

# Jason Scott's favorite 100 - data analysis using NLP

Mini-Project in the ABC of Computational Text Analysis

Josias Bruderer

14. Mai 2021

Abstract

Inhaltsverzeichnis

1	Einleitung	1
2	Datensatz	1
3	Methode	1
4	Resultate	1
5	Diskussion	1
6	Fazit	1

## 1 Einleitung

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 technologische 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.<sup>1</sup> Jason Scott<sup>2</sup> unterhält ein Archiv mit einem Umfang von 58'000 solcher Text-Dateien.

Als Vorbereitung für die Hauptseminararbeit<sup>3</sup>, in welcher dieses Archiv zur Beantwortung der Frage nach dem Einfluss der BBS auf die 1996 formulierte «Declaration of the Independence of Cyberspace»<sup>4</sup> dienen soll, wird in diesem Mini-Project<sup>5</sup> der verkleinerte Datensatz «favorite 100» untersucht. Konkret soll geprüft werden, ob inhaltliche Auffälligkeiten bestehen, die möglicherweise Rückschlüsse auf Scotts Interessen und seine Biografie zulassen. Dieses Mini-Project soll ebenfalls zur Einschätzung dienen, wie aufwändig die Bereinigung des gesamten Textkorpuses ( $N = 58'0000$  Dateien) sein wird.

## 2 Datensatz

- Der Datensatz ist aufrufbar unter: <http://textfiles.com/100/>
- Der Datensatz kann heruntergeladen werden unter: <http://archives.textfiles.com/100.zip>

## 3 Methode

## 4 Resultate

## 5 Diskussion

## 6 Fazit

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

<sup>2</sup> Amerikanischer Archivar, Historiker und Betreiber von textfiles.com sowie Mitwirkender bei archive.org.

<sup>3</sup> Link zum Exposé: <https://ipfs.chixodo.xyz/ipfs/QmcU2Xbwmd7hhRxhMaQPgiCtAhWajfAwbaCREEzschFPE>

<sup>4</sup> Verfasst von der Electronic Frontier Foundation (EFF)

<sup>5</sup> Der Sourcecode ist verfügbar unter: <https://git.makersphere.io/josias/ch.josias.jason-scotts-favorite-100>