

Meta Milestones

15 Jahre Pionierarbeit in Partnerschaft und Innovation

Autor:innen: Jan Paffenholz, Ardit Saliji, Clemens Scheuchenpflug, Felix Stadler, Simon Ulmer, Sebastian Wallner, Anil Yildiz

Universitäten sind darauf bedacht, akademisches Lernen mit praktischer Anwendung zu verbinden. Um das zu erreichen, werden Studierende in praktische Projekte eingebunden, welche häufig zur Erstellung von Seminararbeiten führen.

Diese wissenschaftlichen Arbeiten können wertvolle Informationen für künftige Forschungen oder Anwendungen in institutionellen Repositorien liefern (Fan et al., 2019). Jedoch erschwert die unstrukturierte Form der Forschungsdokumente die Abfrage über mehrere Dokumente hinweg (Fan et al., 2019; Herrera-Urtiaga et al., 2023).

Dieses Projekt befasst sich mit genau diesem Problem und der Lösung durch eine Metadatenextrahierung. Daher war es das Ziel dieser Arbeit, ein Forschungskonzept zur Extraktion und Nutzung von Metadaten aus unstrukturierten Publikationen zu entwickeln. Das entwickelte Konzept sollte ebenfalls fortschrittliche Visualisierungstechniken verwenden, um eine Analyse wissenschaftlicher Arbeiten zu ermöglichen.

Für die Umsetzung wurde im Rahmen dieser Seminararbeit die Design Science Research-Methode laut Peffers et al. (2006) verwendet und folgt dem sechsstufigen Prozess zur Entwicklung und Bewertung eines Artefakts zur Lösung eines spezifischen Problems.

Im Rahmen des Projektes wurden anhand von wissenschaftlicher Literatur relevante Metadaten, eine geeignete Extrahierungsmethode sowie ein Ansatz zur Informationsgewinnung identifiziert. Diese Informationen stellten die Grundlage für den entwickelten Prototypen.

Als Prototyp wurde eine Webapplikation entwickelt, welche es den Lehrenden sowie den Studierenden ermöglicht, Seminararbeiten in eine Web-Oberfläche hochzuladen. Mit Hilfe eines Large Language Models werden relevante Metadaten ausgesehen. Für die Visualisierung wird ebenfalls ein Dashboard zur Verfügung gestellt, welches den Lehrenden Einblick über die relevantesten Kennzahlen über alle Seminararbeiten hinweg verschafft.

Literatur

- Fan, T., Liu, J., Qiu, Y., Jiang, C., Zhang, J., Zhang, W., & Wan, J. (2019). PARDA: A Dataset for Scholarly PDF Document Metadata Extraction Evaluation. In H. Gao, X. Wang, Y. Yin, & M. Iqbal (Hrsg.), *Collaborative Computing: Networking, Applications and Worksharing* (S. 417–431). Springer International Publishing. https://doi.org/10.1007/978-3-030-12981-1_29
- Herrera-Urtiaga, A. P., Torres-Cruz, F., Mendoza-Mollocondo, C. I., Paredes-Quispe, J.-R., & Chambi-Mamani, E. W. (2023). Automatic Extraction of Metadata Based on Natural Language Processing for Research Documents in Institutional Repositories. In Y. Iano, O. Saotome, G. L. Kemper Vásquez, M. T. de Moraes Gomes Rosa, R. Arthur, & G. Gomes de Oliveira (Hrsg.), *Proceedings of the 8th Brazilian Technology Symposium (BTSym'22)* (S. 189–197). Springer International Publishing. https://doi.org/10.1007/978-3-031-31007-2_19
- Peffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2006). The design science research process: A model for producing and presenting information systems research. 83–106.

Meta Milestones

15 Years of Pioneering in Partnership and Innovation

Authors: Jan Paffenholz, Ardit Saliji, Clemens Scheuchenpflug, Felix Stadler, Simon Ulmer, Sebastian Wallner, Anil Yildiz

Universities are aiming to combine academic learning with practical application. To achieve this, students are involved in practical projects, which often lead to the production of seminar papers.

These academic research papers can collectively provide valuable information for future research or applications in institutional repositories (Fan et al., 2019). However, the unstructured form of research documents makes it difficult to query across multiple documents (Fan et al., 2019; Herrera-Urtiaga et al., 2023).

This project deals with exactly this problem and the solution through metadata extraction. Therefore, the aim of this work was to develop a research concept for extracting and using metadata from unstructured publications. The developed concept should also use advanced visualization techniques to enable the analysis of scientific papers.

This seminar paper uses the Design Science Research method according to Peffers et al. (2006) and follows a six-step process to develop and evaluate an artifact to solve a specific problem.

During the project, relevant metadata, a suitable extraction method, and an approach for gaining information were identified based on scientific literature. This information formed the basis for the implemented prototype.

As a prototype, a web application was developed that enables lecturers and students to upload seminar papers to a web interface. Relevant metadata is extracted with the help of a large language model. A dashboard is also provided for visualization, which gives lecturers an insight into the most relevant key figures across all seminar papers.

Literature

- Fan, T., Liu, J., Qiu, Y., Jiang, C., Zhang, J., Zhang, W., & Wan, J. (2019). PARDA: A Dataset for Scholarly PDF Document Metadata Extraction Evaluation. In H. Gao, X. Wang, Y. Yin, & M. Iqbal (Hrsg.), *Collaborative Computing: Networking, Applications and Worksharing* (S. 417–431). Springer International Publishing. https://doi.org/10.1007/978-3-030-12981-1_29
- Herrera-Urtiaga, A. P., Torres-Cruz, F., Mendoza-Mollocondo, C. I., Paredes-Quispe, J.-R., & Chambi-Mamani, E. W. (2023). Automatic Extraction of Metadata Based on Natural Language Processing for Research Documents in Institutional Repositories. In Y. Iano, O. Saotome, G. L. Kemper Vásquez, M. T. de Moraes Gomes Rosa, R. Arthur, & G. Gomes de Oliveira (Hrsg.), *Proceedings of the 8th Brazilian Technology Symposium (BTSym'22)* (S. 189–197). Springer International Publishing. https://doi.org/10.1007/978-3-031-31007-2_19
- Peffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2006). The design science research process: A model for producing and presenting information systems research. 83–106.