

Cloud-basierte Entwicklung eines Topic Modeling-Systems zur Unterstützung der Vokabularkonstruktion für den Taxonomiebau in der Meetings, Incentives, Conferences and Exhibition Domain

Betreuer: M. Sc. Gladys Kouakam, Dr.-Ing. Philippe Tamla, Prof. Dr. Matthias Hemmje

Forschungsbezug

Habilitationsvorhaben von Dr.-Ing. Philippe Tamla sowie Dissertation von M. Sc. Gladys Kouakam mit dem Arbeitstitel: *Supporting Cloud-based Information Extraction Life Cycle from MICE Content as a Basis for Managing Attendee Engagement and Satisfaction as well as Speaker Performance and Reputation in Professional Management of Meetings, Congresses and Events*

Voraussetzungen

Für eine erfolgreiche Bearbeitung der Aufgabenstellungen sind Vorkenntnisse in folgenden Bereichen notwendig:

- Container
- Information Retrieval
- Information Extraction
- Machine Learning
- Microservices
- Named Entity Recognition
- Natural Language Processing
- Unified Modeling Language
- Large Language Model (LLM)

Präambel

Diese Arbeit ist an die Habilitation von Philippe Tamla und die Dissertation von Gladys Kouakam angelehnt. Die Habilitation untersucht die Entwicklung domänenübergreifender, cloud-basierter Referenzmodelle für KI-gestützte Informationsextraktion und Wissensmanagement. Die Dissertation von Gladys Kouakam überträgt diesen Ansatz auf die Domäne der **Meetings, Incentives, Conferences and Exhibitions (MICE)** und untersucht, wie ein cloud-basierter **Information Extraction Life Cycle (IELC)** zur Verarbeitung veranstaltungsbezogener Daten und Inhalte ausgestaltet werden kann. Ziel ist die Entwicklung domänenspezifischer Referenzmodelle, mit denen unstrukturierte Veranstaltungsdaten systematisch gesammelt, aufbereitet, analysiert, in strukturierte Wissensrepräsentationen überführt und für datenbasierte Entscheidungen im professionellen Veranstaltungsmanagement zugänglich gemacht werden können.

Die MICE-Branche umfasst Meetings, Incentive-Veranstaltungen, Kongresse, Konferenzen und Ausstellungen. Sie spielt eine wichtige Rolle bei der Förderung von Zusammenarbeit, Wissensaustausch und professioneller Vernetzung zwischen Wissenschaft, Wirtschaft, öffentlichen Einrichtungen und Fachgesellschaften. In den vergangenen Jahren haben insbesondere Digitalisierung, COVID-19 und Klimawandel die Branche maßgeblich beeinflusst [1,2]. Daraus sind unter anderem hybride und virtuelle Veranstaltungsformate, digitale Interaktionsmöglichkeiten, KI-gestützte Dienste, die verstärkte Nutzung sozialer

Medien sowie neue Anforderungen an Nachhaltigkeit und technische Infrastrukturen entstanden [3]. Aufgrund hoher Teilnehmerzahlen, verschiedener Veranstaltungsformate, zahlreicher beteiligter Akteure und der Integration unterschiedlicher Technologien weisen MICE-Veranstaltungen eine hohe organisatorische und informationstechnische Komplexität auf. Ihre Planung, Durchführung und Auswertung erfordern daher häufig die Expertise von **Professional Congress Organizers (PCOs)** [4]. Zu deren Aufgaben gehören unter anderem die Erstellung von Veranstaltungsprogrammen, die Auswahl von Beiträgen und Vortragenden, die Strukturierung von Sessions, die Betreuung der Teilnehmenden sowie die Bewertung des Veranstaltungserfolgs. Während des gesamten Veranstaltungslebenszyklus entstehen umfangreiche Daten aus unterschiedlichen Quellen. Vor einer Veranstaltung werden beispielsweise Call-for-Papers, Abstracts, wissenschaftliche Beiträge, Angaben zu Autorinnen und Autoren, Begutachtungsergebnisse, Sessions, Tracks und Veranstaltungsprogramme erfasst. Während der Veranstaltung entstehen unter anderem Präsentationen, Sessionaufzeichnungen, Fragen und Antworten, Abstimmungsergebnisse, Teilnahmedaten, Echtzeit-Feedback und Beiträge in sozialen Medien. Nach einer Veranstaltung werden Umfrageantworten, Veranstaltungsberichte, Conference Proceedings und weitere Rückmeldungen erzeugt. Diese Daten liegen in strukturierten, semistrukturierten und unstrukturierten Formaten vor. Hierzu gehören relationale Datenbanken, CSV-, JSON- und XML-Dateien, Programmierschnittstellen, Webseiten, HTML- und PDF-Dokumente, Freitexte sowie audiovisuelle Aufzeichnungen [5, 6]. Zusätzlich sind die Daten häufig auf verschiedene Systeme und Veranstaltungseditionen verteilt und weisen unterschiedliche Schemata, Terminologien, Sprachen, Qualitätsniveaus und Zeitbezüge auf. Die Menge und Heterogenität dieser Daten führen in der MICE-Domäne zu einer Informationsüberflutung. Für PCOs ist es dadurch schwierig, relevante Informationen systematisch zu identifizieren, über verschiedene Quellen hinweg zu verknüpfen und für die Planung oder Bewertung von Veranstaltungen zu nutzen. Dies betrifft beispielsweise die Erkennung aktueller und neu entstehender Themen, die Ermittlung der Interessen und Reaktionen von Teilnehmenden sowie die Bewertung der Performance und Reputation von Vortragenden. Diese Aspekte tragen wesentlich zum Erfolg einer Veranstaltung bei.

Ein zentraler Erfolgsfaktor ist das Engagement der Teilnehmenden [7]. Es wird als Beziehung zwischen den Inhalten einer Veranstaltung, der Beteiligung an der Veranstaltung, der Kongruenz des Event-Brand-Images und der Erfahrung der Teilnehmenden mit den Veranstaltungsinhalten beschrieben [8]. Engagement und Zufriedenheit werden wesentlich durch die Aktualität und Relevanz der angebotenen Themen beeinflusst. Die Identifikation relevanter Inhalte, insbesondere von *Emerging Topics* und *Trending Topics*, kann durch **Topic Modeling (TM)** unterstützt werden und Veranstaltungsorganisatoren dabei helfen, Programme und Sessions zielgerichtet zu gestalten. *Emerging Topics* sind Themen, die neu auftreten, zunehmend Aufmerksamkeit erhalten und das Potenzial besitzen, sich zu etablierten Themen zu entwickeln [9]. *Trending Topics* sind dagegen Themen, die innerhalb einer Fachgemeinschaft bereits eine hohe oder deutlich zunehmende Popularität aufweisen [10]. Die systematische Identifikation dieser Themen erfordert eine zeitbezogene Analyse von Dokumenten aus mehreren Veranstaltungseditionen oder Beobachtungszeiträumen.

Als konzeptionelle Grundlage für die Verarbeitung der heterogenen Daten dient das von Tamla vorgeschlagene IELC-Referenzmodell [11]. Das Modell beschreibt die schrittweise Transformation heterogener Rohdaten in strukturierte und zugängliche Informationen. Es umfasst die Phasen *Data Collection*, *Data Preparation*, *Information Extraction*, *Information & Knowledge Organization* sowie *Information Access & Retrieval*.

In der Phase *Data Collection* werden relevante Datenquellen identifiziert, angebunden und gemeinsam mit Angaben zu Quelle, Format, Erhebungszeitpunkt und Provenienz gespeichert. In der Phase *Data Preparation* werden die Rohdaten durch Bereinigung, Normalisierung, Transformation und Segmentierung in ein einheitliches, maschinenverarbeitbares Datenmodell überführt. In der Phase *Information Extraction* kommen Verfahren wie **Named Entity Recognition (NER)**, **TM**, **Sentiment Analysis (SA)** und **Predictive Analytics (PA)** zum Einsatz. Mit NER werden domänenspezifische Entitäten wie Veranstaltungen, Personen, Organisationen, Sessions oder Themen identifiziert. TM dient der Erkennung latenter thematischer Strukturen in größeren Dokumentkollektionen. Die SA ermöglicht die Extraktion von Meinungen, Bewertungen und emotionalen Reaktionen aus Feedback- und Umfragedaten. PA unterstützt die Ableitung von Scores und Prognosen auf Grundlage historischer und aktueller Veranstaltungsdaten. Die extrahierten Entitäten, Themen und Kennzahlen werden in der Phase *Information & Knowledge Organization* in Vokabularen, Taxonomien, Ontologien oder anderen Wissensrepräsentationen strukturiert. Die Phase *Information Access & Retrieval* stellt schließlich Such-, Browsing- und Navigationsmechanismen bereit, über die die strukturierten Informationen für konkrete Entscheidungs- und Planungsaufgaben genutzt werden können.

Die Übertragbarkeit dieses lebenszyklusorientierten Ansatzes auf unterschiedliche Anwendungsdomänen

wird durch verwandte Forschungs- und Entwicklungsvorhaben verdeutlicht. Das Kooperationsvorhaben **Knowledge-Management Ecosystem-Portal for the National Centre of Excellence for Green Ports and Shipping (KMEP4ECOGPS)** [12] untersucht die Integration und Vermittlung von Informationen über nachhaltige maritime Betriebssysteme und Aktivitäten in grünen Häfen und Schifffahrtsnetzwerken. Es basiert auf dem **Knowledge Management System (KMS) Content and Knowledge Management Ecosystem Portal (KM-EP)**, das vom Lehrgebiet Multimedia und Internetanwendungen der FernUniversität in Hagen gemeinsam mit dem FTK e.V. entwickelt wurde [13,14].

In diesem Kontext wurde das **Thematically Structured Information Access System (TSIAS)** entwickelt [15]. Das System unterstützt einen thematisch strukturierten Informationszugang zu wissenschaftlichen Dokumentkollektionen aus dem Bereich grüner Häfen und nachhaltiger Schifffahrt. Im Rahmen einer Fallstudie wurden 968 wissenschaftliche Artikel aus den Jahren 2000 bis 2024 mithilfe von BERTopic analysiert und daraus 130 Themen identifiziert. Diese Arbeiten zeigen, wie TM zur Erschließung umfangreicher Dokumentkollektionen und zur Untersuchung zeitlicher Themenentwicklungen eingesetzt werden kann.

Ein weiteres verwandtes Forschungsprojekt ist **Artificial Intelligence for Hospitals, Healthcare & Humanity (AI4H3)** [16]. Es untersucht die Unterstützung transparenter und erklärbarer medizinischer Entscheidungen und baut auf den Ergebnissen von **Recommendation Rationalisation (RecomRatio)** auf [17]. Dabei werden medizinische Fachpublikationen analysiert, um Argumente für oder gegen bestimmte Behandlungsmöglichkeiten zu extrahieren, strukturiert bereitzustellen und für medizinische Entscheidungsprozesse nutzbar zu machen. Als technische Grundlage dient ebenfalls das KM-EP [18]. Die zugehörige AI4H3-Hub-Architektur bildet einen Bezugspunkt für die Habilitation von Philippe Tamla und untersucht wiederverwendbare, cloud-basierte Informationsextraktionsprozesse für unterschiedliche Wissensdomänen.

Eine zentrale empirische Grundlage der Dissertation bildet die Zusammenarbeit mit der GLOBIT Global Information Technology GmbH [19]. GLOBIT fungiert als Praxis- und Datenpartner und stellt für Forschungszwecke veranstaltungsbezogene Realweltdaten aus dem von GLOBIT betriebenen Legacy-Konferenzmanagementsystem bereit. Die Daten liegen als kongress- beziehungsweise veranstaltungsspezifische JSON-Dateien vor.

Als primäre Datenquelle wird der Datensatz GLOBIT-Kongressprogrammdaten 2015–2025 verwendet. Nach aktuellem Exportstand umfasst er mehr als 17 594 Beitragsdatensätze und 3 404 Sitzungsdatensätze aus 11 Veranstaltungsausgaben beziehungsweise Konferenzdatensätzen. Die Daten enthalten strukturierte Informationen zu eingereichten wissenschaftlichen Beiträgen und zu den Programmen verschiedener wissenschaftlicher und medizinischer Fachkongresse.

Zu den von GLOBIT betreuten Kongressreihen gehören unter anderem der Deutsche Pflahtag (DPT), der Deutsche Kongress für Parkinson und Bewegungsstörungen (DPG beziehungsweise DPGHD), der European Congress of Pathology (ECP beziehungsweise ESP), die European Conference on Schizophrenia Research (ECSR), der World Congress on ADHD, der Kongress der Deutschen Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde (DGPPN), der European Congress of Child and Adolescent Psychiatry (ESCAP), der Deutsche Suchtkongress (SUCHT), der World Congress of Biological Psychiatry (WFSBP) und der World Congress of Psychiatry (WPA). Welche dieser Kongressreihen in die konkrete Untersuchung einbezogen werden, wird auf Grundlage der verfügbaren JSON-Dateien, der zeitlichen Abdeckung und der Qualität der Textdaten festgelegt.

Für die Untersuchung werden insbesondere folgende Attribute aus den JSON-Dateien verwendet:

- eindeutige beziehungsweise pseudonymisierte Identifikatoren der Sitzungen und Beiträge,
- die Zuordnung zur jeweiligen Kongressreihe und Veranstaltungsausgabe,
- das Veranstaltungsjahr sowie weitere verfügbare zeitliche Metadaten,
- Titel, Beginn und Ende der jeweiligen Sitzung,
- Sessiontyp, Sprache, Raum- und Track-Zuordnung, soweit vorhanden,
- Titel und Abstracttext des jeweiligen Beitrags,
- vorhandene Topic- oder Keyword-Zuordnungen,
- öffentliche Beitrags- beziehungsweise Programmnummern.

Ergänzend können öffentlich zugängliche Veranstaltungsprogramme, Conference Proceedings und bibliografische Metadaten verwendet werden. CrossRef [20] kann zur Ergänzung bibliografischer Angaben über DOIs eingesetzt werden. GROBID [21] dient nicht als eigenständige Datenquelle, sondern als Werkzeug zur Extraktion und Strukturierung von Metadaten und Textinhalten aus PDF-Dokumenten.

Aufbauend auf dem IELC, den beschriebenen Vorarbeiten und der Zusammenarbeit mit GLOBIT untersucht die Dissertation von Gladys Kouakam die Entwicklung domänenspezifischer Referenzmodelle zur Unterstützung von PCOs bei der datengetriebenen Planung, Durchführung und Auswertung von MICE-Veranstaltungen. Im Rahmen der Dissertation wird hierzu die **Information Extraction Life-cycle for Attendee Satisfaction & Speaker Performance (IELC4AS2P)**-Architektur entwickelt (vgl. Abbildung 1). Sie beschreibt den Verarbeitungspfad von veranstaltungsbezogenen Rohdaten über die Informationsextraktion und Wissensorganisation bis hin zu Informationszugangs- und Entscheidungsunterstützungsdiensten.

Die vorgeschlagene Architektur unterscheidet drei zentrale Kategorien von Eingabedaten:

- *Event Metadata*, beispielsweise Veranstaltungsprogramme, Veranstaltungsausgaben, Sessions, Tracks, Themen, Abstracts sowie Angaben zu Teilnehmenden und Vortragenden,
- *Event Content*, beispielsweise wissenschaftliche Publikationen, Präsentationen, WorkshopZusammenfassungen, Conference Proceedings, Veranstaltungsberichte und Aufzeichnungen,
- *Event Execution Data*, beispielsweise Echtzeit-Feedback, Umfrageantworten, Fragen und Antworten, Abstimmungen sowie Anwesenheits- und Interaktionsdaten.

Nach ihrer Sammlung und Aufbereitung werden diese Daten mit unterschiedlichen Information-Extraction-Verfahren untersucht. NER unterstützt die Extraktion eines **Event Vocabulary (EV)**. SA und quantitative Messverfahren dienen der Erzeugung eines **Attendee Interaction and Impact Vocabulary (AIIV)**. TM wird auf zeitlich strukturierte Event-Inhalte angewendet, um ein **Emerging Topic Vocabulary (ETV)** und ein **Trending Topic Vocabulary (TTV)** zu erzeugen. PA-Methoden ergänzen diese Vokabulare um Bewertungen und Metriken zu Teilnehmerengagement, Teilnehmerzufriedenheit, Teilnehmerinteresse, Rednerleistung und Rednerreputation. In der nachfolgenden Phase der Informations- und Wissensorganisation sollen die extrahierten Vokabulare in hierarchische Strukturen überführt werden. Hierzu gehören insbesondere eine **Event Taxonomy (ET)**, eine **Attendee Interaction and Impact Taxonomy (AIIT)**, eine **Emerging Topic Taxonomy (ETT)** und eine **Trending Topic Taxonomy (TTT)**. Die resultierenden Wissensstrukturen bilden gemeinsam das PCO Knowledge und sollen über Such-, Browsing- und Navigationsfunktionen zugänglich gemacht werden. Dies ermöglicht Anwendungsfälle wie die Empfehlung und Planung von Veranstaltungen auf der Grundlage der Interessen und des Engagements der Teilnehmer, der Leistung und des Rufs der Referenten sowie der Entwicklung von Trending und Emerging Topics.

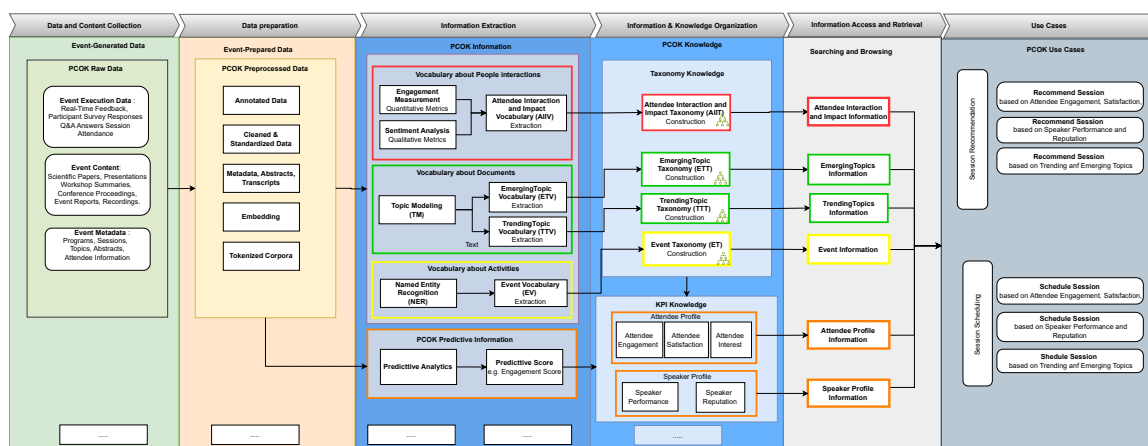


Figure 1: Vorgeschlagene IELC4AS2P-Architektur [22]

Aufgabenbeschreibung

Ziel der Abschlussarbeit ist die Konzeption, prototypische Implementierung und Evaluation eines **cloud-basierten, lifecycle-orientierten Systems** zur automatisierten Durchführung von **Topic Modelings** und zur **Konstruktion von Taxonomien** für **Trending und Emerging Topics** im MICE-Sektor. Im Zentrum steht die systematische Analyse von Event-Inhalten aus **heterogenen strukturierten und unstrukturierten Datenquellen**. Auf Basis dieser Inhalte sollen (i) **Trending- und Emerging-Vokabulare** extrahiert und (ii) daraus **Trending- und Emerging-Taxonomien** abgeleitet werden, die inkrementell über Event-Editionen hinweg aktualisierbar sind. Die vorliegende Arbeit orientiert sich an den Phasen des **IELC** [11] (*Data Collection*, *Data Preparation*, *Information Extraction (TM)*, *Information & Knowledge Organization (Taxonomie)* sowie *Information Access & Retrieval*) sowie an der im Rahmen der Dissertation entwickelten IELC4AS2P-Architektur. Für jede Phase sollen geeignete Verfahren identifiziert, bewertet und in eine konsistente Systemarchitektur integriert werden. Dabei ist es wichtig, Aspekte im Zusammenhang mit einer cloudbasierten Umsetzung zu berücksichtigen.

Die Umsetzung der Abschlussarbeit folgt der Forschungsmethodik nach Nunamaker [23] und gliedert sich in die folgenden Phasen:

Beobachtung (Observation, OP): Zunächst erfolgt eine Recherche zum Stand von Forschung und Technik für jede Phase des IELC im Hinblick auf ihre Anwendung in der MICE-Domäne. Dabei werden bestehende Verfahren und Ansätze zur *Sammlung* und *Aufbereitung* heterogener Veranstaltungsdaten, zur *Extraktion* von TTV und ETV, zur *Organisation* der extrahierten Vokabulare in Taxonomien sowie zum Informationszugang untersucht. Ergänzend werden cloud-basierte Technologien und Architekturansätze betrachtet, die eine skalierbare, modulare und reproduzierbare Umsetzung der einzelnen IELC-Phasen ermöglichen. Darüber hinaus werden geeignete quantitative Evaluationsverfahren und Metriken zur Bewertung der Qualität der erzeugten Vokabulare und Taxonomien recherchiert.

Theoriebildung (Theory Building, TBP): Auf Grundlage der Literatur wird eine Architektur für ein cloud-basiertes Topic-Modeling-System entwickelt. Die Architektur soll sich am IELC orientieren und geeignete Komponenten, Datenflüsse und Schnittstellen für die Verarbeitung, Analyse und Organisation der Daten modellieren. Die Modellierung erfolgt unter Anwendung geeigneter Software-Engineering-Methoden und der **Unified Modeling Language (UML)**.

Systementwicklung (System Development): Auf Basis der entwickelten Architektur wird ein Prototyp implementiert, der die wesentlichen Phasen des IELC unterstützt. Der Prototyp soll relevante Datenquellen integrieren, die Daten für die weitere Analyse aufbereiten, Verfahren des Topic Modelings anwenden, die daraus gewonnenen Informationen in geeignete Wissensstrukturen überführen und Zugangsmechanismen für die erzeugten Informationen bereitstellen. Die Umsetzung soll die Anforderungen einer cloud-basierten Architektur berücksichtigen.

Experimentierung und Evaluation (Experimentation): Abschließend wird der entwickelte Prototyp anhand geeigneter quantitativer Metriken evaluiert. Ergänzend werden qualitative Bewertungsmethoden, insbesondere ein Cognitive Walkthrough [24], durchgeführt, um die Benutzerfreundlichkeit, Nachvollziehbarkeit und Erklärbarkeit des Systems aus der Perspektive eines Veranstaltungsorganisations zu beurteilen. Die Ergebnisse der Evaluation werden zusammengeführt und diskutiert. Abschließend werden die Einschränkungen der entwickelten Lösung sowie mögliche zukünftige Verbesserungen und Weiterentwicklungen aufgezeigt.

Literatur

- [1] Matthias Schultze Wilhelm Bauer, Stefan Rief. Ökosysteme im wandel-zukunftsszenarien für business events im zeitalter grenzenloser kommunikation. In *Future Meeting Space. Fraunhofer-Institut für Arbeitswissenschaft und Organisation IAO [Themenheft]*. Fraunhofer IAO, 2022.
- [2] Hassnah Wee, Nurul Dafiqa Kamarulzaman, and Muhd Saufi Anas. Mice industry survival: A systematic literature review. *International Journal of Academic Research in Business and Social Sciences*, 2022.
- [3] Edgar Göll and Michaela Evers-Wölk. Tagung und kongress der zukunft - management summary, 01 2013.
- [4] Claus Bühnert and Stefan Luppold. *Praxishandbuch Kongress-, tagungs- und konferenzmanagement: Konzeption & Gestaltung, Werbung & PR, Organisation & Finanzierung*. Springer Gabler, 2017.
- [5] J.-D. Kim, T. Ohta, Y. Tateisi, H. Mima, and J. Tsujii. Xml-based linguistic annotation of corpus. In *Proceedings of the first NLP and XML Workshop*, pages 47–53, 2001.
- [6] Inmaculada Martín-Rojo and Ana Gaspar González. The impact of social changes on mice tourism management in the age of digitalization: a bibliometric review. *Review of Managerial Science*, pages 1–24, 02 2024.
- [7] Nurhazwani Abd Latif and Norol Zamzuri. Enhancing attendee engagement: A study on designing successful hybrid events in malaysia. *Information Management and Business Review*, 16:65–72, 09 2024.
- [8] Romi Setiawan, Dermawan Wibisono, and Mustika Purwanegara. Defining event marketing as engagement-driven marketing communication. *Gadjah Mada International Journal of Business*, 24:151, 06 2022.
- [9] Henry Small, Kevin W. Boyack, and Richard Klavans. Identifying emerging topics in science and technology. *Research Policy*, 43(8):1450–1467, 2014.
- [10] Shadi Alzubi, Bilal Hawashin, Ala Mughaid, and Yaser Jararweh. Whats trending? an efficient trending research topics extractor and recommender. In *2020 11th International Conference on Information and Communication Systems (ICICS)*, pages 191–196, 2020.
- [11] Philippe Tamla. Towards a reference model for an information extraction lifecycle supporting data collection, data preparation, information extraction, information and knowledge organization, and information access and retrieval. Technical report, FernUniversität in Hagen, 2023. Available via the institutional repository of FernUniversität in Hagen.
- [12] FTK. Artificial Intelligence for Hospitals, Healthcare & Humanity (AI4H3). R&D White Paper, FTK e.V. Research Institute for Telecommunications and Cooperation, Dortmund, Germany, April 2025.
- [13] Chair of Multimedia and Internet Applications. <http://www.lgmmia.fernuni-hagen.de>, 2023.
- [14] FTK e.V. Research Institute for Telecommunications and Cooperation. <http://www.ftk.de>, 2023.
- [15] Philippe Tamla, Gladys Kouakam, Benedict Hartmann, Matthias Hemmje. Tsias: A framework for thematically structured information access in green ports and green shipping document collections. R&D White Paper, Dortmund, Germany, April 2025.
- [16] FTK. Artificial Intelligence for Hospitals, Healthcare & Humanity (AI4H3). R&D White Paper, Dortmund, Germany, April 2020.
- [17] Bielefeld University. RATIO: Rationalizing Recommendations (RecomRatio), 2017.
- [18] Binh Vu, Yanxin Wu, Haithem Afli, Paul Mc Kevitt, Paul Walsh, Felix Engel, Michael Fuchs, and Matthias Hemmje. A metagenomic content and knowledge management ecosystem platform. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1–8. IEEE, 2019.

- [19] Globit – global information technology gmbh, 2025.
- [20] Rachael Lammey. Using the crossref metadata api to explore publisher content. *Science Editing*, 3(2):109–111, 2016.
- [21] Patrice Lopez. Grobid. <https://github.com/kermitt2/grobid>, 2008–2024.
- [22] FTK. Cloud based event organization and management. R&D White Paper, Dortmund, Germany, April 2025.
- [23] Jay F. Nunamaker Jr, Minder Chen, and Titus D. M. Purdin. Systems Development in Information Systems Research. *Journal of Management Information Systems*, 7(3):89–106, December 1990. Publisher: Routledge .eprint: <https://doi.org/10.1080/07421222.1990.11517898>.
- [24] Peter G. Polson, Clayton Lewis, John Rieman, and Cathleen Wharton. Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. *International Journal of Man-Machine Studies*, 36(5):741–773, May 1992.