

Exposé zur HSA

Seminar: The ABC of Computational Text Analysis

Josias Bruderer

06. Mai 2021

1 Vom Bulletin Board System zur «Declaration of the Independence of Cyberspace»

1.1 Untersuchungsgegenstand

In den frühen Jahren des Internets war der Informationsaustausch und die Kommunikation nicht gleich einfach gestaltet wie heute. Trotzdem wurde in der Vernetzung ein grosses Potential erkannt und technoutopische Communities entstanden. Ein Mittel zum Austausch waren Bulletin Board Systeme (BBS), Server auf denen jede und jeder, mit Zugriff zu einem am Telefonnetz angeschlossenen Computer, Texte hoch- und herunterladen konnten.¹ Diese Hauptseminararbeit soll die BBS als Kommunikationsmittel untersuchen und sich auf Literatur aus der Soziologie und Natural Language Processing (NLP) stützen. Der Fokus soll allerdings auf der empirischen Untersuchung des Datenmaterials liegen.

1.2 Fragestellung

Im Mittelpunkt steht die Frage, *inwiefern die BBS eine Rolle für die Entwicklung und den Austausch des Zeitgeistes spielte, der sich in der Community rund um die 1996 formulierte «Declaration of the Independence of Cyberspace»², wiederfinden lässt*. Diese Frage kann in folgende Teilaspekte gegliedert werden:

- Untersuchung der «Declaration of the Independence of Cyberspace» (und voraussichtlich Texte/Literatur) um geeignete Wörter, Aussagen oder Strukturen zu finden, die den «Zeitgeist» spiegeln
- Finden eines geeigneten Datensatzes, der das Geschehen der BBS spiegelt
- Deskriptive Untersuchung des gewählten Datensatzes um einen Überblick zu erlangen
- Untersuchung des gewählten Datensatzes anhand der Faktoren, die auf den «Zeitgeist» hinweisen

1.3 Relevanz und persönliches Interesse

Die vorgeschlagene Untersuchung hat insofern Relevanz, da die BBS als Vorgänger von Foren, Chatrooms und Mailinglists gesehen werden können. Sie waren eine Kommunikationsplattform die Anfangs aufgrund der Komplexität (und Verfügbarkeit vom Computer) grundsätzlich von «Insidern» verwendet wurde und später, mit Aufkommen des Personal Computers, der breiten Masse zur Verfügung stand. Ich sehe dabei Ähnlichkeiten zu den sozialen Medien, wobei ich vor allem folgende Parallele spannend finde: Die «Declaration of the Independence of Cyberspace» als eine Antwort auf die Kommerzialisierung,

¹ Die Funktionen solcher BBS waren weitreichender. Für die in dieser Arbeit geplante Untersuchung stellt diese Funktion allerdings die wichtigste dar.

² Verfasst von der Electronic Frontier Foundation (EFF)

Monopolisierung und staatliche Kontrolle als ähnliches Ereignis wie die Diskussionen rund um Datenschutz und Privatsphäre.

1.4 Vorabklärung

Zur empirischen Untersuchung kann der Datensatz dienen, der von Jason Scott³ auf textfiles.com bereitgestellt wird. Dieser fokussiert sich auf die Jahre 1980-1995. Der Datensatz umfasst über 58'000 Dateien, alle als ASCII-Text vorliegend. Eine grobe Kategorisierung ist vorhanden. Ansonsten sind die Dateien relativ formlos und bedürfen der Aufbereitung und Bereinigung. Ebenfalls ist eine vorgängige «Durchforstung» des archive.org innerhalb des gewählten Zeitrahmens denkbar, um einen besseren Überblick über mögliche Datenquellen zu erhalten.

Zur technischen Realisation können die im Seminar angeschauten Python-Libraries *spacy* und *textacy* eingesetzt werden. Ebenfalls kann zum Gewinnen eines Überblick über den Datensatz auf den Ansatz des Topic Modellings (z.B. mit *octis*) gesetzt werden. Grundsätzlich soll die Untersuchung mit Python umgesetzt werden.

1.5 Vorwissen

Zu einer ähnlichen Thematik schrieb ich vor einem Jahr bei Markus Unternährer eine Proseminararbeit (Titel: Web 2.0, eine Pleite?). An diese möchte ich in gewisser Weise anschliessen. Die Proseminararbeit behandelte die Jahre nach 1996, Ereignisse verbunden dem Begriff Web 2.0 und die Übernahme Whats-Apps durch Facebook als Fallbeispiel.

³ Amerikanischer Archivar, Historiker und Betreiber von textfiles.com sowie Mitwirkender bei archive.org.