

Vom Bulletin Board System (BBS) zur Declaration of the Independence of Cyberspace

BA Hauptseminararbeit

Josias Bruderer

31. August 2021

Abstract

This is the place where the Abstract goes

Autor	Josias Bruderer (josias.bruderer@stud.unilu.ch)
-------	---

Matrikelnr.	S19-455-070
-------------	-------------

Seminar	The ABC of Computational Text Analysis
---------	--

Semester	Frühlingssemester 2021
----------	------------------------

Studiengang	Universität Luzern, BA Gesellschafts- und Kommunikationswissenschaften
-------------	---

Eingereicht	am 31. August 2021 bei Alex Flückiger und Prof. Dr. Sophie Mützel
-------------	---

Inhaltsverzeichnis

1	Einleitung	1
2	Untersuchungsgegenstand	2
2.1	Bulletin-Board System (BBS)	3
2.2	A Declaration of Intependence of Cyberspace	4
2.3	Forschungslücke	6
3	Methode zur Untersuchung von textfiles.com	7
3.1	Überblick	7
3.2	Datenbereinigung	8
3.3	Analyse mittels NLP	10
4	Ergebnisse	11
4.1	Dokumente pro Jahr	12
4.2	Kategoriengrösse	13
4.3	Worthäufigkeiten	13
4.4	Enitäten	14
4.5	Scattertext	14
4.6	Topic Diversity	15
5	Fazit (not ready)	16
6	Literatur	18
7	Andere Quellen	20
8	Anhang	21
8.1	Anhang 1: Manuelle Untersuchung der Top100	21
8.2	Anhang 2: Sourcecode Top100	21
8.3	Anhang 3: Textfiles.com Analyse	21
8.4	Anhang 4: Sourcecode Textfiles.com	21
8.5	Anhang 5: OCTIS Topic Diversity	22

1 Einleitung

We will create a civilization of the Mind in Cyberspace. May it be more humane and fair than the world your governments have made before.

— John Perry Barlow, Davos, 8. Februar 1996.

Als Antwort auf den *Telecommunications Act* der Vereinigten Staaten von 1996, veröffentlichte John Perry Barlow die Schrift *A Declaration of the Independence of Cyberspace*. Darin erklärt er den Cyberspace zu einem Ort «where anyone, anywhere may express his or her beliefs, no matter how singular, without fear of being coerced into silence or conformity.» (Barlow 1996) Die *Declaration* wurde zum Symbol einer Bewegung, die auf die Befreiung der Internetnutzer_innen hoffte (Turner 2006: 13), gleichzeitig bildete die *Declaration* einen Mythos des unabhängigen Internets, eines utopischen Raumes, wie er von James Bennett (2015) und Daniel Kreiss (2015) beschrieben wird.

Cyberspace ist allerdings nicht ein Begriff, den es seit der Erfindung des Internets in den frühen 90er Jahren gibt. Bereits vor dem Internet existierten Computer Netzwerke, über welche kommuniziert oder Daten ausgetauscht wurden. Eines davon sind die Bulletin Board Systems (BBS), Server auf die über einen am Telefonnetz angeschlossenen Computer, unter anderem Texte hoch- und heruntergeladen werden konnten.

Diese Hauptseminararbeit soll der Frage auf den Grund gehen, inwiefern die BBS eine Rolle für die Entwicklung und den Austausch des Zeitgeistes spielte, der sich in der techno-utopischen Community rund um die 1996 formulierte *Declaration*, wiederfinden lässt. Die Fragestellung ermöglicht eine Annäherung an zwei aktuelle Diskussionen: Erstens galten BBS als dezentrale und unabhängige Kommunikationsmedien, ähnlich wie dies soziale Medien nach dem Platzen der Dotcom-Blase¹ taten oder heute auf Blockchain basierte Technologien tun. Die Entwicklung der BBS und des Internets verlief allerdings nicht nach den Wünschen der frühen Nutzer_innen, worauf im zweiten Kapitel genauer eingegangen wird.

¹ Die Ideologie des Web 2.0 und exemplarisch die Entwicklung WhatsApps vertiefte ich in einer Proseminararbeit, eingereicht im Juli 2020 bei Markus Unternährer. In der vorliegenden Arbeit soll die Zeit vor Web 2.0 behandelt werden.

Auch die aktuell vielversprechenden Technologien erfahren Druck von Industrie, Staat und Gesellschaft - ein gutes Beispiel dazu ist die Entwicklung von Bitcoin.² Zweitens wird die *Declaration* auch heute noch, 25 Jahre nach derer Publikation, reichlich zitiert oder darauf Bezug genommen, wenn es um Meinungsäusserungsfreiheit, die staatliche Regulierung des Internets oder die Monopolstellungen von einigen wenigen Technologieunternehmen geht.³ Auch wenn das Dokument heute keine direkte Anwendung findet, scheint es ein prägendes Element der Internetgeschichte zu sein und nicht als Artefakt in den unzähligen Nullen und Einsen zu verstauben.

Um den Hintergrund der *Declaration* zu verstehen, kann der historische Kontext, welcher im zweiten Kapitel behandelt wird, dienen. Eine weitere Herangehensweise, die in dieser Arbeit behandelt und im Mittelpunkt stehen wird, ist die Analyse bestehender Originaldaten (BBS Textfiles). Das methodische Vorgehen dazu wird im dritten Kapitel beschrieben. Darauf folgen die Ergebnisse dieser Analyse sowie deren Interpretation. Abschliessend wird ein Fazit gezogen, einerseits ob das Vorgehen der Datenanalyse geeignet ist, um der Community rund um die *Declaration* herum den Puls zu fühlen und andererseits welche Erkenntnisse daraus gewonnen werden können, respektive wie an diese Analyse angeschlossen werden kann.

2 Untersuchungsgegenstand

Forschung zum Internet und zu dessen Entwicklung wurde reichlich betrieben und ist in unzähligen Texten aus verschiedensten wissenschaftlichen Disziplinen festgehalten. Auffällig dabei ist, wie in der Einleitung beschrieben, dass auf die *Declaration* ab 1996 häufig Bezug genommen wird, während die BBS, bis auf ei-

² Artikel und Schlagzeilen dazu gibt es reichlich. Eine passende Parallele zu BBS könnte mit der staatlichen Regulierung des Dienste, wie dies in China geschieht, gezogen werden. Wired berichtet über die Machtverhältnisse rund um Bitcoin: <https://www.wired.com/story/opinion-bitcoins-greatest-feature-is-also-its-existential-threat>

³ Ein Beispiel wäre Jochen Steinbickers Kapitel «Überwachung und die Digitalisierung der Lebensführung» (2019), in welchem er Bezug zu Edward Snowden und der NSA nimmt und daran die Vorstellung eines offenen und unabhängigen Internets, wie dies die *Declaration* fordert, destruiert.

nige technische Leitfäden und die Dokumentation von Jascon Scott (2005), nur marginal in der Literatur erwähnt werden. Im folgenden wird ein Überblick über die BBS und die *Declaration* gegeben, sowie eine Skizze derer zugrunde liegenden Utopie zu zeichnen versucht.

2.1 Bulletin-Board System (BBS)

«The Computerized Hobbyist Bulletin Board System is a personal computer based system for message communication among experimenters. People with terminals or computers equipped with modems call in to leave and retrieve messages.» (Christensen und Suess 1978: 150) So wurde die Idee der BBS von deren Erfindern im November 1978 in der Computerzeitschrift *BYTE* erstmals veröffentlicht. Der Zweck war es, eine Plattform für das Sammeln von Newsletter-Artikel des *Chicago Area Computer Hobbyist Exchange*, anstelle eines Anrufbeantworters, zu bieten. So simpel diese Entwicklung auch klingen mag, «[...] the results were definitely impressive. The original home-brewed internet, BBS was primitive but quickly proved revolutionary.» (Gilbertson 2010) Mit dem Aufkommen von leicht bedienbaren Computern Mitte der Neunzigerjahren, stieg die Anzahl der betriebenen BBS in die Zehntausenden (Britannica Academic 2021).

Wichtig zu erwähnen ist, dass die mit der Zeit entstandenen BBS verschiedene Verwendungszwecke hatten: von günstigen E-Mail Plattformen, über Filesharing Plattformen und Softwarepiraterie, bis hin zu grossen Dienstleistungsbüros wie beispielsweise CompuServe. Auch Echtzeitkommunikation im Sinne von Telefonkonferenzen, was wir heute als Chatrooms bezeichnen würden, oder mehrspieler Games wurden angeboten. (Dvorak und Anis 1990: 208–209, 223) Ebenfalls waren BBS nicht eigenständige und völlig neue Erfindungen. Einerseits nutzten Sie bestehende Technologien wie das Telefonnetz oder Computer und auch waren viele derer Funktionen keine völligen Neuheiten. Der grosse Unterschied darin bestand aber in der vergleichsweise niederschweligen Vernetzung von Computer über Modems und mit den BBS standardisierte Datenaustauschplattformen (BBS: *The Documentary* 2005: 00:08:35). Andererseits waren BBS nur ein Teil der Vernet-

zung. Deren Leistung entfalteten Sie durch Entwicklungen wie das FidoNet, mittels welchem verschiedene BBS verbunden wurden und E-Mail Übermittlungen ermöglicht wurden (Quartermann 1990: 254).

Das World Wide Web löste die BBS ab, die Anzahl betriebener Server nahm nach dem Höhepunkt Mitte der Neunzigerjahre drastisch ab. Gewisse BBS passten sich an und sind auch heute noch übers Internet erreichbar. Nichts desto trotz erscheinen die BBS als ein wichtiger Teil der Internetgeschichte oder wie Jason Scott es in seinem Dokumentarfilm formuliert: «The BBS helped bring home computers from kits and novelties to required equipment in homes, work, and schools. It could be argued that BBSes helped speed acceptance and growth of the Internet to the size it has become.» (*BBS: The Documentary* 2005: 00:34:00)

2.2 A Declaration of Intependence of Cyberspace

Die *Declaration* ist einer der Texte, die die Gedanken und Ansichten einer technologischen Bewegung widerspiegeln. Referenzen darauf sind in der Literatur zur Internetgeschichte oder zur Diskussion der digitalen Privatsphäre häufig zu finden. Um deren Entstehung und Relevanz nachvollziehen zu können, werden folgend drei Perspektiven zusammengefasst.

Einerseits kann aus einer Perspektive des Raumverständnisses auf die *Declaration* geblickt werden: Der Cyberspace wird in der Declaration als ein Raum jenseits dessen betrachtet, was Staaten und Autoritäten zu kontrollieren vermögen, respektive dazu befugt wären. Diese neue Raumimagination, wie dies Manfred Fassler formuliert, entstand aufgrund der globalen Telefonnetze und deren «kulturelle[n] Codierung für Fernanwesenheit oder zeiteinheitliche Fernräume» (2008: 188). Fassler spricht von User Generated Spaces und Cyberlocalism und identifiziert Barlows *Declaration* als ein solcher: «Raum hat immer Autoren und Autorinnen, Macher und Macherinnen. Das war aus meiner Sicht auch die wichtige Nachricht John Perry Barlows, als er in seiner *Declaration of Independence in Cyberspace* gegen die alten Raummächte schrieb und in alter Cyberfreak-Manier den greatfull space feierte, in dem die Kybernetik der Freiheit bzw. Freiheit der Kyberne-

tik herrschte.» (Fassler 2008: 191) Diese Haltung wird auch sehr deutlich, in dem Barlow formuliert: «Cyberspace does not lie within your [governmental] borders. Do not think that you can build it, as though it were a public construction project. You cannot. It is an act of nature and it grows itself through our collective actions.» (Barlow 1996)

Diese räumliche Distinktion kann kritisch hinterfragt werden. Damit eröffnet sich eine zweite Perspektive, jene des Mythos der Unabhängigkeit: Entgegen Barlows Haltung, der Cyberspace sei kein öffentliches Projekt, kann argumentiert werden, dass der Cyberspace, und damit exemplarisch auch die BBS, auf Wissen und Infrastruktur basieren und von externen Faktoren beeinflusst werden. «[M]edia spaces are always entangled in economic relations, governmental and regulatory structures, and the workings of institutions.» (Kreiss 2015: 120) Gleichzeitig muss berücksichtigt werden, dass die Ideen der *Declaration* keine neue Vision war. Kreiss nennt exemplarisch die Metapher des «frontiers», die für die 1990 gegründete Electronic Frontiers Foundation (EFF) zentral war. Sie setzte sich für die Stärkung des Schutzes der Freiheiten für die Online-Kommunikation ein, unter anderem auch den Rechten auf Privatsphäre von BBS Benutzenden (Britannica Academic 2016). Barlows *Declaration*, die im Namen der EFF verfasst wurde, kann als Versuch gelesen werden, die Interessen der EFF zu manifestieren und öffentlich bekannt zu machen. Wie Kreiss anmerkt, erreichte die *Declaration* zwar grosse kulturelle Reichweite, stellt aber auch ein äusserst umstrittener Gegenstand dar: «Within the scholarly literature, for instance, the Declaration is now often perfunctorily cited as a footnote to early technology culture and its extreme rhetoric [...] and misguided regulatory thinking [...] during the early boom years of Silicon Valley.» (Kreiss 2015: 124) Der Mythos der Unabhängigkeit im Cyberspace und der Kampf dafür bildet trotzdem nach wie vor einen Rahmen für Projekte wie Wikipedia oder Linux. Dies, in dem er Beteiligten eine Vorstellung über sich selbst und deren Tätigkeit ermöglicht und sie damit motiviert, etwas der Gemeinschaft beizutragen. (Kreiss 2015: 130)

Als dritte Perspektive soll der historische Kontext rund um die *Declaration* betrachtet werden. Wie bereits erwähnt, war sie eine Antwort auf den Telecommuni-

cations Act von 1996 (Britannica Academic 2016; Kreiss 2015: 121; Turner 2006: 13). Damit es dazu kam war eine breite Vorgeschichte notwendig und fast wichtiger noch, die Etablierung einer Utopie, an welcher die frühen Internet-Nutzer_innen festhalten versuchten. Ein Teil der Vorgeschichte ist die Entwicklung der BBS, wie im vorherigen Teilkapitel beschrieben. Vernetzte Computertechnik war allgegenwärtig «and in its shiny array of interlinked devices, pundits, scholars, and investors alike saw the image of an ideal society: decentralized, egalitarian, harmonious, and free.» (Turner 2006: 1) Fred Turner beschreibt in seinem Buch «From Counterculture to Cyberculture» die Entwicklung der Vorstellung einer Bedrohung durch Computer, hin zu einem öffentlichen Verständnis von Computern die Hoffnung auf eine globale Harmonie versprochen (2006: 2, 5–6). So kam es, dass «[t]hirty years later [1996], the same aspects of computing that threatened to dehumanize the students of the Free Speech Movement promised to liberate the users of the Internet.» (Turner 2006: 13)

Diese Art eines unabhängigen Mediums, dass es vermag, die Benutzenden des Internets zu befreien, muss nach James Bennett (2015) als utopisches Ideal verstanden werden. Utopie im Sinne einer Vision die es zu verfolgen gilt und als Ausdruck für eine bessere Lebensweise (Levitas 2010: 1; 2003: 4). Und das ist, was die *Declaration* tat, sie «promised the new digital citizens of the online world a commonweal outside the terrestrial, outmoded structures of state, capitalism and <old media.>» (Bennett 2015: 6)

2.3 Forschungslücke

Die Abschnitte zu BBS und der *Declaration* stellt nur eine Zusammenfassung eines sehr viel grösseren Diskurses dar. Während die BBS vor allem als Werkzeuge beschrieben werden und ausgehend davon auf eine kleine Gruppe von Computer-Enthusiasten geblickt wird, wird die *Declaration* als eines der Symbole für die Utopie einer grösseren sozialen Bewegung dargestellt. Eine Analyse der effektiven Inhalte der BBS ist nicht zu finden, ebenfalls sind Annäherungen an die *Declaration* fast ausschliesslich historischer Art. Mit der im Zentrum dieser Arbeit stehenden

Fragestellung soll ein Einblick in die Inhalte der BBS gewonnen werden und darin nach Auffälligkeiten in Bezug auf die *Declaration* gesucht werden.

3 Methode zur Untersuchung von textfiles.com

Für diese Arbeit konnte ein Archiv mit BBS Text-Dateien ausfindig gemacht werden. Es wird von Jason Scott unterhalten, der nebst Archivar und Historiker Mitwirkender bei archive.org sowie Regisseur vom Film «BBS: The Documentary» ist. Der Datensatz erscheint aufgrund seines Umfangs ($N = 58'000$ Text-Dateien) und Scotts Hintergrund als geeignet, um eine Analyse von BBS Inhalten durchzuführen. Ebenfalls steht ein verkleinerter Datensatz («favorite 100») zur Verfügung, der für die Analyse hilfreich ist.

Die Analyse erfolgt in drei Schritten: Im ersten Schritt wird der verkleinerte Datensatz manuell auf inhaltliche Auffälligkeiten untersucht und daran ein Modell zur Bereinigung des gesamten Datensatzes entwickelt.⁴ Anschliessend wird im zweiten Schritt der gesamte Datensatz bereinigt und gefiltert. Abschliessend wird mithilfe von Techniken der NLP (Natural Language Processing) der Textkorpus untersucht und damit die Fragestellung zu beantworten versucht.

3.1 Überblick

Im verkleinerten Datensatz, «a «best of» collection of one hundred textfiles that [Scott] think[s] capture the spirit of this site and the unique culture that it attempts to preserve» (Jason Scott o. J.), werden Auffälligkeiten zu Jahr, Länge, Struktur und Inhalt festgehalten und fliessen in den folgenden Abschnitt *Datenbereinigung* ein. Wichtig erscheint an dieser Stelle, dass es sich um durchgehend unstrukturierte Texte handelt. Auch sind die Daten trotz derer Menge «Ergebnisse historisch kontingenter Entscheidungen, wie Auswahl, Formatierung, Kombinati-

⁴ Dieser Schritt wurde im Mini-Projekt im Rahmen des Seminars «The ABC of Computational Text Analysis» durchgeführt und ist dokumentiert unter: <https://github.com/josiasbruderer/jason-scotts-favorite-100>

on etc. und als solche weder neutral, transparent oder objektiv, sondern immer konstruiert» (Mützel u. a. 2018: 112). Dies ist für die Auswertung und das Ziehen von Schlüssen relevant. Die folgenden Punkte geben einen besseren Überblick über den Datensatz:⁵

1. Die **Variation** der Textdateien ist relativ gross und reicht von kurzen witzigen Beiträgen und Unterhaltungsverläufen, über ASCII Art bis hin zu detaillierten technischen Instruktionen und Dokumentationen sowie einer Masterarbeit und einem ganzen Buch.
2. In gewissen Textdateien wird ein «**Read X times**» ausgewiesen. Diese Zahlen sind relativ niedrig (meist <100).
3. **Ungültige Zeichen** kommen häufig vor (z.B. «\u1a\u1a\u1a»). Anhäufungen von **Sonderzeichen** dienen zur Formatierung.
4. Verschiedene **Datumsformate** sind zu finden.
5. Es kann nicht davon ausgegangen werden, dass die Textdateien **orthografisch** fehlerfrei sind. Ebenfalls ist mit *Gunk* («*replacing U for You, o for O, Z for S, and similar gunk*») zu rechnen.
6. **Inhaltlich** kommt von sauber recherchierten Artikeln und Facts bis hin zu wilder Fiktion und Ironie alles vor.
7. Textfiles.com führt nebst den Dateien auch Titel (inkl. Jahr wenn vorhanden) und Kategorisierung sowie zum Teil eine Beschreibung auf. Diese **Metadaten** können für die Analyse nützlich sein.

3.2 Datenbereinigung

Aus den aufgelisteten Feststellungen wird nun ein Plan entwickelt, wie die Textfiles bereinigt werden. Das Kapitel «Schöne Daten» von Sophie Mützel et al. (2018) macht deutlich, dass die Datenbereinigung oft nur marginal erwähnt wird, obwohl diese elementar ist. Die Bereinigung soll für diese Arbeit daher genau durchdacht und transparent dokumentiert werden.

⁵ Die Protokollierung der manuellen Untersuchung des Datensatzes ist in Anhang 1: Top100 und der Sourcecode in Anhang 2: Sourcecode Top100 zu finden.

Bereits beim Herunterladen der Textfiles sollen unpassende Dateien (z.B. Bilder oder Audio) exkludiert werden. Gleichzeitig sollen von den verbleibenden Dateien entsprechende Metadaten gespeichert werden. Dann erfolgt die erste Bereinigung, nämlich das Entfernen von Formatierungen und anderen nicht relevante Elemente:

- Bereinigung von ungültigen Zeichen (z.B. \x1a)
- Bereinigung von Zeichen die kein Text repräsentieren
- Bereinigung von Formatierungszeichen (Zeilenumbrüche, Tabulatoren)

Anschliessend werden zusätzliche Metadaten generiert:

- Kategorie, Dateiname
- Zeichenanzahl (roh & bereinigt)
- Durchschnittliche Spaltenbreite (roh)
- Anteil von Fliesstext gegenüber Sonderzeichen (roh)
- Jahr des Textfiles (Annahme: zwischen 1960-1999)

Anhand dieser Metadaten kann eine erste Analyse des Datensets gemacht werden.⁶ Einerseits sollen ungeeignete Kategorien bereits zu diesem Zeitpunkt aus der Analyse ausgeschlossen werden, andererseits sollen geeignete Parameter für die Reduktion des Datensets gefunden werden. Die Analyse kommt zu folgendem Schluss:

- Folgende Kategorien werden ausgeschlossen, da deren durchschnittliches Verhältnis an sinnvollen Zeichen zur Dateilänge (*charratioB*) unter 0.8 liegt. Dies stellen Sammlungen von ASCII-Art, Softwarecode, Bilder oder ähnlichem und nicht Text dar: *tap*, *floppies*, *exhibits*, *artifacts*, *piracy*, *art*, *fidonet-on-the-internet*
- Folgende Kategorien werden aus der Gesamtbetrachtung ausgeschlossen, da deren Charakter jener von Zeitschriften und nicht nutzergenerierter Inhalten entspricht: *magazines*, *digest*

⁶ Dies ist in Anhang 3: Textfiles.com Analysis dokumentiert.

- Dateien, deren Verhältnis von Text (inkl. Satz- und Leerzeichen) zu Dateilänge unter 0.95 sind, werden ausgeschlossen. Dies damit vor allem Fliesstext und Unterhaltungen bei der Analyse erfasst werden.
- Dateien, deren Länge zwischen 300 und 30'000 Zeichen liegen sollen in die Analyse eingeschlossen werden. Kürzere Texte erscheinen kaum aussagekräftig und längere Texte verzerren einerseits das Resultat stark, andererseits erhöhen Sie die Dauer der Berechnungen in der Analyse enorm.

Mit der gewählten Filterung verbleiben 5'510 Textdateien die in den Textkorpus einfließen.⁷ Der Gesamtumfang entspricht ca. 62 Millionen Zeichen. Diese Messwerte sind relevant, da damit die benötigte Rechnerleistung berechnet werden kann. (1GB Arbeitsspeicher pro 1 Million Zeichen) Zusätzlich zu den Textdateien wird die *Declaration* und einige charakteristische Textdateien als Datenset in den Textkorpus aufgenommen, damit mit diesen schliesslich verglichen werden kann.

3.3 Analyse mittels NLP

Unter Natural Language Processing (NLP) werden Methoden zum maschinellen Analysieren, Modellieren und Verstehen von menschlicher Sprache verstanden (Vajjala u. a. 2020). Allerdings ist dies nicht mit der maschinellen Erfassung von Sinn oder Bedeutung zu verwechseln, wie Emily M. Bender und Alexander Koller (2020) argumentieren. Denn um ein solches Modell zu schaffen, wird eine Software mit einer grossen Menge an Daten trainiert und «if the training data is only form, there is not sufficient signal to learn the relation M between that form and the non-linguistic intent of human language users, nor C between form and the standing meaning the linguistic system assigns to each form.» (Bender und Koller 2020: 5187) Für gewisse Anwendungsbereiche von NLP hat dies ernstzu-

⁷ Die Bereinigung und Filterung des Datensatzes ist in Anhang 4: Sourcecode Textfiles.com dokumentiert. Eine genauere Dokumentation inklusive Textkorpus ist online verfügbar: <https://github.com/josiasbruderer/bbs-for-independence/> unter 03_workspace und 03_workspace/states

nehmende Konsequenzen.⁸ Es ist aber gerade diese Eigenschaft, dass das Modell nur den Daten entspricht, mit welchen es gefüttert wird, die in der vorliegenden Arbeit zunutze gemacht wird.

Aus dem bereinigten Datensatz soll ein Textkorpus generiert werden, welcher mithilfe der Python Bibliotheken «textacy», «scattertext» und «OCTIS» untersucht wird. Dabei werden folgende Analysen durchgeführt:

- Worthäufigkeiten absolut und Anzahl vorkommende Dateien: Gesamtdatensatz, Datensatz *Declaration*, pro Textfile-Kategorie
- Kategoriengrösse: Gesamtdatensatz
- Entitäten: Gesamtdatensatz, Datensatz *Declaration*
- Scattertext Analyse: Gesamtdatensatz verglichen mit der *Declaration* und dem Datensatz *Declaration*
- Anzahl Dokumente pro Jahr: Gesamtdatensatz
- Topic Diversity mittels OCTIS (LDA und NeuralLDA): Gesamtdatensatz, Datensatz *Declaration*

Für die Auswertung mittels OCTIS (Terragni u. a. 2021) wird auf Basis des Textkorpus, welcher für die anderen Analysen genutzt wurde, ein separater Datensatz erstellt. Die Tests werden mittels *OCTIS Dashboard* durchgeführt und die Analyse erfolgt mit den Modellen LDA (Srivastava und Sutton 2017) und NeuralLDA (Blei u. a. 2003).⁹

4 Ergebnisse

Im Folgenden werden die Ergebnisse der Analyse dargelegt. Die ersten zwei Analysen *Dokumente pro Jahr* und *Kategoriengrösse* dienen dazu, den Textkorpus zu beschreiben und die Beschaffenheit und Menge an Dateien pro Jahr respektive pro

⁸ Exemplarisch am «BERT neural network» wird das beschrieben im Artikel «Machines Beat Humans on a Reading Test. But Do They Understand?», erschienen im Quanta Magazine (Pavlus 2019).

⁹ Die Testresultate sind online verfügbar unter: https://github.com/josiasbruderer/bbs-for-independence/tree/main/03_workspace/states/state_OCTIS

Kategorie einzuschätzen. Mit den *Worthäufigkeiten* wird eine relativ simple Analyseart gewählt um einen rudimentären Vergleich des Textkorpus mit der Declaration zu ermöglichen. Die *Entitäten* ähnelt den *Worthäufigkeiten*, ermöglichen allerdings eine Analyse die sich auf bestimmbare Einheiten (Personen, Organisationen, Länder etc.) konzentriert. Die *Scattertext* Visualisierung vereinfacht den grafischen Vergleich der *Worthäufigkeiten*. Schliesslich bietet die *Topic Diversity* Analyse eine Methode zur Generierung von Sprachmodellen und innerhalb davon die Clusterbildung von Themen.

4.1 Dokumente pro Jahr

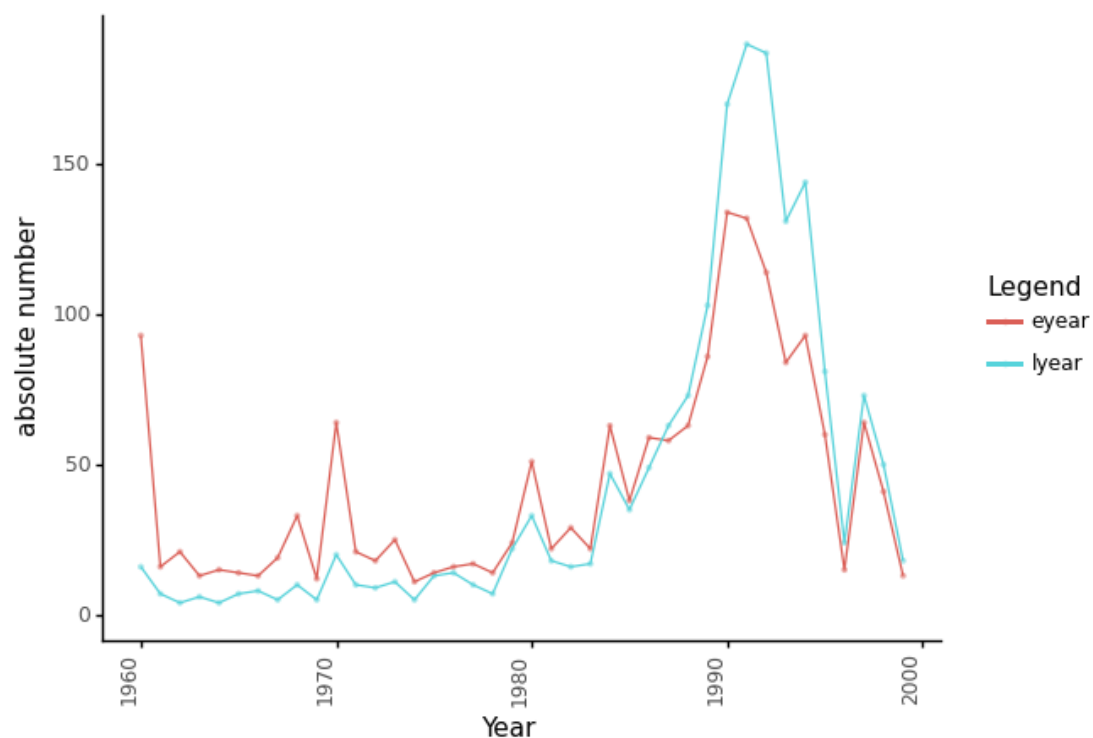


Abbildung 1: Dokumente pro Jahr: Gesamtdatensatz

Die Auswertung der Jahreszahlen der Dokumente spiegelt den Verlauf der Popularität der BBS, analog der in der Literatur beschriebenen Entwicklung. Die beiden Messmethoden korrelieren stark ($r = 0.91$), wobei die Methode *lyear* einen deutlicheren Verlauf zeigt. In der manuellen Untersuchung der Top100 hat sich *lyear*

als genauer herausgestellt, was auch in dieser Auswertung zuzutreffen scheint. Diese Auffallend vielen Textdateien zwischen ca. 1988 und 1995 werden auch die anderen Ergebnisse beeinflussen.

4.2 Kategoriengrösse

Die Kategorisierung der einzelnen Textdateien wurde auf textfiles.com bereits vorgenommen. In diese Kategorien wird auch der ausgewertete Textkorpus gegliedert, wobei die Kategorie «sex» knapp 40% des Textkorpus ausmacht. Auch Politics (14%) und Occult (7%) kommen oft vor. Die restlichen 40% sind auf 36 Kategorien aufgeteilt. Die Betrachtung der Gesamtzeichenanzahl anstelle der Anzahl Dateien hat keinen Einfluss auf die Reihenfolge.

4.3 Worthäufigkeiten

Die Worthäufigkeiten bestätigen, was in der Kategoriengrösse beobachtet wurde. Während im Datensatz der *Declaration* Wörter wie *world, government, people, computer, system, abuse, technology* (Auswahl aus *Frequency* > 10) vorkommen, sind es beim Gesamtdatensatz *cock, body, people, pussy, ass, girl, woman, tongue, suck, fuck* (Auswahl aus *Frequency* > 7000). Wörter, die der Kategorie «sex» zuzuordnen sind, überwiegen. Ausser dem Wort *People* stimmen keine weiteren sinntragenden Wörter in den häufigsten Wörter der beiden Auswertungen überein. Die Zusammenstellung der Worthäufigkeiten des Gesamtdatensatzes ohne die Kategorie «sex» enthält Wörter wie *people, system, world, life, computer, power, government, law, information, change, control, game* (Auswahl aus *Frequency* > 3000). In dieser Auswertung können Wörter gefunden werden, die mit der Auswertung der *Declaration* übereinstimmen: *computer, government, life, people, system, world etc..*

Die Anzahl Dateien in denen ein jeweiliges Wort vorkommt erscheint wenig hilfreich zu sein. Die häufigsten Wörter tragen keinen bestimmaren Sinn, z.B.

time, start, year, danach sind die Resultate auch in dieser Auswertung durch die Kategorie «sex» verzerrt.

4.4 Entitäten

In der Auflistung der Entitäten fällt auf, dass in der Gesamtauswertung Namen sehr hoch in der Rangliste stehen, gefolgt von Bezeichnungen für den amerikanischen Staat und deren Organisationen (FBI, CIA, NSA). Dies kann auch bei der *Declaration* beobachtet werden, allerdings ist ein Vergleich aufgrund der unterschiedlichen Textmengen kaum sinnvoll. Die Analyse ohne die Kategorie «sex» weist deutlich weniger Namen auf. Spannender ist das Verhältnis von ausgewählten Begriffen zu der Häufigkeit mit der FBI (399) oder CIA (486) vorkommt: *Apple* (325), *Constitution* (270), *IBM* (166), *EFF* (80), *Electronic Frontier Foundation* (30), *Fox* (30), *Neidorf* (36), *Radio Shack* (37), *Democratic* (37), *Cyberspace* (29), *Declaration of Independence* (23)

4.5 Scattertext

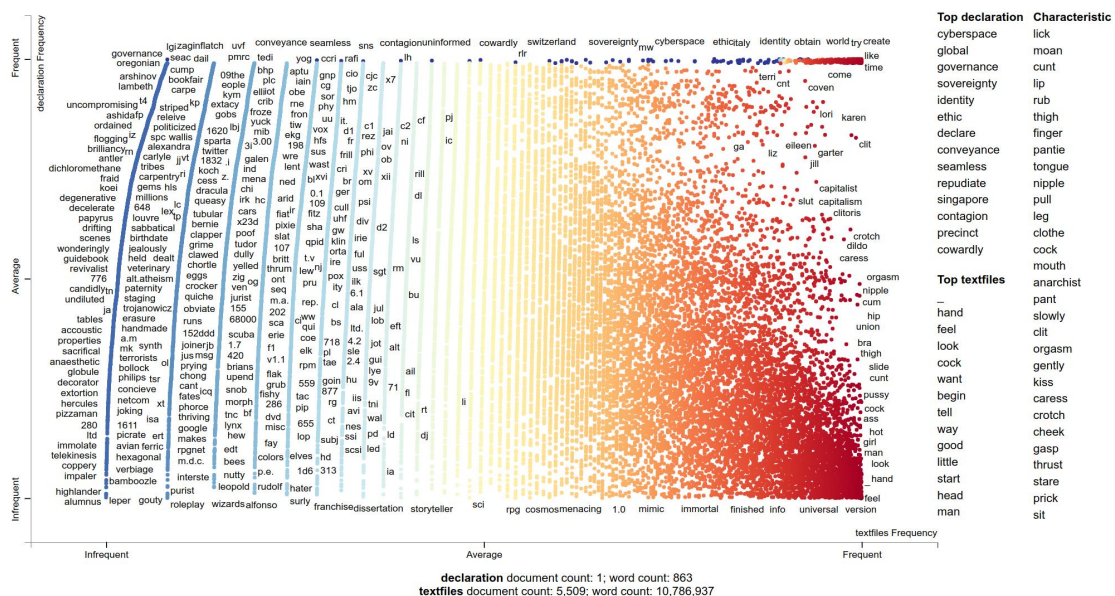


Abbildung 2: Scattertext Visualisierung: Gesamtauswertung zu Declaration

Die Scattertext Visualisierung zeichnet erneut ein recht deutliches Bild. Zwischen der *Declaration* und den BBS-Textfiles ist kein diagonaler Trend und somit keine oder kaum eine Korrelation erkennbar.¹⁰ Eine Auswertung ohne die Kategorie «sex» zeigt ein ähnliches Resultat. Wird die Scattertext Visualisierung für den Gesamtdatensatz, verglichen mit dem Datensatz *Declaration* durchgeführt, kann nach wie vor keine starke Korrelation festgestellt werden, aber immerhin zeigt sich ein leicht diagonaler Trend. : evtl. auch Grafik ersetzen.

4.6 Topic Diversity

Mittels Topic Diversity Analysis wird einerseits der Gesamtdatensatz untersucht und zum Vergleich dazu der Datensatz der *Declaration*. Zur Auswertung wird nach einer Iteration gesucht, bei der die Topic Diversity hoch ist und gleichzeitig die Anzahl Topics ca. 10 beträgt. Insgesamt werden vier Experimente durchgeführt: Je Datensatz (*Declaration* und Gesamtdatensatz) einmal mittels LDA (Srivastava und Sutton 2017) und einmal mittels NeuralLDA (Blei u. a. 2003).

Wie zu erwarten können in den Experimenten des *Declaration* Datensatzes Topics gefunden werden, die sich um die Schwerpunkte der *Declaration* drehen. Die Topic Diversity ist bei LDA eher niedrig (0.3375) während bei NeuralLDA eine hohe (0.8857) festgestellt werden kann. Dies spiegelt sich auch in den Schlagwortwolken.¹¹ Während LDA eine ziemlich wortgetreue Abbildung der *Declaration* erstellt, ist diese bei NeuralLDA zwar sinngemäss aber nicht wortgetreu.

¹⁰ Zur Interpretation von Scattertext siehe: <https://towardsdatascience.com/interpreting-scattertext-a-seductive-tool-for-plotting-text-2e94e5824858>

¹¹ Die Abbildungen der Topics sind zu finden in Anhang 5: OCTIS Topic Diversity



Abbildung 3: Gesamtdatensatz: Topic mittels LDA (Iteration: 3, Run: 0, Topic: 0)

Beim Gesamtdatensatz können mittels LDA ebenfalls Experimente mit eindeutigen Topics gefunden werden (eine davon in Abbildung X). Die Topic Diversity (0.3375) befindet sich in einem ähnlichen Bereich wie bei der *Declaration*. Mittels NeuralLDA können für den Gesamtdatensatz keine Experimente durchgeführt werden, die sinnvolle Topics generieren.

5 Fazit (not ready)

Die Schwierigkeit, dass die grosse Datenmenge kaum auf dem Notebook zu bewältigen ist wurde berücksichtigt. Bei der Programmierung wurde deshalb berücksichtigt, dass Abschnittweise essentielle Variablen zwischengespeichert werden und damit ganze Vorgänge übersprungen werden konnten. So konnten rechenintensive Operationen auf einem Server mit entsprechenden Ressourcen durchgeführt werden.

Dial the system at (312) 528-7141. When you hear the answer tone, connect your modem, and press return several times until the system

determines your transmission rate. If you have a reasonably good telephone line, and a moderately strong modem signal, you should be able to communicate with the system.

Ward Christensen und Randy Suess, November 1978, BYTE Magazin

Evtl. Bezug zu Kapitel «Emergent Social Meanings» in Jones Steven, *Cybersociety 2.0* 1998

«Wir müssen abstrakt-künstlichen Räumen vertrauen, sonst funktionieren diese nicht.» (Fassler 2008)

«Independent media must, of necessity, have an audience. Moreover, and crucially, independent media must find the right kind of audience—one that is committed to the industrial, aesthetic, ethical and sociopolitical ideals of that media and that is enlisted in the discursive struggle over its meaning and value. This often links independent media to particular subcultures, such as gaming, discussed by Hector Postigo in Chapter 9, which demonstrates how independence is often elided with “alternative” media and audiences (discussed further later).» (Bennett 2015: 5)

«Where companies resist, such as Snapchat’s decision to turn down a reported £3bn offer from Facebook, they often find themselves powerless in the face of corporate power’s ability to utilize the very freedoms—as in the malleability and openness of computer code—that underpin digital work cultures: thus Facebook and Twitter quickly built Snapchat-like features into their platforms once their offers had been turned down. And yet, as all these examples also suggest, hope must live on because the utopian desire of media independence is something that cannot be regulated, purchased or otherwise bullied and co-opted out of existence. The desire to create a better way of living through the media is one that motivates many who work in, regulate, finance and study the media. As a utopian promise it remains a vision to be pursued. We hope that this collection helps those interested in studying and making media pursue that vision and that it stimulates further work and debate on this critical area of inquiry. Media independence, after all, matters.» (Bennett 2015: 26)

6 Literatur

BBS: The Documentary, 2005.

Bender, E.M., Koller, A., 2020. Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Gehalten auf der Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, S. 5185–5198. <https://doi.org/10.18653/v1/2020.acl-main.463>

Bennett, J., 2015. Introduction - The Utopia of Independent Media: Independence, Working with Freedom and Working for Free, in: Bennett, J., Strange, N. (Hrsg.), Media independence: working with freedom or working for free?, Routledge research in cultural and media studies. Routledge, Taylor & Francis Group, New York ; London, S. 117–136.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3, 993–1022.

Britannica Academic, 2021. bulletin-board system (BBS) [WWW Document]. *Encyclopædia Britannica*. URL <http://academic.eb.com/levels/collegiate/article/627505> (zugegriffen 21.07.2021).

Britannica Academic, 2016. Electronic Frontier Foundation (EFF) [WWW Document]. *Encyclopædia Britannica*. URL <http://academic.eb.com/levels/collegiate/article/627505> (zugegriffen 21.07.2021).

Christensen, W., Suess, R., 1978. Hobbyist Computerized Bulletin Board. *Byte Magazine* Volume 03.

Dvorak, J., Anis, N., 1990. Dvorak's guide to PC telecommunications. Osborne McGraw-Hill, Berkeley.

Fassler, M., 2008. Cybernetic Localism: Space, Reloaded, in: Döring, J., Thielmann, T. (Hrsg.), *Spatial Turn*. transcript Verlag, S. 185–218. <https://doi.org/10.14361/9783839406830-008>

Gilbertson, S., 2010. Feb. 16, 1978: Bulletin Board Goes Electronic. *Wired*.

- Kreiss, D., 2015. A Vision of and for the Networked World - John Perry Barlow's A Declaration of the Independence of Cyberspace at Twenty, in: Bennett, J., Strange, N. (Hrsg.), *Media independence: working with freedom or working for free?*, Routledge research in cultural and media studies. Routledge, Taylor & Francis Group, New York ; London, S. 117–136.
- Levitas, R., 2010. *The concept of utopia*. Peter Lang, Oxford.
- Levitas, R., 2003. Introduction: The Elusive Idea of Utopia. *History of the Human Sciences* 16, 1–10. <https://doi.org/10.1177/0952695103016001002>
- Mützel, S., Saner, P., Unternährer, M., 2018. Schöne Daten! Konstruktion und Verarbeitung von digitalen Daten, in: Houben, D., Prietl, B. (Hrsg.), *Datengesellschaft*. transcript Verlag, S. 111–132.
- Pavlus, J., 2019. Machines Beat Humans on a Reading Test. But Do They Understand? [WWW Document]. *Quanta Magazine*. URL <https://www.quantamagazine.org/machines-beat-humans-on-a-reading-test-but-do-they-understand-20191017/> (zugegriffen 25.08.2021).
- Quartermann, J.S., 1990. *The matrix: computer networks and conferencing systems worldwide*. Digital Press, Bedford, Mass.
- Srivastava, A., Sutton, C., 2017. Autoencoding Variational Inference For Topic Models. arXiv:1703.01488 [stat].
- Steinbicker, J., 2019. Überwachung und die Digitalisierung der Lebensführung, in: Stempfhuber, M., Wagner, E. (Hrsg.), *Praktiken der Überwachen: Öffentlichkeit und Privatheit im Web 2.0*. Springer VS, Wiesbaden, S. 79–96.
- Terragni, S., Fersini, E., Galuzzi, B.G., Tropeano, P., Candelieri, A., 2021. OCTIS: Comparing and Optimizing Topic models is Simple!, in: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*. Association for Computational Linguistics, Online, S. 263–270.
- Turner, F., 2006. *From counterculture to cyberculture: Stewart Brand, the Whole Earth Network, and the rise of digital utopianism*. University of Chicago Press, Chicago.

Vajjala, S., Majumder, B., Gupta, A., Surana, H., 2020. Practical natural language processing: a comprehensive guide to building real-world NLP systems, First edition. ed. O'Reilly, Beijing Boston Farnham Sebastopol Tokyo.

7 Andere Quellen

Barlow, J.P., 1996. A Declaration of the Independence of Cyberspace [WWW Document]. Electronic Frontier Foundation. URL <https://www.eff.org/cyberspace-independence> (zugegriffen 21.07.2021).

Jason Scott, o. J. Jason Scott's Top 100 Textfiles [WWW Document]. T E X T F I L E S. URL <http://textfiles.com/100/> (zugegriffen 16.05.2021).

8 Anhang

8.1 Anhang 1: Manuelle Untersuchung der Top100

8.2 Anhang 2: Sourcecode Top100

8.3 Anhang 3: Textfiles.com Analyse

8.4 Anhang 4: Sourcecode Textfiles.com

8.5 Anhang 5: OCTIS Topic Diversity



Abbildung 4: Declaration_T_LDA_Iteration4-Run1-Topic1.png



Abbildung 5: Declaration_T_NeuralLDA_Iteration3-Run2-Topic11.png



Abbildung 6: Categories_T_LDA_Iteration3-Run0-Topic0.png



Abbildung 7: Categories_T_LDA_Iteration3-Run1-Topic1.png