

KI-gestützter Wordcloud-Generator für Beteiligungsprozesse

Carsten Oldenburg, Axel Häusler

(MSc. Carsten Oldenburg, Technische Hochschule Ostwestfalen-Lippe, nextPlaceLab, carsten.oldenburg@th-owl.de)
(Prof. Dr. Axel Häusler, Technische Hochschule Ostwestfalen-Lippe, nextPlaceLab, axel.haeusler@th-owl.de)

1 ABSTRACT

Es liegt auf der Hand, dass eine digitale Unterstützung von Planungs- und Beteiligungsverfahren in vielfacher Hinsicht enorme Vorteile bietet. So können mittels moderner, digitaler Partizipationsplattformen Prozessbeteiligte orts- und zeitunabhängig an städtebaulichen Ideenfindungs- und Bewertungsverfahren teilnehmen und ihre Gedanken, Meinungen und Vorschläge mit anderen teilen und diskutieren. Seit einigen Jahren stehen hierfür eine Reihe adaptierbarer Softwareprodukte zur Verfügung, z. B. Consul, ein community-basiertes Opensource-Projekt der Consul Democracy Foundation auf GitHub, die proprietäre Software citizenLab des gleichnamigen belgischen Unternehmens, oder dem neuseeländischen Pendant Loomio der Loomio Cooperative Ltd. und viele weitere. Der webbasierte Zugang ermöglicht dabei nicht nur eine potenzielle Reichweitensteigerung an Teilnehmenden und die schnelle Verlinkung zu anderen digitalen Inhalten bzw. Medien, sondern erleichtert auch die statistische Informationsauswertung und die mediale wie inhaltliche Dokumentation des Gesamtprozesses. Aktuelle Softwarelösungen sind dabei als anwenderfreundliches Baukastensystem konzipiert, das je nach Anwendungsfall individuell, modular und ohne Programmierkenntnisse zusammengesetzt werden kann. Die zuschaltbaren Module reichen von einfachen Formularmasken über interaktive Karten-Tools, MindMaps und Umfragen bis hin zu integrierten Video-Chat-Funktionen und kollaborativen Whiteboards.

Zukünftig ist davon auszugehen, dass die modulare Struktur und die enorm vielfältigen Einsatzgebiete dieser Softwarelösungen zunehmend auch KI-gestützte Funktionen als neue Features enthalten werden bzw. im Baukasten bestehende Module optimieren oder ablösen werden. Die Gründe hierfür liegen größtenteils im disruptiven Fortschritt der Softwareentwicklung. Andererseits darf aber auch erwogen werden, ob nicht doch häufig beobachtete Hemmnisse oder Probleme bisheriger Partizipationsverfahren ggf. durch den unterstützenden Einsatz von KI auch abgebaut oder verringert werden könnten.

Beide Perspektiven stellen für sich genommen schon sehr breite Grundlagenforschungsfelder dar, die insbesondere durch die noch hinzukommenden Aspekte der Technologieakzeptanz enorm komplex werden können. Da aber die technologische Hürde zur Umsetzung einfacher Software-Prototypen durch die Vielzahl zur Verfügung stehender Opensource-Tools sehr niedrig ist, entwickelte der Forschungsschwerpunkt nextPlace der Technischen Hochschule Ostwestfalen-Lippe zunächst eine allererste, prototypische Hardware-Software-Applikation, um - im Sinne eines Proof-of-Concept - die Relevanz und Aufwände tiefergehender Forschungs- und Entwicklungsarbeiten abschätzen zu können. Folglich stellen die nachfolgenden Ausführungen einen technischen Erfahrungsbericht der ersten Entwicklungsschritte dar, um einen einfachen, kostengünstigen und experimentellen Zugang in dieses noch recht junge Forschungsfeld nachvollziehbar zu machen.

Keywords: Data Visualisation, Participation, Speech Recognition, Artificial Intelligence, Internet of Things

2 HINTERGRUND UND MOTIVATION

Wordclouds sind ein bekanntes Werkzeug der Informationsvisualisierung. Sie bieten den Vorteil, dass sich Textinhalte prägnant und zusammenfassend schnell und visuell erfassen lassen. In der Regel begegnet man Wordclouds am Anfang oder Ende von Folien-Präsentationen, um die inhaltliche Bandbreite grafisch darzustellen, die wichtigsten Aspekte der enthaltenen Information hervorzuheben oder einfach als gestalterisches Element in jeglichen Publikationsformen. Ein derartiger Nutzen kann aber auch für, in Präsenz oder digital stattfindenden Workshops oder Besprechungen von Vorteil sein, um aktuelle Diskussionsinhalte für Teilnehmerinnen, Teilnehmer und Außenstehende strukturiert und medial publizierbar widerzuspiegeln. Hierfür wäre es dann allerdings von Vorteil, wenn die Wordcloud einerseits dem Stand der Diskussion in Echtzeit folgen und auch vorhergehende Diskussionstände archivieren könnte, um inhaltliche Redundanzen aufzuzeigen oder ggf. auch Vergleiche zwischen unterschiedlichen Diskussionen herstellen zu können. Damit dies möglich wird, gilt es, die Diskussionsinhalte mehrerer Teilnehmerinnen und Teilnehmer als Tonaufnahme mitzuschneiden und daraus mittels KI-gestützter

Spracherkennung automatisch, in Echtzeit, die wichtigsten Begriffe zu extrahieren und in Form einer digitalen Schlagwortwolke grafisch aufzubereiten.

Technisch gesehen, sind auf der Eingabeseite hierfür einfache Textformate, z. B. Begriffslisten, notwendig, die üblicherweise am häufigsten vorkommenden Schlagwörter vereinen. Neben dem hier beschriebenen Testeinsatz der Spracherkennungs-KI im Rahmen zweier Veranstaltungen gehen die denkbaren Verwendungsmöglichkeiten der Textvisualisierung aber weit darüber hinaus. So unterstützen die Schlagwörter nicht nur externe zugeschaltete oder verspätet eintreffende Teilnehmer beim schnellen Auffassen des aktuellen Diskussionsstands oder der begleitenden Dokumentation, sondern ermöglichen auch zeitliche Veränderungen des Gesprächsablaufs oder temporäre Änderungen der inhaltlichen Prioritäten objektiv messbar zu machen.

In Verbindung mit einer dazu verknüpften, semantischen Netzwerkstruktur ließe sich so, mittels Machine-Learning-Algorithmen, eine kontinuierliche Informations- & Wissensstruktur aufbauen, die einerseits durch fortwährende Diskussionsbeiträge wächst und andererseits auch redundante Diskussionsphasen erkennt oder an geeigneter Stelle neue inhaltliche Verknüpfungen oder Inspirationen aus anderen Beteiligungsprozessen einbringen kann. Als allererster Schritt in dieses mächtige Anwendungsfeld wurde in diesem Technologieprojekt zunächst auf experimentelle Weise mit der Verknüpfung einfacher Hardware- und offen verfügbarer Softwarekomponenten begonnen, wobei einschränkend hinzugefügt werden muss, dass eine weitreichende Erprobung und qualitative Bewertung allerdings noch ausstehen.

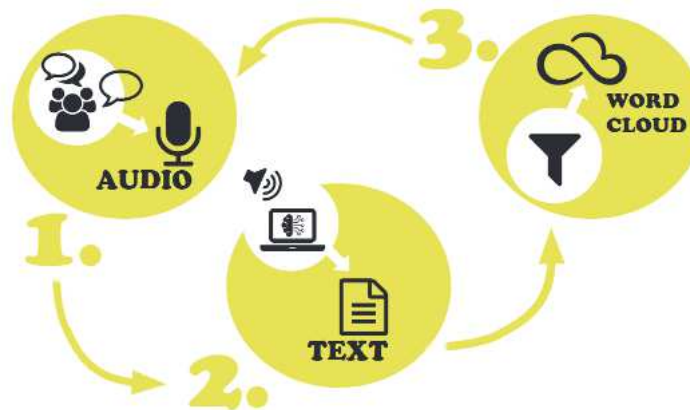


Abb. 1: Grundprinzip der Applikation [Jacobi 2020]

2.1 Stand der technischen Forschung

Erste technisch vergleichbare Konzepte wurden schon vor einigen Jahren u.a. von Haubold (2004) oder auch Flamary et al. (2011) beschrieben. Verwandte Konzepte aus jüngster Zeit finden sich u.a. bei Padmanandam et al. (2021) oder Wu et al. (2020). Bestehende Software-Unternehmen wie IBM, Google oder auch Facebook bieten bereits niederschwellige Zugänge zu intelligenten Spracherkennungssystemen und kombinierbaren Textanalyse-Tools. Das vermutlich prominenteste Beispiel ist die KI-Plattform Watson von IBM, in der die Spracherkennungsfunktionen online als Software-as-a-Service prozessiert werden. Neben diesen proprietären Diensten bietet aber z.B. auch die Mozilla Foundation ein opensource-basiertes Ökosystem zur Spracherkennung und Sprachsynthese mittels diverser Deep-Learning Algorithmen an (vgl. Ardila et al. 2019). Das Projekt ist in sofern von besonderem Interesse, da es im Rahmen der Mozilla-CommonVoice-Kampagne, die Community anregt, ihre eigenen Sprachaufnahmen an ein offenes Repository zu spenden, um die Spracherkennungsbibliotheken fortwährend trainieren zu können.

Prinzipiell ist der Nutzen von Speech-to-Text-Services schon vielfach beschrieben worden und grundsätzlich seit langem erprobt. Der Einsatz im Rahmen von Gruppengesprächen oder Workshops benötigt aber wesentlich performantere Serverkapazitäten, die erst in jüngster Zeit einer breiten Anwenderschicht zur Verfügung stehen. Neuere Konzepte zum Einsatz von Echtzeit-Spracherkennung finden sich daher vornehmlich im Bildungskontext, zur ergänzenden Verständnisunterstützung Beeinträchtigter, wie Google's Project Euphonia oder auch als Dokumentationsfunktion, gekoppelt an Videokonferenzsysteme, wie beispielsweise innerhalb der Software Zoom.

In Hinblick auf den Einsatz KI-gestützter Verfahren innerhalb öffentlicher Beteiligungsprozesse, lieferte in jüngster Zeit insbesondere auch das Verbundprojekt Civitas Digitalis, der Universitäten Hamburg und Kassel mit der Fortiss GmbH und Stadt Kassel wichtige Erkenntnisse (Balta et al. 2019). Insbesondere konnten hier interessante Grundlagen zur Nutzerakzeptanz digitaler Beteiligungsformate und daraus abgeleitete Handlungsempfehlungen zusammengestellt werden. Allerdings wurde in diesem Projekt durch die Verwendung eines textbasierten Chat-Bots ein verwandter aber etwas anderer technologischer KI-Ansatz gewählt.

3 UMSETZUNG

Im hier beschriebenen Projekt lag der Fokus vor allen Dingen auf einer einfachen Do-It-Yourself-Technologie, möglichst geringem Kostenaufwand und darauf, mittels OpenSource-basierter Software eine erste Schnittstelle zur Nutzbarmachung weiterführender KI-Bausteine herzustellen. Auf technischer Seite wurde ein kostengünstiger Einplatinen-Computer (Raspberry-Pi) verwendet, der die gesprochenen Workshop-Inhalte mittels einer zusätzlich notwendigen Sensorplatine aus acht Mikrofonen (MatrixCreator-Shield) über die GPIO-Schnittstelle (general purpose input/output) aufnimmt. Dies ermöglicht Aufnahmen mit bis zu acht Kanälen. Zu beachten ist allerdings, dass Nebengeräusche und Lärmquellen die Qualität der Aufnahme erheblich reduzieren können. Die Programmierung dieses DIY-Aufnahmegeräts erfolgte durch eigene Skripte auf Basis passender Python-Bibliotheken. Zur Umwandlung der Sprachaufnahmen in Text kam die Open-Source- Spracherkennungsbibliothek Kaldi (Povey et al. 2011) zum Einsatz. Diese Bibliothek bietet umfangreiche Softwarewerkzeuge zur Spracherkennung und Signalverarbeitung.

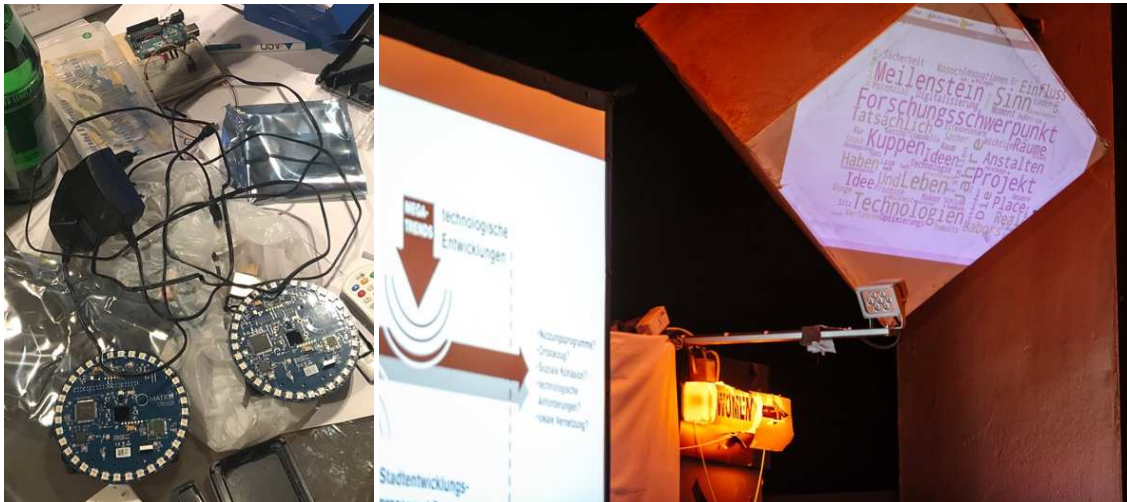


Abb. 2: Bauelemente des DIY-Aufnahmegeräts und Einsatz im Rahmen der Workshop-Veranstaltung 02/2020

Kaldi selbst ist zwar in C++ geschrieben, jedoch ist eine Einbindung in Python durch eine Wrapperbibliothek problemlos möglich. Da es sich bei Kaldi um freie Software handelt, kann diese ohne weitere Kosten auf einem eigenen Server aufgesetzt werden. Dies ist insbesondere datenschutzrechtlich von Vorteil, da auf diesem Wege nicht die Notwendigkeit besteht, Serverinfrastrukturen im Ausland zu verwenden, wie es bei den Cloudtechnologien von Microsoft (Azure) oder IBM (Watson) der Fall wäre. Da die Rechenleistung des Raspberry Pis nicht ausreichend ist für eine zeitnahe Spracherkennung, wird die Spracherkennungssoftware Kaldi unter einem eigens für das Projekt aufgesetzten Linux-Cloudserver betrieben.

Aufgrund beschränkter Performanceressourcen werden die Audioaufnahmen derzeit nicht in unmittelbarer Echtzeit, verarbeitet. Sondern Aufnahmen von 20 Sekunden Länge werden vom Raspberry Pi packetweise an den Kaldi-Server geschickt. Vor der eigentlichen Spracherkennung müssen die Daten noch vorprozessiert werden. Genauer gesagt gilt es, die Audioaufnahmen auf die von Kaldi verwendete Frequenz von 16 kHz herunterzusampeln und die Audiokanäle auf ein Mono-Format zusammenzufassen. Im Anschluss werden die Sprachaufnahmen durch Kaldi prozessiert und die Sprache in Text umgewandelt.

Eine wichtige Limitierung bei diesem Prozess stellt das Sprachmodell dar, dessen Qualität einen hohen Einfluss auf das Ergebnis der Spracherkennung hat. Da es nur sehr wenige, frei verfügbare Modelle für die deutsche Sprache gibt, wurde ein Opensource-Sprachmodell des Fachbereichs Informatik der Universität Hamburg verwendet (Milde, Köhn 2018). Prinzipiell wäre auch die Erzeugung eines eigenen Modells

denkbar, allerdings würden hierfür, neben den sehr zeitraubenden Trainingsdurchläufen (viele Tage bis Wochen) auch mehrere hundert Stunden an Basis-Trainingsdaten benötigt werden.

3.1 Schlagwortidentifizierung

Nach erfolgreicher Datenumwandlung in Text besteht jedes Prozesspaket aus mehr oder weniger vollständigen Sätzen, die sich insbesondere durch eine besondere Häufung von nicht-brauchbaren Wortelelementen, wie Artikel, Pronomen, Präpositionen, etc. auszeichnen. Mittels eines einfachen Textparsers in Python wurde diese Teile aus den Textdaten eliminiert, um im Anschluss die wichtigsten Hauptwörter abfragen und in das Skript zur Erstellung der WordCloud implementieren und dort nach Häufigkeit gewichten zu können. Im Abstand von 20 Sekunden erzeugte das Skript folglich eine neue, dem letzten Sprachverlauf folgende Wordcloud, die ihrerseits wieder durch Beamerprojektion in den Veranstaltungsraum zurückgespiegelt wurde.

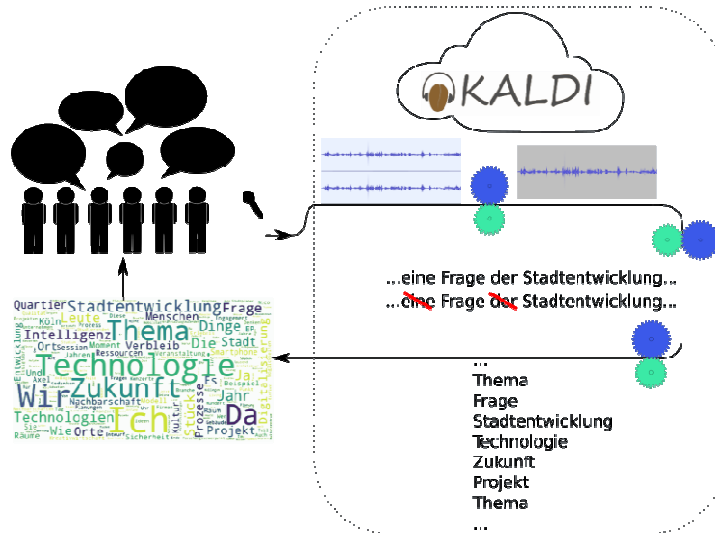


Abb. 3: Schematische Darstellung der Schlagwortidentifizierung und Resultat als Beamerprojektion (eigene Darstellung)

Gleichermaßen wurde in erster rudimentärer Form während Probedurchläufen versucht, thematisch unpassende Begriffe ebenfalls auszusortieren, um die visuelle Informationsdichte der WordCloud möglichst nah am thematischen Schwerpunkt des Veranstaltungsthemas zu halten. Allerdings wurde genau an dieser Stelle relativ schnell klar, dass exakt hier entweder die Grenzen im nutzbringenden Einsatz der Technologie oder die Potenziale zur inhaltlichen Weiterentwicklung liegen. Hierauf soll im Kap.4 noch weiter eingegangen werden.

3.2 Bislang beobachtete Hemmnisse

Während unserer Testworkshops wurden mit der hier beschriebenen Konfiguration gleichzeitig die Diskussionen von zwei räumlich getrennten Workshopgruppen aufgezeichnet, um diese später inhaltlich und zeitlich analysieren zu können. Leider stellte sich im Anschluss der Diskussionen heraus, dass die Aufnahmequalität der Sprachaufnahmen mit der Matrix Creator Shield für die weitere Analyse zu schlecht waren und in Folge zu einer sehr geringen Worterkennung resultieren. Folglich eignet sich das hier besprochene Setting für eine einfache Raumaufnahme ohne individuelle Mikrofone nur bedingt.

Für einen weiteren Veranstaltungstermin mit größerer Besucherzahl wurde an dem Raspberry Pi eine externe Soundkarte per USB angeschlossen, die direkt mit dem Mischpult für eingesetzte Funkmikrofone zur Audioaufnahme verbunden wurde. Die direkte Verbindung der Soundkarte mit dem Audiosignal aus dem Veranstaltungsmischpult ermöglichte es, Nebengeräusche weitestgehend zu eliminieren, was wiederum in einer weitaus höheren Worterkennung resultierte.

Derzeit wird die Wordcloud noch auf dem Raspberry Pi generiert. Eine Generierung auf dem Server, der bereits die Spracherkennung durchführt, würde den Raspberry Pi weiter entlasten. Weiterhin wäre es möglich, die Spracherkennung online durchzuführen, d.h. die Audioaufnahmen werden nicht in Paketen mit einer Aufnahmelänge von 20 Sekunden an den Server geschickt, sondern direkt, ohne diese zwischenspeichern, an den Server weitergeleitet. Dies würde gewährleisten, dass Wörter während der

Aufnahme nicht auseinandergeschnitten werden und so von der Spracherkennungssoftware nicht korrekt erkannt werden würden.

4 KONTEXTERWEITERUNG

Wie unter 2.1 und 3.1. bereits angedeutet, stellt zwar die reine Spiegelung des kurzfristig Gesagten ein interessantes und kurzweiliges Ergänzungsfeature klassischer Veranstaltungsformate dar. Allerdings muss auch attestiert werden, um mittels Echtzeit-Spracherkennung der Veranstaltung eine neue und inhaltlich weiterführende Informationsebene anzudienen, reicht die rein nach Häufigkeit gewichtete Textanalyse nicht aus. Wesentlich interessanter wäre es, wenn die besprochenen Inhalte in ihrem Kontext automatisch interpretiert bzw. bestimmten (stadtentwicklungs-)methodischen Analyseebenen zugeordnet werden könnten. Folglich muss der bislang reine Textdatensatz um eine gewisse Semantik ergänzt oder in diese eingebettet werden, um in der Lage zu sein, aus den Häufigkeiten bestimmte Bedeutungsattribute abzuleiten und folglich auch wiederkehrende Verknüpfungen in der Diskussion oder Argumentation zu erkennen. Auch wären hierdurch besonders interessante Vergleichsanalysen zwischen verschiedenen Beteiligungsprozessen oder auch unterschiedlichen Stadtentwicklungsdiskussionen denkbar.

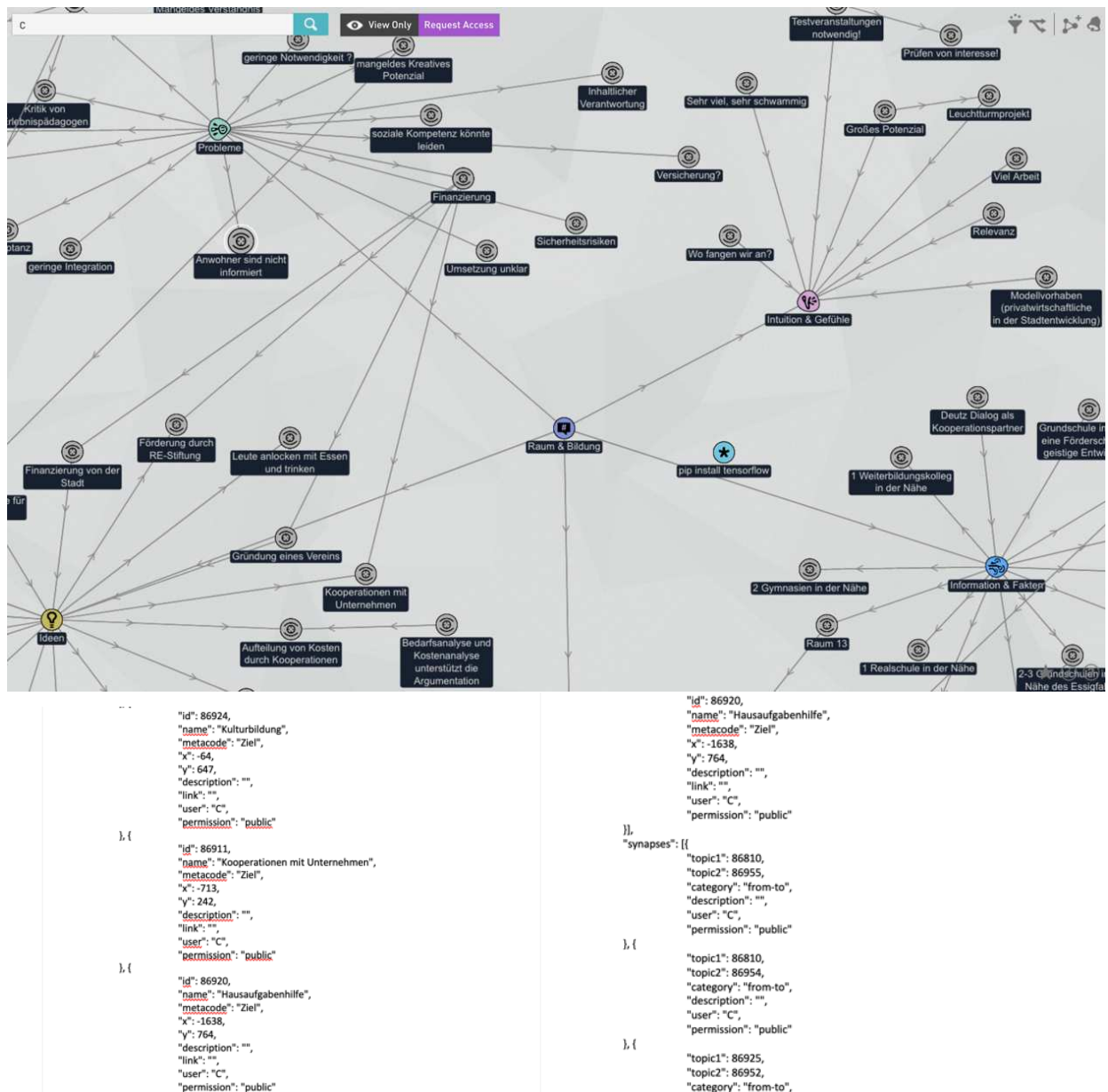


Abb. 4: Visualisierung der interaktiven MindMap und Auszug des Exports im JSON-Format (eigene Darstellung)

Um sich diesem, doch komplexen Thema in einem ersten Schritt nähern zu können, wurden in den gleichen Workshopveranstaltungen, in denen die Sprachaufnahmen als Speech-to-Wordcloud prozessiert wurden, zunächst interaktiv digitale MindMaps erzeugt, die relevante Begriffe quasi „per Hand“ einer

Kontextkategorie zugeordnet haben. Ausgangspunkt dieses sehr rudimentären Kontextmodells bildete das Koordinatensystem der bewährten SWOT-Analyse aus Stärken, Schwächen, Potenzialen und Hemmnissen. Um das inhaltliche Interpretationsspektrum des Kontextmodells noch zu erweitern, ergänzten wir die vier Bedeutungskategorien um weitere Kontextrubriken, wie z.B. Idee, Problem, Fakt, gesetzliche Vorgabe, Location, Ressource, Partner, etc.. Zur digital-interaktiven Erstellung der MindMaps kam die freie Software MetaMaps zum Einsatz, mit der sich eine klassifizierte Verknüpfung der diskutierten Inhalte kollaborativ und web-basiert erzeugen lässt. Diese Softwarebibliothek wurde auch deshalb gewählt, da die erstellten MindMaps als JSON-Datei exportiert werden können und damit folglich die Grundlage für ein maschinenlesbares und routingfähiges Knoten-Kanten-Modell darstellen. Dieses wird als Basis für das noch zu erstellende, semantische Netz dienen, mit dem die Sprachinhalte später weiter ausgewertet sollen.

Da sich viele Gesprächsinhalte auch mit konkreten Orten im diskutierten Stadtgebiet beschäftigten, wurde eine zweite, räumlich orientierte Kontextebene als kartografische Verknüpfung relevanter Orte hinzugefügt. Auf einem großen TouchTable wurde die kollaborative, auf den Softwareframeworks Django und Leaflet basierende Python/JavaScript-Bibliothek uMap, die eine einfache, der MindMap-ähnliche, aber georeferenzierte Zeichnung von Verbindungslinien zwischen relevanten Kartenpositionen ermöglicht. Die so erzeugten Geodaten können ebenfalls mit zusätzlichen Attributen verknüpft werden und wiederum im JSON-Format abgespeichert und exportiert werden.



Abb. 5: Diskussion der räumlichen Wechselwirkungen via WebMapping-Tool auf dem Touchscreen

Im weiteren Verlauf gilt es nun die beiden JSON-Datensätze durch ein geeinigtes Feature-Schema inhaltlich und datentechnisch so zu verknüpfen, dass damit ein routing-fähiges Knoten-Kanten-Netz aus relevanten, wiederkehrend Bedeutungskategorien, einschließlich einer ersten Verbindungsstruktur erzeugt werden kann. Diese Arbeiten, einschließlich der Generierung des semantischen bis neuronalen Netzwerks und der automatisierten Abfrage der Begriffe aus der Spracherkennung sind allerdings noch gänzlich zu erbringen.

5 AUSBLICK

Der hier beschriebene Prototyp bildet ein Teilprojekt eines mehrjährigen EFRE-Forschungsvorhabens zur Erforschung KI- & AR-gestützter Technologien in Stadtentwicklungsprozessen. Zum aktuellen Zeitpunkt liegen die Komponenten des hier beschriebenen KI-Konzepts nur in prototypischer Form und als unverknüpfte Softwareskripte vor. Im weiteren Verlauf ist vorgesehen, die jeweils analysierten

Schlagworteinhalte so in die künftige Semantik zu implementieren, dass diskutierte Chancen und Risiken verglichen, statistisch analysiert und qualitativ bewertet werden können. Ziel ist es, diskutierte Wechselwirkungen und Verflechtungen städtebaulicher Potenziale und Hemmnisse zu dokumentieren, zu untersuchen und ggf. Scheinargumente oder Self-fulfilling Prophecies frühzeitig durch lernende Vergleichsanalysen erkennen zu können.

Neben den beschriebenen softwaretechnischen Aufgaben sind aber darüber hinaus auch noch wichtige datenschutzrechtliche Rahmenbedingungen abzuklären und generell der individuelle Umgang von Projektbeteiligten mit der KI oder auch die daran geknüpften Erwartungen zu studieren.

6 REFERENCES

- Ardila R., Branson M., Davis K., Henretty M., Kohler M., Meyer J., Morais R., Saunders L., Tyers F.M., and Weber G., "Common voice: A massively-multilingual speech corpus," arXiv preprint arXiv:1912.06670, 2019
- Balta D., Krcmar H., Kuhn P., Kulus D., Sellami M.: "Digitalgestützte Bürgerbeteiligung & KI – Beispiele, Chancen, Herausforderungen" in PLAERIN; 2019(1), pp.19-22; 02/2019
- Flamary R., Anguera X. and Oliver N., "Spoken WordCloud: Clustering recurrent patterns in speech," 2011 9th International Workshop on Content-Based Multimedia Indexing (CBMI), 2011, pp. 133-138, doi: 10.1109/CBMI.2011.5972534.
- Haubold A., "Analysis and visualization of index words from audio transcripts of instructional videos," IEEE Sixth International Symposium on Multimedia Software Engineering, Miami, FL, USA, 2004, pp. 570-573, doi: 10.1109/MMSE.2004.27.
- Milde, B., Köhn, A.: Open Source Automatic Speech Recognition for German, Language Technology and Natural language Systems2group, FB Informatik, Universität Hamburg, Proceedings of 13th ITG Conference on Speech Communication, Oldenburg, 2018; urn:nbn:de:gbv:18-228-7-2433
- Mucha, H., Jacobi, R. & Robert, S., (2019). Partizipation und Künstliche Intelligenz. Mensch und Computer 2019 - Workshopband. Bonn: Gesellschaft für Informatik e.V.. DOI: 10.18420/muc2019-ws-413
- Padmanandam K., Bheri S. P. V. D. S., Vegesna L. and Sruthi K., "A Speech Recognized Dynamic Word Cloud Visualization for Text Summarization," 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021, pp. 609-613, doi: 10.1109/ICICT50816.2021.9358693.
- Povey D., Ghoshal A., Boulianne, G., Burget L., Glembek O., Goel N., Hannemann M., Motlíček P., Qian Y., Schwarz P., Silovský J., Stemmer G., Vesel K. (2011). The Kaldi speech recognition toolkit. IEEE 2011 Workshop on Automatic Speech Recognition and Understanding.
- Wu, T.H., Zhao, Y., & Amiruzzaman, M., Interactive Visualization of AI-based Speech Recognition Texts. EuroVA@Eurographics/EuroVis., 2020 doi:10.2312/eurova.20201091
- Weblinks:
 IBM Watson Speech-to-Text: <https://www.ibm.com/cloud/watson-speech-to-text>; abgerufen 24.07.2021
 Mozilla CommonVoice Project: <https://commonvoice.mozilla.org/de>; abgerufen 24.07.2021
 Google Project Euphonia: <https://sites.research.google/euphonia/getting-involved/>; abgerufen 24.07.2021
 Zoom Audiotranscription: <https://support.zoom.us/hc/en-us/articles/115004794983-Automatically-Transcribe-Cloud-Recordings> abgerufen 24.07.2021
 metamaps Github Project: <https://github.com/metamaps/metamaps>; abgerufen 24.07.2021
 uMap Github Project: <https://github.com/umap-project/umap>; abgerufen 24.07.2021
 EFRE.NRW LivingLab Essigfabrik: <https://livinglab-essigfabrik.eu/>; abgerufen 24.07.2021