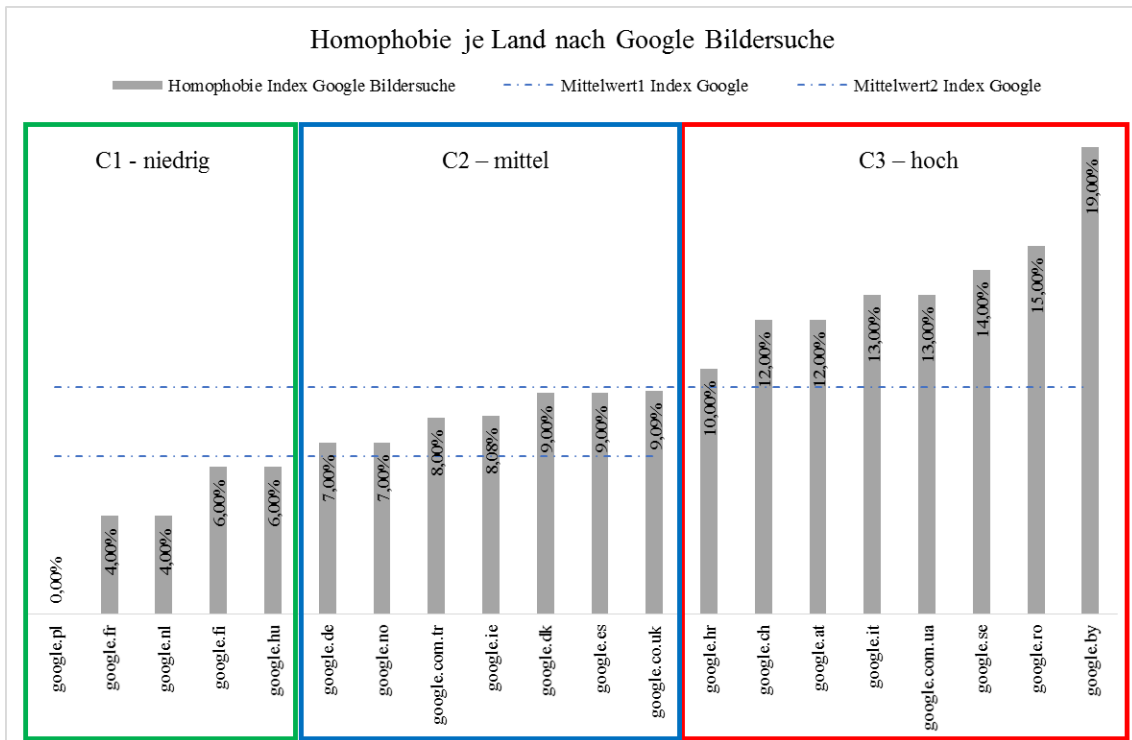


Abbildung 5-1 - Gesamtindex für Homophobie je Land nach Google Bildersuche (eigene Darstellung)

Die Länder können anhand der Google Bildersuche über die Mittelwert-Betrachtung wieder in 3 Cluster zugeordnet werden. C1 beinhaltet 5 Länder, C2 7 und C3 8 Länder. Dies entspricht, mit einer leichten Abweichung, der Zuordnung der Länder durch die EVS. (Tabelle 5-2) Nach Pearson ist im Abgleich der erhobenen Google Daten zu den EVS Daten dennoch lediglich eine schwache Korrelation ($r = 0,246$) gegeben. (Tabelle 5-1)

Korrelationen

		HomophobieIndexEVS	BILDSUCHE
HomophobieIndexEVS	Korrelation nach Pearson	1	,246
	Signifikanz (2-seitig)		,295
	N	20	20
BILDSUCHE	Korrelation nach Pearson	,246	1
	Signifikanz (2-seitig)	,295	
	N	20	20

Tabelle 5-1 - Korrelation der Google Indices der Bildersuche zum EVS Index Homophobie (SPSS)

Auch der direkte Abgleich der Cluster lässt mit einer Quote von 50% auf eine eher zufällige Zuteilung der Länder zu den jeweiligen Clustern schließen. (Tabelle 5-2) Zudem liegt auch Cohen's Kappa ($k=0,242$, Tabelle A 10) mit einer Datenreliabilität von 4-15% eher in einem geringen Bereich. (Tabelle A 2)

eingestuft. Ein hoher Anteil an negativ gewerteten Bildern spiegelt über Google daher einen hohen Grad an Ausländerfeindlichkeit im jeweiligen Land wieder.

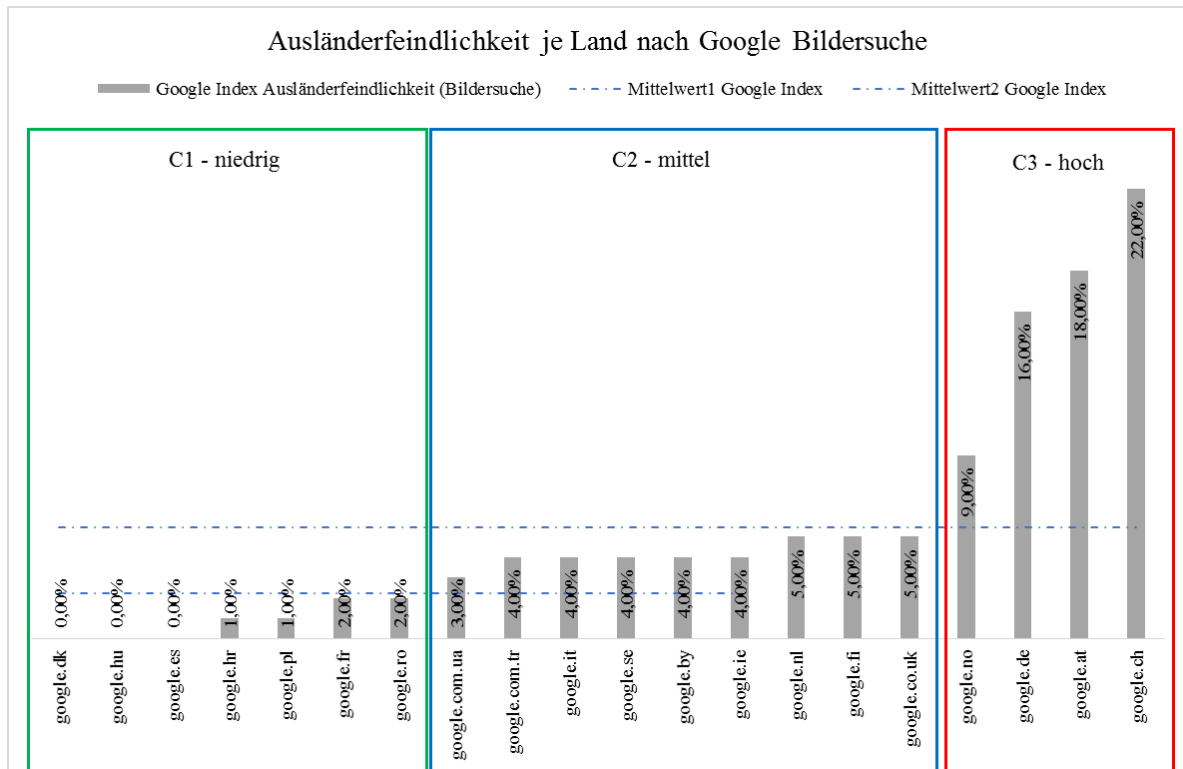


Abbildung 5-2- Gesamtindex für Ausländerfeindlichkeit je Land nach Google Bildersuche (eigene Darstellung)

Bei der Bildersuche bestehen weiterhin die bereits in Kapitel 4.2.3 diskutierten Probleme der Definition, bzw. der Mehrdeutigkeit des Wortes „Ausländer“ in diversen anderen Landessprachen. Auch in der Bildersuche lässt sich dieser Effekt wiederfinden. So werden im angelsächsischen Raum (UK, Irland) hauptsächlich Bilder einer Band (*Foreigner*) zum Suchbegriff geliefert. Auch in anderen Ländern (Polen, Türkei, Rumänien) sind thematisch besonders häufig Filme und Bücher mit dem Suchbegriff „Ausländer“ zu finden.

Korrelationen

		EVS	BILDSUCHE
EVS	Korrelation nach Pearson	1	-,271
	Signifikanz (2-seitig)		,248
	N	20	20
BILDSUCHE	Korrelation nach Pearson	-,271	1
	Signifikanz (2-seitig)	,248	
	N	20	20

Tabelle 5-3 - Korrelation der Google Indices der Bildersuche zum EVS Index Ausländerfeindlichkeit (SPSS)

Auch die Korrelation nach Pearson zeigt für die Google Indices der Bildersuche im Vergleich zum EVS Index zum Thema Ausländerfeindlichkeit eine schwache Übereinstimmung ($r=-0,271$). (Tabelle 5-3)

Abgleich C1	0	Abgleich C2	1	Abgleich C3	0
	1		1		0
	1		1		0
	1		1		0
	0		0		0
	0		0		0
					1
					0
	50,00%		66,67%		12,50%
TOTAL	20				
Falsch zugeordnet	12	60,00%			
Richtig zugeordnet	8	40,00%			

Tabelle 5-4 - Zuordnungsübersicht Cluster 1-3 EVS vs. Google Bildersuche - Ausländerfeindlichkeit (eigene Darstellung)

Die detaillierte Zuordnungsübersicht zeigt, dass vor allem in Cluster 3 lediglich nur ein einziges Land richtig zugeordnet wird. (Tabelle 5-4 - Zuordnungsübersicht Cluster 1-3 EVS vs. Google Bildersuche - Ausländerfeindlichkeit (eigene Darstellung)) Hierdurch wird die Genauigkeit der Gesamtzuordnung massiv beeinflusst. Demnach ist die gesamte Datenreliabilität nach Cohen ($k=0,118$) in einem sehr geringen Bereich (0-4%). (Tabelle A 2, Tabelle A 11) Insgesamt bildet die Bildersuche das Thema Ausländerfeindlichkeit aber differenzierter ab, als die Suchervollständigkeit zum gleichen Thema. Dennoch ist auch bei der Bildersuche ein sehr geringes Spektrum an Intervalllängen in den jeweiligen Clustern zu erwähnen. (Abbildung 5-2) Ein direkter Abgleich mit den EVS Daten zeigt zudem, dass die Zuordnung der Länder zu den Clustern mit 40% eher ungenau stattfindet.

6 Ergebnisse: (K)Eine Abbildung der Gesellschaft mit Hilfe von Google

Eine Abbildung der gesellschaftlichen Einstellung zu den Themen *Homophobie*, *Ausländerfeindlichkeit* und *Toleranz von Rauschmittel* analog zu den Daten der EVS / WHO / EMCDDA ist mit Hilfe der Suchmaschine Google nicht einheitlich möglich. Einzig der Index, auf Basis der mittels Google Textsuche (explizites Vorgehen) generierten Daten zum Thema Homophobie korreliert im Vergleich der 20 Länder stark und stimmt größtenteils mit den Daten der EVS überein. Eine hohe Korrelation der Google Daten mit anderen Themen (Ausländerfeindlichkeit, Toleranz von Drogen), bzw. mit anderen Vorgehensweisen (Suchvervollständigung, Bildersuche) ist nicht – zumindest nicht in guter Übereinstimmung mit den jeweiligen Studien-Daten – gegeben. Dennoch bleibt zu diskutieren, inwiefern *self-reported data*⁵⁵ und Web Daten eine signifikante Korrelation aufweisen. Boase und Ling untersuchten z.B. im Jahr 2013 die Nutzungsdauer von mobilen Geräten über die Log-Daten der jeweiligen Mobilfunkanbieter. Im Abgleich mit den von den Nutzern selbst eingeschätzten Daten zur Mobilfunknutzung zeigte sich, dass viele Nutzer nicht richtig einschätzen können, wie lange sie ihr mobiles Gerät wirklich am Tag nutzen. Eine Diskrepanz zwischen Selbsteinschätzung und Web Daten ist keine Seltenheit. (Boase & Ling, 2013) Auch die Ergebnisse zur Toleranz von Rauschmitteln, auf Basis einer impliziten Suche, bilden die untersuchten Gesellschaften anhand der Google Landesversion nicht analog zu den jeweiligen Studienergebnissen der WHO bzw. der EMCDDA ab. Im Abgleich der Studiendaten mit den Google Daten bleibt somit festzuhalten, dass anhand der untersuchten Vorgehensweisen zur Datengenerierung mittels Google nur bedingt ein Abbild der Gesellschaft erstellt werden kann, bzw. sichtbar wird. Dennoch bilden die mittels Google generierten Daten in allen Fällen Unterschiede zwischen den Ländern im europäischen Raum ab und ermöglichen somit einen interkulturellen Vergleich. Die generierten Daten zeigen, welche Länder im Vergleich der untersuchten Länder homophob(er), ausländerfeindlich(er) oder tolerant(er) beim Konsum von Rauschmittel sind. So zeigt sich, dass beispielsweise Länder wie die Türkei, Ungarn, Polen oder die Ukraine eine im Vergleich zu den anderen Ländern geringere Akzeptanz für das Thema Homosexualität aufbringen. In Länder wie Frankreich, Großbritannien, Norwegen und den Niederlanden ist Homophobie in Folge der Google (Textsuche) Indices hingegen vergleichsweise weniger

⁵⁵ *Self-reported data* im Sinne von Selbsteinschätzung, Befragung über eigene Fertigkeiten

verbreitet. Zudem konnten relevante Themenkreise identifiziert werden, die die Basis für Homophobie in den jeweiligen Ländern bilden. Vor allem der Bereich Religion (Homosexualität als Strafe Gottes / als Verstoß gegen die Bibel / den Koran, etc.), der Bereich Gesellschaft (Homosexualität als Hindernis für gesellschaftliche Entwicklung / Adoption von Kindern, etc.) und das Thema Krankheit (Heilung von Homosexualität, Psychische Erkrankungen, etc.) spielen wichtige Rollen in der Entstehung von Homophobie in den verschiedenen Kulturkreisen.

Im Themenbereich Ausländerfeindlichkeit zeigt sich nach Google vor allem, dass Länder wie Spanien, Polen, Österreich, die Schweiz oder die Ukraine im Vergleich der 20 Länder stärkere ausländerfeindliche Tendenzen aufweisen, als die anderen untersuchten Länder. Länder wie Dänemark, Rumänien, die Türkei, Italien und Kroatien hingegen sind nach Google vergleichsweise in einem niedrigen Segment, bezogen auf das Thema Ausländerfeindlichkeit, anzuordnen. Bei der Entstehung von Ausländerfeindlichkeit spielen nach Google vor allem die Ebenen gesellschaftliche Ausländerfeindlichkeit (negative Einflüsse durch Ausländer auf die eigene Kultur, etc.) und politische Ausländerfeindlichkeit (rechtsradikale Parteien, etc.) eine wichtige Rolle.

Für das Thema Toleranz von Rauschmitteln konnte keine Übersicht zu Themenfeldern erstellt werden, da hier das implizite Suchverfahren angewandt wurde. Nach Google sind vor allem die Länder Kroatien, Rumänien, Deutschland und die Schweiz intolerant im Umgang mit dem Konsum von Alkohol. Auch die Toleranz von Marihuana ist nach Google vor allem in Spanien, Rumänien, den Niederlanden und Österreich eher gering. Eine vergleichsweise hohe Akzeptanz für den Konsum von Alkohol findet sich nach Google vor allem in Ländern wie Finnland, Großbritannien, Weißrussland, Ukraine, Polen und Dänemark. Während Marihuana vergleichsweise vor allem in der Ukraine, Deutschland, Irland, Ungarn, Italien, Norwegen, Polen, Schweden und Finnland eine relativ hohe Akzeptanz in der Gesellschaft erfährt. Die Ergebnisse der Google Indices stimmen insgesamt allerdings oftmals nicht mit denen der EVS / WHO / EMCDDA überein. Daher stellt sich vor allem die Frage, welche demographische Gruppe durch die Indices der Google Suche in den jeweiligen Ländern repräsentiert wird. Sind die Werte zur z.B. Ausländerfeindlichkeit nicht repräsentativ für die Gesamtbevölkerung, können sie dennoch Validität und Reliabilität für einen Teil der jeweiligen Bevölkerung besitzen. Die teils mangelnde Übereinstimmung zwischen Google Daten und Studiendaten kann also evtl. auf einen Selektionseffekt zurückgeführt werden. Dieser könnte darin begründet sein, dass manche Gruppen, durch gewisse Dispositionen, eher dazu geneigt sind Google stärker zu

nutzen, als andere. Die Frage, die sich somit im Zusammenhang mit der Validität der Daten stellt, ist: Wer nutzt Google in den jeweiligen Ländern? Und weiter: Sind die generierten Daten repräsentativ für die Nutzer der jeweiligen Google Landesversionen, bzw. der Länder? Zudem gilt festzuhalten, dass vor allem die Google Textsuche Ergebnisse liefert, die eine differenzierte Darstellung der Länder ermöglicht. Darüber hinaus liefert die Analyse der mittels Textsuche generierten Daten (zum Thema Homophobie) die höchste Validität im Bezug zum Abgleich mit den EVS Daten. Auch die Korrelation nach Pearson ist im Teilbereich Homophobie signifikant. Die Google Suchvervollständigung hingegen bietet alleinstehend sehr wenig Aussagekraft (weder manuell, noch mit Hilfe des Tools *Google Autocomplete*). Zum einen besteht nach Pearson oftmals eine sehr geringe, oder keine Korrelationen zwischen den mittels Google Suchvervollständigung generierten Daten und den Studiendaten, was sich nachfolgend auch durch eine schlechte Zuordnungseinteilungen zu den Clustern im Vergleich zur jeweiligen Studie ausdrückt. Zum anderen verhindert die geringe Anzahl an Analyseeinheiten eine exakte Differenzierung zwischen den Ländern. Generell bleibt für die Suchvervollständigung auch festzuhalten, dass teils keine Auswertung möglich war, weil keine automatische Vervollständigung zu einem Suchbegriff von Google ausgegeben wurde. Dieses Manko war weder bei der Textsuche, noch im Exkurs bei der Bildersuche der Fall, weshalb diese beiden Verfahren der Suchvervollständigung vorzuziehen sind. Ein erster Ansatz zur Evaluierung der Google Bildersuche im Exkurs zeigt im Vergleich zur Suchvervollständigung insgesamt auch eine Verbesserung der Pearson Korrelation (Homosexualität). Durch die Zuordnung der Länder zu den Clustern ist auch mittels Bildersuche eine Abgrenzung der verschiedenen europäischen Kulturen zu bestimmten Themeninhalten möglich – allerdings fehlt auch hier eine hohe Güte (Korrelation) im Vergleich zu den Daten der EVS. Wobei aber aufgrund der teils sehr textlastigen Analyseeinheiten (Thema Homosexualität mit sehr vielen Memes) in manchen Ländern eine Verzerrung durch eine falsche Kodierung aufgrund einer mangelhaften Übersetzung in diesem Bereich nicht auszuschließen ist. Insgesamt zeigt die Datenauswertung, dass bei einer hohen Korrelation nach Pearson auch eine valide Zuteilung der Google Daten zu den gleichen Clustern, wie in der Abgleichs-Studie, zu erwarten ist. Bei einer geringen Korrelation nach Pearson zeigen die Google Daten hingegen auch eine geringe Validität im Abgleich mit den Daten der jeweiligen Studie.

Neben der generellen Validität der Daten ist der verwendete Suchbegriff von besonderer Bedeutung für die Datengenerierung. Ein allein möglichst neutraler Suchbegriff ist für einen sozio-kulturellen Vergleich mittels Google nicht der einzige verlässliche Schlüssel für eine

erfolgreiche Datengenerierung. Wichtig bei der Auswahl des Suchbegriffs ist vor allem auch eine länderübergreifende, einheitliche Definition. Darüber hinaus muss beachtet werden, dass der Suchbegriff länderübergreifend keine Synonyme, bzw. andere namentlich Zusammenhänge aufweist (siehe Suche Google UK: foreigner führt zur Band *Foreigner*). Daher bleibt festzuhalten, dass die Ergebnisse der Google Suche in Ihrer Güte wesentlich durch die verwendeten Suchbegriffe, und weniger durch die Methode selbst determiniert werden. Eine weitere Datenerhebung mittels Google Suche zum Thema Ausländerfeindlichkeit, z.B. mit den Suchbegriffen „Immigrant“ oder „Flüchtling“, könnte eine differenzierte Darstellung liefern. Aufgrund der derzeitigen europaweiten Debatte um Flüchtlingshilfe ist jedoch auch bei diesen Suchbegriffen von einem *visual bias*⁵⁶ auszugehen, so dass die Ergebnisse ebenso verzerren, wie der Suchbegriff „Ausländer“ selbst es z.B. im angelsächsischen Raum tut. Zudem ist gerade bei Themen wie Homosexualität und Ausländerfeindlichkeit fraglich, wie sehr der Google Algorithmus auf die Suchergebnisse einwirkt, um rassistische, beleidigende, oder sonstige verunglimpfenden Suchergebnisse zu filtern.

⁵⁶ Übersetzung durch den Autor: *visuelle Verzerrung*. Visual bias = überproportional häufiges Auftauchen von (Bild-)Inhalten in den Massenmedien in Zusammenhang mit einem bestimmten Thema; Entstehung von „visual framing“, d.h. es entsteht „ein mehrstufiger Prozess, in dem bestimmte Bilder oder Bildaspekte zu einem Thema ausgewählt, andere hingegen vernachlässigt werden, was den Rezipierenden eine bestimmte Bedeutung oder Interpretation implizieren kann“ (Geise, 2013)

7 Fazit

Da die Daten von zwei der drei via Google untersuchten Themen nicht mit den EVS Validierungsdaten korrelieren, bzw. bei der Zuteilung in die jeweiligen Cluster übereinstimmen, kann die Forschungsfrage, ob Google eine Gesellschaft als Ganzes abbildet, nicht eindeutig beantwortet werden. Tendenziell muss die Frage nach aktuellem Stand eher mit nein beantwortet werden und widerspricht somit der Theorie Richard Rogers, dass Google eine direkte Spiegelung der Gesellschaft darstellt. Es bleibt jedoch festzuhalten, dass die Google Daten zum einen durch die Wahl des Suchbegriffs determiniert sind und somit das Ergebnis beeinflusst wird – andere Suchbegriffe könnten valide und reliable Daten liefern. Zum anderen ist auch die Diskrepanz zwischen *self-reported data* gegenüber Web Daten anzumerken. Zudem können die Ergebnisse aber auch wie diskutiert durch einen Selektionseffekt durch Google selbst beeinflusst sein. Es gilt jedoch auch festzuhalten, dass mit Hilfe der Suchmaschine Google Daten generiert werden konnten, die Unterschiede zwischen verschiedenen Kulturen innerhalb Europas sichtbar machen. Somit ist es insgesamt möglich, mittels Google Suche eine kulturvergleichende Forschung durchzuführen. Eine Abbildung der Gesellschaft analog zu den anderen Studienergebnissen ist allerdings mit Hilfe der Google Suche nur bedingt möglich (nur EVS, Homophobie ähnlich). Letztlich bleibt für die erarbeitete Methode somit die Frage ungeklärt, für wen, also für welche demographische Gruppe die mittels Google generierten Daten im jeweiligen Land Repräsentativität besitzen. Auch wenn die mittels Google generierten Daten nicht mit den Daten der anderen Studien übereinstimmen, zeigt sich durch die (meist) differenzierte Darstellung der Daten auf Basis der Google Textsuche, dass Auswertung zu kulturellen Unterschieden zwischen den verschiedenen via Google abgebildeten Ländern erstellt werden können. Im Ergebnis ist es somit gelungen, eine kulturvergleichende Studie zu den Themengebieten *Homophobie*, *Ausländerfeindlichkeit* und *Toleranz von Rauschmitteln* durchzuführen, bzw. Unterschiede zwischen den Ländern mit Hilfe der Google Suche sichtbar zu machen. Dennoch entsteht durch den Abgleich mit den Studiendaten der EVS / WHO / EMCDDA in diesem Rahmen keine Validität und Reliabilität für die erhobenen Google Daten. Hierfür müsste ggf. ein weiterer Abgleich zwischen den generierten Daten (repräsentativ für die Nutzergruppen von Google) und einem bestimmten demographischen Cluster auf Basis der Validierungsdaten (Google Nutzer des jeweiligen Landes) erfolgen. Darüber hinaus wurde im Verlauf der Arbeit deutlich, dass mittels Google Web Suche (im expliziten Verfahren) nicht lediglich der kulturvergleichende Aspekt zu den Themen

Homophobie und Ausländerfeindlichkeit über die Bildung eines einzelnen Index besteht. Die Methode der expliziten Suche bietet viel mehr auch die Möglichkeit, themenrelevante Kreise zu identifizieren, die die gesellschaftlichen Einstellungen und Denkweisen zu bestimmten Themen in unterschiedlichen Kulturen widerspiegeln. Daher bietet Google nicht allein eine eindimensionale Darstellung eines Themas (z.B. *Wie homophob ist die Gesellschaft eines Landes in Vergleich zu einem anderen?*), sondern vielmehr eine vielschichtige Ansicht zu einem Thema (z.B. *Warum besteht Homophobie in der Gesellschaft eines Landes? Wie unterscheiden sich die Länder / Kulturen dabei untereinander, bzw. voneinander?*). Die Methode eröffnet für zukünftige Studien daher verschiedene Möglichkeiten, wie z.B. relevante Themenfelder vor Beginn einer Feldphase zu identifizieren, um bestimmte Themenbereiche gezielt in Umfragen ansprechen zu können.

Um die angewandten Methoden der Google Suche zu verbessern, sind zudem weitere Untersuchungen der Google Suche selbst notwendig. In Folgestudien sollte vor allem ein Prozess zur Entwicklung einheitlicher Suchbegriffe im Vordergrund stehen. Alleine in Europa besteht auf Grund der verschiedenen Sprachfamilien (Germanisch, Slawisch, Romanisch, etc.) in der Natur der Sprachwelten eine unterschiedliche Wort- und Meinungsdefinitionen in den jeweiligen Kulturen / Ländern. In zukünftigen Untersuchungen stellt sich somit die Frage, wie die einheitliche Definition eines Suchbegriffs sichergestellt werden kann – dies ist in der vorliegenden Arbeit vor allem in Zusammenhang mit dem Suchbegriff „Ausländer“ aufgefallen. Der kulturübergreifende, gleich definierte Suchbegriff ist einer der Kernpunkte für eine valide Datengenerierung per Google. Zukünftig könnte zum Beispiel ein Pre-Test auf Basis der Bildersuche, oder der Suchvervollständigung, als ein erster Indikator für die Nutzbarkeit eines Suchbegriffs dienen. Dennoch ist auch hierdurch keine eindeutige Absicherung einer gleichen Definition zu gewährleisten. Somit stellt sich die Frage, wie eine einheitliche, kulturübergreifende Definition eines Suchbegriffs in den untersuchten Ländern sichergestellt werden kann. Zudem ist das Vorgehen beim angewandten Verfahren der Bildersuche (Exkurs) verbesserungswürdig. Zum einen besteht im Bereich der Bild-Inhaltsanalyse selbst weiterhin großer Forschungsbedarf. Von diesem Bedarf losgelöst, besteht zum anderen die Notwendigkeit weiterer Untersuchungen der Google Bildersuche selbst. Aufgrund des hohen Text-Anteils sollten bei zukünftigen Untersuchungen der Bildersuche z.B. stets Muttersprachler als Übersetzer in den Prozess eingebunden werden. Vor allem der große Anteil an Memes kann verhindern, dass ein Bild ohne den zugehörigen Text richtig verstanden und somit sinnvoll ausgewertet werden kann.

Ferner könnten weitere Themen und ggf. andere Ansätze (z.B. die implizite Suche) in Kombination mit der Bildersuche angewandt werden. Außerdem bietet die Anzahl der Analyseeinheiten weiteres Optimierungspotential. Weitere Tests könnten sichtbar machen, ob eine Erhöhung der Anzahl an Analyseeinheiten (Text und Bildersuche) die Datenkorrelation verbessert. Es erscheint aber auch möglich, dass die Menge der Analyseeinheiten reduziert werden kann, ohne Verlust der Datenqualität zu verursachen.

Auch wenn die mittels Google erhobenen Daten, in Anlehnung an die Validierungsdaten der Studien keine eindeutige Abbildung der Gesellschaft (nach EVS, WHO, EMCDDA) zeigen, bietet die Google Suche dennoch die Möglichkeit, Unterschiede zwischen verschiedenen Kulturen abzubilden. Per Google generierte Daten können somit zurzeit nicht als Ersatz für vergleichende Forschung auf den unterschiedlichen Ebenen des gesellschaftlichen Verständnisses im europäischen Raum dienen. Dennoch können die mit Hilfe der Suchmaschine Google generierten Daten zur Unterstützung genutzt werden, um zum Beispiel verschiedene kulturell relevante Themenkreise auszuarbeiten oder die Erstellung von Fragebögen effizienter zu gestalten. Auch wenn Google sich somit derzeit nicht zwangsläufig als generelle Methode zur Erhebung von gesellschaftlichen Unterschieden anbietet, ermöglicht es dennoch einen ersten Eindruck zur Grundeinstellung einer Gesellschaft zu einem bestimmten Thema. Die Nutzung des Verfahrens als unterstützende Maßnahme bietet somit weitreichende Möglichkeiten. Unter seiner Zuhilfenahme könnten z.B. Fragebögen effizienter erstellt / gestaltet werden, mit der Zielsetzung, den Befragten konkrete Anhaltspunkte zur Identifizierung und Herausarbeitung relevanter Themenkreise (als Hilfestellung) an die Hand zu geben.

Zusammenfassend lässt sich somit festhalten, dass Webinstrumente wie Google, nach erster empirischer Beweisführung, nur bedingt ähnliche Ergebnisse wie repräsentative Befragungen liefern und die Gesellschaft somit nicht analog zu den Ergebnissen der genannten Studien abbilden. Es kann daher darauf geschlossen werden, dass Webinstrumente wie Google zurzeit nur bedingt repräsentative Befragungen ab- bzw. nachbilden können. Dennoch liefert die Masterarbeit den Erkenntnisgewinn, dass die Webwissenschaften (hier vertreten durch die Methode *Search as Research*) komplementär eingesetzt werden können, um z.B. kulturvergleichende Studien und Befragungen zu unterstützen. Hierbei liegt das Hauptaugenmerk vor allem auf der Erarbeitung von Themenkreisen, bzw. der Extraktion von gesellschaftlichen Einstellungen, die sich auf ein zu untersuchendes Thema beziehen. Zudem konnten im Verlauf der Arbeit weitere Forschungsfelder, die vor der empirischen Untersuchung eher unbekannt waren, aufgedeckt

werden. So bietet u.a. auch die Google Bilder Suche eine Möglichkeit vertiefend in das Thema *Search as Research* einzusteigen.

Literaturverzeichnis

- World Health Organization. (2011). *Global status report on alcohol and health*. Schweiz: WHO Library Cataloguing-in-Publication Data.
- Andelfinger, V. P., & Hänisch, T. (2015). *Internet der Dinge*. Wiesbaden: Springer Fachmedien .
- Baun, C. (2012). *Computernetze kompakt* (2. Auflage ed.). Frankfurt: Springer Vieweg.
- Boase, J., & Ling, R. (2013, July). Measuring Mobile Phone Use: Self-Report Versus Log Data. *Journal of Computer-Mediated Communication*(4), pp. 508–519.
- Döring, N. (2008). Online-Forschung. In U. Sander, F. von Gross, & K.-U. Hugger, *Handbuch Medienpädagogik* (pp. 357-363). Wiesbaden: VS Verlag für Sozialwissenschaften.
- EMCDDA. (2014). *12-Monats-Prävalenz des Cannabiskonsums von Jugendlichen und jungen Erwachsenen nach europäischen Ländern im Jahr 2014*. 2015: DHS.
- European Values Study. (2014a). *About EVS*. Retrieved August 04, 2015, from <http://www.europeanvaluesstudy.eu/>:
http://www.europeanvaluesstudy.eu/frmShowpage?v_page_id=4386315781860116
- European Values Study. (2014b). *Methods and sample - Survey 2008*. Retrieved August 04, 2015, from <http://www.europeanvaluesstudy.eu/>:
http://www.europeanvaluesstudy.eu/frmShowpage?v_page_id=1680373125305663
- EVS. (2015). *European Values Study Longitudinal Data File 1981-2008 (EVS 1981-2008)*. *GESIS Data Archive, Cologne*. ZA4804 Data file Version ZA4804_v2-0-0_2012_09-07_patch_1, doi:10.4232/1.11005. Retrieved from <https://dbk.gesis.org/dbksearch/sdesc2.asp?no=4804&db=e&doi=10.4232/1.11005>
- Geise, S. (2013). *Visual Framing*. (S. Geise, & K. Lobinger, Editors) Retrieved Juli 24, 2015, from <http://www.halem-verlag.de>: <http://www.halem-verlag.de/2013/visual-framing/>
- GESIS – Leibniz-Institut für Sozialwissenschaften e.V. (2015, Juni 08). *European Values Study*. Retrieved August 04, 2015, from <http://www.gesis.org/>:
<http://www.gesis.org/unser-angebot/daten-analysieren/umfragedaten/european-values-study/>
- Google Inc. (2015a). *Algorithmen*. Retrieved Juli 23, 2015, from <http://www.google.de>:
<http://www.google.com/insidesearch/howsearchworks/algorithms.html?hl=de>

- Google Inc. (2015b). *Automatische Vervollständigung*. Retrieved Juli 23, 2015, from <http://www.google.de: https://support.google.com/websearch/answer/106230?hl=de>
- Google Inc. (2015c). *Crawling und Indexierung*, c&i. Retrieved Juli 23, 2015, from <http://www.google.de: https://www.google.com/intl/de/search/about/insidesearch/howsearchworks/crawling-indexing.html>
- Google Inc. (2015e). *Standort bei Google ändern*. Retrieved Juli 24, 2015, from <http://www.google.de: https://support.google.com/websearch/answer/179386?hl=de>
- Hair, J. F. (2006). *Multivariate data analysis* (6 ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Kozinets, R. V. (2013). *Netnography - Doing Ethnographic Research Online*. Los Angeles [u.a.]: SAGE.
- Krause, E., & Golkowsky, C. (2014). Analyse der Social Media-Aktivitäten. In C. Rogge, & R. Karabasz, *Social Media im Unternehmen - Ruhm oder Ruin* (pp. 143-151). Wiesbaden: Springer Vieweg.
- Lewandowski, D. (2015). *Suchmaschinen verstehen*. Hamburg: Springer Vieweg.
- Lipinski, K. (2015, August 11). *GUI (graphical user interface)*. (D. B. GmbH, Editor) Retrieved August 11, 2015, from <http://www.itwissen.info: http://www.itwissen.info/definition/lexikon/graphical-user-interface-GUI-Grafische-Benutzeroberflaeche.html>
- Lipp, M. (2006). *VPN - Virtuelle Private Netzwerke: Aufbau und Sicherheit*. München: Addison-Wesley.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. (D. o. National University, Ed.) *Biochemia Medica*, 22(3), pp. 276-282.
- Meulemann, H. (2013). *Soziologie von Anfang an - Eine Einführung in Themen, Ergebnisse und Literatur* (3. Auflage ed.). (H. Sahner, R. Sackmann, & M. Bayer, Eds.) Wiesbaden: Springer Fachmedien.
- Microsoft Corporation. (2015). Internetoptionen - Allgemein - Browserverlauf. Retrieved Oktober 06, 2015
- Miller, D., & Slater, D. (2000). *The Internet: An Ethnographic Approach*. Oxford: Berg.
- Mühlichen, A. (2014). Informationelle Selbstbestimmung. In N. Baur, & J. Blasius, *Handbuch Methoden der empirischen Sozialforschung* (pp. 93-102). Wiesbaden: Springer VS.

- O'Hara, K., & Hall, W. (2008). *Web Science*. Oxford: Association of Learning Technologies Newsletter. Retrieved from <http://eprints.soton.ac.uk/265682/1/OHara-Hall-ALT-N-Web-Science.pdf>
- Privax LTD (GB). (2015, Oktober 06). *hidemyass.com/servers*. Retrieved Oktober 06, 2015, from hidemyass.com: <https://www.hidemyass.com/servers>
- Rampl, H. (2014). *Definition Usability*. Retrieved August 06, 2015, from <http://www.handbuch-usability.de>: <http://www.handbuch-usability.de/begriffsdefinition.html>
- Reinecke, J. (2014). Standardisierte Befragung. In N. Baur, & J. Blasius, *Handbuch Methoden der empirischen Sozialforschung* (pp. 601-798). Wiesbaden: Springer VS.
- Rippel, S., & Reipel, C. (2015). *Methoden kulturvergleichender Sozialforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Rogers, R. (2009, Mai 8). The End of the Virtual – Digital Methods. *Inaugural Speech, Chair, New Media & Digital Culture*, (pp. 1-20). Amsterdam.
- Rogers, R. (2013). *Digital Methods*. Cambridge: MIT Press.
- Schäfers, B. (2013). *Einführung in die Soziologie*. Wiesbaden: Springer Fachmedien.
- Scherfer, K., & Volpers, H. (2013). Eine Einführung der Herausgeber. In K. Scherfer, & H. Volpers (Eds.), *Methoden der Webwissenschaften* (pp. 7-11). Berlin: LIT Verlag.
- Schmitz, M. (2014, November 28). Was sind generische Webmethoden? Eine Einführung in das Methodenfeld . Köln: Fachhochschule Köln. Retrieved Juli 06, 2015, from <https://www.youtube.com/watch?v=n5VjuGrnWeg&feature=youtu.be>
- Seibold, B. (2002, Januar). Die flüchtigen Web-Informationen einfangen: Lösungsansätze für die Online-Inhaltsanalyse bei dynamischen Inhalten im Internet. *Publizistik*, pp. 45-56.
- Strutz, H., & Wagner, U. (1980). *Soziologie - Einführung für Angehörige der öffentlichen Verwaltung* (Vol. 16). Köln, Stuttgart, Berlin, Hannover, Kiel, Mainz, München: Deutscher Gemeindeverlag und Verlag W. Kohlhammer GmbH.
- Volpers, H. (2008). Warum und zu welchem Zweck benötigen wir einen Webwissenschaft? In *Webwissenschaft - Eine Einführung* (pp. 31-51). Berlin: LIT Verlag.
- Volpers, H. (2013). Online-Inhaltsanalyse. In K. Scherfer, & H. Volpers (Eds.), *Methoden der Webwissenschaft* (pp. 15-42). Berlin: LIT Verlag.
- Wagner, P., & Hering, L. (2014). Online-Befragung. In N. Baur, & J. Blasius, *Handbuch Methoden der empirischen Sozialforschung* (pp. 661-674). Wiesbaden: Springer VS.

Weiss, S. (2010). Werbung im Web. Der Stellenwert der klassischen Online-Werbung in der Werbekommunikation. *Inaugural-Dissertation*. Münster (Westf.): Westfälischen Wilhelms-Universität.

World Health Organization. (2011). *Global status report on alcohol and health*. Schweiz: WHO Library Cataloguing-in-Publication Data.

Anhang I

Country/Region	1981	1990	1999	2008
USA	1982	1990		
Canada	1982	1990		
Belgium	1981	1990	1999	2009
Denmark	1981	1990	1999	2008
France	1981	1990	1999	2008
Germany*	1981	1990	1999	2008/2009
Great Britain	1981	1990	1999	2009/2010
Iceland	1984	1990	1999	2009/2010
Ireland	1981	1990	1999/2000	2008
Italy	1981	1990	1999	2009
Malta	1983/1984	1991	1999	2008
Netherlands	1981	1990	1999	2008
Northern Ireland	1981	1990	1999	2008
Spain	1981	1990	1999	2008
Sweden	1982	1990	1999/2000	2009/2010
Norway	1982	1990		2008
Austria		1990	1999	2008
Bulgaria		1990/1991	1999	2008
Czech Republic		1991	1999	2008
Estonia		1990	1999	2008
Finland		1990	2000	2009
Hungary		1991	1999	2008/2009
Latvia		1990	1999	2008
Lithuania		1990	1999	2008
Poland		1990	1999	2008
Portugal		1990	1999	2008
Romania		1993	1999	2008
Slovak Republic		1991	1999	2008
Slovenia		1992	1999	2008
Belarus			2000	2008
Croatia			1999	2008
Greece			1999	2008
Luxembourg			1999	2008
Russian Federation			1999	2008
Turkey			2001	2008/2009
Ukraine			1999	2008
Albania				2008
Armenia				2008
Azerbaijan				2008
Bosnia and Herzegovina				2008
Cyprus				2008
Northern Cyprus				2008
Georgia				2008
Kosovo				2008
Macedonia, Republic of				2008
Moldova, Republic of				2008
Montenegro, Republic of				2008
Serbia				2008
Switzerland				2008

*1981 only West-Germany

Abbildung A 1 - European Values Study 1981 - 2008: Participating countries (GESIS – Leibniz-Institut für Sozialwissenschaften e.V., 2015)

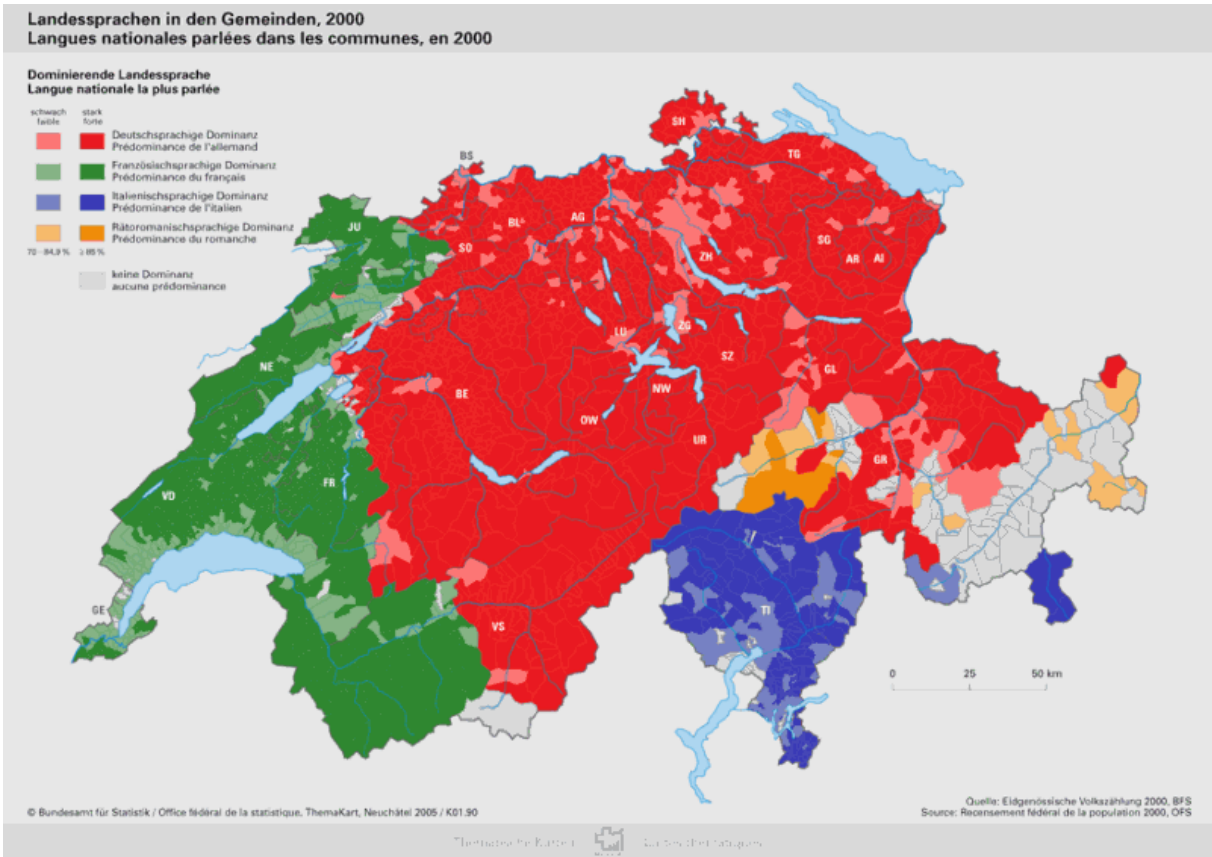


Abbildung A 2 - Landessprachen in den Gemeinden, 2000 (Schweiz); Bundesamt für Statistik Schweiz, abgerufen am 06.10.2015, von: http://www.bfs.admin.ch/bfs/portal/de/index/regionen/thematische_karten/maps/bevoelkerung/sprachen_religionen.html

Google Autocomplete

Input

Input words to retrieve autocomplete for, one per line

Homosexualität

Select one or more languages:

- Estonian
- Filipino
- Finnish
- French
- German
- Greek
- Hebrew
- Hindi
- Hungarian
- Icelandic

Select one or more countries:

- French Southern Territories
- Gabon
- Gambia
- Georgia
- Germany
- Ghana
- Gibraltar
- Greece
- Greenland
- Grenada

Autocomplete

Google Autocomplete, an Introduction

Retrieves autocomplete suggestions from Google

This tool retrieves Google autocomplete suggestions (normally found in the search box).

Enter one or more words and select languages and countries. Each combination of word-language-country will be queried after which the results will be displayed in a table.

If you would like the number of results per query suggestion you will have to use the [Google Scraper](#). You then leave the top box (url inputs) empty, put the suggestions in the bottom box, and specify the language and local google version in the advanced settings.

[More information about Google autocomplete suggestions.](#)

word	country	language	suggestion
Homosexualität	Germany	German	homosexualität
Homosexualität	Germany	German	homosexualität im islam
Homosexualität	Germany	German	homosexualität bei tieren
Homosexualität	Germany	German	homosexualitäten
Homosexualität	Germany	German	homosexualität ursachen
Homosexualität	Germany	German	homosexualität in der bibel
Homosexualität	Germany	German	homosexualität krankheit
Homosexualität	Germany	German	homosexualität definition
Homosexualität	Germany	German	homosexualität in russland
Homosexualität	Germany	German	homosexualität deutschland

Abbildung A 3 - Setup Google Autocomplete (DMI), Google Landesversion Deutschland (Screenshot, von: <https://tools.digitalmethods.net/beta/scrapeGoogle/autocomplete.php>)

Staat	Serverstandort (VPN)	Besonderheit	ccTLD (Google Landesversion)	Google Suchbegriffe
Dänemark*	Kopenhagen	Landeshauptstadt	google.dk	Homoseksualitet, udlænding, ulovlige stoffer
Deutschland	Berlin	Landeshauptstadt	google.de	Homosexualität, Ausländer, illegale Drogen
England	London	Landeshauptstadt	google.co.uk	Homosexuality, foreigner, illegal drugs
Finnland*	Helsinki	Landeshauptstadt	google.fi	Homoseksuaalisuus, ulkomaalainen, laittomia huumeita
Frankreich*	Paris	Landeshauptstadt	google.fr	Homosexualité, étranger, drogues illégales
Irland	Dublin	Landeshauptstadt	google.ie	Homosexuality, foreigner, illegal drugs
Italien*	Mailand	Rom nicht gegeben, Mailand größter wählbarer Standort	google.it	Omosessualità, straniero, droghe illegali
Kroatien*	Zagreb	Landeshauptstadt	google.hr	Homoseksualnost, stranac, ilegalne droge
Niederlande*	Amsterdam	Landeshauptstadt	google.nl	Homoseksualiteit, buitenland, illegale drugs
Norwegen*	Oslo	Landeshauptstadt	google.no	Homofili, utlending, illegale rusmidler
Österreich	Vienna	Landeshauptstadt	google.at	Homosexualität, Ausländer, illegale Drogen

Polen*	Warschau	Landeshauptstadt	google.pl	Homoseksualizm, cudzoziemiec nielegalny narkotyki
Rumänien*	Bukarest	Landeshauptstadt	google.ro	Homosexualitate, străin, droguri ilegale
Schweden*	Stockholm	Landeshauptstadt	google.se	Homosexualitet, utlänning, illegala droger
Schweiz	Zürich	Deutschsprachige Schweiz, repräsentativ – deutschsprachige Schweiz vergleichsweise groß	google.ch	Homosexualität, Ausländer, illegale Drogen
Spanien*	Madrid	Landeshauptstadt	google.es	Homosexualidad, extranjero, drogas ilegales
Türkei*	Istanbul	Landeshauptstadt	google.com.tr	Eşcinsellik, ecnebi, yasadışı uyuşturucu
Ukraine*	Odessa	Kein anderer Serverstandort via HMA in der Ukraine gegeben	google.com.ua	Гомосексуализм, иностранец, незаконные наркотики
Ungarn*	Budapest	Landeshauptstadt	google.hu	Homoszexualitás, külföldi, illegális drogok
Weißrussland*	Minsk	Landeshauptstadt	google.by	Гомосексуализм, иностранец, незаконные наркотики

Tabelle A 1 - Suchbegriffe Google für Indices (eigene Darstellung)

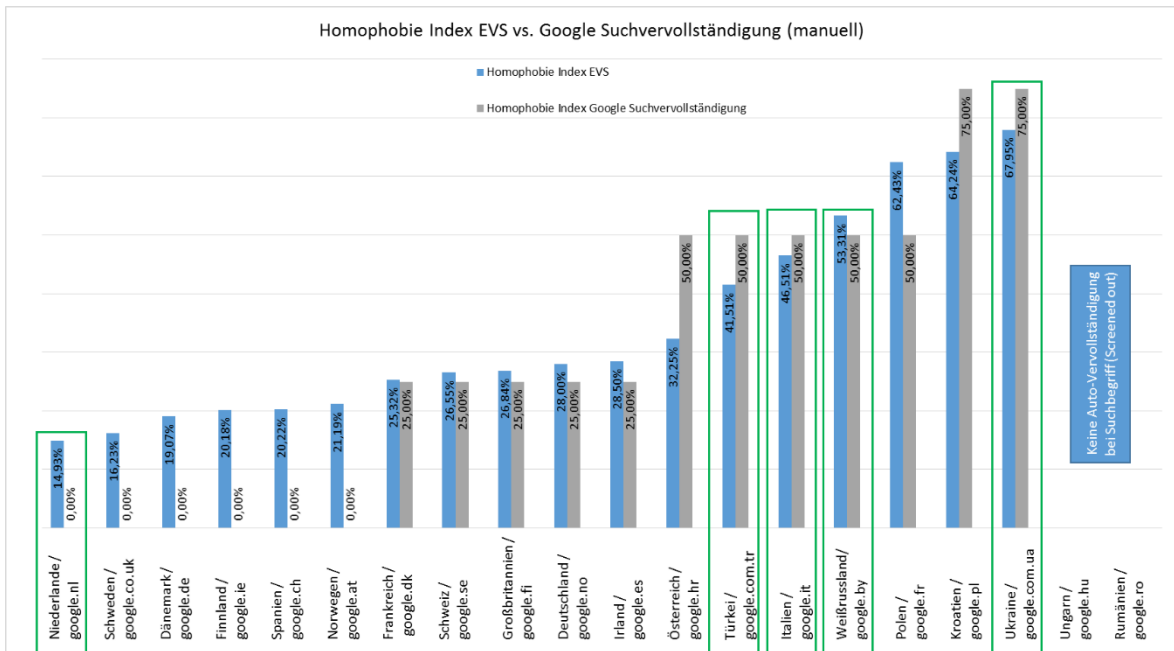


Abbildung A 4- Homophobie Index EVS vs. Google Suchvervollständigung (manuell) (eigene Darstellung)

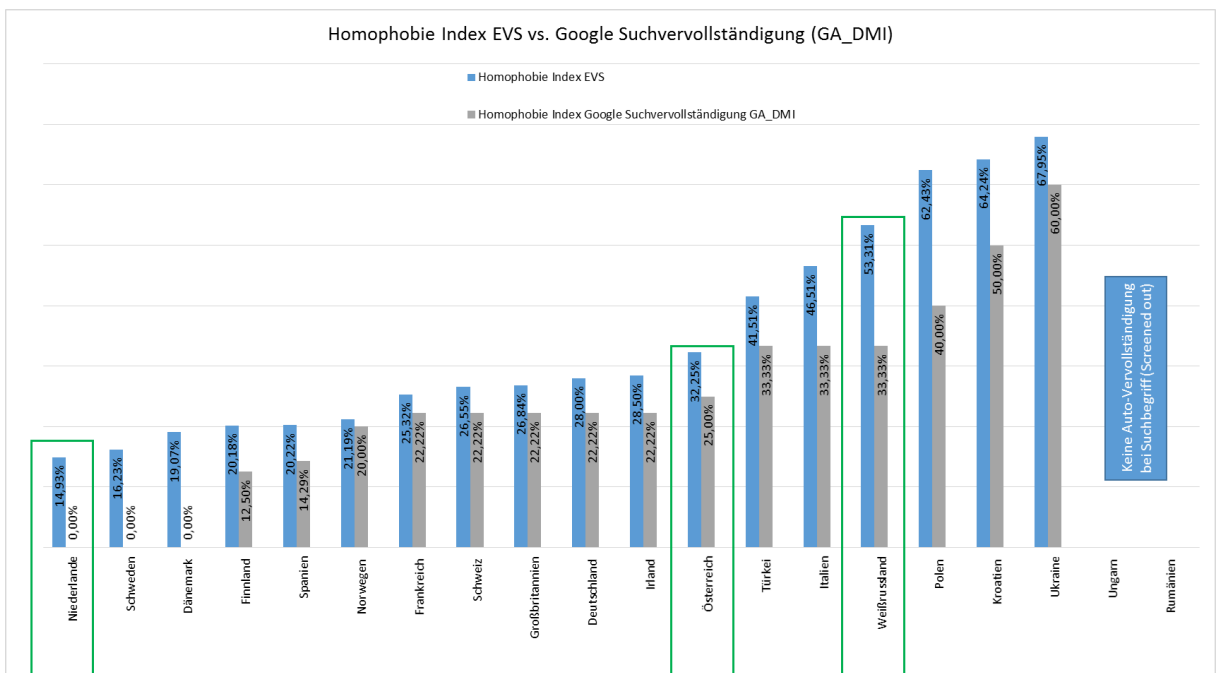


Abbildung A 5 - Homophobie Index EVS vs. Google Suchvervollständigung (Google Autocomplete) (eigene Darstellung)

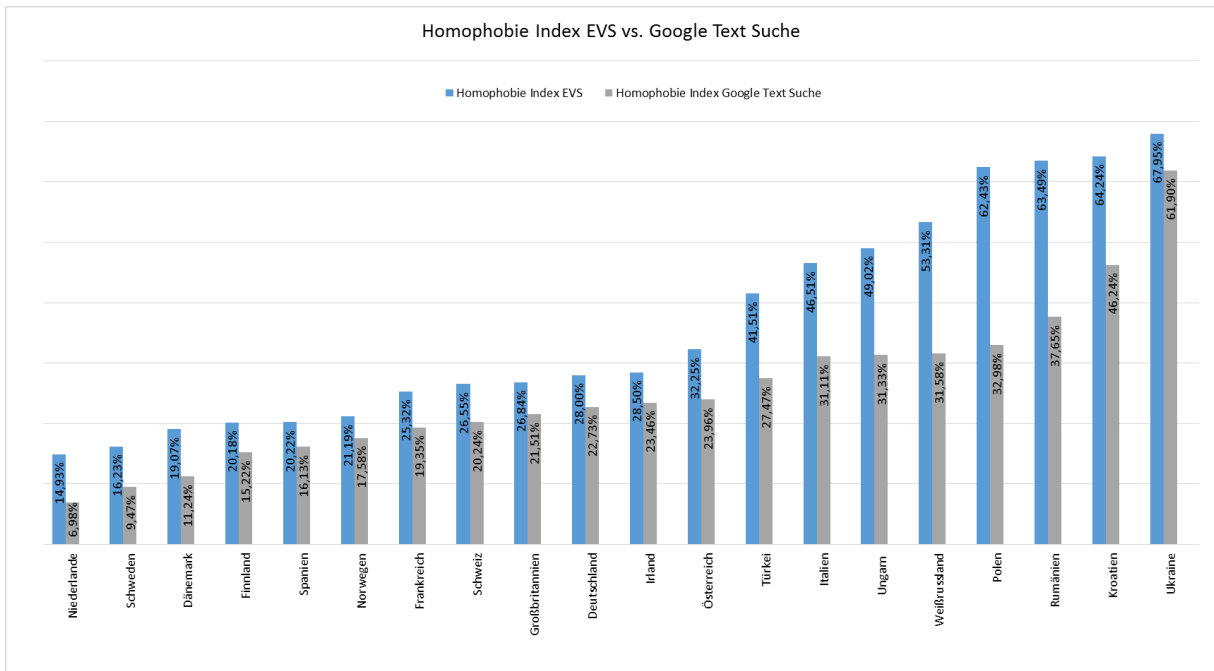


Abbildung A 6 - Homophobie Index EVS vs. Google Textsuche (eigene Darstellung)

Value of Kappa	Level of Agreement	% of Data that are Reliable
0-.20	None	0-4%
.21-.39	Minimal	4-15%
.40-.59	Weak	15-35%
.60-.79	Moderate	35-63%
.80-.90	Strong	64-81%
Above .90	Almost Perfect	82-100%

Tabelle A 2 - Abgleichstabelle Kappa-Wert zu Datenreliabilität (McHugh, 2012)

EVS * Suchvervollständigung Manuell Kreuztabelle (Homophobie)

Anzahl

	SuchverVollMan			Gesamt
	1,00	2,00	3,00	
1	1	5	0	6
EVS 2	5	0	1	6
3	0	0	6	6
Gesamt	6	5	7	18

Symmetrische Maße

	Wert	Asymptotischer Standardfehler ^a	Näherungsweise T ^b	Näherungsweise Signifikanz
Maß der Übereinstimmung	Kappa	,083	,160	,502
Anzahl der gültigen Fälle	ppa	18		,615

a. Die Null-Hyphothese wird nicht angenommen.

b. Unter Annahme der Null-Hyphothese wird der asymptotische Standardfehler verwendet.

Tabelle A 3 - Kreuztabelle & Cohen's Kappa EVS vs. Index Google Homophobie: Suchvervollständigung manuell (SPSS)

EVS * Suchvervollständigung DMI Kreuztabelle (Homophobie)

Anzahl

	SuchverVollDMI			Gesamt
	1	2	3	
1	2	3	1	6
EVS 2	3	1	2	6
3	0	2	4	6
Gesamt	5	6	7	18

Symmetrische Maße

	Wert	Asymptotischer Standardfehler ^a	Näherungsweise T ^b	Näherungsweise Signifikanz
Maß der Übereinstimmung	,083	,169	,502	,615
Anzahl der gültigen Fälle	18			

a. Die Null-Hyphothese wird nicht angenommen.

b. Unter Annahme der Null-Hyphothese wird der asymptotische Standardfehler verwendet.

Tabelle A 4 - Kreuztabelle & Cohen's Kappa EVS vs. Index Google Homophobie: Suchvervollständigung DMI Tool (SPSS)

EVS * Textsuche Kreuztabelle (Homophobie)

Anzahl

	Textsuche			Gesamt
	1	2	3	
1	3	2	1	6
EVS 2	3	3	0	6
3	0	1	7	8
Gesamt	6	6	8	20

Symmetrische Maße

	Wert	Asymptotischer Standardfehler ^a	Näherungsweise T ^b	Näherungsweise Signifikanz
Maß der Übereinstimmung Kappa	,470	,156	2,958	,003
Anzahl der gültigen Fälle	20			

a. Die Null-Hyphothese wird nicht angenommen.

b. Unter Annahme der Null-Hyphothese wird der asymptotische Standardfehler verwendet.

Tabelle A 5 - Kreuztabelle & Cohen's Kappa EVS vs. Index Google Textsuche Homophobie (SPSS)

EVS * Google TS&SV Kreuztabelle (Homophobie)

Anzahl

	GoogleIG			Gesamt
	1,00	2,00	3,00	
1,00	2	3	1	6
EVS 2,00	1	3	2	6
3,00	2	1	5	8
Gesamt	5	7	8	20

Symmetrische Maße

	Wert	Asymptotischer Standardfehler ^a	Näherungsweise T ^b	Näherungsweise Signifikanz
Maß der Übereinstimmung Kappa	,242	,165	1,532	,125
Anzahl der gültigen Fälle	20			

a. Die Null-Hyphothese wird nicht angenommen.

b. Unter Annahme der Null-Hyphothese wird der asymptotische Standardfehler verwendet.

Tabelle A 6 - Kreuztabelle & Cohen's Kappa EVS vs. Index aus Google Textsuche & Suchvervollständigung Homophobie (SPSS)

EVS * Textsuche Kreuztabelle (Ausländerfeindlichkeit)

Anzahl

	Textsuche			Gesamt
	1	2	3	
1	2	1	3	6
EVS 2	2	3	1	6
3	3	2	3	8
Gesamt	7	6	7	20

Symmetrische Maße

	Wert	Asymptotischer Standardfehler ^a	Näherungsweise T ^b	Näherungsweise Signifikanz
Maß der Übereinstimmung Kappa	,098	,166	,620	,535
Anzahl der gültigen Fälle	20			

a. Die Null-Hypothese wird nicht angenommen.

b. Unter Annahme der Null-Hypothese wird der asymptotische Standardfehler verwendet.

Tabelle A 7 – Kreuztabelle & Cohen's Kappa EVS vs. Index Ausländerfeindlichkeit Google Textsuche (SPSS)

WHO * TS_ALK Kreuztabelle

Anzahl

	TS_ALK			Gesamt
	1	2	3	
1	0	2	0	2
WHO 2	1	3	2	6
3	4	1	7	12
Gesamt	5	6	9	20

Symmetrische Maße

	Wert	Asymptotischer Standardfehler ^a	Näherungsweise T ^b	Näherungsweise Signifikanz
Maß der Übereinstimmung Kappa	,187	,149	1,189	,234
Anzahl der gültigen Fälle	20			

a. Die Null-Hypothese wird nicht angenommen.

b. Unter Annahme der Null-Hyphothese wird der asymptotische Standardfehler verwendet.

Tabelle A 8 - Kreuztabelle & Cohen's Kappa EVS vs. Index Toleranz von Alkohol Google Textsuche (SPSS)

EMCDDA * TS_MAR Kreuztabelle

Anzahl

	TS_MAR			Gesamt
	1	2	3	
1	1	1	0	2
EMCDDA 2	1	1	4	6
3	2	4	5	11
Gesamt	4	6	9	19

Symmetrische Maße

	Wert	Asymptotischer Standardfehler ^a	Näherungsweise T ^b	Näherungsweise Signifikanz
Maß der Übereinstimmung Kappa	-,046	,175	-,274	,784
Anzahl der gültigen Fälle	19			

a. Die Null-Hyphothese wird nicht angenommen.

b. Unter Annahme der Null-Hyphothese wird der asymptotische Standardfehler verwendet.

Tabelle A 9 – Kreuztabelle & Cohen's Kappa EVS vs. Index Toleranz von Marihuana Google Textsuche (SPSS)

EVS * Bildersuche Kreuztabelle

Anzahl

	Bildersuche			Gesamt
	1	2	3	
1	2	3	1	6
EVS 2	1	3	2	6
3	2	1	5	8
Gesamt	5	7	8	20

Symmetrische Maße

	Wert	Asymptotischer Standardfehler ^a	Näherungsweise T ^b	Näherungsweise Signifikanz
Maß der Übereinstimmung Kappa	,242	,165	1,532	,125
Anzahl der gültigen Fälle	20			

a. Die Null-Hyphothese wird nicht angenommen.

b. Unter Annahme der Null-Hyphothese wird der asymptotische Standardfehler verwendet.

Tabelle A 10 - Kreuztabelle & Cohen's Kappa EVS vs. Index Google Bildersuche Homophobie (SPSS)

EVS * BILDSUCHE_AUSL Kreuztabelle

Anzahl

	BILDSUCHE_AUSL			Gesamt
	1	2	3	
1	3	0	3	6
EVS 2	2	4	0	6
3	2	5	1	8
Gesamt	7	9	4	20

Symmetrische Maße

	Wert	Asymptotischer Standardfehler ^a	Näherungsweise T ^b	Näherungsweise Signifikanz
Maß der Übereinstimmung Kappa	,118	,151	,790	,429
Anzahl der gültigen Fälle	20			

a. Die Null-Hyphothese wird nicht angenommen.

b. Unter Annahme der Null-Hyphothese wird der asymptotische Standardfehler verwendet.

Tabelle A 11 - Kreuztabelle & Cohen's Kappa EVS vs. Index Google Bildersuche Ausländerfeindlichkeit (SPSS)

Anhang II - Codebuch

Codeplan Top 100

Inhalte Google TOP 100						
Für fehlende Werte → 99 coden (z.B. wenn V1 = 0, dann V2 = 99)						
Variable	Variablenname	Wert	Label	Erläuterung	Operationalisierung/Definitionen	Hinweise/Beispiel
Block 1: Homophobie (Suchbegriff: <i>Homosexualität</i> – in der jeweiligen Landessprache)						
Fragestellung: Wie <i>Homophob</i> ist ein Land nach Google Index?						
#	Nummer	Nummerierung	1-100	Nummer des verkodeten Suchergebnisses in Reihenfolge der Rangreihe nach Google von 1 (erstes Suchergebnis) bis 100 (letztes verkodetes Suchergebnis: 100)	Suchergebnisse in der Rangreihenfolge nach Google durch nummerieren von 1-100	<ol style="list-style-type: none"> 1. Suchergebnis = 1 2. Suchergebnis = 2 3. Suchergebnis = 3 4. Suchergebnis = 4 5. Usw.
V1	Thematische Relevanz	bivariat	0= nein; 1=ja	Ist das einzelne Suchergebnis relevant für die Erhebung in Zusammenhang mit dem Thema? Zudem wenn kein kultureller Bezug (kein Zusammenhang zwischen Suchergebnis mit kodiertem Land) = keine Relevanz.	Ein Suchergebnis hat keine Relevanz wenn es nicht in den Themenkreis fällt der verkodet wird. D.h. z.B. ein Bericht darüber, dass Homosexuelle in Deutschland im Sport nicht akzeptiert werden ist relevant, wenn Deutschland verkodet wird. Ein Bericht darüber, dass Homosexuelle in Italien im Sport nicht	Das Thema lautet Homophobie. Wenn der Sucheintrag nicht in Zusammenhang mit dem Thema „Homosexualität“ steht, ist er nicht relevant.

					akzeptiert werden ist nicht relevant, wenn Deutschland kodiert wird! Der Bezug zum jeweiligen Land muss gegeben sein!	
V2	Tendenz des Inhalts		1 = neutral oder positiv	Suchergebnis, die eine neutrale oder positive Aussage oder Behauptung über das Thema selbst beinhalten. (→ jeweils die Primärintention coden) Kein kultureller Bezug (kein Zusammenhang zwischen Suchergebnis mit kodiertem Land) = keine Relevanz.	Suchergebnisse, deren Inhalte neutrale Aussagen zum „Homosexualität“ Thema beinhalten und Suchergebnisse, deren Inhalte positive Aussagen zum Thema „Homosexualität“ beinhalten	Generelle, neutrale Aussagen zum Thema. Bsp.: Es gibt homosexuelle Menschen, die im Job erfolgreich sind. Positive Aussagen zum Thema. Bsp.: Es ist gut , dass homosexuelle Menschen heute erfolgreich im Job sind. Positive Verben: gut, toll, erfreulich, etc.
			2 = negativ	Suchergebnis, in dem eine negative Aussage oder Behauptung über das Thema selbst gemacht wird (→ jeweils die Primärintention coden)	Suchergebnisse, deren Inhalte negative Aussagen zum Thema beinhalten – Homophobe Aussagen. Negative Aussagen zum Thema, die sich auf das jeweilige Land beziehen, das verkodet wird.	Negative Aussagen zum Thema. Bsp.: Es ist schade , dass homosexuelle Menschen heute erfolgreich im Job sind. Negative Verben: schlecht, ungewollt, heilen, etc., Aussagen über Abneigung der Gesellschaft, generelle Verunglimpfungen von homosexuellen Menschen, Kommentare die schwule verletzen, Aussagen die Missachtung

						ausdrücken, die Homosexualität als etwas schlechts darstellen (z.B. als Krankheit), Gewalt, Rechte (Ehe), etc.
			90 = anderes Land	Wenn kein Bezug zum Land gegeben ist. Kein Zusammenhang zwischen Suchergebnis mit kodiertem Land = anderes Land relevant	Suchergebnisse, die jedwede Inhalte und Aussagen (positive, neutrale oder negative) zum Thema beinhalten, sich aber nicht auf das Land beziehen, das verkodet wird	Bsp.: Bei der Google Suche auf der ccTLD google.de mit deutscher Sprache enthält der Textkörper die Aussage: „In den USA ist das Thema Homosexualität noch immer ein Tabuthema.“ -> diese Aussage bezieht sich auf die Homophobie in den USA, sagt aber nicht eindeutig etwas über die Homophobie in Deutschland aus
V2.5	Inhaltsebene	Nummerierung	3-8	Nur wenn V2 = 2 oder =90 Auf welcher Ebene / wo wird Homophobie sichtbar? Auf welche Inhaltsebene ist Homophobie zu erkennen?	3 = Religion 4 = Krankheit 5 = Gesellschaft 6 = Gewalt 7 = anderes Land 8 = nicht zuzuordnen	Beispiele: Homosexualität ist eine Sünde = 3 Homosexualität ist heilbar = 4 Homosexuelle sollen keine Kinder adoptieren dürfen = 5 Homosexuelle gehören an den Pranger = 6 Homosexualität in den USA... = 7 Wenn nicht eindeutig, dann = 8 (z.B. gegen Homosexualität)

Block 2: Ausländerfeindlichkeit (Suchbegriff: *Ausländer* – in der jeweiligen Landessprache)
Fragestellung: Wie *Ausländerfeindlich* ist ein Land nach Google Index?

#	Nummer	Nummerierung	1-100	Nummer des verkodeten Suchergebnisses in Reihenfolge der Rangreihe nach Google von 1 (erstes Suchergebnis) bis 100 (letztes verkodetes Suchergebnis: 100)	Suchergebnisse in der Rangreihenfolge nach Google durch nummerieren von 1-100	<ol style="list-style-type: none"> 1. Suchergebnis = 1 2. Suchergebnis = 2 3. Suchergebnis = 3 4. Suchergebnis = 4 5. Usw.
V3	Thematische Relevanz	bivariat	0= nein; 1=ja	Ist das einzelne Suchergebnis relevant für die Erhebung in Zusammenhang mit dem Thema? Zudem wenn kein kultureller Bezug (kein Zusammenhang zwischen Suchergebnis mit kodiertem Land) = keine Relevanz.	Ein Suchergebnis hat keine Relevanz wenn es nicht in den Themenkreis fällt der verkodet wird. D.h. z.B. ein Bericht darüber, dass Ausländer in Deutschland im Sport nicht akzeptiert werden ist relevant, wenn Deutschland verkodet wird. Ein Bericht darüber, dass Ausländer in Italien im Sport nicht akzeptiert werden ist nicht relevant, wenn Deutschland kodiert wird! Der Bezug zum jeweiligen Land muss gegeben sein!	Das Thema lautet Ausländerfeindlichkeit. Wenn der Sucheintrag nicht in Zusammenhang mit dem Thema „Ausländer / Einwanderer / Flüchtlinge“ steht, ist er nicht relevant.

V4	Tendenz des Inhalts		1 = neutral oder positiv	Suchergebnis, in dem eine neutrale oder positive Aussage oder Behauptung über das Thema selbst gemacht wird (→ jeweils die Primärintention coden) Kein kultureller Bezug (kein Zusammenhang zwischen Suchergebnis mit kodiertem Land) = keine Relevanz.	Suchergebnisse, deren Inhalte neutrale Aussagen zum Thema „Ausländer / Einwanderer / Flüchtlinge“ beinhalten und Suchergebnisse, deren Inhalte positive Aussagen zum Thema „Ausländer / Einwanderer / Flüchtlinge“ beinhalten	Generelle, neutrale Aussagen zum Thema. Bsp.: Flüchtlinge sind Menschen, die in Sicherheit suchen. Positive Aussagen zum Thema. Bsp.: Ausländer / Einwanderer / Flüchtlinge: Es ist gut, dass Deutschland Ausländer / Einwanderer / Flüchtlinge aufnimmt. Zuwanderung ist gut. Etc. Positive Verben: gut, toll, erfreulich, etc.
			2 = negativ	Suchergebnis, in dem eine negative Aussage oder Behauptung über das Thema selbst gemacht wird (→ jeweils die Primärintention coden)	Suchergebnisse, deren Inhalte negative Aussagen zum Thema beinhalten – rassistische Aussagen. Negative Aussagen zum Thema, die sich auf das jeweilige Land beziehen, das verkodet wird.	Bsp.: Ausländer / Einwanderer / Flüchtlinge: Ausländer nehmen uns die Jobs weg. Ausländer sind kriminell, etc. Negative Verben: schlecht, ungewollt, etc. Generelle verunglimpfungen von Ausländern / Einwanderern / Flüchtlingen, Kommentare die in ihrer Art verletzend sind, Aussagen die Missachtung ausdrücken, Gewalt, etc.

			90 = anderes Land	Wenn kein Bezug zum Land gegeben ist. Kein Zusammenhang zwischen Suchergebnis mit kodiertem Land = anderes Land relevant	Suchergebnisse, die jedwede Inhalte und Aussagen (positive, neutrale oder negative) zum Thema beinhalten, sich aber nicht auf das Land beziehen, das verkodet wird	Bsp.: Bei der Google Suche auf der ccTLD google.de mit deutscher Sprache enthält der Textkörper die Aussage: „In den USA ist das Thema Ausländerfeindlichkeit noch immer ein Tabuthema.“ -> diese Aussage bezieht sich auf die Ausländerfeindlichkeit in den USA, sagt aber nicht eindeutig etwas über die Ausländerfeindlichkeit in Deutschland aus
V4.5	Inhaltsebene	Nummerierung	3-8	Nur wenn V2 = 2 oder =90 Auf welcher Ebene / wo wird Homophobie sichtbar? Auf welche Inhaltsebene ist Homophobie zu erkennen?	3 = Religion 4 = Krankheit 5 = Gesellschaft 6 = Gewalt 7 = anderes Land 8 = nicht zuzuordnen	Beispiele: Gott will keine Ausländer bei uns = 3 Ausländer bringen Krankheiten = 4 Ausländer sollen ihre Kinder woanders bekommen = 5 Ausländer muss man verprügeln = 6 Ausländer in den USA... = 7 Wenn nicht eindeutig, dann = 8 (z.B. Ausländer raus)

Block 3: Drogenkonsum (Suchbegriff: *illegale Drogen* – in der jeweiligen Landessprache)
Fragestellung: Ist Marihuana als Rauschmittel akzeptiert, oder nicht?

#	Nummer	Nummerierung	1-100	Nummer des verkodeten Suchergebnisses in Reihenfolge der Rangreihe nach Google von 1 (erstes Suchergebnis) bis 100 (letztes verkodetes Suchergebnis: 100)	Suchergebnisse in der Rangreihenfolge nach Google durch nummerieren von 1-100	6. Suchergebnis = 1 7. Suchergebnis = 2 8. Suchergebnis = 3 9. Suchergebnis = 4 Usw.
V5	Anzahl der Nennungen	Anzahl der Nennungen auf 100 Suchergebnisse	x	Suchergebnis, in dem das Wort Marihuana, Haschisch, Marihuana, oder ein anderes Synonym als Ergebnis in Zusammenhang mit dem Suchbegriff „illegale Drogen“ (je Landessprache) auftaucht. (→ jeweils pro Ergebnis die Anzahl der Nennungen auf 100 Suchergebnisse zählen)	Suchergebnisse, die Marihuana, Haschisch, Marihuana, oder ein anderem Synonym in Zusammenhang mit dem Thema illegale Drogen stellen. Drogen implizieren = illegale Drogen Substanzen bezeichnet, welche als nicht verkehrsfähig in Anlage I des Betäubungsmittelgesetzes (BtMG) aufgeführt sind. Das heißt: Je mehr Nennungen von Marihuana in Zusammenhang mit dem Thema illegale Drogen, desto geringer ist die Akzeptanz der Droge je Land / Kultur.	Alle Aussagen zum Thema. Bsp.: Auch Marihuana ist eine Droge.

Codeplan Auto-Vervollständigung

Inhalte Google Auto-Vervollständigung (Suchvervollständigung)						
Für fehlende Werte → 99 Coden (z.B. wenn V6 = 0, dann V2 = 99)						
Variable	Variablenname	Wert	Label	Erläuterung	Operationalisierung/Definitionen	Hinweise/Beispiel
Block 1: Homophobie (Suchbegriff: <i>Homosexualität</i> – in der jeweiligen Landessprache)						
Fragestellung: Wie <i>Homophob</i> ist ein Land nach Google Auto-Vervollständigung?						
#	Zahlencode	Alphabet	1 - 4 / 10	Nummer des verkodeten Suchvervollständigung Ergebnis in Reihenfolge der Rangreihe nach Google von: 1 (erstes Suchvervollständigung) bis max. 10	Suchergebnisse in der Rangreihenfolge nach Google Auto-Vervollständigung durch nummerieren 1-5	<ol style="list-style-type: none"> 1. Suchvervollständigung-Ergebnis = 1 2. Suchvervollständigung-Ergebnis = 2 3. Suchvervollständigung-Ergebnis = 3 4. Suchvervollständigung-Ergebnis = 4 5. Usw.
V6	Thematische Relevanz	bivariat	0= nein; 1=ja	Ist das einzelne Suchergebnis relevant für die Erhebung in Zusammenhang mit dem Thema? Zudem wenn kein kultureller Bezug (kein Zusammenhang zwischen Suchergebnis mit kodiertem Land) = keine Relevanz.	Ein Suchvervollständigung-Ergebnis hat keine Relevanz wenn es nicht in den Themenkreis fällt der verkodet wird.	Das Thema lautet Homophobie. Wenn das Suchvervollständigung-Ergebnis nicht in Zusammenhang mit dem Thema „Homosexualität“ steht, ist es nicht relevant.

V7	Tendenz des Inhalts		1 = neutral oder positiv	Suchvervollständigung-Ergebnisse, in denen eine neutrale oder positive Aussage oder Behauptung über das Thema selbst gemacht wird (→ jeweils die Primärintention coden)	Suchvervollständigung-Ergebnisse, deren Inhalte neutrale oder positive Aussagen zum „Homosexualität“ Thema beinhalten und	Generelle, neutrale Aussagen zum Thema. Bsp. Ergebnis : „Homosexualität Definition“ Positive Aussagen zum Thema. Bsp. Ergebnis: „Homosexualität willkommen“
			2 = negativ	Suchvervollständigung-Ergebnisse, in denen eine negative Aussage oder Behauptung über das Thema selbst gemacht wird (→ jeweils die Primärintention coden)	Suchvervollständigung-Ergebnisse, deren Inhalte negative Aussagen zum Thema beinhalten – Homophobe Aussagen	Negative Aussagen zum Thema. Bsp. Ergebnis: „Homosexualität Krankheit“
			90 = anderes Land	Wenn kein Bezug zum Land gegeben ist. Kein Zusammenhang zwischen Suchergebnis mit kodiertem Land = anderes Land relevant	Suchergebnisse, die jedwede Inhalte und Aussagen (positive, neutrale oder negative) zum Thema beinhalten, sich aber nicht auf das Land beziehen, das verkodet wird	Bsp.: Bei der Google Suche auf der ccTLD google.de mit deutscher Sprache enthält die Vervollständigung: „Homosexualität USA“ Dies bezieht sich auf die USA, sagt aber nicht eindeutig etwas über die Homophobie in Deutschland aus
V7.5	Inhaltsebene	Nummerierung	3-8	Nur wenn V2 = 2 oder =90 Auf welcher Ebene / wo wird Homophobie sichtbar? Auf welche Inhaltsebene ist Homophobie zu erkennen?	3 = Religion 4 = Krankheit 5 = Gesellschaft 6 = Gewalt 7 = anderes Land	Beispiele: Homosexualität Sünde = 3 Homosexualität heilbar = 4 Gesellschaft gegen Homosexuelle = 5 Homosexuelle Foltern = 6

					8 = nicht zuzuordnen	Homosexualität USA = 7 Wenn nicht eindeutig, dann = 8 (z.B. gegen Homosexualität)
--	--	--	--	--	----------------------	---

Block 2: Ausländerfeindlichkeit (Suchbegriff: *Ausländer* – in der jeweiligen Landessprache)

Fragestellung: Wie *Ausländerfeindlich* ist ein Land nach Google Index?

#	Zahlencode	Alphabet	1 - 100	Nummer des verkodeten Suchvollständigkeit Ergebnis in Reihenfolge der Rangreihe nach Google von: 1 (erstes Suchvollständigkeit) bis 100 (letztes verkodetes Suchvollständigkeit)	Suchergebnisse in der Rangreihenfolge nach Google Suchvollständigkeit durch nummerieren 1-100	<ol style="list-style-type: none"> 1. Suchvollständigkeit-Ergebnis = 1 2. Suchvollständigkeit-Ergebnis = 2 3. Suchvollständigkeit-Ergebnis = 3 4. Suchvollständigkeit-Ergebnis = 4 5. Usw.
V8	Thematische Relevanz	bivariat	0= nein; 1=ja	Ist das einzelne Suchergebnis relevant für die Erhebung in Zusammenhang mit dem Thema? Zudem wenn kein kultureller Bezug (kein Zusammenhang zwischen Suchergebnis mit kodiertem Land) = keine Relevanz.	Ein Suchvollständigkeit-Ergebnis hat keine Relevanz wenn es nicht in den Themenkreis fällt der verkodet wird.	Das Thema lautet Ausländerfeindlichkeit. Wenn das Suchvollständigkeit-Ergebnis nicht in Zusammenhang mit dem Thema „Ausländerfeindlichkeit“ steht, ist es nicht relevant.
V9	Tendenz des Inhalts		1 = neutral oder positiv	Suchvollständigkeit-Ergebnisse, in denen eine neutrale oder positive Aussage oder Behauptung über das Thema selbst gemacht wird (→ jeweils die Primärintention coden)	Suchvollständigkeit-Ergebnisse, deren Inhalte neutrale oder positive Aussagen zum „Ausländer / Migranten / Flüchtlinge“ Thema beinhalten	<p>Generelle, neutrale Aussagen zum Thema.</p> <p>Bsp. Ergebnis : „Ausländer / Migranten / Flüchtlinge Definition“</p> <p>Positive Aussagen zum Thema.</p> <p>Bsp. Ergebnis:</p>

						„Ausländer / Migranten / Flüchtlinge Hilfe“
			2 = negativ	Suchvervollständigung-Ergebnisse, in denen eine negative Aussage oder Behauptung über das Thema selbst gemacht wird (→ jeweils die Primärintention coden)	Suchvervollständigung-Ergebnisse, deren Inhalte negative Aussagen zum Thema beinhalten – Rassistische Aussagen	Negative Aussagen zum Thema. Bsp. Ergebnis: „Ausländer / Migranten / Flüchtlinge raus“
			90 = anderes Land	Wenn kein Bezug zum Land gegeben ist. Kein Zusammenhang zwischen Suchergebnis mit kodiertem Land = anderes Land relevant	Suchergebnisse, die jedwede Inhalte und Aussagen (positive, neutrale oder negative) zum Thema beinhalten, sich aber nicht auf das Land beziehen, das verkodet wird	Bsp.: Bei der Google Suche auf der ccTLD google.de mit deutscher Sprache enthält die Suchvervollständigung : „Ausländer USA“ Dies bezieht sich auf die USA, sagt aber nicht eindeutig etwas über die Fremdenfeindlichkeit in Deutschland aus
V7.5	Inhaltsebene	Nummerierung	3-8	Nur wenn V2 = 2 oder =90 Auf welcher Ebene / wo wird Homophobie sichtbar? Auf welche Inhaltsebene ist Homophobie zu erkennen?	3 = Religion 4 = Krankheit 5 = Gesellschaft 6 = Gewalt 7 = anderes Land	Beispiele: Ausländer Sünder = 3 Ausländer Krankheit = 4 Gesellschaft gegen Ausländer = 5 Ausländer Hass = 6

					8 = nicht zuzuordnen	Ausländer USA = 7 Wenn nicht eindeutig, dann = 8 (z.B. Ausländer raus)
--	--	--	--	--	----------------------	---

Block 3: Drogenkonsum (Suchbegriff: *illegale Drogen* – in der jeweiligen Landessprache)

Fragestellung: Ist Marihuana als Rauschmittel akzeptiert, oder nicht?

V10	Anzahl der Nennungen	Anzahl der Nennungen auf 5 Suchvollständigun- g-Ergebnisse	x	Suchvollständigkeit, in dem das Wort Marihuana, Haschisch, Marihuana, oder ein anderes Synonym als Ergebnis in Zusammenhang mit dem Suchbegriff „illegale Drogen“ (je Landessprache) auftaucht. (→ jeweils pro Ergebnis die Anzahl der Nennungen auf 5 Suchvollständigun- g-Ergebnisse zählen)	Suchvollständigun- g-Ergebnisse, die Marihuana, Haschisch, Marihuana, oder ein anderes Synonym in Zusammenhang mit dem Thema illegale Drogen stellen. Das heißt: Je mehr Nennungen von Marihuana in Zusammenhang mit dem Thema illegale Drogen, desto geringer ist die Akzeptanz der Droge je Land / Kultur.	Aussagen zum Thema. Bsp. Ergebnis: „Illegale Drogen Marihuana“
------------	-----------------------------	--	---	--	---	--

Eidesstattliche Erklärung

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

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 geistigem Eigentum überprüft werden kann.

Köln, den 20.11.2015

pers. Unterschrift