

Visualisierung von Daten aus Online-Partizipationsverfahren

Philipp Grawe
Institut für Informatik
Heinrich-Heine-Universität Düsseldorf
Universitätsstr. 1
40225 Düsseldorf
philipp.grawe@hhu.de

ABSTRACT

Online-Partizipation nutzt das Internet, um Menschen über Entscheidungen diskutieren zu lassen und ihnen die Möglichkeit der Teilhabe an Prozessen zu gewähren. In dieser Arbeit sollen verschiedene Methoden zur Visualisierung solcher Online-Partizipationsverfahren vorgestellt werden, die dabei helfen sollen Verfahren zu überblicken und mit ihrem Ergebnis zu arbeiten. Dabei können sowohl extrahierte Themen als auch Metadaten visualisiert werden.

1. EINLEITUNG

Um Menschen an Entscheidungen oder Prozessen teilhaben zu lassen, bietet sich die Möglichkeit online Plattformen anzubieten, auf denen Diskussionen stattfinden, Vorschläge eingereicht und über eben diese abgestimmt werden können. Genutzt werden diese Möglichkeiten am prominentesten in der Kommunalpolitik, wo Bürgerinnen und Bürger online an Debatten wie z.B. Bürgerhaushalten teilhaben können. Aber auch Institutionen und Firmen setzen Online-Partizipation ein, um Menschen zu beteiligen. Die Teilhabe von Betroffenen oder Bürgern an Entscheidungen schafft, auch durch die Bereitstellung im Internet, Transparenz und Vertrauen. Vorschläge können neue, bisher nicht bedachte, Sichtweisen offenbaren und Abstimmungen dieser Vorschläge ein unverbindliches Stimmungsbild liefern. Durch die so entstandene Partizipation wird erhofft, mehr Blickwinkel der Betroffenen bei der Problemlösung einzubeziehen und eine höhere Akzeptanz der Entscheidungen zu erzielen.

Konkret heißt dies meist, dass Teilnehmer Vorschläge erstellen, kommentieren und zustimmend bzw. ablehnend bewerten können. Außerdem gibt es Verfahren bei denen Bürgerinnen und Bürger über Vorschläge der Verwaltung diskutieren und abstimmen können. Insgesamt gibt es sowohl auf einen Zeitraum begrenzte Verfahren wie z.B. Bürgerhaushalte, als auch dauerhafte Plattformen, wie z.B. Mängelmelder für den kommunalen Raum. Meistens haben die Online-Partizipationsverfahren konkrete Themen als Bezug, beispielsweise die

Verwendung des Tempelhoferfelds in Berlin.

Da die Verfahren oft einem großen Personenkreis offenstehen und nicht selten auf viel Engagement stoßen, kann so eine erhebliche Anzahl von Vorschlägen und Textbeiträgen entstehen. Eine manuelle Auswertung all dieser Beiträge ist mit einem nicht zu unterschätzenden Ressourcenaufwand verbunden. Deshalb bietet sich hier eine maschinell unterstützte Analyse an.

Weitere Hintergründe und eine gute \ddot{A} bersicht zum Einsatz von maschinellen Analyseverfahren bei Online-Partizipationsverfahren bieten Liebeck et al. [11].

Diese Thematik wird viel diskutiert und erforscht, z.B. durch das *NRW-Forschungskolleg Online-Partizipation*¹ welches auch Veranstaltungen abhalt die den Austausch von Forschung und Praxis fordern soll. Jahrlich findet dazu das Praxis-symposium \ddot{A} Online-Partizipation in Kommunen \ddot{A} statt. Auerdem bietet der Monitor Online-Partizipation² eine \ddot{A} bersicht ber Verfahren in NRW der letzten Jahre.

In dieser Arbeit wird die Visualisierung extrahierter Themen und anderer Analysedaten von Online-Partizipationsverfahren betrachtet und dabei auf die Besonderheiten von Online-Partizipationsverfahren eingegangen. Zum Extrahieren dieser Themen wird Topic Modeling eingesetzt, dessen Modelle dann als Grundlage zur Visualisierung dienen konnen. Als Datengrundlage werden in dieser Arbeit oft Online-Partizipationsverfahrenherangezogen, welche sich mit der Verbesserung des Radverkehrs beschaftigt. Diese wurden parallel in drei Gebieten durchgefuhrt und dabei wissenschaftlich begleitet [4].

Ein Projekt mit Webcrawlern von kommunalen Online-Partizipationsverfahren aus NRW ist ffentlich auf *GitHub*³ verfugbar und kann dazu verwendet werden Daten von laufenden, als auch abgeschlossenen Verfahren zu erlangen. Daten dieser Verfahren wurden als Grundlage fur diese Arbeit verwendet.

1.1 Problematiken bei Online-Partizipationsverfahren

Generell ist es sinnvoll vor dem Einsatz solcher Verfahren die Datengrundlage zu begutachten. Online-Partizipationsverfahren konnen, wie beispielsweise die Raddialoge, sehr zielgerichtet sein. Dies macht es notwendig zu hinterfragen, ob die Themen in den Dokumenten gut genug trennbar sind.

¹<https://www.fortschrittskolleg.de>

²<http://www.monitor-online-partizipation.de>

³<https://github.com/Liebeck/OnlineParticipationDatasets>

31st GI-Workshop on Foundations of Databases (Grundlagen von Datenbanken), 11.06.2019 - 14.06.2019, Saarburg, Germany.
Copyright is held by the author/owner(s).

Es ist schwierig semantische Strukturen in Dokumenten zu finden, wenn diese wenig verschiedene semantische Strukturen aufweisen.

Ein weiter zu bedenkender Aspekt liegt in der Form der Online-Partizipationsverfahren. Mitunter sind Vorschläge, und damit die Dokumente, sehr kurz und beinhalten nur wenige Wörter. Damit diesem Dokument die richtigen Themen zugewiesen werden können, muss das eingesetzte Verfahren sehr präzise arbeiten. Das Problem Topic Modeling auf kurzen Texten zu verwenden wird unter anderem von Jin et al. [7] oder Yan et al. [26] behandelt.

Auf der anderen Seite stehen lange Vorschläge mit vielen Kommentaren. Diese können gleich mehrere Themen behandeln. Oder aber der Diskurs in den Kommentaren wird nur über einen Aspekt geführt, was ebenfalls die Themen dieses Dokuments verfälscht.

Die üblichen in der Forschung eingesetzten Datensätze zum Topic Modeling entstammen Zeitungen oder Wikipedia. Meistens sind Dokumente aus diesen beiden Spektren nicht nur länger, sondern viel themenbezogener als bei Online-Partizipationsverfahren. Neben der Tatsache, dass diese Dokumente formaler geschrieben sind, unterliegen sie in der Regel einer Systematik der Kategorisierung. So steht schon das Erstellen im Zusammenhang mit beispielsweise einer Kategorie bei Wikipedia oder Rubrik einer Zeitung. Zusätzlich unterliegen diese Texte einer strengen Qualitätskontrolle, anders als Vorschläge oder Kommentare bei Online-Partizipationsverfahren, die diese Kategorisierung einigermaßen sicherstellt.

2. VISUALISIERUNG

Daten wurden schon visualisiert, als es noch keine Computer gab. Eine Visualisierung kann zum einen eine ÄJbersicht über große Datenmengen schaffen, zum anderen können gefundene Strukturen dargestellt werden. Auch kann eine Visualisierung Menschen ermöglichen, Strukturen oder Besonderheiten der Daten zu erkennen. Die Arten der Visualisierung und deren Möglichkeiten sind vielfältig und richten sich nach Art und Struktur der Daten, Anforderungen der Benutzer und nicht selten Ästhetik. Dabei sollte die Verwendung durch den Menschen im Vordergrund stehen und zu dessen Verständnis beitragen.

Visualisierungen können interdisziplinär unter verschiedensten Aspekten betrachtet werden, wobei diese Arbeit nur einen kleinen Ausschnitt über die Möglichkeiten bieten kann. Der Hauptfokus der zu visualisierenden Daten liegt auf Texten, aber auch statistische Daten sind verwendbar. Kucher und Kerren [10] bieten eine interessante ÄJbersicht über Techniken, die zur Visualisierung von Text verwendet werden können und stellen ein Online-Tool dafür bereit⁴.

Die Daten die bei Online-Partizipation anfallen, sind hauptsächlich Textbeiträge, aber etwa auch Abstimmungsdaten. Dazu gibt es auch erhobene Zeitpunkte, etwa wann ein Beitrag erstellt wurde. Je nachdem wie viel Einsicht in ein System vorliegt, können auch Zeitpunkte von Abstimmungen erhoben oder das Suchverhalten protokolliert werden. Analog zu Cao und Cui. [1] können so folgende Ebenen der Visualisierung bei der Online-Partizipation identifiziert werden: Gesamtheit aller Dokumente, Dokumentenebene, Wortebene und Themenebene. Außerdem können zeitliche Ver-

läufe und Suchanfragen visualisiert werden.

Im folgenden werden verschiedene Techniken vorgestellt und diskutiert, auf welchen Ebenen diese im Bezug auf Online-Partizipationsverfahren verwendet werden können.

2.1 Histogramme

Histogramme bilden eine recht einfache, aber übersichtliche Visualisierung. Dabei werden Datenpunkte eines kontinuierlichen Wertebereiches in Intervalle unterteilt, dessen Wahrscheinlichkeitsverteilungen über eine Variable dargestellt werden Velleman und Hoaglin [25]. Diese Intervalle, Klassen genannt, werden über den gesamten Wertebereich der Datenpunkte gebildet. Die Addition der Häufigkeiten ergibt dann die Gesamtanzahl aller Datenpunkte. Dabei können entweder die absoluten oder die relativen Häufigkeiten betrachtet werden.

Im Zusammenhang mit Online-Partizipation können Histogramme eine ÄJbersicht über numerische Werte bieten. Etwa Abstimmungsdaten von Vorschlägen, Anzahl an Kommentaren der Vorschläge, Anzahl der Wörter in einem Dokument oder auch Nutzungsdaten von Benutzern, falls diese erhoben werden. Ein Beispiel solcher Histogramme zeigt Abbildung 1. Mit Hilfe dieser Abbildung kann beurteilt werden, wie die Anzahlen der Zustimmungen und Kommentare bei den einzelnen Vorschlägen verteilt ist, was Aufschluss über die Beteiligung gibt.

2.2 Word Clouds

Eine relativ neue Visualisierung sind Word Clouds [5], die Wörter in einer Art Wolke darstellen. Die Größe der Worte wird durch ein Gewicht bestimmt. Dies kann die Häufigkeit eines Wortes sein. Es bietet sich etwa $tfidf$ an, oder tf unter Verwendung des Logarithmus, damit häufige Wörter nicht zu stark dominieren. Außerdem können die Worte auch verschiedene Farben haben, welche hauptsächlich der einfacheren Unterscheidung dienen.

Wesentlich ist die Anordnung der Worte innerhalb der Wolke, da dies die Wahrnehmung der Wolke beeinflusst. Lohmann et al. [12] haben verschiedene Ansätze evaluiert und kommen zu dem Ergebnis, dass die Anordnung in einer Word Cloud von ihrem Verwendungszweck abhängt. Soll diese dazu dienen die Suche nach einem Wort zu vereinfachen, empfehlen sie eine sequentielle, alphabetisch geordnete Anordnung. Wird eine Word Cloud hingegen verwendet, um die häufigsten Begriffe zu visualisieren, sei eine zirkuläre Ausrichtung mit den häufigsten Begriffen im Zentrum zu bevorzugen. Schließlich zeigen Lohmann et al. [12] auch die Möglichkeit auf, Word Clouds nach Themen zu strukturieren.

Für Online-Partizipationsverfahren können Word Clouds die Häufigkeit von Wörtern auf verschiedenen Ebenen visualisieren. Word Clouds können das gesamte Verfahren oder Beiträge, aber auch gefundene Themen darstellen. Denkbar ist auch die Darstellung von Suchergebnissen, die aber wenig sinnvoll erscheint da der Suchende an konkreten Ergebnissen interessiert ist. Bei dem Einsatz von Word Clouds sollte die Zweckmäßigkeit im Auge behalten werden. Abbildung 2 zeigt ein Beispiel einer Word Cloud, deren enthaltenen Wörter zufällig angeordnet sind.

2.3 Visualisierungen mittels nicht-linearer Dimensionsreduzierung

Der nächste Teilbereich von Visualisierungen ist vergleichs-

⁴<http://textvis.lnu.se/>

ren Verfahrens gesetzt werden. Denkbar ist auch, dass zwei Korpora verglichen werden die jeweils mehrere Online-Partizipationsverfahren zusammenfassen.

Neben der Visualisierung von nicht-linearer Dimensionsreduktion, werden Streudiagramme eingesetzt, um Korrelationen zwischen zwei numerischen Features zu visualisieren.

2.5 Netzdiagramm

Netzdiagramme stellen die Ausprägungen verschiedener Dimensionen in einer runden Form da. Dabei wird jede Dimension auf einer eigenen Achse um den Mittelpunkt eines Kreises herum aufgetragen [24]. Zwar werden die Punkte auf den einzelnen Achsen verbunden, jedoch stehen die benachbarten Achsen nicht beabsichtigt nebeneinander. Sollen in einem Diagramm mehrere Objekte dargestellt werden, sollten diese visuell unterscheidbar sein, etwa durch verschiedene Farben [13]. Allerdings sollten damit nicht zu viele Objekte verglichen werden, weshalb häufig auch Diagramme für verschiedene Objekte nebeneinander gezeigt werden, wobei die Position der Achsen in jedem Diagramm gleich sein sollte.

Daten von Online-Partizipationsverfahren können damit in vielfältiger Hinsicht verglichen werden. So können Netzdiagramme dazu genutzt werden einzelne Aspekte des Verfahrens zu visualisieren, beispielsweise die Verteilung von Vorschlägen oder Kommentaren auf die einzelnen Wochentage (siehe Abbildungen 4). Die Information wann sich beteiligt wird, kann etwa für sozialwissenschaftliche Auswertungen interessant sein.

Diese zu vergleichenden Aspekte, also Dimensionen, werden anders als bei Histogrammen jedoch vorher ausgewählt. Verwendet werden Netzdiagramme auch, um die Themenverteilung eines Dokuments oder einer Suchanfrage zu visualisieren [20]. Dabei wird an jede Achse ein für das Thema aussagekräftiger Begriff vermerkt. Diese aussagekräftigen Begriffe können auch mit *Topic Labeling* [16] gefunden werden. Schließlich können mit Netzdiagrammen auch verschiedene Dokumente, Dokumentenmengen oder gesamte Online-Partizipationen visuell verglichen werden.

2.6 Visualisierung von extrahierten Themen

Extrahierte Themen zu visualisieren hat vor dem Hintergrund dieser Arbeit einen besonderen Stellenwert. Es gibt auch Arbeiten, die sich hauptsächlich damit beschäftigen Themen in unterschiedlicher Weise visualisieren, visualisierend vergleichen oder damit Interaktion visualisierend zu unterstützen. Davon werden hier einige vorgestellt.

Murdock und Allen [17] stellen eine Visualisierung von Ähnlichkeitssuchen unter Verwendung eines Topic Models vor. Dies ermöglicht dem Benutzer ähnliche Dokumente zu einem von ihm ausgewählten zu finden. Visualisiert werden die ähnlichsten Dokumente mit einem nach Ähnlichkeit sortierten Balkendiagramm. Der Charakter dieser Visualisierung ist mehr von unterstützender Natur und eher für Korpora geeignet, wo Dokumente leicht anhand ihres Titels identifiziert werden können. Somit erscheint eine Verwendung der Visualisierung sinnvoll, die dem Benutzer ähnliche Dokumente zu dem aktuell betrachteten aufzeigt und dabei gefundene Themen mit darstellt. Denkbar ist auch die Ähnlichkeit einer Anfrage zu Dokumenten in dieser Weise zu visualisieren.

LDavis, welches von Sievert und Shirley [21] entwickelt wurde, kombiniert im wesentlichen zwei Darstellungformen. Ei-

ne davon ist, die Themen anhand ihrer Ähnlichkeit im zweidimensionalen Raum darzustellen. Dazu wird eine Distanzmatrix der Dokumente mit einem dimensionsreduzierenden Verfahren, z.B. PCA, auf zwei Dimensionen reduziert [3]. Dargestellt wird jedes Thema in diesen zwei Dimensionen als Kreis, dessen Fläche sich nach dem Einfluss des Themas auf den Korpus bezieht. Kreise können sich überlappen und auch komplett in anderen Kreisen liegen, wobei kein Zusammenhang zwischen Einfluss des Themas und der Lage erkennbar ist. Themen können so andere überlagern, auch wenn sie sich wenig ähneln, nur falls eines einen großen Einfluss hat. Dieses Überlagern kann Hierarchie vortäuschen, die so nicht aus der Berechnung abgeleitet werden kann. Grundsätzlich ist *LDavis* auch für Online-Partizipationsverfahren geeignet, jedoch sollte die Art der Verwendung auch unerfahrene Benutzer nicht überfordern. In Abbildung 5 ist ein Beispiel von *LDavis* zu sehen.

Die Autoren schlagen ebenfalls Clustering innerhalb dieses zweidimensionalen Raums vor.

Die andere Darstellungsform zeigt ein Balkendiagramm der Top-Words eines ausgewählten Themas an. Jedes Wort hat zwei Balken, einen der den Anteil am Thema und einen der die Häufigkeit im Korpus wiedergibt. Bewegt man die Maus über einen Begriff, wird im Graph der Einfluss auf die Themen durch Veränderung der Fläche der Kreise angezeigt.

Ein Verfahren Begriffe innerhalb eines Themas zu visualisieren, haben Smith et al. [23] vorgestellt. Dazu werden kräftebasierte Zeichenverfahren [6] unter Einbeziehung von statistischen Daten und Daten der Themen verwendet. Jedes Thema wird als eigener Graph dargestellt, wo eine ausgewählte Anzahl von Begriffen die Knoten darstellen. Kanten verbinden zwei Knoten nur, wenn der Kookorenzwert der dazugehörigen Begriffe hoch genug ist, sie also häufig genug zusammen auftreten. Zusätzlich beschreibt die Fläche der Knoten den Zugehörigkeitswert eines Begriffs zu einem Thema. Die Graphen der Themen werden dann nach Kovarianz zwischen den Themen angeordnet.

Diese Art der Visualisierung ist nur begrenzt sinnvoll, da sie schon bei einer begrenzten Anzahl an Themen und kräftebasierten Graphen mit vielen Kanten unübersichtlich werden können.

2.7 Simple Formen der Visualisierung

Sinnvoll sein kann der Einsatz von simplen Formen der Visualisierung. Diese sind nicht selten intuitiver oder werden auch ohne viel Vorwissen verstanden.

Balkendiagramme sind nicht die schlichteste Visualisierungsform, jedoch lassen sie sich analog zu Netzdiagrammen verwenden, so dass sich ähnliche Überlegungen ergeben. Anders als Histogramme, visualisieren Balkendiagramme feste Variablen, dessen Ausprägungen dargestellt werden. Auch müssen Balkendiagramme nicht die Häufigkeit darstellen. Wie bereits erörtert, können Balkendiagramme auch horizontal angewendet werden und die Balken aus mehreren Segmenten bestehen.

Eine viel verwendete und leicht zugängliche Methode extrahierte Themen zu visualisieren, sind die jeweiligen Top-Wörter als Liste darzustellen, absteigend nach Zugehörigkeit zum Thema. Solche Wortlisten müssen nicht weiter erklärt werden.

2.8 Schlussfolgerung

Betrachtet man die vorhandenen Methoden zur Visuali-

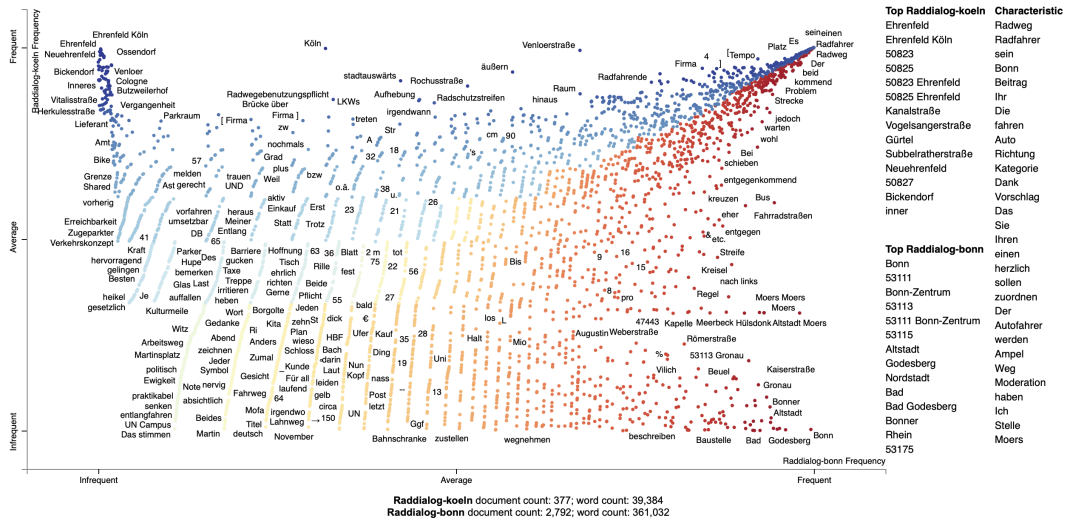


Abbildung 3: Screenshot von *Scattertext*: Vergleich der Online-Partizipationsverfahren der Raddialoge in Bonn und Köln-Ehrenfeld.

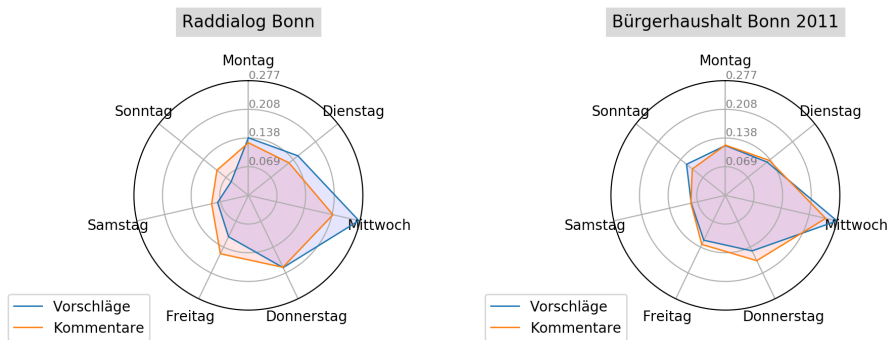


Abbildung 4: Relative Häufigkeiten der Vorschläge und Kommentare des Raddialog Bonn und des Bonner Bürgerhaushalt 2011 an den Wochentagen.

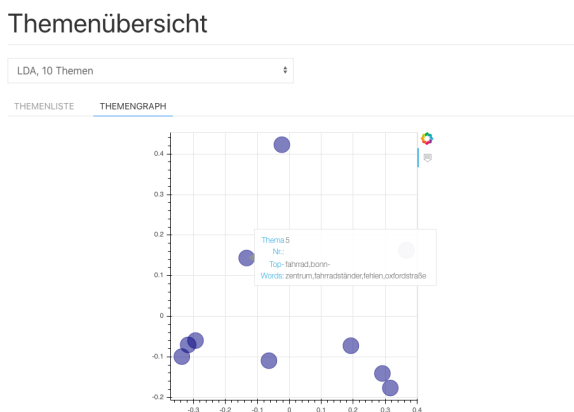


Abbildung 5: Screenshot einer 2D-Darstellung des Modells von extrahierten Themen

sierung fällt ins Auge, dass vor allem Metadaten visualisiert werden können. Numerische Daten wie Abstimmungszahlen, Anzahl an Kommentaren oder zeitliche Daten können mit gängigen Methoden visualisiert werden. Dazu gehören Graphen, Diagramme und Histogramme.

Texte zu visualisieren ist dagegen eine Herausforderung. Einerseits können simple Wortlisten oder Wortwolken verwendet werden, andererseits können Methoden angewendet werden, um extrahierte Themen zu visualisieren. Diese Visualisierungen sind oft komplex oder bieten wenig Vorteile gegenüber einer textbasierten Darstellung.

Schließlich kann eine gute Visualisierung auch bedeuten, Suchergebnisse oder Beiträge verständlich und übersichtlich darzustellen.

3. FAZIT UND AUSBLICK

Online-Partizipationsverfahren maschinell zu analysieren, kann Teilnehmenden und Betreuenden einen Einblick in und eine AIBersicht über das Verfahren geben. Beides kann helfen Verfahren so zu gestalten, dass sich mehr Menschen beteiligen und diese Beteiligung besser ausgewertet wird. Eine zentraler Wunsch bei Online-Partizipation ist, die Akzeptanz von Entscheidungen bei den Betroffenen zu steigern.

Zur Themenextraktion bieten sich neben der state-of-the-art Methode LDA ebenfalls NMF und LSI an. Diese extrahierten Themen werden mit einer sortierten Menge von Wörtern assoziiert. Damit kann eine ÄJbersicht über die Beiträge gegeben werden. Die Ergebnisse, also gefundenen Themen sind jedoch stark subjektiv [2]. Da Wortlisten extrahierter Themen nicht alle Aspekte eines Online-Partizipationsverfahrens abdecken können, wurden Visualisierungen vorgestellt. Neben Darstellungen für Metadaten, gibt es auch Visualisierungen für extrahierte Themen oder Modelle. Diese sind leider oft kompliziert oder zeigen selten neue Aspekte auf. Weiterhin ist die Beurteilung einer Visualisierung stark subjektiv. Visualisierungen sollten intuitiv, informativ und leicht zugänglich sein.

In Zukunft bietet sich weitere Forschung an, um extrahierte Themen intuitiv und interaktiv darzustellen. Außerdem können diese Visualisierungen in Studien evaluiert werden, um die Verständlichkeit zu beurteilen.

4. REFERENCES

- [1] N. Cao and W. Cui. Introduction to text visualization. In *Atlantis Briefs in Artificial Intelligence*, 2016.
- [2] J. Chang, J. L. Boyd-Graber, S. Gerrish, C. Wang, and D. M. Blei. Reading tea leaves: How humans interpret topic models. In *Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems 2009. Proceedings of a meeting held 7-10 December 2009, Vancouver, British Columbia, Canada.*, pages 288–296, 2009.
- [3] J. Chuang, D. Ramage, C. D. Manning, and J. Heer. Interpretation and trust: designing model-driven visualizations for text analysis. In *CHI*, 2012.
- [4] T. Escher and B. Rottinghaus. Local government platforms for citizen participation and their effects on legitimacy. 2018.
- [5] J. Feinberg. Wordle. In J. Steele and N. Iliinsky, editors, *Beautiful visualization: looking at data through the eyes of experts*, chapter 3. O’Reilly Media, Inc., 2010.
- [6] T. M. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164, 1991.
- [7] O. Jin, N. N. Liu, K. Zhao, Y. Yu, and Q. Yang. Transferring topical knowledge from auxiliary long texts for short text clustering. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 775–784. ACM, 2011.
- [8] J. S. Kessler. Scattertext: a browser-based tool for visualizing how corpora differ. *CoRR*, abs/1703.00565, 2017.
- [9] G. Klir and B. Yuan. *Fuzzy sets and fuzzy logic*, volume 4. Prentice hall New Jersey, 1995.
- [10] K. Kucher and A. Kerren. Text visualization techniques: Taxonomy, visual survey, and community insights. In *2015 IEEE Pacific Visualization Symposium (PacificVis)*, pages 117–121, 2015.
- [11] M. Liebeck, K. Esau, and S. Conrad. Text Mining für Online-Partizipationsverfahren: Die Notwendigkeit einer maschinell unterstützten Auswertung. *HMD Praxis der Wirtschaftsinformatik*, 54(4):544–562, 2017.
- [12] S. Lohmann, J. Ziegler, and L. Tetzlaff. Comparison of tag cloud layouts: Task-related performance and visual exploration. In *IFIP Conference on Human-Computer Interaction*, pages 392–404. Springer, 2009.
- [13] R. B. Lydiard, K. Rickels, B. Herman, and D. E. Feltner. Comparative efficacy of pregabalin and benzodiazepines in treating the psychic and somatic symptoms of generalized anxiety disorder. *International Journal of Neuropsychopharmacology*, 13(2):229–241, 2010.
- [14] L. v. d. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.
- [15] L. McInnes and J. Healy. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv e-prints*, Feb. 2018.
- [16] Q. Mei, X. Shen, and C. Zhai. Automatic labeling of multinomial topic models. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 490–499. ACM, 2007.
- [17] J. Murdock and C. Allen. Visualization techniques for topic model checking. In *AAAI*, pages 4284–4285, 2015.
- [18] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *science*, 290(5500):2323–2326, 2000.
- [19] J. W. Sammon. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, C-18:401–409, 1969.
- [20] S. Sasaki, K. Yoshii, T. Nakano, M. Goto, and S. Morishima. Lyricsradar: A lyrics retrieval system based on latent topics of lyrics. In *Ismir*, pages 585–590, 2014.
- [21] C. Sievert and K. Shirley. Ldavis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, pages 63–70, 2014.
- [22] V. D. Silva and J. B. Tenenbaum. Global versus local methods in nonlinear dimensionality reduction. In *Advances in neural information processing systems*, pages 721–728, 2003.
- [23] A. Smith, J. Chuang, Y. Hu, J. L. Boyd-Graber, and L. Findlater. Concurrent visualization of relationships between words and topics in topic models. 2014.
- [24] N. R. Tague et al. *The quality toolbox*, volume 600. ASQ Quality Press Milwaukee, WI, 2005.
- [25] P. F. Velleman and D. C. Hoaglin. *Applications, basics, and computing of exploratory data analysis*. Duxbury Press, 1981.
- [26] X. Yan, J. Guo, Y. Lan, and X. Cheng. A biterm topic model for short texts. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1445–1456. ACM, 2013.