

Artikel aus:

Zeitschrift für digitale Geisteswissenschaften

Titel:

Vorstellung eines (teil-)automatisierten Verfahrens zur Analyse der Multimodalität von Webseiten

Autor*in:

Thomas Jurczyk

Kontakt:

thomas.jurczyk-q88@rub.de

Institution:

Ruhr-Universität Bochum, Centrum für Religionswissenschaftliche Studien (CERES)

GND:

[106744744X](#)

ORCID:

[0000-0002-5943-2305](#)

DOI des Artikels:

[10.17175/2021_003](#)

Nachweis im OPAC der Herzog August Bibliothek:

[ppn 1749189682](#)

Erstveröffentlichung:

09.09.2021

Lizenz:

Sofern nicht anders angegeben



Medienlizenzen:

Medienrechte liegen bei den Autor*innen

Letzte Überprüfung aller Verweise:

23.08.2021

GND-Verschlagwortung:

[Cluster-Analyse](#) | [Maschinelles Lernen](#) | [Multimodalität](#) | [Sozialwissenschaften](#) | [Website](#) |

Zitierweise:

Thomas Jurczyk: Vorstellung eines (teil-)automatisierten Verfahrens zur Analyse der Multimodalität von Webseiten. In: Zeitschrift für digitale Geisteswissenschaften. Wolfenbüttel 2021. PDF Format ohne Paginierung. Als text/html abrufbar unter DOI: [10.17175/2021_003](#).

Thomas Jurczyk

Vorstellung eines (teil-)automatisierten Verfahrens zur Analyse der Multimodalität von Webseiten

Abstracts

Der vorliegende Artikel möchte ein Verfahren zur (teil-)automatisierten Analyse der Multimodalität von Webseiten vorstellen und diskutieren. Dabei steht im Fokus, unbekannte Webseiten auf deren Multimodalität hin zu untersuchen, ohne dass diese vorher annotiert oder sonst anderweitig in Bezug auf ihre Multimodalität analysiert worden wären. Zusätzlich zur Etablierung eines (teil-)automatisierten Verfahrens möchte dieser Artikel die Frage diskutieren, inwiefern die Klassifizierung der Multimodalität von Webseiten im Zusammenhang mit den gesellschaftlichen Domänen steht, aus denen die Webseiten stammen (Politik, Wirtschaft, Religion etc.).

This article presents and discusses a method for the (partially) automated analysis of the multimodality of web pages. The focus lies on analyzing unknown web pages for their multimodality without annotating them beforehand. In addition to establishing a (partially) automated procedure, this article wants to discuss the question to what extent the classification of the multimodality of web pages is related to the social domains from which the web pages originate (politics, economy, religion, etc.).

1. Einleitung

Unter Multimodalität wird in diesem Beitrag die Anordnung und das Zusammenwirken verschiedener *modes* wie Bilder, Ton und Schrift im Prozess der Generierung von Zeichen verstanden.¹ Im Falle von Webseiten werden insbesondere Bilder, verschiedentlich formatierte Textelemente sowie audio-visuelle Elemente wie Videos untersucht.

Aufgrund der Komplexität multimodaler Strukturen, die teils sehr diverse modes miteinander kombinieren, konzentriert sich die Forschung zum Beispiel im Bereich der Bildlinguistik oder im Kontext der Analyse von Webseiten zumeist auf eine qualitativ-manuelle Analyse des Zusammenspiels der verschiedenen modes, wobei die Ergebnisse der Einzelanalysen dann durchaus quantitativ synthetisiert für allgemeinere Aussagen herangezogen werden.² Um die Komplexität der Multimodalität für den Prozess der Zeichengenerierung beispielsweise auf Webseiten adäquat zu besprechen, ist eine manuell-qualitative Untersuchung durchaus sinnvoll, da deren Automatisierung nur sehr schwer realisierbar ist; nicht zuletzt, weil die menschliche Perzeption und Einschätzung in diesem Prozess eine zentrale Rolle spielen und Aspekte wie der initiale Eindruck einer Sehfäche stark vom jeweiligen Betrachter abhängen.³

Das hier vorzustellende Analyseverfahren möchte trotz dieser Schwierigkeiten eine Möglichkeit aufzeigen, wie zumindest grobe Strukturmerkmale der Multimodalität von Webseiten (teil-)automatisiert untersucht und zusammengefasst werden können. Ein solches quantitatives, (teil-)automatisiertes Verfahren steht dabei keinesfalls im Widerspruch zu den etablierten manuell-qualitativen Zugängen, sondern möchte diese lediglich um eine neue Perspektive ergänzen sowie in ihrer Durchführung unterstützen. Beispielsweise könnte ein etabliertes automatisiertes Verfahren künftig dazu genutzt werden, größere Datenmengen (in diesem Falle von Webseiten) hinsichtlich ihrer multimodalen Struktur vorzufiltern, um damit eine gezieltere qualitativ-manuelle Analyse einzelner Fallbeispiele aus den unterschiedlichen Clustern zu ermöglichen. Darüber hinaus kann eine automatisierte Analyse und Kategorisierung auch dazu verwendet werden, bestehende Klassifizierungen aus der qualitativ-manuellen Forschung zu kontrastieren, empirisch zu untermauern oder zu erweitern.⁴

Versuche eines computergestützten Vorgehens bei der Analyse multimodaler Strukturen finden sich bei O'Halloran et al.⁵ Allerdings fokussieren sich die dort vorgestellten Ansätze auf die (automatisierte) Analyse kürzerer Sequenzen (beispielsweise eines Interviewausschnitts) oder sie arbeiten mit bereits annotierten oder anderweitig vorstrukturierten Daten (siehe auch Kapitel 2).

Zusätzlich zur Etablierung eines (teil-)automatisierten Verfahrens möchte dieser Artikel die Frage diskutieren, inwiefern die Klassifizierung der Multimodalität von Webseiten im Zusammenhang mit den gesellschaftlichen Domänen steht, aus denen die Webseiten stammen (Politik, Wirtschaft, Religion, Universität, Journalismus, Unterhaltung). Es wäre zu fragen, ob bestimmte multimodale Strukturen von Webseiten bereits unabhängig von der konkreten Semantik ihrer verwendeten modes (beispielsweise eines dargestellten Kreuzes als bildliches Element) zur domänenspezifischen Zeichengenerierung genutzt werden können. Zugespielt gefragt: Gibt es politische, wirtschaftliche oder religiöse multimodale Strukturen auf Webseiten, die voneinander unterscheidbar sind, ohne die Semantik der jeweils konkret verwendeten modes einzubeziehen?

¹ Lyons 2016, 268f.; Stöckl 2016, S. 4f. In Bezug auf modes vgl. Kress 2017.

² Vgl. Stöckl 2011; Meer / Pick 2019; Jewitt (Hg.) 2017. Für Webseiten vgl. Meier 2016; Freudenberg / Sharbat Dar [im Erscheinen].

³ Vgl. die komplexe Terminologie einer visuellen Grammatik in Stöckl 2011, S. 52. Für die erste Wahrnehmungsebene einer Sehfäche vgl. Meer / Pick 2019, S. 60.

⁴ Vgl. Meier 2016, S. 429f.

⁵ O'Halloran et al. 2017.

Im Folgenden wird in einem ersten Schritt (Kapitel 2) nochmals detaillierter auf das Forschungsinteresse dieses Artikels eingegangen. Hierbei richtet sich der Blick auf aktuelle Forschungen im Bereich Multimodalität von Webseiten und deren potentielle Ergänzung durch ein (teil-)automatisiertes Verfahren zur Analyse unbekannter Datensätze. Anschließend werden die Methode und der Aufbau des Programms beschrieben, das die (teil-)automatisierte Analyse und Kategorisierung der Webseiten in diesem Artikel durchführen soll (Kapitel 3). Auf Basis dieses Programms wird im Analyseteil (Kapitel 4) ein erster Testdatensatz von Webseiten automatisiert analysiert und kategorisiert, wobei die Untersuchung und der Nachvollzug der Ergebnisse im Fokus stehen. Im abschließenden letzten Kapitel (5) werden die Ergebnisse des (teil-)automatisierten Analysevorgangs diskutiert und evaluiert. In diesem Zusammenhang stehen vor allem die Fragen nach einer zukünftigen Ausweitung der Untersuchung auf größere Datenmengen sowie der Mehrwert einer (teil-)automatisierten Vorgehensweise gegenüber klassischen manuell-qualitativen Analysemethoden im Zentrum.

2. Forschungsfrage und Forschungsstand

Die Forschungsfrage dieses Artikels betrifft die Möglichkeit einer automatisierten Auswertung und Klassifizierung unbekannter Webseiten auf Basis ihrer multimodalen Eigenschaften. Ansätze in Richtung einer automatisierten Auswertung multimodaler Strukturen finden sich unter anderem bei O'Halloran et al.⁶ und auf der Webseite des [Multimodal Analysis Lab](#).⁷

Obwohl sich die Multimodalitätsforschung nicht auf Webseiten beschränkt, sondern an unterschiedlichsten Medien und Formen der Multimodalität interessiert ist,⁸ konzentriert sich dieser Artikel auf Webseiten. Webseiten haben gegenüber Werbeplakaten, gedruckten Zeitungen oder Filmen den Vorteil, dass sie bereits in einer digitalen und in Bezug auf ihre Multimodalität vorstrukturierten Form vorliegen. So sind beispielsweise eingebundene Bilder auf Webseiten meist durch die Nutzung von entsprechenden Tags wie `` gekennzeichnet, die sich relativ leicht automatisiert identifizieren und mit entsprechenden Parsern auslesen lassen. Selbiges gilt für Texte oder die Einbindung von Videos auf Webseiten. Somit entfallen im Falle von Webseiten die ansonsten notwendigen und zeitaufwendigen Schritte der Digitalisierung und anschließenden Vorstrukturierung der Digitalisate, beispielsweise durch Annotationen.

Darüber hinaus sind Webseiten sehr populär und finden sich in fast allen gesellschaftlichen Bereichen (Wirtschaft, Religion, Wissenschaft, private Seiten usw.). Diese Tatsache erlaubt es, im Kontext der Analyse und Kategorisierung der Multimodalität von Webseiten die Frage zu stellen, inwieweit Cluster von Webseiten mit ähnlichen multimodalen Strukturen mit den gesellschaftlichen Feldern korrelieren, aus denen sie stammen. Gibt es beispielsweise eine bestimmte Form von Multimodalität auf politischen Webseiten, die sich von der religiösen Webseiten unterscheidet, sodass diese nicht nur auf inhaltlicher, sondern auch auf formaler Ebene unterscheidbar sind? Oder lassen sich multimodal-strukturelle Muster über die Grenzen gesellschaftlicher Domänen hinweg verfolgen, sodass etwaige inhaltliche Unterschiede zwischen den Seiten (Semantik) von einer gemeinsamen multimodalen Struktur (Form) überdeckt werden, die einem übergreifenden Webseiten-Typ entsprechen? Um diese Frage(n) zu beantworten, wurden die Webseiten aus dem Testdatensatz zum einen hinsichtlich der gesellschaftlichen Domäne, aus der sie stammen, sowie auch des ihnen zuweisbaren Webseiten-Typs nach Stefan Meier⁹ vorannotiert.¹⁰

Im Folgenden soll eine Methode vorgestellt werden, wie ein solches Programm zur (teil-)automatisierten Analyse von Multimodalität aussehen könnte. Die (teil-)automatisierte Analyse wird im Kontext dieses Artikels mit der Programmiersprache Python und durch eine Mischung aus selbstgeschriebenen Klassen und interaktiver Arbeit mit *Jupyter Notebooks* implementiert. Hierbei handelt es sich jedoch nur um ein Implementierungsbeispiel. Im Fokus steht die übergreifende Methode und nicht deren konkrete Realisierung. Die hinter dem Programm stehende Methode sollte auch ohne jede Programmierkenntnisse verständlich werden und entsprechende Passagen in diesem Artikel, die Code-Beispiele enthalten, können ohne große Verluste übersprungen werden. Auch ist das hier vorgestellte Python-Programm nicht als Prototyp einer zukünftigen Softwarelösung misszuverstehen, sondern wird nur im Kontext dieses Artikels verwendet, um die Analysen durchzuführen, und zu Dokumentationszwecken auf meinem [GitHub Account](#) bereitgestellt.¹¹

3. Methode und Implementierung

Die Methode der (teil-)automatisierten Analyse multimodaler Strukturen von Webseiten besteht aus folgenden Schritten, die in Form eines in Python geschriebenen Programms implementiert und ausgeführt werden.¹²

⁶ Vgl. O'Halloran et al. 2017; Wignell et al. 2021; O'Halloran et al. 2020.

⁷ Multimodal Analysis Lab (Hg.) 2013.

⁸ Siehe die Beispiele in Jewitt (Hg.) 2017.

⁹ Meier 2016, S. 429f.

¹⁰ Eine solche Vorannotation scheint auf den ersten Blick dem gerade geforderten Prinzip zuwider zu laufen, dass das Verfahren auf die Analyse unbekannter und eben gerade nicht vorannotierter Daten konzentriert ist. Dies ist aber nicht der Fall, denn bei der Domänen-/Typen-Analyse handelt es sich um eine zusätzliche Fragestellung, die unabhängig vom dem eigentlichen Verfahren der Teilautomatisierung zu betrachten ist und als Beispiel dazu dienen soll, wozu (teil-)automatisierte Verfahren konkret genutzt werden können.

¹¹ Jurczyk 2021a.

¹² Siehe das [GitHub Repository](#), Jurczyk 2021a.

Die hier vorgeschlagene Methode zur Analyse der Multimodalität von Webseiten beschränkt sich auf die Startseiten der Webseiten (sogenannte *landing pages*). Eine solche Beschränkung ist sinnvoll, weil sie nicht nur die Operationalisierbarkeit stark vereinfacht, sondern in Anlehnung an die Sehflächen aus dem Bereich der Bildlinguistik¹³ den Schwerpunkt der Analysen auf den multimodalen Ersteindruck der Seiten legt. Es geht also nicht darum, die Komplexität einzelner Webseiten unter Einbezug ihrer Unterseiten und deren multimodaler Struktur zu untersuchen, sondern den Bereich in den Fokus zu stellen, der sich den Nutzer*innen beim erstmaligen Betreten der Seite eröffnet.

Weil diese Art der Erstwahrnehmung der Webseiten wiederum stark von den verwendeten Endgeräten abhängt, mit denen sie geöffnet werden, ist es wichtig, zu Beginn die Analysemodalitäten anzugeben. Die folgenden Untersuchungen wurden mit einem Desktop Computer auf einem relativ großen Bildschirm durchgeführt (32 Zoll Bildschirmdiagonale). Somit muss während des Nachvollzugs der hier vorgenommenen Auswertungen beachtet werden, dass sich die multimodalen Strukturen der Webseiten bei der Verwendung anderer Endgeräte (beispielsweise Mobiltelefone) durchaus von denen auf dem hier verwendeten Desktop Computer unterscheiden können.

Die Definition der ersten Wahrnehmungsebene einer Sehfläche aus der Bildlinguistik wird in diesem Artikel dahingehend ausgeweitet, dass der automatisiert zu untersuchende Bereich den gesamten Inhalt der Startseiten einbezieht. Diese Entscheidung wurde getroffen, weil eine Beschränkung auf den beim Öffnen der Webseiten sichtbaren Bereich, der im Browser dargestellt wird (und der meist deutlich kleiner ist als der Bereich der gesamten Startseite), nur schwer automatisiert vorzunehmen ist und von Bildschirm zu Bildschirm (beziehungsweise Fenstergröße zu Fenstergröße) unterschiedlich ausfallen kann. Somit bezieht die automatisierte Analyse nicht nur den anfänglichen Moment des Öffnens der Seite mit ein, sondern inkludiert auch das Scrollen und somit die Wahrnehmung der gesamten Startseite, was zwar über den Moment des Ersteindrucks hinausgeht, aber den multimodalen Charakter der Webseite dafür adäquater repräsentiert.

Die Methode lässt sich in drei Schritte unterteilen: 1) Corpuserstellung, 2) Datensammlung und -aufbereitung sowie 3) Analyse. Die ersten beiden Schritte der Corpuserstellung und der Sammlung sowie Aufbereitung der multimodalen Daten müssen dabei nur einmal durchgeführt werden, um das Corpus zu initialisieren. Eine erneute Initialisierung des Corpus ist nur dann notwendig, wenn es sich ändert, beispielsweise durch das Hinzufügen weiterer Webseiten, oder eine aktuelle Version der Daten benötigt wird.

Diese drei Schritte werden im Folgenden detailliert diskutiert und anhand einer exemplarischen Implementierung in Python umgesetzt.

3.1 Corpuserstellung

Der erste Schritt besteht aus der Erstellung eines Webseiten-Corpus mit den zu untersuchenden Webseiten. Das Corpus muss in der hier vorzustellenden Realisierung von den Nutzer*innen als CSV-Datei (>websites.csv<) im Ordner des Programms angelegt werden. Im Kontext dieses Artikels enthält die betreffende CSV-Datei die folgenden drei Spalten:¹⁴

URL	Domäne	Typen nach Meier
https://marginalie.hypotheses.org/	UNI	News

Tab. 1: Beispiel eines Eintrags einer Webseite in der Datei >websites.csv<. [Jurczyk 2021]

Die erste Spalte enthält die URLs der Webseiten, die zweite deren gesellschaftliche Domäne (in der hier gezeigten Zeile UNI für Universität)¹⁵ und die Einordnung der Webseite nach den von Stefan Meier¹⁶ vorgeschlagenen Webseiten-Typen in:

- Newsorientierte Webseiten
- Imageorientierte Webseiten
- Kampagnenorientierte Webseiten
- Plattformorientierte Webseiten

Die URLs aus der Datei >websites.csv< werden in das in Python geschriebene und mit einem *Jupyter Notebook* ausgeführte Programm geladen.

¹³ Vgl. Meer / Pick 2019, S. 59f.

¹⁴ Wie bereits angemerkt, erfolgt die Auswertung unabhängig von den Domänen und Typen. Diese können also in einer anderen Implementierung durchaus weggelassen werden, ohne dass das eigentliche Verfahren davon beeinflusst würde. Sie wurden in diesem Artikel hinzugefügt, um die zweite Forschungsfrage (siehe Einleitung Kapitel 2) besser beantworten zu können.

¹⁵ Die gesellschaftlichen Domänen sind Universität / Wissenschaft (UNI), Politik (POL), Religion (REL), Unterhaltung (ENT), Wirtschaft (ECO) sowie Journalismus (JOU) und wurden aus der Sichtung des Materials induktiv generiert. Diese können bei Erweiterung des Datensatzes natürlich ebenfalls erweitert werden, beispielsweise durch Kunst (ART) oder einer allgemeineren Kategorie Bildung (EDU).

¹⁶ Vgl. Meier 2016, S. 429f.

```

1. from scraper import Corp
2.
3. cp = Corpus()
4. cp.initCorpus()

```

Hierzu wird die Corpus-Klasse importiert und als Corpus-Objekt initialisiert. Das Corpus wird daraufhin mit der Methode `initCorpus()` initialisiert. Es reicht aus, das Corpus einmal zu initialisieren. Eine erneute Initialisierung ist nur dann notwendig, wenn das Corpus verändert wurde oder wenn die Nutzer*innen es für sinnvoll erachten, die Daten neu zu erheben. Entsprechend muss dieser Teil des Programms nur einmal ausgeführt werden.

Die `initCorpus()`-Methode iteriert über die in `>websites.csv<` befindlichen URLs und öffnet diese mit *Selenium*¹⁷ in einem Browserfenster auf Bildschirmgröße.¹⁸ Die User*innen sind daraufhin aufgefordert, etwaige aufpoppende Banner beziehungsweise Abfragen manuell zu bestätigen. Sobald dies erledigt ist, kann via Tastendruck das Speichern des Codes der Webseite im *Jupyter Notebook* bestätigt werden. Nachdem automatisiert bis zum Ende der Webseite gescrollt wurde,¹⁹ speichert das Programm für jede Seite den beim Aufruf der Webseite aktuellen HTML-Code in einem separaten Ordner namens `>CorpusData<`. Der Ordner enthält neben dem HTML-Cod

Die gespeicherten Daten dienen erstens der

Dokumentation und werden zweitens im weiteren Programmverlauf für die Extraktion bestimmter multimodaler Eigenschaften wie der verwendeten Textmenge genutzt. Wenn das Corpus neu initialisiert wird, werden alle Daten im Ordner `>CorpusData<` überschrieben.

Warum wird dieser Vorgang lediglich teil- und nicht vollautomatisiert durchgeführt? Der Hauptgrund hierfür liegt im Anspruch des Programms, sehr diverse Webseiten in das Corpus aufnehmen und analysieren zu können. Komplexe Seiten, die Inhalte dynamisch nachladen oder überhaupt erst durch eine Bestätigung der User*innen aufgerufen werden können,²⁰ sind nur sehr schwer ohne Kenntnisse des Webseitenaufbaus automatisiert abrufbar. *Selenium* wird in diesem Artikel gegenüber anderen Bibliotheken wie `requests` der Vorzug gegeben, weil mit *Selenium* die Webseiten in einem Browser geöffnet werden, was das Erscheinungsbild, wie es sich auch den User*innen beim Aufruf der Seite offenbart, akkurat widerspiegelt. Dieser Aspekt ist für die Analyse multimodaler Strukturen äußerst wichtig, da sich beispielsweise die Darstellungen von Bildern abhängig von dem genutzten Device und der Fenstergröße stark unterscheiden können.

3.2 Bilder, Videos und Texte

Nachdem das Corpus initialisiert wurde, muss im nächsten Schritt die `DataPreparation()`-Klasse importiert werden. Diese Klasse dient dazu, um mit Hilfe der Methoden `getImages()` und `createAnalyzerDict()` eine Datei zu erstellen, die alle notwendigen Informationen über den multimodalen Aufbau der Webseiten aus dem Webseiten-Sample enthält (diese Datei wird als `>merged_data_dict.pickle<` in demselben Verzeichnis wie das *Jupyter Notebook* gespeichert). Ähnlich wie im Falle der Corpus-Initialisierung ist es ausreichend, diese Analyse einmalig für jedes Corpus durchzuführen.

```

1. from data_preparation import DataPreparation
2.
3. dp = DataPreparation()
4. dp.getImages()
5. data_dict = dp.createAnalyzerDict()

```

Genau wie im ersten Schritt der Corpus-Initialisierung arbeitet die `getImages()`-Methode ebenfalls mit *Selenium* und erfordert die manuelle Bestätigung der User*innen. Die `getimages()` Methode ruft die einzelnen Webseiten auf und speichert die Bild- und Videoinformationen in einem Byte-File (`>image_data.pickle<`). Die Datei `>image_data.pickle<` muss solange nicht neu erzeugt werden, wie sich das Corpus nicht ändert oder die Daten veraltet sind. Der Vorteil in der Nutzung von *Selenium* liegt erneut darin, dass mit *Selenium* die Maße der aktuellen Darstellung eines Bildes beziehungsweise Videos im Browser abgefragt werden können. Somit erlaubt die Nutzung von *Selenium* die Erstellung einer präzisen Momentaufnahme der multimodalen Struktur einer Webseite.

Im abschließenden Schritt des Aufrufs der `createAnalyzerDict()`-Methode wird das für die Analyse zentrale `>merged_data_dict.pickle<` erstellt. Die Generierung der in dieser Datei enthaltenen Informationen werden vollautomatisiert und auf Basis der in den vorangegangenen Schritten gesammelten Daten durchgeführt. Die Datei `>merged_data_dict.pickle<` enthält die folgenden Informationen (im Datenformat eines Python Dictionary):

¹⁷ Bei Selenium handelt es sich hier um ein Framework, das eine automatisierte Steuerung verschiedener Browser ermöglicht. Vgl. auch die offizielle Webseite unter [SeleniumHQ Browser Automation](#).

¹⁸ Im Falle der hier vorgestellten Implementierung wird der *Edge Browser* verwendet.

¹⁹ Dies ist notwendig, falls gewisse Bereiche der Seite erst nach und nach geladen werden, sobald diese von den User*innen erreicht werden.

²⁰ Beispielsweise, weil zuerst von den Nutzer*innen Einverständniserklärungen zur Nutzung von Cookies usw. abgegeben werden müssen.

- Anzahl der Bilder auf der Webseite²¹
- Anzahl der großen, mittleren, kleinen und sehr kleinen Bilder²²
- Anzahl der Hintergrundbilder
- Anzahl der Videos²³
- Anzahl der großen und kleinen Videos²⁴
- Die Gesamtlänge textlicher Elemente auf der Seite²⁵
- Die Anzahl der Überschriften
- Die Anzahl der Hyperlinks
- Die Anzahl der internen Hyperlinks
- Die Anzahl der externen Hyperlinks

Darüber hinaus enthält die Datei zu Dokumentationszwecken und für die weitere Analyse die vollständigen Daten, darunter die URLs zu den Bildern und Videos, die Domänen und Webseiten-Typen, sowie den kompletten Text der Webseiten, welcher der Berechnung der Textlänge zugrunde liegt.

3.3 Analyse

Der Analyseteil der hier vorzustellenden Methode wurde in der Datei ›analyzer.py‹ realisiert. Die Analyse basiert auf einem *k-Means-Clustering*, das es erlaubt, die Datensätze auf Basis ihrer *Features*²⁶ zu Gruppen mit ähnlichen Eigenschaften zusammenzufassen.²⁷ Die Klasse `Analyzer()` arbeitet mit den Daten, die während des vorangegangenen Schrittes in der Datei ›merged_data_dict.pickle‹ abgespeichert wurden. Auf Basis der in ›merged_data_dict.pickle‹ vorhandenen Daten generiert die Klasse `Analyzer()` außerdem folgende relationale Werte:

Jeweils das Verhältnis zwischen großen / mittleren / kleinen / sehr kleinen Bildern und der Gesamtzahl der Bilder auf einer Webseite

- Das Verhältnis zwischen der Gesamtzahl der Bilder auf einer Webseite und der Textlänge
- Das Verhältnis zwischen großen und mittleren Bildern und der Textlänge
- Das Verhältnis zwischen internen und externen Links
- Das Verhältnis zwischen der Anzahl der Überschriften und der Textlänge

Der Ablauf der Analyse besteht aus folgenden Schritten:

1. In einem ersten Schritt werden die in die Analyse einzubeziehenden Features festgelegt. Standardmäßig sind alle Features ausgewählt. Die Auswahl der Features geschieht über die Methode `setColumnSelection()`. Die aktuell ausgewählten Features können mit der Methode `getColumnSelection()` abgerufen werden.
2. Der zweite Schritt besteht in der Standardisierung der Daten mit Hilfe der scikit-learn `StandardScaler()`-Klasse, die eine Standardisierung der Daten durch Berechnung der *z-scores* ermöglicht.²⁸ Obwohl dieser Schritt optional ist und die Analyse auch mit den nicht standardisierten Daten erfolgen kann, ist eine solche Standardisierung anzuraten, weil es ansonsten zu Cluster-Bildungen während der Analyse kommen kann, die vor allem aus den unterschiedlichen Scales zwischen den Features resultieren.²⁹
3. Anschließend können die ausgewählten (standardisierten) Features mit der Methode `clusterDataKMeans()` analysiert werden. Hierbei ist es möglich, der Methode die Anzahl *k* der gewünschten Cluster zu übergeben (der *default*-Wert ist *k*=3). Die Methode visualisiert anschließend tabellarisch die Zuordnung der Webseiten zu den einzelnen Clustern innerhalb des *Jupyter Notebook*.
4. Mit der Methode `createElbowPlot()` ist es außerdem möglich, sich auf Basis der *inertia* des Clustering-Algorithmus einen sogenannten ›Elbow-Plot‹ anzeigen zu lassen, der dabei helfen kann, die adäquate Anzahl von Clustern für den Datensatz und die gewählte Feature-Auswahl zu ermitteln. Die ideale Anzahl der Cluster lässt sich dadurch ermitteln, das in dem Plot nach der Anzahl *k*-Cluster gesucht wird (*X*-Achse), bei der die Kurve des Plots einen Knick hat und merklich abflacht.³⁰

²¹ SVG-Dateien wurden bewusst nicht mitgezählt, da diese häufig lediglich für sehr kleine graphische Elemente genutzt werden und damit möglicherweise ein fälschliches Bild der Gesamtzahl an Bildern auf einer Webseite vermitteln.

²² Groß: Größer als 700px (Höhe oder Breite). Mittel: Größer als 348px (Höhe oder Breite). Klein: Größer als 35px (Höhe oder Breite). Sehr klein: Größer als 1px (Höhe oder Breite).

²³ Hier wurden nur die auf der Seite befindlichen Video-Tags gezählt und analysiert. Es gibt Webseiten, auf denen Videos anders integriert sind – zumeist werden diese dann aber auch nicht live auf der Seite abgespielt, sondern sind als Links implementiert.

²⁴ Groß: Größer als 700px (Höhe oder Breite). Alle anderen Videos werden als ›klein‹ deklariert.

²⁵ Genauso wie die meisten anderen Abfragen ist auch dieser Wert mit Vorsicht zu behandeln, da nicht immer zuverlässig feststellbar ist, welche Texte wirklich für die Nutzer*innen sichtbar sind. Manche sind eventuell in ausklappbaren Menüs versteckt. Für die Details, welche Maßnahmen getroffen wurden, um möglichst nur den dargestellten Text einzubeziehen, verweise ich auf den Code in meinem [GitHub Repository](#).

²⁶ Unter Features versteht man im Kontext des maschinellen Lernens die Eigenschaften oder Attribute, die für bestimmte Phänomene erhoben werden (hier also Anzahl der Bilder, Textlänge, Anzahl Videos etc.), und deren konkrete Werte (also zum Beispiel 10 Bilder, 2.540 Wörter, 2 Videos etc.) in der Analyse genutzt werden. Die Begrifflichkeiten werden allerdings nicht immer einheitlich verwendet. Vgl. Géron 2019, S. 8.

²⁷ Ein weiteres Beispiel, in dem *k*-Means-Clustering für die Analyse von Multimodalität verwendet wurde, findet sich in O'Halloran et al. 2017. Zur Funktionsweise von *k*-Means-Clustering vgl. u. a. Géron 2019 und mein Clustering Tutorial »Clustering with Scikit-Learn« auf der Webseite [The Programming Historian](#), Jurczyk 2021b.

²⁸ Für die Details der Klasse und Implementierung vgl. die offizielle scikit-learn-Dokumentation auf [Sklearn.preprocessing.StandardScaler](#), scikit-learn developers (Hg.) 2007.

²⁹ Beispielsweise kann es passieren, dass Unterschiede in den Textlängen aufgrund ihrer größeren Differenz plötzlich deutlich ausschlaggebender für die Clusterbildung werden als Unterschiede in der Anzahl der Bilder, obwohl letzteres Feature für das multimodale Arrangement der Webseiten vielleicht deutlich wichtiger ist als die absolute Differenz in der Textmenge.

³⁰ Genauere Erläuterungen zu dieser Methode der Ermittlung der Anzahl an *k*-Clustern für einen Datensatz finden sich in Géron 2019, S. 245f. und in dem bereits erwähnten Tutorial »Clustering with Scikit-Learn« auf der Webseite [The Programming Historian](#), Jurczyk 2021b. Es sei an dieser Stelle erwähnt, dass neben den

5. Anschließend erlaubt es die Methode `getScreenshotsFromClusters()` Screenshots von den in den jeweiligen Clustern enthaltenen Webseiten zu erstellen und diese in separaten Ordnern abzuspeichern. Die Screenshots werden dabei wie in Kapitel 3.1 teilautomatisiert mit *Selenium* erstellt, wobei die User*innen dazu aufgefordert sind, etwaige erscheinende Warnungen und Abfragen manuell zu entfernen. Die Sammlung von Screenshots kann dazu dienen, die Cluster-Bildungen manuell visuell nachzuvollziehen, indem die multimodalen Strukturen der Webseiten innerhalb eines Clusters von den Nutzer*innen manuell miteinander verglichen werden.³¹

Der Code zur Realisierung der hier beschriebenen Analyse sieht in der hier vorgestellten Implementierung wie folgt aus:

```
1. from analyzer import Analyzer
2.
3. # Initialisierung des Analyzer() Objekts
4. ana = Analyzer()
5.
6. # Übersicht über die default Spalten
7. ana.getColumnSelection()
8.
9. # In diesem Beispiel wird sich auf die Analyse der Textlänge ('total_length') & der Bildanzahl beschränkt
10. ana.setColumnSelection(['total_length', 'total_images'])
11.
12. # Reduzierung des DataFrames auf die ausgewählten Spalten via Indexing mit pandas
13. df_small = ana.data_df[ana.column_selection]
14.
15. # Standardisierung der Daten
16. standardized_data = ana.standardizeData(df_small)
17.
18. # Clustering (k=3)
19. clustered_data = ana.clusterDataKMeans(standardized_data)
20.
21. # Optional: Ermittlung Anzahl k-Clusters für die Analyse (danach evtl. erneut clustern)
22. ana.createElbowPlot(clustered_data)
23.
24. # Screenshots der Webseiten in den jeweiligen Clustern
25. ana.getScreenshotsFromClusters(clustered_data)
```

Die auf diese Weise entstandenen Cluster können nun von den User*innen dazu genutzt werden, weitere Untersuchungen durchzuführen, wobei ein erster Schritt darin bestehen sollte, sich einen Überblick über die jeweils in den Clustern befindlichen Webseiten zu verschaffen und sich mit den dieser Einteilung zugrunde liegenden Werten, die von den für die Analyse ausgewählten Features abhängen, vertraut zu machen. Die so entstandenen Cluster können außerdem für Fragestellungen wie die nach dem Zusammenhang zwischen Webseiten-Typen, gesellschaftlicher Domäne und der multimodalen Struktur genutzt werden.

Die hier vorgestellte Methode und Implementierung sollen im Folgenden anhand eines ausgewählten Webseiten-Samples und dessen Analyse exemplifiziert werden.

4. Analyse eines Webseiten-Samples

Das in diesem Artikel verwendete Webseiten-Sample ist verhältnismäßig klein gewählt, um eine Überprüfung der Analyseergebnisse zu gewährleisten. Prinzipiell ist die Größe des Webseiten-Samples nicht limitiert. Es muss jedoch in die Überlegungen während der Corpus-Erstellung mit einbezogen werden, dass zumindest in der hier verwendeten Implementierung ein manuelles Eingreifen der User*innen an zahlreichen Stellen notwendig ist. Eine zukünftig zu entwickelnde Verbesserung des Programms könnte jedoch versuchen, den derzeit notwendigen manuellen Teil ebenfalls zu automatisieren oder zumindest stark zu reduzieren. Auch ist erneut darauf hinzuweisen, dass die hier vorgestellte Methode eher ergänzenden Charakter hat und als Vorstufe einer zusätzlichen qualitativen Überprüfung oder Fortsetzung der Analyse anzusehen ist. Eine qualitativ-manuelle Überprüfung ist ebenfalls ab einer bestimmten Größe des Datensatzes nur noch schwer möglich.

sogenannten ›Elbow-Plots‹ weitere Verfahren wie die Analyse der Silhouette Scores existieren, die ebenfalls dazu beitragen können, eine adäquate Anzahl k-Cluster für einen Datensatz zu ermitteln. Im Falle einer Ausweitung des zu untersuchenden Webseiten-Corpus bzw. der Ausarbeitung der hier vorgeschlagenen Methode wäre es entsprechend sinnvoll, diese Verfahren ebenfalls einzubinden. Dies ist aus Gründen des Umfangs und aus der Tatsache heraus, dass die ›Elbow-Plots‹ im Kontext des hier zu betrachtenden Corpus bereits gute Ergebnisse geliefert haben, noch nicht geschehen.

³¹ Wobei die Einbindung von Videoelementen nur schwer darstellbar ist.

Bei der Auswahl der Webseiten wurde darauf geachtet, Webseiten aus möglichst diversen gesellschaftlichen Domänen sowie verschiedener Webseiten-Typen nach Meier (siehe Kapitel 3.1) miteinander zu kombinieren. Dies soll dabei helfen, die eingangs aufgeworfene Frage zu untersuchen, ob sich Webseiten aus derselben gesellschaftlichen Domäne beziehungsweise solche desselben Webseiten-Typs hinsichtlich ihres multimodalen Arrangements ähneln. Da die Startseiten von sozialen Netzwerken wie Facebook oder Twitter zumeist das Einloggen der User*innen erfordern und kampagnenorientierte Webseiten nicht immer leicht von imageorientierten Webseiten zu unterscheiden sind, wurde die Auswahl der Webseiten-Typen der Einfachheit halber in dieser Analyse auf die zwei Typen der newsorientierten und imageorientierten Webseiten beschränkt. Eine spätere Integration der beiden anderen Typen kann jedoch leicht nachträglich vorgenommen werden und wäre für eine Ausweitung der in diesem Artikel beispielhaft behandelten Fragestellung interessant.

4.1 Das Webseiten-Sample und Corpus-Erstellung

Das in diesem Artikel gewählte Webseiten-Corpus enthält die folgenden zwanzig Webseiten:

URL	DOMAIN	TYPE
https://marginalie.hypotheses.org/	UNI	News
https://www.ruhr-uni-bochum.de/de	UNI	Image
https://www.uni-bielefeld.de/	UNI	Image
https://www.faz.net/	JOU	News
https://www.heise.de/	JOU	News
https://www.nytimes.com/	JOU	News
https://netzpolitik.org/	JOU	News
https://www.nike.com/	ECO	Image
https://www.adidas.com/	ECO	Image
https://www.volkswagen.de/	ECO	Image
https://www.lufthansa.com/	ECO	Image
https://www.ekd.de/	REL	Image
https://ditib.de/	REL	Image
https://www.katholisch.de/	REL	News
https://www.gamestar.de/	ENT	News
https://www.prosieben.de/	ENT	Image
https://de.ign.com/	ENT	News
https://www.spd.de/	POL	Image
https://www.cdu.de/	POL	Image
https://www.fdp.de/	POL	Image

Tab. 2: Das zu untersuchende Webseiten-Sample. *katholisch.de* wurde als News-Webseite charakterisiert, da es sich der Selbstbeschreibung nach um eine (katholische) Nachrichtenseite handelt, die besonders auf tagesaktuelle Neuigkeiten fokussiert ist. Siehe für diese problematische Einteilung auch das Kapitel 4.2.4. [Jurczyk 2021]

In einem ersten Schritt wird das Corpus wie in Kapitel 3.1 beschrieben initialisiert.³² Daraufhin werden die Bilder und Videos der einzelnen Webseiten mit Hilfe der `DataPreparation()`-Klasse analysiert und die so gewonnenen Daten in `>merged_data_dict.pickle<` gespeichert (siehe Kapitel 3.2).

4.2 Analyse

Die Analyse mit Hilfe der `Analyzer()`-Klasse wird mit unterschiedlichen Selektionen der Features der Corpus-Webseiten durchgeführt, wobei alle Features zuerst via z-score standardisiert werden (siehe Kapitel 3.3):

³² Die im Folgenden beschriebenen Daten basieren auf den Corpus-Webseiten, die am 15.02.2021 initialisiert wurden. Siehe auch die Datei `>info.txt<` im entsprechenden Ordner des [GitHub Repository](#).

1. In einem ersten Schritt werden alle erhobenen Features in die Analyse einbezogen (Vollständige Features; vgl. auch die Übersicht der Features in Kapitel 3.2)
2. In einem zweiten Schritt werden nur die Bild- und Video-Features als Basis für das Clustering genutzt (Bild- und Video-Features)³³
3. In einem dritten Schritt werden nur die Textlänge sowie die Anzahl der Überschriften einbezogen (Text-Features)
4. Der letzte Schritt enthält eine Kombination der Text-, Video- und Bild-Features (Text- und Bild-Features)³⁴

Im Folgenden sollen die im Kontext der jeweilig ausgewählten Features entstandenen Cluster vorgestellt und diskutiert werden. Dabei geht es während der Besprechung zum einen darum, wie nachvollziehbar die Cluster in Bezug auf das multimodale Erscheinungsbild der Webseiten sind. Als Vergleichswerte werden die manuell vom Autor des Artikels aufgerufenen Webseiten und die angelegten Screenshot-Sammlungen herangezogen und qualitativ analysiert. Zum anderen sollen auch die anderen beiden Kategorien (Domänen und Webseiten-Typen) mit in die Diskussion einbezogen werden, indem gefragt wird, inwieweit sich auch hier Clusterbildungen erkennen lassen, die mit der Einteilung auf Basis der multimodalen Features der Seiten korrelieren.

Jede Teilanalyse beginnt mit der Erstellung eines ›Elbow-Plot‹, um einen Eindruck von der potentiell ›richtigen‹ Auswahl an Clustern zu erlangen. Getestet werden dabei k-Means Durchläufe mit k-Werten zwischen zwei und acht. Anschließend werden die Daten mit dem aus der Analyse des ›Elbow-Plot‹ resultierenden k-Wert mit einem k-Means-Clustering gruppiert und die Ergebnisse wie oben beschrieben besprochen.

4.2.1 Vollständige Features

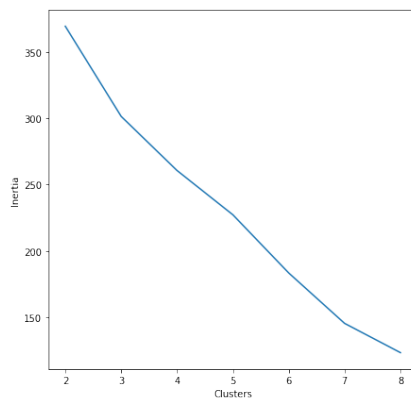


Abb. 1: Elbow-Plot der vollständigen Feature-Auswahl. [Jurczyk 2021]

Wie an dem ›Elbow-Plot‹ erkennbar ist, bietet es sich im Kontext der vollständigen Features an, einen möglichst hohen k-Wert für die Anzahl der Cluster zu wählen. In diesem Falle wurden sieben Cluster gewählt. Die k-Means Clusterbildung sieht mit k=7 Clustern wie folgt aus:

URL	Cluster	Domäne	Typ
ditib.de	0	REL	Image
marginalie.hypotheses.org	0	UNI	News
de.ign.com	1	ENT	News
ekd.de	1	REL	Image
faz.net	1	JOU	News
gamestar.de	1	ENT	News
heise.de	1	JOU	News
katholisch.de	1	REL	News
netzpolitik.org	2	JOU	News
cdu.de	2	POL	Image
lufthansa.com	2	ECO	Image

Tab. 3: Tabelle mit farblich markiertem k-Means Clustering des Webseiten-Samples unter Einbezug aller Features. [Jurczyk 2021]

³³ In die Analyse einbezogen werden die folgenden Features: Hintergrundbilder, Bilder gesamt, große / mittlere / kleine / sehr kleine Bilder, Verhältnisse zwischen großen / mittleren / kleinen / sehr kleinen Bildern und der Gesamtzahl der Bilder, Anzahl großer und kleiner Videos.

³⁴ In die Analyse einbezogen werden die folgenden Features: Verhältnisse zwischen Gesamtzahl der Bilder und der Textlänge sowie zwischen großen und mittleren Bildern und der Textlänge. Außerdem wurde die Anzahl der großen und kleinen Videos einbezogen.

nike.com	2	ECO	Image
spd.de	2	POL	Image
uni-bielefeld.de	2	UNI	Image
volkswagen.de	2	ECO	Image
nytimes.com	3	JOU	News
adidas.com	4	ECO	Image
prosieben.de	5	ENT	Image
fdp.de	6	POL	Image
ruhr-uni-bochum.de	6	UNI	Image

Tab. 3: Tabelle mit farblich markiertem k-Means Clustering des Webseiten-Samples unter Einbezug aller Features. [Jurczyk 2021]

Das Clustering unter Einbezug aller Features macht es auf den ersten Blick schwierig, eine Struktur beziehungsweise ein Muster in der Clusterbildung zu erkennen, was unter anderem durch die hohe Anzahl an Clustern für eine relativ geringe Anzahl an Webseiten bedingt ist. Dies kann jedoch als Hinweis gewertet werden, dass die Webseiten in der Tat relativ divers sind, was gut an den Single-Webseiten-Clustern 3, 4 und 5 deutlich wird.

Neben der Feststellung, dass die Webseiten unter Einbezug aller Features relativ divers sind, stechen zwei Cluster besonders hervor (Cluster 1 und Cluster 2).

Cluster 1 enthält vor allem Webseiten vom Typ ›News‹, die sich allerdings über ganz unterschiedliche Domänen erstrecken und von Entertainment Seiten (ign.com) über klassische journalistische Angebote (faz.net) bis hin zur religiösen Newsportalen (katholisch.de) reichen. Es gibt jedoch zwei Webseiten vom Typ ›News‹, die nicht in Cluster 2 enthalten sind. Der erste *outlier* ist die Webseite netzpolitik.org, deren multimodale Struktur sich von der anderer News-Webseiten in der Form absetzt, dass netzpolitik.org verhältnismäßig wenige Überschriften (netzpolitik.org: 70, faz.net: 167, heise.de: 164), dafür allerdings viele große Bildern beinhaltet (netzpolitik.org: 11, faz.net: 3, heise.de: 0)³⁵, was die Seite zusammen mit der Anordnung der News in nur einer Spalte deutlich aufgeräumter wirken lässt.³⁶ Der zweite *outlier* ist die amerikanische News-Webseite nytimes.com. Diese fällt in eine gesonderte Kategorie, weil nytimes.com als einzige der hier aufgeführten News-Webseiten direkt abgespielte Videos auf der Startseite eingebaut hat, die sie in der Tat aus multimodaler Perspektive von den anderen Seiten abhebt.

Cluster 2 enthält vor allem Webseiten vom Typ ›Image‹, die ebenfalls aus verschiedenen gesellschaftlichen Bereichen stammen, wobei der Bereich der ›Ökonomie‹ (ECO) dominiert. Ein Blick auf die Werte der Features zeigt, dass im Falle die Webseiten vom Type ›Image‹ vor allem das Verhältnis zwischen großen Bildern und der Anzahl der Gesamtbilder deutlich größer und umgekehrt das Verhältnis zwischen kleinen Bildern und der Gesamtzahl der Bilder deutlich kleiner ist als im Falle von Cluster 1. Auch unterscheiden sich die Webseiten in Cluster 1 und Cluster 2 hinsichtlich ihrer Textmenge und Anzahl der Überschriften.

Zusammenfassend kann festgehalten werden, dass der Einbezug aller Features bereits in ersten sinnvollen Clusterbildungen des multimodalen Arrangements mündet, die insbesondere mit dem Typ der Webseiten korrelieren und unter anderem zwischen Webseiten mit (wenigen) großen Bildern (imageorientierte Webseiten) und Webseiten mit (vielen) kleineren Bildern und einer größeren Textmenge (newsorientierte Webseiten) unterscheiden.

³⁵ Dies liegt allerdings bei heise.de nur daran, dass viele der Bilder nur ganz knapp an der Grenze für große Bilder (700px) gescheitert sind.

³⁶ Die absoluten Werte der Features für das hier verwendete Webseiten-Sample können in der CSV-Datei ›data_abs_values.csv‹ im GitHub Repository nachvollzogen werden, vgl. Jurczyk 2021a.

4.2.2 Bild- und Video-Features

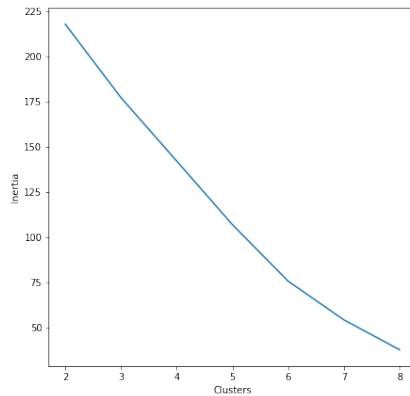


Abb. 2: ›Elbow-Plot‹ der Bilder- und Video-Feature Auswahl. [Jurczyk 2021]

Im Falle der Bild- und Video-Feature-Auswahl wird im ›Elbow-Plot‹ deutlich, dass sich eine Clusteranzahl von $k=6$ für die Analyse anbietet. Ein durchgeführtes k-Means-Clustering mit $k=6$ Clustern ergibt die folgende Einteilung des Webseiten-Samples:

URL	Cluster	Domäne	Typ
de.ign.com	0	ENT	News
ruhr-uni-bochum.de	1	UNI	Image
fdp.de	1	POL	Image
adidas.com	1	ECO	Image
ekd.de	1	REL	Image
faz.net	1	JOU	News
gamestar.de	1	ENT	News
heise.de	1	JOU	News
katholisch.de	1	REL	News
cdu.de	1	POL	Image
lufthansa.com	1	ECO	Image
spd.de	1	POL	Image
netzpolitik.org	1	JOU	News
uni-bielefeld.de	2	UNI	Image
nike.com	2	ECO	Image
volkswagen.de	2	ECO	Image
ditib.de	3	REL	Image
marginalie.hypotheses.org	3	UNI	News
prosieben.de	4	ENT	Image
nytimes.com	5	JOU	News

Tab. 4: Tabelle mit farblich markiertem k-Means Clustering des Webseiten-Samples unter Einbezug der Bild- und Video-Features. [Jurczyk 2021]

Die Analyse basierend auf den Bild- und Video-Features ist in Bezug auf die Domänen und Webseiten-Typen deutlich schwieriger zu interpretieren als im Falle der Inklusion aller Features. Ein Blick auf die Werte der Features in der Tabelle ›data_abs_values.csv‹ macht deutlich, dass die Cluster vor allem aus der unterschiedlichen Verteilung von großen (Cluster 2), mittleren (Cluster 1) und sehr kleinen (Cluster 0) Bildern innerhalb des Webseiten-Samples resultieren. Hinzu kommt die unterschiedliche Verwendung von Videos. Durch die nur sehr geringe und optisch kaum wahrnehmbare Differenz zwischen mittleren Bildern, die oft nur knapp unter 700px groß sind (heise.de), und großen Bildern, die teils nur knapp über 700px groß sind (netzpolitik.org), ist ein Clustering auf Basis dieser Daten zumindest im Kontext dieses Webseiten-Samples kaum geeignet, um das multimodale Arrangement der Webseiten zu klassifizieren. Hinzu kommen *outlier* wie prosieben.de. Diese unterscheidet sich optisch nur geringfügig von anderen News-Webseiten, die Klassifizierung in einem separaten Cluster erfolgt in diesem Falle aufgrund der ausgiebigen Nutzung von Hintergrundbildern, die allerdings in <figure>-Tags verwendet werden und sich dort sowohl in Sachen Größe als auch Erscheinungsbild kaum von regulären Bildern unterscheiden. Dies weist erneut auf die angesprochene Schwierigkeit hin, diverse und idealiter unbekannte Webseiten vollautomatisiert miteinander zu vergleichen.

Insgesamt ist die Einschränkung der Features auf Bilder und Videos zwar geeignet, eine allgemeine Einschätzung der Distribution von Elementen zu erlangen, die durch ihre vielfache Verwendung und oftmals sehr präzise Positionierung auf Webseiten eine wichtige Rolle während der multimodalen Zeichengenerierung spielen, aber um sie als alleinige Merkmale für das Clustering heranzuziehen, scheinen sie in ihrer diversen Verwendung zumindest im derzeitigen Zustand des Programms ohne eine Ausarbeitung der Bildklassifizierungsstrategien während der Datengenerierung nicht geeignet.

4.2.3 Text-Features

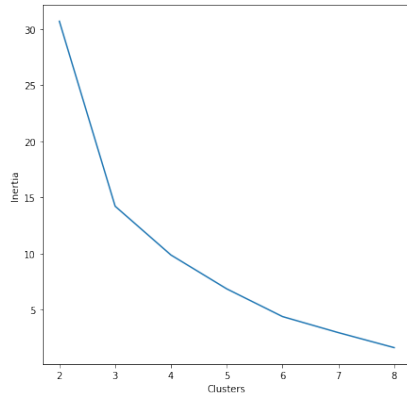


Abb. 3: ›Elbow-Plot‹ der Text-Feature-Auswahl. [Jurczyk 2021]

Im Falle der Text-Features lässt sich der ›Elbow‹ bei $k=3$ oder $k=4$ Clustern erkennen, wobei in diesem Beispiel $k=4$ Cluster ausgewählt wurden. Das k -Means-Clustering mit $k=4$ Clustern resultiert in der folgenden Einteilung des Webseiten-Samples:

URL	Cluster	Domäne	Typ
volkswagen.de	0	ECO	Image
nike.com	0	ECO	Image
katholisch.de	0	REL	News
gamestar.de	0	ENT	News
ekd.de	0	REL	Image
netzpolitik.org	0	JOU	News
marginalie.hypotheses.org	0	UNI	News
prosieben.de	1	ENT	Image
nytimes.com	1	JOU	News
faz.net	2	JOU	News
heise.de	2	JOU	News
de.ign.com	2	ENT	News
ruhr-uni-bochum.de	3	UNI	Image
fdp.de	3	POL	Image
adidas.com	3	ECO	Image
cdu.de	3	POL	Image
lufthansa.com	3	ECO	Image
spd.de	3	POL	Image
uni-bielefeld.de	3	UNI	Image
ditib.de	3	REL	Image

Tab. 5: Tabelle mit farblich markiertem k -Means Clustering des Webseiten-Samples unter Einbezug der Text-Features. [Jurczyk 2021]

Im Falle des Text-Clustering lassen sich eindeutige Trends erkennen, die vor allem mit den Webseiten-Typen korrelieren (Cluster 1, Cluster 2 sowie Cluster 3). Cluster 0 enthält Webseiten, die ungefähr in der Mitte liegen, was die Textmenge und die Verwendung von Überschriften anbetrifft. Cluster 1 umfasst Webseiten, die ebenfalls eine mittlere Menge an Text umfassen, aber einen ausgiebigen Gebrauch von Überschriften machen.

Besonders im Falle von nytimes.com muss jedoch darauf aufmerksam gemacht werden, dass beispielsweise die Verwendung von `<h2>`-Tags im unteren Bereich der Webseite von der Darstellung her kaum von einem regulären Text zu unterscheiden ist, weshalb nytimes.com eher in Cluster 2 einzuordnen ist.³⁷

Cluster 2 enthält News-Webseiten, die eine große Anzahl an Wörtern und viele Überschriften aufweisen und die damit als Idealtypen von News-Webseiten gelten können. Cluster 3 hingegen enthält Webseiten (imageorientierte Webseiten nach Meier), die im Vergleich zu den anderen Webseiten eine geringe Textmenge und wenige Überschriften beinhalten.

Insgesamt kann das auf den Textmengen basierende Clustering als durchaus sinnvoll angesehen werden, um den Typ einer Webseite zu klassifizieren. Auch deuten die Unterschiede in der Verwendung von Text, trotz der offensichtlich vorhandenen Schwierigkeiten (siehe nytimes.com), auch auf eine unterschiedliche Verwendung eines wichtigen multimodalen Elements hin (Schrift und deren verschiedentliche Verwendung, beispielsweise als Fließtext oder als Überschrift).

Der letzte Analyseschritt führt die Text-Features mit den Video- und Bild-Features zusammen. Im Gegensatz zum ersten Analyseschritt, der alle Features einbezogen hat, werden im folgenden Schritt die Hyperlinks und die Hintergrundbilder ignoriert, da letztere zumindest in diesem Webseiten-Sample zu falschen Aussagen geführt haben (siehe Kapitel 4.2.2, Diskussion um prosieben.de) und außerdem zu fragen ist, inwieweit Hyperlinks einen Einfluss auf die Erstwahrnehmung der Sehfläche einer Webseite haben, deren Analyse hier im Fokus steht.

4.2.4 Text-, Video- und Bild-Features

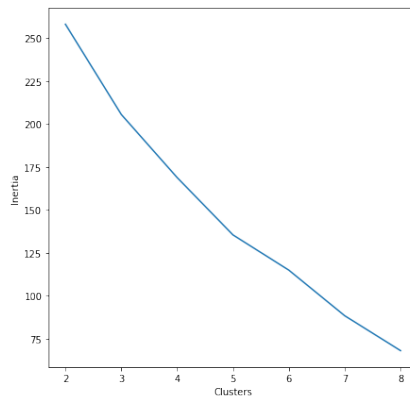


Abb. 4: ›Elbow-Plot‹ der Bild-, Video- und Text-Feature-Auswahl. [Jurczyk 2021]

Im Falle einer Kombination der Text-, Bild- und Video-Features ist die Wahl der Clusteranzahl auf Basis des ›Elbow-Plot‹ erneut schwierig und erinnert an die Analyse unter Einbezug aller Features in Kapitel 4.2.1. In diesem Falle habe ich mich dazu entschieden, das k-Means-Clustering mit k=8 Clustern durchzuführen, da ein ›Elbow‹ innerhalb des Plots nur schwer erkennbar ist. Die Einteilung der Webseiten des Webseiten-Samples sieht mit k=8 Clustern wie folgt aus:

URL	Cluster	Domäne	Typ
nytimes.com	0	JOU	News
spd.de	1	POL	Image
volkswagen.de	1	ECO	Image
uni-bielefeld.de	1	UNI	Image
nike.com	1	ECO	Image
cdu.de	1	POL	Image
ekd.de	2	REL	Image
gamestar.de	2	ENT	News
lufthansa.com	2	ECO	Image
heise.de	2	JOU	News

Tab. 6: Tabelle mit farblich markiertem k-Means Clustering des Webseiten-Samples unter Einbezug der Bild-, Video- und Text-Features. [Jurczyk 2021]

³⁷ Siehe auch die Diskussion in Kapitel 4.2.2 betreffs der Verwendung von Hintergrundbildern im Falle von prosieben.de. Die hier vorhandene Verwendung von Überschriften-Tags, die nicht wirklich als Überschriften deutlich werden, geht bezüglich ihrer Problematik für die automatisierte multimodale Analyse von Webseiten in eine ähnliche Richtung.

katholisch.de	2	REL	News
faz.net	2	JOU	News
netzpolitik.org	2	JOU	News
ditib.de	3	REL	Image
marginalie.hypothesen.org	3	UNI	News
prosieben.de	4	ENT	Image
ruhr-uni-bochum.de	5	UNI	Image
fdp.de	5	POL	Image
de.ign.com	6	ENT	News
adidas.com	7	ECO	Image

Tab. 6: Tabelle mit farblich markiertem k-Means Clustering des Webseiten-Samples unter Einbezug der Bild-, Video- und Text-Features. [Jurczyk 2021]

Das hier ersichtliche feingliedrige Clustering zeichnet ein realistisches Bild der multimodalen Struktur der im Webseiten-Sample enthaltenen Webseiten. Besonders hervorzuheben sind Cluster 1 und Cluster 2.

Cluster 1 kann als ›Werbe-Cluster‹ bezeichnet werden, da es ausschließlich Webseiten vom Typ ›Image‹ enthält, die sich durch eine relativ geringe Textmenge und eine geringe Anzahl an Bildern auszeichnen, die dafür aber meist sehr groß sind. Die Ersteindruck der Seiten erinnert damit an Werbeplakate oder Reklame in gedruckten Magazinen.³⁸ Die Seite adidas.de (Cluster 7) fällt nur deshalb aus diesem Cluster heraus, weil sie ein (sehr großes) Video auf der Startseite platziert hat, und damit ein anderes multimodales Arrangement aufweist. Ein vergleichender Blick auf zwei Webseiten aus Cluster 1 (spd.de und volkswagen.de) unterstreicht die Kohärenz des Clustering in Bezug auf deren multimodale Struktur und den Ersteindruck beim Öffnen der Webseiten. Beide Webseiten werden eindeutig von großen Bildern und wenig Text, der eher den Bildern zuzuarbeiten scheint, dominiert.³⁹



Abb. 5: Ansicht Startseite spd.de (Zoom-Stufe 50 %, 17.02.2021). [Jurczyk 2021]



Abb. 6: Ansicht Startseite volkswagen.de (Zoom-Stufe 50 %, 17.02.2021). [Jurczyk 2021]

Cluster 2 umfasst Webseiten mit, im Vergleich zu Cluster 1, viel Text, der außerdem durch zahlreiche Überschriften unterteilt wird. Webseiten in Cluster 2 enthalten darüber hinaus relativ viele Bilder. Cluster 2 steht damit idealtypisch für News-Webseiten. Die Strukturähnlichkeiten der Webseiten dieses Clusters lassen sich erneut durch einen manuellen Nachvollzug des Ersteindrucks der Webseiten bestätigen, der sich deutlich von dem der Webseiten in Cluster 1 unterscheidet. Als Beispiele für Cluster 2 wurden die Webseiten faz.net und katholisch.de ausgewählt.

³⁸ Meer / Pick 2019, S. 65–68.

³⁹ Die folgende Auswahl zeigt immer nur zwei Screenshot-Beispiele pro Cluster. Die restlichen Screenshots finden sich im GitHub Repository. Die Webseiten aus den einzelnen Clustern wurden mit einer sehr geringen Zoomstufe (50 %) vorgenommen, um so einen besseren Eindruck von dem gesamten Aufbau der Webseite zu erhalten. Dies ist beispielsweise im Falle von ign.com wichtig, da sich der Kopfbereich kaum von dem anderer Nachrichtenseiten unterscheidet. Sobald aber gescrollt wird, werden die Unterschiede deutlich, die hier in der sehr großen Anzahl kleiner Bilder gepaart mit kurzen Teaser-Texten liegen, die in nur einer Spalte angelegt sind. Die Screenshots der Webseiten wurden am 17.02.2021 erstellt (und damit einige Tage nach Initialisierung des Corpus).



Abb. 7: Ansicht Startseite faz.de (Zoom-Stufe 50 %, 17.02.2021). [Jurczyk 2021]

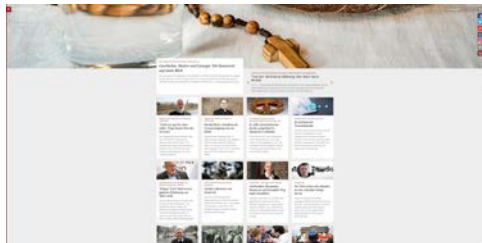


Abb. 8: Ansicht Startseite katholisch.de (Zoom-Stufe 50 %, 17.02.2021). [Jurczyk 2021]

Dass in Cluster 2 auch zwei Webseiten vom Typ ›Image‹ erscheinen, zeigt, dass sich auch imageorientierte Webseiten durchaus als Newsportale multimodal strukturieren können, was beispielhaft an der Webseite ekd.de deutlich wird. Die Webseite ekd.de orientiert sich ebenfalls in Form klassischer Artikelstrukturen (Teaser-Text kombiniert mit kleinen bis mittleren Teaser-Bildern im Kachelformat) an den News-Webseiten, allerdings überwiegt in der thematischen Ausrichtung der Hauptseite im Gegensatz zu bspw. katholisch.de noch die allgemeine Funktion eines digitalen Informationsportals der EKD, das nicht primär auf tagesaktuelle Geschehnisse ausgerichtet ist. Dahingegen ist katholisch.de zwar ebenfalls eine kirchlich orientierte Webseite, versteht sich aber laut Impressum ausdrücklich als News-Webseite, die Neuigkeiten aus dem Bereich der Katholischen Kirche berichtet bzw. in den Vordergrund stellt. Dennoch bleibt die Einteilung bezüglich des Webseiten-Typs in diesem Falle problematisch, und das gemeinsame Erscheinen von sowohl ekd.de als auch katholisch.de in einem News-orientierten Cluster weist darauf hin, dass beide Seiten in ihrer multimodalen Struktur mehr Gemeinsamkeiten als Unterschiede aufweisen, weshalb die Einteilung in unterschiedliche Webseiten-Typen noch einmal zu überdenken wäre (was durchaus als positives bzw. erkenntnisförderndes Ergebnis des Clusterings zu werten ist).

Die Webseite nytimes.com fällt aufgrund eines zentral eingebundenen, wenn auch relativ klein ausfallenden Videos aus diesem Cluster heraus (und bildet mit Cluster 0 in Bezug auf die multimodale Struktur überzeugend ein separates Cluster).

Besonders hervorzuheben ist außerdem das Cluster 3, das mit der Seite marginalien.hypothesis.org und ditib.de zwei Webseiten enthält, die gewissermaßen Außenseiter innerhalb des Webseiten-Samples darstellen.⁴⁰ Die Webseite marginalien.hypothesis.org repräsentiert einen klassischen Blog, der nur wenige Bilder, dafür aber relativ viel Text enthält, und sich entsprechend von sowohl den imageorientierten Webseiten in Cluster 1 und den News-Webseiten in Cluster 2 absetzt. Die Seite ditib.de ist spartanisch gehalten, insbesondere was ihre Größe anbetrifft – so handelt es sich um die einzige Webseite des Samples, bei der von den User*innen nicht gescrollt werden muss, um das Ende der Seite zu erreichen.

5. Fazit

Dieser Artikel hat es sich zur Aufgabe gemacht, ein (teil-)automatisiertes Verfahren zur Analyse des multimodalen Arrangements von Webseiten vorzustellen, anzuwenden und kritisch zu diskutieren. Im Fokus stand dabei die erste Wahrnehmungsebene der multimodalen Struktur der Webseiten, die während des initialen Aufrufs der Webseiten eine zentrale Rolle spielt (erste Wahrnehmungsebene der Sehfläche). Im Zentrum stand dabei die Analyse der Video-, Bild-, Hyperlink- und Textelemente der Webseiten, die anhand eines in Python geschriebenen Programms gesammelt, verarbeitet und schließlich mit Hilfe eines k-Means-Clustering analysiert wurden.

Die Analyse in Kapitel 4 hat gezeigt, dass unterschiedliche Selektionen von Features zu unterschiedlichen Clusterbildungen des Webseiten-Samples führen, die sich gegenseitig ergänzen. Als besonders geeignet für die Analyse des multimodalen Arrangements hat sich eine Kombination aus Video-, Bild- und Text-Features (Kapitel 4.2.4) herausgestellt, die anhand von acht Clustern überzeugend in der Lage war, den Testdatensatz auf Basis des multimodalen Erscheinungsbildes der Webseiten zu strukturieren. Dies macht Hoffnung für die Möglichkeit eines zukünftigen Ausbaus und einer Weiterführung der hier vorgeschlagenen Methode, die dazu dienen könnte, deutlich größere Webseiten-Samples vorzustrukturieren, um so unter anderem eine gezieltere qualitative Analyse ihrer Multimodalität zu ermöglichen, die durch das hier aufgezeigte quantitative Verfahren keineswegs ersetzt, sondern lediglich ergänzt und vereinfacht werden soll.

⁴⁰ Diese beiden Webseiten sind auch in den vorangegangenen k-Means-Clustering Verfahren mit anderen Feature-Selektionen häufig als Cluster aufgetreten.

Neben der Vorstellung dieses (teil-)automatisierten Verfahrens hat es sich dieser Artikel ebenfalls zur Aufgabe gemacht zu fragen, inwieweit die auf Basis der multimodalen Features gruppierten Webseiten Hinweise auf mögliche Domänen-Cluster liefern beziehungsweise inwieweit die von Meier vorgeschlagenen Webseiten-Typen mit den gefundenen Clustern korrelieren.

Insbesondere in Bezug auf die letztgenannten Webseiten-Typen kann nach Durchführung der Analyse gesagt werden, dass entsprechende Parallelen erkennbar waren, obwohl einschränkend hinzugefügt werden muss, dass sich dieser Artikel lediglich auf zwei der vier von Meier vorgeschlagenen Typen beschränkt hat. So korrelierten verschiedene Cluster (wie in Kapitel 4.2.4) mit Meiers News- oder Image-Typen, wobei das hier vorgeschlagene Analyseverfahren dazu in der Lage war, innerhalb der jeweiligen Webseiten-Typen noch einmal feiner zu unterscheiden (beispielsweise, ob News-Webseiten Videos eingebunden hatten). Dabei traten Überlappungen auf, indem auch imageorientierte Webseiten (ekd.de) die multimodale Form einer News-Webseite annehmen können.

Keine Korrelation wurde hingegen zwischen den Clustern der multimodalen Analyse und den gesellschaftlichen Domänen ersichtlich, aus denen die jeweiligen Webseiten stammen. Vielmehr wurde deutlich, dass sich Webseiten aus verschiedenen gesellschaftlichen Bereichen durchaus auf ein ›gemeinsames‹ multimodales Arrangement einigen können, was im Falle des hier vorliegenden Webseiten-Samples besonders an den Domänen Politik, Wirtschaft und Universität (sowie teilweise Religion und journalistische Webseiten) deutlich wurde, die allesamt eine ähnliche multimodale Struktur aufweisen. Somit scheint es zumindest in Bezug auf das multimodale Arrangement erst einmal keine von der Semantik der einzelnen modes unabhängige übergreifende Semantik zu geben, die aus der Struktur der Seiten resultiert und die beispielsweise einen spezifischen politischen oder religiösen Wert hätte. Vielmehr orientiert sich der Gebrauch multimodaler Elemente und deren Arrangement aus Sicht der Domänen an der intendierten Wirkung der Webseite, die sich gut anhand der von Meier vorgeschlagenen Kategorien klassifizieren lässt. So können beispielsweise religiöse Webseiten sowohl als News-Webseiten (katholisch.de) oder auch als imageorientierte Seiten (ditib.de) mit unterschiedlichsten multimodalen Arrangements auftreten. Dasselbe gilt für die Domäne der Universität, die sowohl als professionell gestaltete imageorientierte Webseiten oder aber in Form eines Blogs (marginalien.hypothesis.de) in Erscheinung treten kann.

Bibliographische Angaben

- Maren Freudenberg / Dunja Sharbat Dar: Popular Cultural Representations of Femininity in the International Christian Fellowship: An Analysis of the »Ladies Lounge 2021« Webpage. In: Zeitschrift für Religion, Gesellschaft und Politik. [im Erscheinen] [[Nachweis im GBV](#)]
- Aurélien Géron: Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. Concepts, tools, and techniques to build intelligent systems. Beijing u. a. 2019. [[Nachweis im GBV](#)]
- Multimodal Analysis Lab. Hg. von Interactive & Digital Media Institute (IDMI). 2013. [[online](#)]
- The Routledge handbook of multimodal analysis. Hg. von Carey Jewitt. 2. Auflage. London u. a. 2017. (= Routledge handbooks) [[Nachweis im GBV](#)]
- Thomas Jurczyk (2021a): ZfdG Website Scraper. In: GitHub. 22.02.2021. [[online](#)]
- Thomas Jurczyk (2021b): Clustering with Scikit-Learn in Python. In: The Programming Historian (2021). [im Erscheinen]
- Gunther Rolf Kress: What is mode? In: The Routledge handbook of multimodal analysis. Hg. von Carey Jewitt. 2. Auflage. London u. a. 2017, S. 60–76. (= Routledge handbooks) [[Nachweis im GBV](#)]
- Agnieszka Lyons: Multimodality. In: Research Methods in Intercultural Communication. Hg. von Zhu Hua. Wiley-Blackwell, NJ 2016, S. 268–280. Siehe auch [[Nachweis im GBV](#)]
- Dorothee Meer / Ina Pick: Einführung in die Angewandte Linguistik. Gespräche, Texte, Medienformate analysieren. Stuttgart 2019. [[Nachweis im GBV](#)]
- Stefan Meier: Websites als multimodale digitale Texte. In: Handbuch Sprache im multimodalen Kontext. Hg. von Hartmut Stöckl / Nina-Maria Klug. Berlin u. a. 2016, S. 410–436. (= Handbücher Sprachwissen, 7) [[Nachweis im GBV](#)]
- Kay O'Halloran / Sabine Tan / Peter Wignell / Rui Wang / Kevin Chai / Rebecca Lange: Towards a Discipline of Multimodality: Parallels to Mathematics and Linguistics and New Ways Forward. In: Multimodality. Disciplinary Thoughts and the Challenge of Diversity. Hg. von Janina Wildfeuer. 1. Auflage. Boston MA 2020, S. 93–112. [[Nachweis im GBV](#)]
- Kay L. O'Halloran / Marissa K. L. E / Sabine Tan: Multimodal analytics. Software and visualization techniques for analyzing and interpreting multimodal data. In: The Routledge handbook of multimodal analysis. Hg. von Carey Jewitt. 2. Auflage. London u. a. 2017, S. 386–396. (= Routledge handbooks) [[Nachweis im GBV](#)]
- SeleniumHQ Browser Automation. In: selenium.dev. Hg. von Software Freedom Conservancy. 2004–2021. [[online](#)]
- Sklearn.preprocessing. StandardScaler. In: scikit learn.org. Hg. von scikit-learn developers. 2007–2021. [[online](#)]
- Hartmut Stöckl: Sprache-Bild-Texte lesen. Bausteine zur Methodik einer Grundkompetenz. In: Bildlinguistik. Theorien - Methoden - Fallbeispiele. Hg. von Hans-Joachim Diekmannshenke / Michael Klemm / Hartmut Stöckl. Berlin 2011, S. 45–70. (= Philologische Studien und Quellen, 228) [[Nachweis im GBV](#)]
- Hartmut Stöckl: Multimodalität. Semiotische und textlinguistische Grundlagen. In: Handbuch Sprache im multimodalen Kontext. Hg. von Hartmut Stöckl / Nina-Maria Klug. Berlin u. a. 2016, S. 3–35. (= Handbücher Sprachwissen, 7) [[Nachweis im GBV](#)]
- Peter Wignell / Kevin Chai / Sabine Tan / Kay O'Halloran / Rebecca Lange: Natural Language Understanding and Multimodal Discourse Analysis for Interpreting Extremist Communications and the Re-Use of These Materials Online. In: Terrorism and Political Violence 33 (2021), H. 1, S. 71–95. [[Nachweis im GBV](#)]

Abbildungslegenden und -nachweise

Tab. 1: Beispiel eines Eintrags einer Webseite in der Datei ›websites.csv‹. [Jurczyk 2021]

Tab. 2: Das zu untersuchende Webseiten-Sample. katholisch.de wurde als News-Webseite charakterisiert, da es sich der eigenen Ausrichtung nach um eine (katholische) Nachrichtenseite handelt. Siehe für diese problematische Einteilung auch das Kapitel 4.2.4). [Jurczyk 2021]

Abb. 1: Elbow-Plot der vollständigen Feature-Auswahl. [Jurczyk 2021]

Tab. 3: Tabelle mit farblich markiertem k-Means Clustering des Webseiten-Samples unter Einbezug aller Features. [Jurczyk 2021]

Abb. 2: ›Elbow-Plot‹ der Bilder- und Video-Feature Auswahl. [Jurczyk 2021]

Tab. 4: Tabelle mit farblich markiertem k-Means Clustering des Webseiten-Samples unter Einbezug der Bild- und Video-Features. [Jurczyk 2021]

Abb. 3: Elbow-Plot der Text-Feature Auswahl. [Jurczyk 2021]

Tab. 5: Tabelle mit farblich markiertem k-Means Clustering des Webseiten-Samples unter Einbezug der Text-Features. [Jurczyk 2021]

Abb. 4: ›Elbow-Plot‹ der Bild-, Video- und Text-Feature-Auswahl. [Jurczyk 2021]

Tab. 6: Tabelle mit farblich markiertem k-Means Clustering des Webseiten-Samples unter Einbezug der Bild-, Video- und Text-Features. [Jurczyk 2021]

Abb. 5: Ansicht Startseite spd.de (Zoom-Stufe 50 %, 17.02.2021). [Jurczyk 2021]

Abb. 6: Ansicht Startseite volkswagen.de (Zoom-Stufe 50 %, 17.02.2021). [Jurczyk 2021]

Abb. 7: Ansicht Startseite faz.de (Zoom-Stufe 50 %, 17.02.2021). [Jurczyk 2021]

Abb. 8: Ansicht Startseite katholisch.de (Zoom-Stufe 50 %, 17.02.2021). [Jurczyk 2021]