iDOKS: Ein integriertes Dokumentationssystem zur Zusammenfassung von Gesprächen und Meetings

Robert Wardenga[1,*] *Daniel Vogel*[2,*] *Felix Gräßer*[3,*] *Mira Schielke*[2], *Leonard Starke*[2], *Rico Petrick*[4], *Torsten Rex*[2], *Jens Lehmann*[1,†]

[1]*Institut für Angewandte Informatik e.V.,* [2]*MediaInterface GmbH,* [3]*Linguwerk GmbH,* [4]*LASA – Lausitz Advanced Scientific Applications*

Kurzfassung: Dieser Beitrag stellt einen Zwischenstand des durch das BMBF-Programm „Mensch-Technik-Interaktion: Technik zum Menschen bringen“ geförderten Projektes „iDOKS“ vor. Ziel ist es, Spracherkennung sowie Sprechertrennung und -identifikation mit Textzusammenfassung (sog. Verdichtung, engl. Summarization) zu verbinden. Dies soll die aufwändige Dokumentation von Multi-Sprecher-Szenarien verbessern, die beispielsweise nach Meetings anfällt. Für die ESSV demonstriert ein einfacher Dialog aus der FAZ die Funktionalität (siehe Anhang A). Hier werden eine WER von 7,4 % und eine DER von 1,4 % erreicht. Der ROUGE-L-Wert liegt bei 13. Die Werte dieses Beispiels bewegen sich in der Größenordnung aktueller wissenschaftlicher Publikationen und praktischer Erfahrungswerte.

1 Einleitung

Der technologische Fortschritt der letzten Jahre erlaubt die Entwicklung immer größerer Modelle für die Spracherkennung und andere NLP-Bereiche. Aktuell erreicht das Open-Source-Spracherkennungssystem Whisper mit multilingual trainierten Modellen eine WER von 6,4 % bzw. 5,5 % auf den deutschen Korpora von Common-Voice bzw. Libri-Speech [1]. In einem Multi-Sprecher-Szenario wird dann zusätzlich eine Sprechertrennung und -identifikation nötig [2]. Dies erlaubt dann eine automatische 1:1-Transkription. Anschließend werden manuell wichtige Informationen aus den Gesprächsinhalten für die weitere Dokumentation extrahiert und abstrahiert [3]. Im Projekt „iDOKS“ werden diese Schritte zu einem System verknüpft. Als Datengrundlage dienen verfügbare Audio- und Textdaten sowie Protokolle der Ratssitzungen des Dresdner Stadtrats.

2 Grundlagen und Stand der Wissenschaft

2.1 Spracherkennung

Spracherkennungssysteme (kurz ASR für Automatic Speech Recognition) haben in den vergangenen Jahren durch Deep Neural Networks (DNN) sowie dazugehörige Deep-Learning-Technologien einen enormen Qualitätssprung erfahren. Anfänglich wurde der traditionell hierarchische Aufbau von akustischer, lexikalischer und sprachlicher Modellierung (kurz AM, LexM

[*] bezeichnet gleichrangige Beiträge
[†] Arbeit außerhalb der Amazon-Tätigkeit

und LM) durch DNNs unterstützt, mittlerweile kommen jedoch auch hier Transformer-Technologien zum Einsatz. Diese benötigen weniger Hierarchieebenen und ermöglichen End-to-End-Ansätze, die hervorragende Ergebnisse erzielen [4, 5].

Da ASR-Systeme mittlerweile sehr gute Erkennungsleistungen erreichen, verbleibt als einflussreichste Größe die Störung durch Raumhall $s_{\mathrm{R}}(t)$ (R für Reverberation) sowie Rauschen $n(t)$ (N für Noise). Beide addieren sich zum Schall des Sprachsignals $s_{\mathrm{D}}(t)$ und werden als gestörtes Mikrofonsignal $x(t) = s_{\mathrm{D}}(t) + s_{\mathrm{R}}(t) + n(t)$ empfangen. Um dem entgegenzuwirken, befasst sich die Wissenschaft mit dem Gebiet der Robustheit von „Distant Talking ASR“, wobei zahlreiche Ansätze auf Mikrofon-, Signal-, Merkmal- oder Modellebene erforscht wurden. Erfolgreich ist u.a. Multi-Condition-Training, gepaart mit Beamforming-Mikrofonarrays. [6, 7].

2.2 Sprechertrennung und -identifikation

Sprechertrennung und -identifikation (kurz SDI) sind Klassifizierungsprobleme, die darauf abzielen, Audiomaterial in zeitliche Segmente aufzuteilen und mit dem entsprechenden Sprecher zu versehen. Die Sprechertrennung unterteilt Audiomaterial in Segmente, welche Sprecher repräsentieren, ohne dass a priori Informationen über die Sprecher bekannt sind. Im Gegensatz dazu zielt die Sprecheridentifikation darauf ab, den erkannten Sprechersegmenten bekannte Sprecher anhand von Sprecherprofilen zuzuordnen [8].

Anwendungen von SDI erfordern typischerweise Informationen über die Anzahl der Sprecher, die in einem Audiomaterial insgesamt vorhanden sind, sowie deren Variabilität. Typische Einsatzszenarien sind z.B. die Verschriftung von Interviews, Besprechungen und vielen weiteren Gesprächssituationen mit Dialogcharakter.

Der aktuelle Stand der Technik konzentriert sich auf End-to-End-Lösungen auf Grundlage von DNNs [2]. Diese lassen sich unterteilen in solche Architekturen, die sich (i) ausschließlich auf SDI fokussieren (Region Proposal Networks [9], Encoder-Decoder-Based Attractors [10]) und Architekturen, die (ii) SDI sowie automatische Spracherkennung gemeinsam ausführen (Sequence Transduction Architecture [11]).

2.3 Textzusammenfassung

Die automatische Zusammenfassung von Texten hat das Ziel, aus einem Ursprungstext eine verdichtete Version zu erzeugen, wobei die relevanten Informationen erhalten bleiben. Die Zusammenfassung von Texten kann verschiedenartig typisiert werden [3]. Abstraktive neuronale Ansätze sind dabei zurzeit am relevantesten und führen seit Aufkommen der Attention-Architektur [12] und den darauf aufbauenden Transformer-Architekturen (siehe z.B. BERT [13]) in den Benchmarks zur Zusammenfassung von Texten. Den aktuellen Stand der Technik bilden dabei in Englisch Z-Code++ [14] und PEGASUS [15], welche 470 bzw. 710 Millionen Parameter vorweisen. In Deutsch erreichen multilinguale Modelle aufgrund der schlechteren Datenlage die besten Resultate. Dabei wird eine multilinguale Datengrundlage benutzt, um die Qualität der Modelle per Sprachtransfer zu verbessern [16].

Für die Evaluierung von Modellen zur Zusammenfassung von Texten werden verschiedene Metriken verwendet: Von besonderer Bedeutung sind dabei ROUGE, BLEU und METEOR, welche die Hypothese mit der Grundwahrheit aus einem Testdatensatz auf lexikalischer Ebene vergleichen [17]. Weitere Metriken wurden entwickelt, um darüber hinaus die semantische Ähnlichkeit (z.B. BERTScore [18]) sowie die Wahrheitstreue und faktische Richtigkeit zu bewerten [19].

Aktuelle Studien deuten darauf hin, dass die Halluzination von Fakten sowie deren Veränderung ein Problem in Modellen darstellt [20]. Diesem Problem kann sich u.a. mit der Integration von Faktextraktionsmethoden genähert werden [21].

2.4 Wissensgraphen

Wissengraphen bezeichnen ein auf Graphen basierendes Datenmodell, in dem Entitäten als Knoten und Beziehungen zwischen ihnen als Kanten auftreten. Entitäten können darüber hinaus auch durch Beziehungen mit konkreten Werten verbunden werden. Damit ist ein Wissensgraph ein multi-relationaler Graph, den wir formal durch $\mathscr{G} = (\mathscr{E}, \mathscr{R}, \mathscr{L}, \mathscr{F})$ beschreiben. $\mathscr{E}$ bezeichnet die Menge der Entitäten, $\mathscr{R}$ die Relationen, $\mathscr{L}$ eine Menge aus Werten und $\mathscr{F}$ die Menge der Fakten, also $(h, r, t) \in \mathscr{F}$ mit $h \in \mathscr{E}$, $t \in \mathscr{E} \cup \mathscr{L}$ und $r \in \mathscr{R}$ [22]. So könnte ein Wissensgraph über Fußball die folgenden Fakten enthalten: (Lena Sophie Oberdorf, Land der Staatsangehörigkeit, Deutschland), (Lena Sophie Oberdorf, spielt für, VfL Wolfsburg), (Lena Sophie Oberdorf, spielt für, Nationalmannschaft Deutschland).

Mit dieser flexiblen Struktur lässt sich Wissen aus verschiedenen Domänen bereitstellen und verknüpfen, um es in Anwendungen gezielt einsetzen zu können. Große Wissensgraphen wie DBpedia [23] und Wikidata [24] werden z.B. in Suchmaschinen, für Frage-Antwort-Systeme oder zur Schlussfolgerung von neuen Fakten aus gegebenen Daten verwendet.

3 Methoden

3.1 Systemarchitektur

Das System wird mithilfe einer Micro-Service-Architektur modular, skalierbar und plattformunabhängig in ASP .NET Core 6.0 implementiert. Die einzelnen Micro-Services werden jeweils innerhalb eines Docker-Containers instanziiert und kommunizieren miteinander über definierte REST-Schnittstellen. Mit dieser Art der umgebungsagnostischen sowie virtualisierten Implementation wird die Plattformunabhängigkeit gewährleistet [25].

Im Transkriptions- und Sprechertrennungsmodul werden Audiodateien transkribiert, segmentiert und den entsprechenden Sprechern zugeordnet. Vorangehend werden Sprecherprofile mithilfe einer Referenzdatei angelegt. Konfigurationen und Ergebnisse werden im JSON-Format kommuniziert. Die Audiodatei wird als HTML-Formulardatei versendet. Im Verdichtungsmodul werden die zuvor generierten Transkriptionen in einem XML-Schema verarbeitet und verdichtet zurückgegeben.

In einem Administrationsmodul werden Nutzerdaten, Ergebnisse sowie Sprecherinformationen in einem Datenbankmodul gespeichert. Das Modul kann als Attached Service nach Vorbild des 12-Factor-Paradigmas betrachtet werden [26]. Im einem Aggregationsmodul werden die Funktionsmodulschnittstellen miteinander vereint und dem Frontend bereitgestellt. Das Modul operiert nach dem Backend-for-Frontend-Prinzip [27].

3.2 Transkriptions- und Sprechertrennungsmodul

Das Modul zur ASR und SDI ermöglicht die Verarbeitung von Audiodateien im sog. Batch-Processing. Die Modulschnittstelle gibt die erkannten Tokens (Wörter, Satzzeichen) zusammen mit Zeitstempeln (Anfang, Ende) sowie einem zugeordneten Sprecher (Sprecher-ID) zurück.

Zur Adaption der ASR an die Akustik des Zielszenarios (z.B. Hall sowie sonstige Nebengeräusche bei Redebeiträgen im Stadtrat) kommt das Multi-Condition-Training zur Anwendung, das eine Vielzahl von praktisch relevanten akustischen Umgebungsbedingungen einbezieht.

Das für das Projekt entwickelte System zur SDI ermöglicht eine robuste Sprechertrennung sowie die Wiedererkennung bekannter Sprecher anhand von Sprecherprofilen. Für die Sprecheridentifkation werden dazu Sprechermerkmale, d.h. Sprecherprofile, aus gegebenen Audiosequenzen erzeugt. Die einzelnen Verarbeitungsschritte zur SDI sowie zur Integration der ASR sind nachfolgend beschrieben.

Das Audiosignal wird in überlappende Segmente unterteilt und DNN-basierte Sprechermerkmale aus jedem Segment in Form von d-Vektoren extrahiert [2]. Anschließend werden Sprecherwechsel durch die Berechnung der Cosinus-Distanz zwischen aufeinanderfolgenden Segmenten erkannt, mit einem konfigurierbaren Distanz-Schwellwert abgeglichen und die Zeitinformationen der Sprecherwechsel extrahiert. Es folgt die Extraktion von Merkmalen aus Sprechersegmenten, die Gruppierung (Clustering) ähnlicher Merkmalsvektoren und die Zuordnung dieser Cluster zu Sprechersegmenten (= Sprechertrennung). Merkmalsvektoren aus Sprechersegmenten werden mit den Profilen bekannter Sprecher abgeglichen und diesen zugeordnet (= Sprecheridentifikation). Im Falle unbekannter Sprecher wird eine neue Sprecher-ID vergeben. Anhand der Zeitstempel der ASR-Ausgabe und der Zeitpunkte der Sprecherwechsel lassen sich die Tokens den Sprecher-IDs zuordnen.

3.3 Verdichtungsmodul

Um die Ursprungstexte wahrheitsgetreu zusammenzufassen, wird ein neuronaler Ansatz auf Basis der Transformer-Architektur verfolgt, welcher mithilfe von Faktextraktionsmethoden modifiziert wird. Als Basis für das Modell wird mT5 verwendet – ein auf multilingualen Daten vortrainiertes Encoder-Decoder-Modell, das frei verfügbar ist.

Diesem Modell wird ein Modul zur Extraktion von Koreferenzen und ein Modul zur Extraktion von Fakten hinzugefügt (nachfolgend mT5-mod). Das Koreferenzmodul verarbeitet die Texteinbettung in einem hochdimensionalen Raum durch den mT5-Encoder, um Wörter und Wortgruppen, welche die gleiche Entität beschreiben, zu gruppieren. Das Faktextraktionsmodul verarbeitet ebenfalls die Texteinbettung aus dem mT5-Encoder, allerdings mit dem Ziel, eine Relation zwischen zwei gegebenen Entitäten aus dem Text zu extrahieren. Die Ausgaben beider Module werden in einem zweiten Schritt vom Encoder genutzt, um eine neue Einbettung zu erzeugen, welche schließlich vom Decoder zur Erzeugung der Zusammenfassung genutzt wird. Die Module zur Koreferenz- und Faktextraktion geben darüber hinaus die im Text vorhandenen Koreferenzcluster und Fakten aus. Abbildung 1 gibt einen Überblick über die beschriebenen Modifikationen der Architektur.

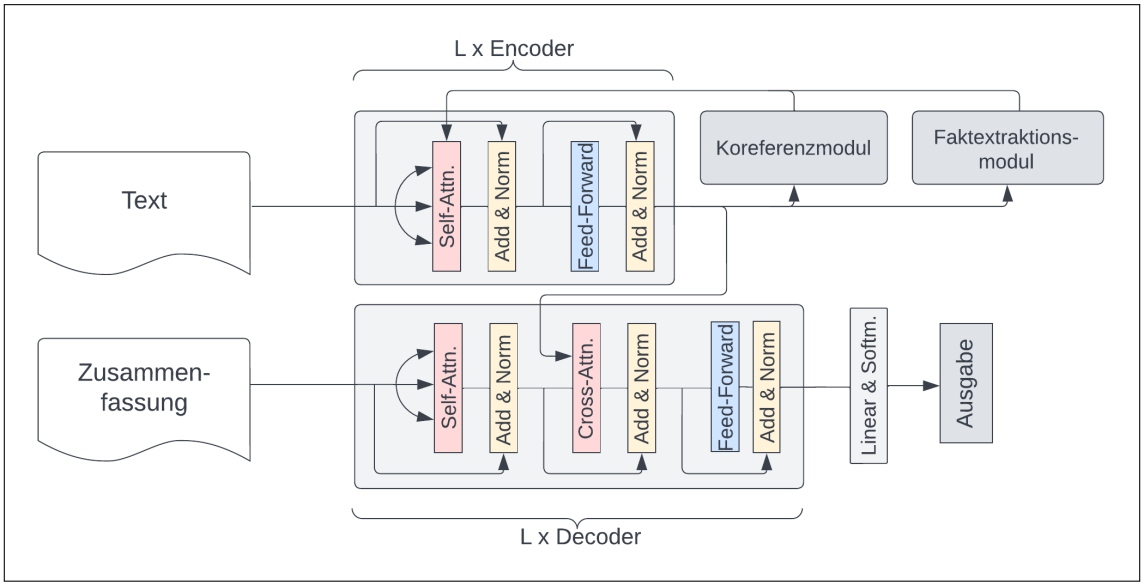

Abbildung 1 – Modifikationen der Architektur des Encoder-Decoder-Modells.

Zur Feinabstimmung des Gesamtmodells wird es gleichzeitig auf Datensätzen zur Koreferenzextraktion, Faktextraktion und Zusammenfassung trainiert, wobei dem Training auf Zu-

sammenfassungen das größte Gewicht zukommt. Das Training der beiden Untermodule dient dabei der Leitung des Gesamtmoduls hin zu vertrauenswürdigen Zusammenfassungen.

4 Ergebnisse

Im Rahmen einer ersten Evaluation wurde eine Beispieldatei von den beiden in den Abschnitten 3.2 und 3.3 beschriebenen Modulen getrennt verarbeitet und ausgewertet. Es handelt sich um ein nachgesprochenes Interview mit zwei Sprechern (männlich und weiblich, je 30 Jahre alt, siehe Anhang A). Der Dialog ist charakterisiert durch eine sequenzielle Sprechweise mit nur vereinzelten Unterbrechungen, Überlappungen oder Interjektionen. Es wurde manuell eine Referenztranskription und eine Referenz für die Sprechertrennung erstellt. Für die Quantifizierung der Qualität des ASR- und SDI-Systems wurden gemäß State-of-the-Art die Metriken „Word Error Rate“ (WER) und „Diarization Error Rate“ (DER) verwendet (vgl. Tabelle 1).

Tabelle 1 – Ergebnisse des Transkriptions- und Sprechertrennungsmoduls für den Beispieldialog.

Bezeichnung	WER %	DER %
FAZ_Dialog	7.35	1.51

Wesentliche Fehlerquellen bei der Transkription sind (i) falsch erkannte Wörter bei überlappenden Sprechern, (ii) fehlende Wörter im Vokabular (sog. out-of-vocabulary) sowie (iii) Abweichungen in der Interpunktion. Die Sprechertrennung funktioniert im Beispieldialog sehr robust. Nur eng aufeinanderfolgende Sprecherwechsel, wie sie z.B. bei Unterbrechungen auftreten, führen zu geringfügigen Abweichungen von der manuell erstellten Referenz. Interjektionen und Verzögerungslaute werden sowohl durch die Spracherkennung als auch durch die Sprechertrennung sehr zuverlässig unterdrückt.

Tabelle 2 zeigt die quantitativen Verdichtungsergebnisse. Diese erreichen für den Beispieldialog den State-of-the-Art für die semantische Ähnlichkeit (BERTScore), sind durch den geringen Umfang allerdings nicht repräsentativ.

Tabelle 2 – Ergebnisse des Verdichtungsmoduls für den Beispieldialog.

Modell	ROUGE-1	ROUGE-2	ROUGE-L	METEOR	BERTScore
mT5-mod	25,30	3,22	13,59	21,88	68,01

5 Ausblick

In diesem Beitrag wurde der erste Prototyp des Projekts „iDOKS“ beschrieben. Im weiteren Projektverlauf liegt der Schwerpunkt für das Transkriptions- und Sprechertrennungsmodul auf der Optimierung der Robustheit der ASR und SDI unter praxisrelevanten akustischen Bedingungen und der Integration weiterer algorithmischer Ansätze. Zudem wird untersucht, inwieweit die Spracherkennung von der Trennung des Eingangssignals nach Sprechern profitieren kann. Für das Verdichtungmodul werden Methoden untersucht, um den erzeugten Wissensgraphen noch direkter in die Generierung von Zusammenfassungen einzubinden. Zudem wird der Wissengraph zur Extraktion von Schlüsselpunkten in sehr langen Texten verwendet, um den logischen Fluss eines Textes in der Zusammenfassung auch dann wiedergeben zu können, wenn die Länge die Kapazität des Transformer-Modells übersteigt. Neben dem Einsatz in der parlamentarischen Gremienarbeit erscheint ein Einsatz in klassischen Konferenzsystemen oder auch im klinischen Bereich bspw. in Tumorboards sinnvoll.

Literatur

[1] RADFORD, A., J. W. KIM, T. XU, G. BROCKMAN, C. MCLEAVEY, und I. SUTSKEVER: *Robust Speech Recognition via Large-Scale Weak Supervision*. 2022.

[2] PARK, T. J., N. KANDA, D. DIMITRIADIS, K. J. HAN, S. WATANABE, und S. NARAYANAN: *A Review of Speaker Diarization: Recent Advances with Deep Learning*. 2021.

[3] GAMBHIR, M. und V. GUPTA: *Recent Automatic Text Summarization Techniques: A Survey*. 47, S. 1–66, 2017.

[4] HINTON, G. ET AL.: *Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups*. *IEEE Signal Processing Magazine*, 29(6), S. 82–97, 2012.

[5] WANG, Y., Y. SHI, F. ZHANG, C. WU, J. CHAN, C.-F. YEH, und A. XIAO: *Transformer in action: a comparative study of transformer-based acoustic models for large scale speech recognition applications*. 2020.

[6] TANG, H., W.-N. HSU, F. GRONDIN, und J. GLASS: *A Study of Enhancement, Augmentation, and Autoencoder Methods for Domain Adaptation in Distant Speech Recognition*. 2018.

[7] QIAN, Y., T. TAN, und D. YU: *An Investigation Into Using Parallel Data for Far-Field Speech Recognition*. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, S. 5725–5729. 2016.

[8] FRIEDLAND, G. und D. VAN LEEUWEN: *Speaker Recognition and Diarization*, Kap. 7, S. 115–129. John Wiley Sons, Ltd, 2010.

[9] HUANG, Z., S. WATANABE, Y. FUJITA, P. GARCIA, Y. SHAO, D. POVEY, und S. KHUDANPUR: *Speaker Diarization with Region Proposal Network*. 2020.

[10] HORIGUCHI, S., Y. FUJITA, S. WATANABE, Y. XUE, und P. GARCIA: *Encoder-Decoder Based Attractors for End-to-End Neural Diarization*. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30, S. 1493–1507, 2022.

[11] SHAFEY, L. E., H. SOLTAU, und I. SHAFRAN: *Joint Speech Recognition and Speaker Diarization via Sequence Transduction*. 2019.

[12] VASWANI, A., N. SHAZEER, N. PARMAR, J. USZKOREIT, L. JONES, A. N. GOMEZ, L. U. KAISER, und I. POLOSUKHIN: *Attention is all you need*. In I. GUYON, U. V. LUXBURG, S. BENGIO, H. WALLACH, R. FERGUS, S. VISHWANATHAN, und R. GARNETT (Hrsg.), *Advances in Neural Information Processing Systems*, Bd. 30. Curran Associates, Inc., 2017.

[13] DEVLIN, J., M. CHANG, K. LEE, und K. TOUTANOVA: *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. *CoRR*, abs/1810.04805, 2018.

[14] HE, P., B. PENG, L. LU, S. WANG, J. MEI, Y. LIU, R. XU, H. H. AWADALLA, Y. SHI, C. ZHU, W. XIONG, M. ZENG, J. GAO, und X. HUANG: *Z-Code++: A Pre-Trained Language Model Optimized for Abstractive Summarization*. 2022.

[15] ZHANG, J., Y. ZHAO, M. SALEH, und P. LIU: *PEGASUS: Pre-Training with Extracted Gap-Sentences for Abstractive Summarization*. In H. DAUMÉ und S. AARTI (Hrsg.), *Proceedings of the 37th International Conference on Machine Learning*, Bd. 119, S. 11328–11339. PMLR, 2020.

[16] XUE, L., N. CONSTANT, A. ROBERTS, M. KALE, R. AL-RFOU, A. SIDDHANT, A. BARUA, und C. RAFFEL: *mT5: A Massively Multilingual Pre-Trained Text-to-Text Transformer*. In *Proceedings of the 2021 Conference of the NAACL: Human Language Technologies*, S. 483–498. Association for Computational Linguistics, 2021.

[17] FABBRI, A. R., W. KRYŚCIŃSKI, B. MCCANN, C. XIONG, R. SOCHER, und D. RADEV: *SummEval: Re-Evaluating Summarization Evaluation. Transactions of the Association for Computational Linguistics*, 9, S. 391–409, 2021.

[18] ZHANG*, T., V. KISHORE*, F. WU*, K. Q. WEINBERGER, und Y. ARTZI: *BERTScore: Evaluating Text Generation with BERT*. In *International Conference on Learning Representations*. 2020.

[19] KRYŚCIŃSKI, W., B. MCCANN, C. XIONG, und R. SOCHER: *Evaluating the Factual Consistency of Abstractive Text Summarization. arXiv preprint arXiv:1910.12840*, 2019.

[20] JI, Z., N. LEE, R. FRIESKE, T. YU, D. SU, Y. XU, E. ISHII, Y. BANG, A. MADOTTO, und P. FUNG: *Survey of Hallucination in Natural Language Generation. ACM Computing Surveys*, 2022.

[21] ZHU, C., W. HINTHORN, R. XU, Q. ZENG, M. ZENG, X. HUANG, und M. JIANG: *Enhancing Factual Consistency of Abstractive Summarization*. In *Proceedings of the 2021 Conference of the NAACL: Human Language Technologies*, S. 718–733. Association for Computational Linguistics, 2021.

[22] JI, S., S. PAN, E. CAMBRIA, P. MARTTINEN, und P. S. YU: *A Survey on Knowledge Graphs: Representation, Acquisition and Applications. IEEE Transactions on Neural Networks and Learning Systems*, 33(2), S. 494–514, 2022.

[23] LEHMANN, J., R. ISELE, M. JAKOB, A. JENTZSCH, D. KONTOKOSTAS, P. N. MENDES, S. HELLMANN, M. MORSEY, P. VAN KLEEF, S. AUER, und C. BIZER: *DBpedia - A Large-Scale, Multilingual Knowledge Base Extracted from Wikipedia. Semantic Web Journal*, 6, S. 167–195, 2015.

[24] VRANDEČIĆ, D. und M. KRÖTZSCH: *Wikidata: A Free Collaborative Knowledgebase. Commun. ACM*, 57, S. 78–85, 2014.

[25] DE LA TORRE, C., B. WAGNER, und M. ROUSOS: *.NET Microservices: Architecture for Containerized .NET Applications*. Microsoft Developer Division, .NET and Visual Studio Product Teams, 2022.

[26] WIGGINS, A.: *The Twelve-Factor App*. 2017. URL `https://12factor.net/12factor.epub`.

[27] NEWMAN, S.: *Building Microservices*. Bd. 2. O'Reilly Media, Incorporated, 2021.

[28] FAZ: *Sven Plöger über Klimaschutz*. 2021. URL `https://www.faz.net/aktuell/rhein-main/sven-ploeger-ueber-klimawandel-17609967.html`.

A Demo

Originaltext [28]

FAZ: Herr Plöger, Corona hatte die Klimadebatte in den vergangenen anderthalb Jahren ganz schön in den Hintergrund gedrängt, oder?
Plöger: Das stimmt aber nur für die Schlagzeilen, die ganz oben stehen. Die Debatte über das Klima ist weitergeführt worden – zum Glück. Und Fridays for Future hatte in den Jahren 2018 und 2019 wirklich enormen Zulauf. Das war natürlich auch mit der Dürre des heißen Sommers im Jahr 2019 verknüpft. Der Klimawandel wird jetzt spürbar...
FAZ: Und dann kam die Flut im Ahrtal.
Plöger: Und damit der abermalige Hinweis darauf, dass extreme Wetterlagen mit all ihren Gefahren immer häufiger werden. Wie viele Menschen im Anschluss helfen wollten und gespendet haben, hat mich sehr berührt.
FAZ: Sie verfolgen die Klimadebatte durch Ihren Beruf, in dem es längst um mehr als nur um das Wetter geht, ja schon sehr lange. Haben Sie in der Zeit einen Wandel in der Art festgestellt, wie die Diskussion geführt wird?
Plöger: Das Thema wird viel emotionaler besprochen als früher. Und das liegt nicht nur an der Fridays-for-Future-Bewegung. Ich gebe Ihnen ein Beispiel: Wenn sie heute durch den Wald gehen und sehen, wie es den Bäumen geht, wie viel Holz aus dem Wald gefahren wird, weil die Bäume nicht mehr gesund waren, dann berührt das die Menschen mehr als früher. Sie überlegen dann bewusster, was sie besser machen können. Diesen Wandel nehme ich ganz eindeutig wahr.
FAZ: Stimmt, ein solcher Eindruck ist einer der Gründe dafür, warum die F.A.Z. jetzt mit ihrer Aufforstungsaktion beginnt.
Plöger: Und es gilt für immer mehr Unternehmen. Erst vor ein paar Tagen hat sich eine Initiative mehrerer Dutzend Großkonzerne und auch mittelständischer Firmen zusammengeschlossen, um ein Zeichen zu setzen, dass mehr gegen den menschengemachten Klimawandel getan werden muss. Fridays for Future wiederum muss aus meiner Sicht an einer Weiterentwicklung gelegen sein. Möchte man einfach weiter vor dem Römer protestieren oder im Römer Kommunalpolitik machen?
FAZ: Kommen wir noch einmal zu den Unternehmen zurück, zum Beispiel auch zur F.A.Z. – was halten Sie von dem Vorwurf, mit solchen Aktionen werde nichts anderes als „Greenwashing“ betrieben?
Plöger: Das wäre ja nicht fair. „Greenwashing“ bedeutet ja, man behauptet, dass man sich nachhaltiger verhält und sich um die Umwelt kümmert – und dann ist es gar nicht so. Hier sagt die F.A.Z., dass sie etwas tut und was sie tut. Allenfalls könnte man daran zweifeln, ob es ausreichend ist, was getan wird ...
FAZ: Aber helfen punktuelle Verhaltensänderungen eines Individuums oder eines kleineren Unternehmens überhaupt etwas angesichts der globalen Herausforderung?
Plöger: Selbstverständlich! Man verändert ein bisschen etwas von der weiteren Entwicklung, die wir nicht wollen. Man spürt vielleicht keine sofortigen Auswirkungen, aber natürlich gibt es sie. Und die Gegenfrage ist doch auch erlaubt. Ist es besser, wenn wir nichts tun?

Zusammenfassung

Der Interviewte, Herr Plöger, sagt, dass die Debatte über den Klimawandel trotz der Aufmerksamkeit, die der Coronavirus-Pandemie geschenkt wurde, weitergeführt wurde. Er erwähnt auch, dass extreme Wetterverhältnisse immer häufiger werden und dass die Diskussion über den Klimawandel in den letzten Jahren emotionaler geworden ist. Er erwähnt auch, dass Unternehmen und Initiativen mehr gegen den Klimawandel unternehmen, gibt aber auch zu bedenken, dass es Bedenken bezüglich der Ausreichendheit dieser Maßnahmen gibt. Schließlich geht er auf den Vorwurf des 'Greenwashing' ein und sagt, dass es ungerecht wäre, Unternehmen vorzuwerfen, nur vorzugeben, Maßnahmen zu ergreifen, gibt aber zu, dass einige die Wirksamkeit der ergriffenen Maßnahmen in Frage stellen könnten.

Tabelle 3 – Beispieldialog mit Zusammenfassung, generiert durch mT5-mod.