

Vorhaben: CORONA – QURATOR - PANQURA – eine Technologieplattform für mehr Informationstransparenz in Krisenzeiten

Teilprojekt 3: Validierung von Quellen, Inhalten, Fakten

FKZ: 03COV03C

Laufzeit: 01.01.2021 - 31.12.2022

Condat AG, Alt-Moabit 91d, 10559 Berlin

Schlussbericht - Teil I

Ursprüngliche Aufgabenstellung sowie wissenschaftlicher und technischen Stand, an den angeknüpft wurde

Die Corona Pandemie hatte gezeigt, dass mit dem Aufkommen einer Krise auch immer eine überbordende Menge neuer Informationen zu Auswirkungen, Betroffenen, Regelungen, Medikamenten und Maßnahmen publiziert wird, deren Wahrheitsgehalt und Relevanz nur schwer einzuschätzen ist. Dies liegt unter anderem daran, dass sich vor allem in Sozialen Netzen zu neu aufkommenden Themen schnell verschiedene Foren mit divergierenden Meinungen bilden, für die es weder etablierte zuverlässige Quellen gibt, noch ein anerkannter Kreis von Experten verfügbar ist. Dies führt dazu, dass im Medienbereich aufwändige, manuelle Analysen durchzuführen sind, um Informationen aus dem Web und Sozialen Medien zu bewerten, bevor sie verwendet werden können. Vor diesem Hintergrund entwickelte Condat im vorliegenden Teilvorhaben neue Services zur Analyse und Validierung von Quellen, Inhalten und Fakten auf Basis von KI-Technologien, mit denen Nutzer schneller und sicherer Informationen bewerten können.

Das Projekt knüpfte an den neuesten wissenschaftlichen Stand an, vor allem an zahlreiche internationale Initiativen wie First Draft News¹, European Journalism Center² oder BMBF³, die Methoden und Richtlinien entwickelt haben, mit denen Redaktionen und Journalisten den Wahrheitsgehalt von Informationen bewerten und validieren können. An diesen Ansätzen ist immer eine Redaktion beteiligt, die mit Hilfe von Tools entscheidet, ob eine Nachricht veröffentlicht werden kann. Da auf der anderen Seite auch immer wieder Bedenken hinsichtlich der Wahrung von Persönlichkeitsrechten bei der Nutzung von KI-Methoden gesehen wurden, hat die European Consumer Consultative Group der EU 2018 eine entsprechende Policy Recommendation herausgegeben⁴.

¹ <https://de.firstdraftnews.org/>

² <https://www.ejc.net/>

³ <https://www.bmbf.de/de/linkliste-diesen-quellen-koennen-sie-in-der-corona-krise-trauen-11260.html> und <https://www.bmbf.de/de/faktencheck-zum-coronavirus-11162.html>

⁴ European Consumer Consultative Group (2018). Policy Recommendations For A Safe And Secure Use Of Artificial Intelligence, Automated Decision-Making, Robotics And Connected Devices In A Modern Consumer World. Abgerufen Am 15. Juli 2019, Von https://ec.europa.eu/info/sites/info/files/ECcg-Recommendation-On-Ai_May2018_En.Pdf.

Ablauf des Vorhabens

Da bei der Antragstellung die Maßnahmen zur Eindämmung der Covid-19-Pandemie mit ihren Reisebeschränkungen und Abstandsregeln bereits bekannt waren, konnte das Projekt diese bereits in der Planung berücksichtigen. Deshalb konnte das Panqura Projekt auch weitgehend dem Plan entsprechend durchgeführt werden. Die im Vorhaben entwickelten Services setzen auf die im Qurator Projekt entwickelte KI und ML/DL basierte Plattform auf. Dabei wurde die Basisplattform allem vom DFKI und der FHG bereitgestellt, während von allen Partnern zahlreiche Technologie-Komponenten eingebracht wurden, die zur Entwicklung der Services für die unterschiedlichen Anwendungsbereiche von allen Partnern genutzt werden konnten. Die Schritte zur Entwicklung der Panqura-Services wurden vom Projektteam wie folgt umgesetzt:

- Entwicklung einer Architektur,
- Integration und Koordination der Verfahren und Konzepte,
- separate Entwicklung der einzelnen Komponenten,
- dynamisches Zusammenfügen der Bausteine durch klar definierte Schnittstellen,
- Integration der zu entwickelnden Komponenten,
- Durchführung des Testbetriebs und
- Validierung durch Use Case Anwendungen,

Im Rahmen des Projektes wurden die Komponenten von Condat und den Partnern integriert und zusammengeführt und ihr Zusammenwirken getestet. Die dabei entstandenen Prototypen wurden auf verschiedenen öffentlichen Online- und Präsenz-Präsentationen vorgestellt, insbesondere auf der Qurator Konferenz im Oktober 2022 und zum Abschlusstreffen mit dem Projektträger PTJ im Dezember 2022.

Die wesentlichen Ergebnisse des Vorhabens und Zusammenarbeit mit anderen Stellen

Condat hat in seinem Teilprojekt folgende innovative Funktionen entwickelt, die über den State of the Art hinausgehen:

- **Validierung von Quellen und Autoren:** es wurden eine Reihe neuer Services zur Analyse, Identifikation und Validierung von Referenzen zu Quellen und Autoren im Web und in Sozialen Medien entwickelt. Die entwickelten Methoden beruhen auf den W3C Signals und verwenden Daten wie Verleger, Herkunft und Reichweite zur Einschätzung der Authentizität und Reputation von Quellen.
- **Analyse und Bewertung von Inhalten:** es wurde eine verfeinerte Bewertung von Texten auf Basis von Sprache, Grammatik, Fehlerhaftigkeit und Konsistenz (z.B. von Titel, Zusammenfassung und Inhalt) entwickelt.
- **Validierung von Fakten:** Hierzu wurden Methoden entwickelt, um relevante Fakten aus Texten zu extrahieren und auf Basis vorhandener Ressourcen zu Fakten und Falschinformation zu bewerten.

Über die Präsentation des Projektes auf Veranstaltungen (u.a. IBC, IPTC, FK TG, Xinnovations, Semantic Media Web, Meetups) online und vor-Ort gab es viele Begegnungen mit anderen Firmen und Organisationen, die innovative Konzepte im Bereich KI und ML/DL nachfragen. Dabei hatte Condat u.a. Erfahrungsaustausch hinsichtlich der Konzepte und Ausnutzung weiterer Potentiale mit Firmen, die ihre Themen- und domänenspezifische Expertise einbringen konnten. Dazu gab es Gespräche mit der FK TG, DAI-Labor/Berlin, dem W3C-Büro, der EBU und dem IRT.

Vorhaben: CORONA – QURATOR - PANQURA – eine Technologieplattform für mehr Informationstransparenz in Krisenzeiten

Teilprojekt 3: Validierung von Quellen, Inhalten, Fakten

FKZ: 03COV03C

Laufzeit: 01.01.2021 - 31.12.2022

Condat AG, Alt-Moabit 91d, 10559 Berlin

Schlussbericht - Teil II

1. Im Vorhaben durchgeführte Arbeiten, insbesondere im Vergleich zur ursprünglichen Vorhabensbeschreibung

Im Projekt wurde zunächst eine übergreifende Architektur für die Lösungen aller Partner entwickelt, auf deren Basis die Condat Services konzipiert, entwickelt und erprobt werden konnten. Folgende Skizze aus der Gesamt-Vorhabensbeschreibung zeigt die wesentlichen Themenschwerpunkte der gesamten Technologieplattform mit den von Condat realisierten drei Komponenten:



Abb. 1: Verortung der Arbeitspakete des Teilprojekts in der PANQURA - Informationsplattform

Die technologischen Entwicklungsziele für die geplanten Arbeitspakete (APs) sowie die zu entwickelnden Technologiekomponenten sind bereits in der Verbundvorhabenbeschreibung gegliedert nach den 5 Layern der Informationsplattform (vgl. Abbildung 1), im Detail beschrieben. Condat hat im Themenschwerpunkt „3. Verifizierung“ die Komponenten „Validierung von Quellen und Autoren“, „Analyse und Bewertung von Inhalten“ und „Validierung von Fakten- und Verschwörungstheorien“ konzipiert und entwickelt. Alle im Antrag geplanten Anforderungen und Ziele wurden abgedeckt, es gab lediglich einige Anpassungen durch das Aufkommen neuer ML/DL Technologien.

Die im Projekt entstandene Lösung umfasst für die einzelnen APs die nachfolgenden Funktionen.

AP C1: Validierung von Quellen und Autoren

In diesem AP wurden Methoden entwickelt, mit denen Referenzen zu Quellen und Autoren aus Web und Sozialen Medien ermittelt werden können. Es wurde ein Konzept und eine Methode konzipiert, spezifiziert und entwickelt, mit der Informationen zu Quellen und Autoren aus dem Web und Sozialen Medien extrahiert werden. Da die Bewertungen von Verlegern und Multiplikatoren auch für die Validierung von Quellen und Autoren relevant sein können, werden möglichst viele Informationen dazu gesammelt. Um die Quellen und Autoren eindeutig zu identifizieren ist eine Normalisierung von Namen und Orten erforderlich, da sowohl innerhalb einer Sprache wie auch sprachübergreifend verschiedene Schreibweisen verwendet werden können.

Es gibt eine Reihe von Informationen im Web und Sozialen Netzen, die als Indikatoren für die Reputation und Glaubwürdigkeit von Quellen verwendbar sind. Allerdings liefern sie zusammengenommen lediglich Anhaltspunkte, die in einzelnen Fällen jedoch nicht immer zutreffen müssen, z.B. kann auch eine recht unbekannte Webseite mit wenig Publikum seriös sein und gute Ergebnisse liefern. Trotzdem sind die Ergebnisse statistisch meist zutreffend und helfen dem Endnutzer sich ein Urteil zu bilden.

Es werden Daten ermittelt wie z.B. die Reichweite: dazu gibt es eine Reihe von Informationen im Web und Sozialen Netzen, die einfach abrufbar sind: z.B. zu Followern (Twitter, LinkedIn, Facebook, etc.) wie auch Verzeichnisse der Zuschauer/Leser-Zahlen für Medien und verifizierte Accounts, die z.B. einen blauen Haken haben. Da sich in Krisensituationen die Gültigkeit von Informationen wie z.B. Regelungen oder Öffnungszeiten relativ schnell ändern können, wird zunächst ermittelt, wie aktuell eine Information ist. Dies kann z.B. anhand eines angegebenen Datums oder zumindest auf Basis des Erscheinungsdatums einer Publikation erfolgen.

AP C2: Analyse und Bewertung von Inhalten

Zur Bewertung von Texten durch Metadaten wurden Auffälligkeiten in der Sprache, Inhalten und Form ermittelt. Zur Analyse relevanter Texte und Metadaten aus Web-Inhalten wurden die unterschiedlichen Elemente wie Meta-Daten, Formatierung, Titel, Zusammenfassungen, Inhalten und Links extrahiert und klassifiziert. Dazu wird auch deren Relevanz unterschieden, z.B. ob Meta-Daten in Überschriften oder wiederholt auftreten. In dem Unter-AP C2.2 wurden Grammatik, Rechtschreibung und Sprache analysiert. Zudem wurden auch Lösungen entwickelt, um Auffälligkeiten zu erkennen, die auf eine verminderte Glaubwürdigkeit hinweisen. Dies umfasst die Erkennung von grammatikalischen Defiziten, signifikante Fehler in der Rechtschreibung und auffällige Sprache durch Schlüsselwörter. Zur Identifikation kritischer Inhalte auf Basis von Schlüsselwörtern wurden Methoden entwickelt, um Wörter und Ausdrücke zu Waffen oder Rauschgift herauszufiltern. Bei der Erkennung

auffälliger Sprache wird auch versucht auf Basis kritischer Schlüsselwörter zu bestimmten Themen zu erkennen, ob der Text auf eine bestimmte Szene oder Gruppe hinweist. Häufig kann dies durch Prüfung der Quelle geklärt werden. Ansonsten wird versucht auf Basis von KI-Methoden und von bekannten Indizes mit Schlüsselwörtern von Kommissionen des Bundes und der Länder versucht, sie einer der bekannten Quellen, Szenen oder Gruppen zuzuweisen. Da sich die Themen und Schlüsselwörter immer wieder verändern, erfolgt auf Basis von ML und den Nutzeraktivitäten ein permanentes Update des im Projekt aufgebauten Thesaurus.

AP C3: Validierung von Fakten und Verschwörungstheorien

In diesem AP wurden relevante Fakten aus Texten extrahiert die dann auf Basis von KI-Technologien in vorhandenen Ressourcen identifiziert wurden. Da in der Regel nur kleinere Bereiche eines Dokuments für die Analyse relevant sind, werden diese nach bestimmten Kriterien in unterschiedlichen Arten von Texten erkannt und extrahiert. In diesen werden dann relevante Fakten identifiziert. Dabei wurden die Texte nicht vollständig parsiert, sondern sich auf wesentliche grammatikalische und lexikalische Elemente konzentriert, die für die Formulierung von Fakten relevant sind. Die Extraktion erfolgt auf Basis einfacher allgemeiner und domänenabhängiger Regeln. Die Texte müssen nicht notwendigerweise bzgl. Grammatik und Rechtschreibung korrekt sein. Die Erkennung von Eigennamen (Personen, Orte, etc.) spielt für den Faktencheck eine wichtige Rolle und erfolgt über Wikidata. Es wurde ein KI basierten Algorithmus zur Bewertung der Übereinstimmung von Fakten entwickelt, der auf verschiedene vorhandene Methoden aufsetzt. Dieser neue Algorithmus kann einen neuen Fakt gegen eine gesicherte Referenz prüfen. Weiterhin wurde ein formales Bewertungsschema entwickelt, das auch Berechnungen, Vergleiche und Weitergabe von Werten erlaubt. Die Analyse und Erkennung von Verschwörungstheorien erfolgt analog zur Faktenvalidierung, allerdings sind eine Reihe von erweiterten Konzepten erforderlich, da eine Verschwörungstheorie mehrere Elemente enthält, die in einer bestimmten Reihenfolge auftreten. Die Identifikations-Algorithmen basieren nicht nur auf einem Fakt, sondern einer Reihe von Fakten.

2. Die wichtigsten Positionen des zahlenmäßigen Nachweises

Es gab auf Grund der Corona-Richtlinien geringfügige Veränderungen im Projektablauf, z.B. führten die fehlenden vor-Ort Termine zu einigen Kommunikations-Hindernissen. Letztlich gab es jedoch keine Probleme bzgl. der kooperativen Konzeption und Entwicklung mit den Partnern. Darüber hinaus gab es lediglich einige Änderungen im Arbeitsplan, die sich aus dem Aufkommen neuer KI-Technologien ergeben hatten. Diese Anpassungen konnten jedoch ohne Probleme erfolgen, da die FuE Projektpartner des Konsortiums das neue Know-How zu den neuen KI Methoden kontinuierlich in das Projekt transferiert hatten.

Dem zahlenmäßigen Nachweis kann man entnehmen, dass der Arbeitsplan über die 24 Monate hinweg hinsichtlich geleisteter Personen-Monate, der Kosten und geplanten Gesamtaufwände weitgehend eingehalten wurde. Die größte Position sind die Aufwände für die Mitarbeiter, die sich relativ gleichmäßig über die Projektlaufzeit verteilen.

3. Die Notwendigkeit und Angemessenheit der geleisteten Arbeit

Aufgrund der hoch innovativen KI-, ML- und Medien-Technologien bestand ein technisches und wirtschaftliches Risiko, das Condat alleine aus seinen Kundenprojekterlösen nicht tragen konnte und somit die für das Vorhaben beantragten Fördermittel benötigte. Die Entwicklung der Komponenten zur Analyse, Klassifikation und Validierung, waren notwendig, da solche Funktionen und Komponenten derzeit noch nicht von anderen System-Anbietern verfügbar sind. Dies liegt zum einen daran, dass es sich bei den KI Methoden um eine innovative Technologie handelt, wie auch die verwendeten ML-Methoden in allen APs in diesem Projekt unter den für diese Lösung vorliegenden Bedingungen noch nicht erprobt waren.

3. Der voraussichtliche Nutzen, Verwertbarkeit, fortgeschriebener Verwertungsplan

Allgemeine Trends

Es waren einige generelle Trends zu beobachten, die zum Teil auch Auswirkungen der Covid-19 Pandemie sind, die Folgen für die Verwertung der Projektergebnisse haben könnten:

- wachsende Anforderungen an Datenschutz und -Sicherheit
- Aufkommen neuer Technologien im ML/DL Bereich, wie die Transformer Technologien
- steigende Patent/Copyright Risiken und
- erhöhte Handelsbarrieren.

Wirtschaftliche Verwertbarkeit

Condat strebt an, die entstandenen Ergebnisse individuell sowie auch mit einzelnen Partnern zu verwerten. Dabei ist nachfolgend nur die individuelle Verwertung dargestellt. Eine Verwertung mit anderen Projektpartnern (z.B. 3PC, Semtation) ist durch die sich ergänzenden IT-Lösungen und Kompetenzen sinnvoll. Konsortium übergreifend sind bereits gemeinsame Veranstaltungen wie die Qurator-Konferenzen und Xinnovations-Events erfolgt. Condat hat bereits damit begonnen, die entstandenen Lösungen bis hin zur Erreichung der Produktreife weiterzuentwickeln. Damit kommt zu den Kosten des geförderten Projektes noch einmal ein beträchtlicher Aufwand hinzu, und zudem sind weitere Ausgaben für Marketing- und Vertriebsaufwendungen kalkuliert.

Verwertungsmöglichkeiten und perspektivisch aussichtsreiche Märkte

Da sich der Markt für die entwickelten Analyse- und Validierungs-Technologien immer weiter entwickelt, haben wir die verschiedenen Teilbereiche wie KI, ML, und CMS Lösungen betrachtet. Nach Einschätzung vieler Studien befindet sich alle diese Bereiche der KI Transformation in einer weiteren Phase der Beschleunigung ¹. Nach einer Analyse von Gartner² wächst der KI Markt derzeit jährlich um über 20% auf über 70 Mrd € in 2023 weltweit. Die größten Wachstumsraten hat dabei Knowledge Management mit über 30% und der Digital Workplace mit mehr als 20%, beides Bereiche, die

¹ <https://www2.deloitte.com/de/de/pages/trends/ki-studie-2022.html>

² <https://www.gartner.com/en/newsroom/press-releases/2021-11-22-gartner-forecasts-worldwide-artificial-intelligence-software-market-to-reach-62-billion-in-2022>

von den Panqura Lösungen adressiert werden. Einer Studie von IDC³ aus 2021 mit einer etwas anderen Marktsegmentierung erwartet sogar noch höhere Wachstumsraten für die Segmente AI Software Platforms (33%) und AI Software Development (26%). Dabei handelt es sich vor allem um die Einführung von Komponenten auf Basis von KI, Semantisches Web, Verifikation, Wikis und Wissensdatenbanken. Wenn man die Zahlen von weltweit ca. 70 Mrd € in 2023 herunterbricht für Deutschland, erhält man ca. 6 Mrd € für Deutschland, von denen nach unserer Abschätzung ca. 300 Millionen € für die von Panqura entwickelten Systeme und Tools ausgegeben werden.

Zielmärkte

Die entwickelte Lösung bietet mit seinen zahlreichen Komponenten eine gute Grundlage für die Einbindung und Nutzung im Medien / TV Bereich. Das im Projekt entstandene System ist zunächst für die Vermarktung im TV / Medienbereich konzipiert, kann jedoch auch für andere Bereiche wie Publishing, Journalismus, Mediatheken oder andere kommerzielle Websites verwendet werden. Damit ergänzt die entstandene Lösung die von Condat bereits erfolgreich vermarktete Smart Media Engine⁴ und kann wie vorgesehen verwertet werden. Aufgrund unserer Marktbeobachtung planen wir für das vorliegende Projekt nachfolgende Märkte für Mitarbeiter- und Kunden-Portale zu adressieren, in denen wir für die nächsten 10 Jahre eine hohe Nachfrage nach innovativen Funktionen zur Recherche, Analyse und Validierung erwarten:

- **Nachrichtenagenturen oder Redaktionen im Medien/TV- Bereich mit Newsrooms** für die Erstellung von Nachrichten, wobei diese ihre Inhalte in der Regel auch zusätzlich über das Internet, mobil und in der Tagespresse verbreiten. Bei der Bereitstellung der Funktionen für Newsrooms ist zu beachten, dass dort rund um die Uhr sehr schnelle Reaktionszeiten erforderlich sind und oft eine Ad-hoc Nutzung zur Gewinnung eines Überblicks über die bereitstehenden Quellen angeboten werden sollten.
- **Redaktionen von TV-/Video-Produktionen oder weitere ePublishing Anbieter**, die TV-Sendungen oder tiefergehende Publikationen planen, erstellen oder aufbereiten. Hier werden in Zeiträumen von 3 Tagen bis 5 Monaten Analysen und Reportagen für verschiedenste Bereiche erstellt, in denen eine hohe Qualität und Korrektheit hinsichtlich der dargestellten Informationen Priorität haben. Dazu werden von der entstandenen Panqura-Lösung umfangreichere Funktionen mit mehr Parametern und besseren Erläuterungen der Ergebnisse angeboten.
- **AV Inhalte-Anbieter wie WebTV, Mediatheken und mobile Portale**, die über Internet oder mobile Distributionskanäle AV Materialien verbreiten. Dort werden sowohl Nachrichten wie auch kurze und tiefergehende Berichte erstellt. In der Regel werden dort einfachere Workflows verwendet und mehr „user generated content“ (z.B. aus Twitter, Facebook, TikTok) nutzen, so dass vor allem eine Validierung der Korrektheit von Informationen bzgl. Ort, Zeit und Quellen sowie eine Rechte- und Kostenanalyse unterstützt wird.

³ <https://www.idc.com/getdoc.jsp?containerId=prUS47482321>

⁴ siehe Condat Portfolio: <https://www.condat.de/loesungen/smart-media-engine/>

Wettbewerb

Auch wenn es keine direkten Wettbewerber gibt, die Lösungen mit vergleichbarer Funktionalität anbieten, wäre es schon möglich, dass auf Basis von KI- und CMS-Software Systemen eine Lösung entwickelt wird, die einen ähnlichen Funktionsumfang enthält. Selbst wenn auf diesem Weg kaum ein durchgängiges System und eher Einzellösungen entstehen würden, können wir die KI- und CMS Anbieter als Wettbewerber ansehen. Einige der wichtigsten sind hierbei Microsoft, Google, Nvidia oder Adobe. Viele dieser Wettbewerber beginnen auch mit der Integration von innovativen KI Technologien in ihre Systeme, keines kann jedoch den ganzen Umfang der im vorliegenden Teilprojekt entwickelten Funktionen anbieten.

Verwertung durch Condat

Das aus der Projektentwicklung entstandene Produkt gliedert sich nahtlos in die Produktlinie Smart Media Engine ein, die in verwandten Bereichen bereits erfolgreich im Einsatz ist. Sie ergänzt somit unsere Kernprodukte und Kompetenzen. Aktuell hat Condat ca. 18 Mitarbeiter im Bereich TV / Medien beschäftigt, die einen jährlichen Umsatz von 2 Mio € erwirtschaften. Dieser Bereich soll u.a. gestützt durch das Panqura-Vorhaben ausgebaut werden, wobei insbesondere folgende Effekte zum Tragen kommen:

- das anwachsende KI- und ML/DL Know-How führt zu einem breiteren Leistungsspektrum und einer Ausweitung der Condat Aktivitäten
- durch einen erhöhten Bekanntheitsgrad und besseren Sichtbarkeit von Condat als Dienstleister für KI und ML/DL basierte Systeme (z.B. durch die Publikationen und Veranstaltungen) ergeben sich zusätzliche Consulting Projekte und
- durch die enge Kooperation und Synergie mit den Projektpartnern wie 3PC, Semtation und Ubermetrics und ergeben sich vermehrte Kontakte zu neuen Zielgruppen und in der Folge zusätzliche Projekte.

Für die Entwicklung der Anzahl der Mitarbeiter und des Umsatzes ist für die folgenden 5 Jahre nach Projektabschluss eine jährliche Steigerung von ca. 5 % geplant, die sich zu einem großen Teil aus dem vorliegenden Panqura Projekt und den darin entwickelten KI und ML/DL-gestützten Verfahren ergeben.

Akquisitions- und Informationsaktivitäten

Condat wird über sein gutes Netzwerk, das auf langjährige Projekte, Partner und Organisationen aufbaut, die Verwertung in folgenden Bereichen vorantreiben:

- Medien mit TV/ Radio und Film, TV/Radio-Sender und Medien-Unternehmen: Projekte und Kontakte bei RBB, LFM, ZDF, Bayerischer Rundfunk, WDR, NDR, MDR, N-TV, DVB, Deutsche Welle
- ePublishing, Medien-Produktionen, Verlage: Projekte und Kontakte bei Cornelsen, Klett, Westermann, Springer und Holtzbrinck.

4. Der während der Durchführung des Vorhabens dem Zuwendungsempfänger bekannt gewordene Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Aus unserer kontinuierlichen Marktbeobachtung ergaben sich keine neuen Wettbewerber oder Lösungen, die die Durchführung des Vorhabens beeinträchtigen könnten. In unserer Analyse konnten

wir feststellen, dass führende Anbieter von Content Management- und Redaktions-Systemen neue Funktionen zum automatisierten Journalismus, Sprachassistenten, Storytelling und Social-Media Zugriff integrieren. Allerdings werden die dazu gehörigen KI Methoden meist nur separat angeboten, die Lösungen sind nur begrenzt für Medien-Redaktionen verwendbar, sind nicht durchgängig und es gibt noch kaum ausreichende Datenmengen zum Training der KI Funktionen.

Aufgrund der dynamischen Entwicklung im TV-Medienbereich, wie z.B. dass unsere Kunden LFM (Landesmedienanstalten), ZDF, WDR, Deutsche Welle und rbb immer mehr Inhalte und Angebote auch im Web anbieten, besteht ein großes Interesse an den im vorliegenden Projekt entstandenen Know-how sowie den KI Methoden. Zudem spielt auch das Auftauchen neuer Technologien, wie z.B. Transformer Methoden im Bereich von ML/DL, deren Relevanz zu Projektbeginn noch nicht erkennbar war, eine wichtige Rolle. Dies führte dazu, dass die wichtigsten dieser neuen Technologien aufgegriffen und in den von Condat entwickelten Komponenten integriert wurden.

5. Die erfolgten oder geplanten Veröffentlichungen des Ergebnisses

Wenn zur Wahrung berechtigter Interessen des ZE oder Dritter oder aus anderen sachlichen Gesichtspunkten bestimmte Einzelheiten aus dem Bericht vertraulich zu behandeln sind (z. B. Wahrung der Priorität bei Schutzrechtsanmeldungen), so hat der ZE den ZG ausdrücklich darauf hinzuweisen.

Folgende Veröffentlichungen bzgl. des Panqura Projektes wurden entweder von Condat initiiert oder mit Condats Beteiligung erstellt:

- Home-Page: <https://qurator.ai/>
- Condat Home-Page, <http://www.condat.de/forschung/>
- <https://qurator.ai/qurator-conference-2022/>
- <https://qurator.ai/conference-qurator-2021/>
- BMBF Unternehmen Region, <https://www.innovation-strukturwandel.de/strukturwandel/de/unternehmen-region/die-initiativen/wachstumskerne/wachstumskerne-2018-2021/qurator/qurator.html>
- https://www.innovation-strukturwandel.de/strukturwandel/de/report/_documents/artikel/n-g/nuetzliche-quellen-statt-datenflut.html?nn=451418
- <https://www.xinnovations.org/projekte/wachstums-kern-qurator>
- <https://www.condat.de/condat-auf-ebu-online-workshop/>
- <https://tech.ebu.ch/publications/ai-quration-technologies-for-newsroom-systems>
- Computational Analysis of Storylines - Making Sense of Events (Studies in Natural Language Processing), Tommaso Caselli, Jan Thomsen, Rolf Fricke, et al, Cambridge University Press, 25. November 2021; <https://www.cambridge.org/core/books/computational-analysis-of-storylines/D1384A4B404350654614A8222EA8DE6D>

Dazu kommen ca. 15 Posts auf LinkedIn und Twitter.

6. Angabe der verwendeten Fachliteratur sowie der benutzten Informations- und Dokumentationsdienste

- [1] E. Kuzey, J. Strötgen, V. Setty, and G. Weikum, “Temponym Tagging,” presented at the the 25th International Conference Companion, New York, New York, USA, 2016, pp. 841–842.
- [2] F. Mahdisoltani, J. Biega, and F. M. Suchanek, “Yago3: A knowledge base from multilingual wikipedias,” presented at the 7th Biennial Conference on Innovative Data Systems Research, 2014.
- [3] D. Wang, T. Li, and M. Ogihara, “Generating Pictorial Storylines Via MinimumWeight Connected Dominating Set Approximation in Multi-View Graphs.,” *AAAI*, 2012.
- [4] F. Corcoglioniti, M. Dragoni, M. Rospocher, and A. P. Aprosio, “Knowledge Extraction for Information Retrieval.,” *ESWC*, vol. 9678, no. 2, pp. 317–333, 2016.
- [5] M. Kejriwal and P. Szekely, “Information Extraction in Illicit Web Domains,” presented at the the 26th International Conference, New York, New York, USA, 2017, pp. 997–1006.
- [6] G. Weikum, J. Hoffart, and F. M. Suchanek, “Ten Years of Knowledge Harvesting - Lessons and Challenges.,” *IEEE Data Eng. Bull.*, 2016.
- [7] S. Zhang, L. Yao, and A. Sun, “Deep Learning based Recommender System: A Survey and New Perspectives,” *arXiv.org*, vol. cs.IR. 24-Jul-2017.
- [8] J. Neidhardt, D. R. Fesenmaier, T. Kuflik, and W. Wörndl, “RecTour 2017,” presented at the the Eleventh ACM Conference, New York, New York, USA, 2017, pp. 386–387.