

Verknüpfungen und Kontextualisierung durch Annotationen - Forschen mit multimodalen Daten.

Kröber, Cindy

cindy.kroeber@uni-jena.de
FSU Jena, Deutschland

Bruschke, Jonas

jonas.bruschke@uni-wuerzburg.de
Uni Würzburg, Deutschland

Utescher, Ronja

ronja.utescher@uni-jena.de
FSU Jena, Deutschland

Maiwald, Ferdinand

ferdinand.maiwald@uni-jena.de
FSU Jena, Deutschland

Pattee, Aaron

aaron.pattee@lmu.de
LMU München, Deutschland

Abstract

In der kunsthistorischen Forschung ist die eingehende und umfassende Auseinandersetzung mit Quellenmaterial verschiedenster Natur ein aufwendiger und anspruchsvoller Schritt, um ein ganzheitliches Verständnis des Forschungsgegenstandes zu erlangen. Dabei ist neben dem Inhalt der Quellen auch ihre Verlässlichkeit und ursprüngliche Intention relevant. Annotationen können dabei behilflich sein, verhältnismäßig schnell einen Überblick über den Inhalt einer Quelle zu bekommen, sowie Widersprüche und Zusammenhänge hervorzuheben.

Dieser Beitrag zeigt auf, welches Potential die Verknüpfung von multimodalen Daten mittels Annotationen bietet. Dazu wird skizziert welche Arbeitsschritte für die Annotierung von 3D-Modellen, Bildern und Texten nötig sind und welche Möglichkeiten der Automatisierung sich abzeichnen. Verschiedene Anwendungsszenarien für die Forschungsplattform 4D Browser verdeutlichen Einsatzmöglichkeiten für die Forschung in der Kunst- und Architekturgeschichte.

Einleitung

Texte wie auch Fotografien und Bilder sind wichtiges Quellenmaterial für die geschichtswissenschaftliche Forschung und bilden die Grundlage für themen- und theorieorientierte Untersuchungen u.a. in der Architekturwissenschaft, der Kunstgeschichte und den Kulturwissenschaften. Die Quellen ermöglichen eine (digitale) Rekonstruktion oder Untersuchung verschiedener Gebäude, ihrer Baugeschichte und möglicher Effekte auf die Stadtentwicklung. Um ein Gebäude umfassend zu analysieren, ist sehr viel Quellenmaterial erforderlich. Die Suche nach Bildern und Texten, aber auch deren Kontextualisierung und Auswertung ist herausfordernd (Beaudoin and Brady 2011). Einerseits gibt es eine riesige Menge an online verfügbaren Daten, andererseits ist eine Filterung der Ergebnisse in der Regel nicht zufriedenstellend und wird teilweise sogar vermieden.

Im BMBF-geförderten Projekt HistKI wird an einer Lösung gearbeitet, die unterschiedliche Daten in einer Forschungsplattform namens 4D Browser (<https://4d-browser.urbanhistory4d.org/>) verknüpft und somit die Verarbeitung von multimodalen Daten für die Forschung unterstützt. Die prototypische Webanwendung bietet eine Vielzahl an Funktionalitäten und wird stetig weiterentwickelt (Dewitz et al. 2019). Historische Bild- und Textquellen sind in einem virtuellen 3D Stadtmodell von Dresden verortet (siehe Abbildung 1). Eine Zeitschiene kann neben einer Suchleiste für die Filterung der Quellen genutzt werden. Die Daten können neben der altbekannten Metadatenabfrage auch durch die Projektion auf das 3D-Modell gefiltert werden. Somit können Bilder von Gebäuden gefunden werden, ohne dass der Objektname Bestandteil der Metadaten ist. Zusätzlich stehen Visualisierungswerkzeuge für die Analyse von Kamerastandpunkten zur Verfügung, die Forschung im Bereich der Stadtentwicklung und -wahrnehmung unterstützen.



Abbildung 1. Benutzeroberfläche des 4D Browsers mit räumlich und zeitlichen verorteten Fotos in einem 3D-Stadtmodell.

Annotieren der Daten

Nachdem in einem ersten Schritt in der Forschungsplattform 4D Browser Bilder, 3D Modelle und Texte zur Ver-

fügung gestellt wurden, sollen diese Daten nun im Sinn der Multimodalität miteinander verknüpft werden. Zur Verknüpfung der drei Datentypen werden Wörter bzw. Begriffe wie architektonische Elemente und Bezeichnungen aus der Stadt- und Architekturgeschichte als Label bzw. Annotation verwendet. Im Bereich der digitalen Geisteswissenschaften ist der Getty Art & Architecture Thesaurus (AAT) etabliert und bietet eine Hierarchie für architektonische Elemente (Baca and Gill 2015). Des Weiteren wird auch Wikidata mit einer Vielzahl an Entitäten und Klassen sowie entsprechenden semantischen Beziehungen untereinander für u.a. Kunst und architektonische Elemente zunehmend im Bereich des kulturellen Erbes verwendet (Schmidt, Thiery, and Trognitz 2022). Relevante Begriffe wurden mit Hilfe einer Textanalyse von Beiträgen zum Dresdner Zwinger bzw. barocker Architektur, Baugeschichte und Bauforschung identifiziert (Utescher et al. 2022). Schließlich konnte eine Liste mit ca. 400 Einträgen - darunter 140 Architekturelemente - zusammengestellt werden, die sowohl AAT- wie auch Wikidata IDs enthält.

Texte

Bisher stehen mehrere Wikipedia-Artikel, wissenschaftliche Publikationen und populärwissenschaftliche Beiträge im 4D Browser zur Verfügung. Teilweise können Dokumente mit Hilfe eines PDF-Viewers betrachtet werden, der auch weitere Werkzeuge bereitstellt. Einige Artikel wurden bereits annotiert. Für diese steht zusätzlich ein ‚Plain Text‘ bereit, in dem die Annotationen eingeblendet werden können (siehe Abbildung 2). Die Annotierung von Text wurde mit Hilfe von KI-Ansätzen realisiert. Für semantische Textannotationen basierend auf der Identifikation von Wortähnlichkeiten kommt fastText zu Einsatz (Bruschke et al. 2023). In den Texten werden Wörter oder Wortgruppen aus der zusammengestellten Begriffsliste annotiert sowie Begrifflichkeiten mit einer hohen semantischen Ähnlichkeit. Named Entity Recognition (spaCy’s NER) als Transformer-basiertes Model kommt zur Eigennamenerkennung zum Einsatz. So können Personen und Orte im Text automatisch annotiert werden.

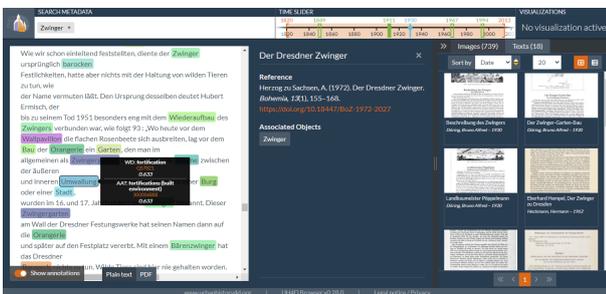


Abbildung 2. Beispiel eines annotierten Textes im 4D Browser.

3D Modelle

Im Projekt wurde für Testzwecke ein detailliertes 3D-Modell des Kronentors vom Dresdner Zwinger in einer Modellierungsumgebung manuell annotiert. In diesem Modell sind die meisten architektonischen Elemente separate Objekte, die neben der Bezeichnung des Elements noch die entsprechende AAT- oder Wikidata-ID tragen. Einige Objekte sind gruppiert und bilden eine Einheit im Sinn einer Hierarchie von Elementen.

Fotografien

Bisher wurden im 4D Browser ca. 4000 meist historische Fotografien von Dresden manuell oder mittels eines semi-automatischen Ansatzes verortet (Maiwald 2022, Maiwald et al. 2023). Neu hinzugekommen ist ein kleiner, manuell segmentierter Bilddatensatz (siehe Abbildung 3). Die Segmentierung und Annotierung erfolgte mit Label Studio (https://labelstud.io/), einer Anwendung zur Beschriftung und Annotation von Daten, welches ein standardisiertes Ausgabeformat bietet. Jede Benennung folgt dem gleichen Schema wie bei den 3D-Modelle und enthält sowohl AAT- als auch Wikidata-IDs. Eine automatische Übertragung der Annotationen vom 3D-Modell auf die verorteten Fotografien ist geplant, allerdings erreichten bisherige Ansätze noch nicht den angestrebten Grad an Zuverlässigkeit bzw. Genauigkeit.

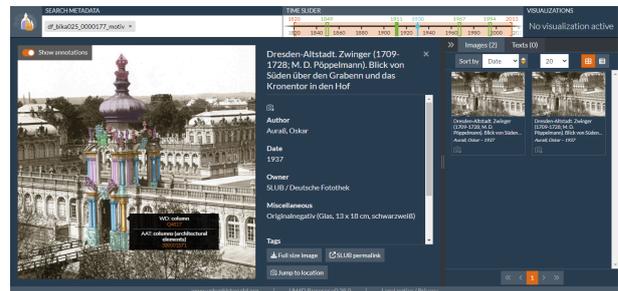


Abbildung 3. Beispiel eines segmentierten und annotierten Fotos vom Dresdner Zwinger im 4D Browser. Foto: Oskar Aurb, https://www.deutschefototek.de/documents/obj/72018745

Mit den neuen Entwicklungen in Bereich der KI können viele der bisher manuell bearbeiteten Schritte automatisiert werden. Jedoch existierte bisher kein Trainingsdatensatz, der die domänenspezifischen architektonischen Elemente enthält, die für die Forschungsszenarien relevant sind. Daher sind unsere annotierten Daten Teil eines öffentlich-zugänglichen Datensatzes und sollen perspektivisch als Trainingsdatensatz genutzt werden (Bruschke et al. 2023).

Verknüpfung der verschiedenen Daten

Neben einer Visualisierung der Annotationen in Text, 3D und Bild bietet der 4D Browser noch eine Übersicht aller Annotationen eines Objektes in einer Wordcloud (siehe Abbildung 4). Wordclouds gibt es für alle drei Datentypen.

chitektur und architektonischen Elementen eingegangen. Bei der Untersuchung eines bestimmten architektonischen Elementes wie einer Skulptur stellen sich folgende Fragen: Wo kann man die Skulptur finden? Wer hat sie geschaffen und wann? Was oder wer ist abgebildet? Warum wurden Motiv und Standort gewählt? Wie wurde oder wird die Skulptur wahrgenommen und wie wirkt sie im Kontext des Gebäudes und der Umgebung? Gibt es ähnliche oder andere Skulpturen, die es wert sind, untersucht zu werden, z.B. innerhalb des Gebäudes, mit einem ähnlichen Motiv, oder von demselben Schöpfer? Ist die aktuelle Skulptur das Original oder eine Kopie? Stand sie schon immer an diesem Ort oder wurde sie versetzt? Ausgangspunkt für die Untersuchung ist das 3D-Modell an der Position der Skulptur. Verortete Fotos geben Hinweise zum Aussehen der Figur. Relevante Texte automatisch verknüpft durch Annotationen können helfen weitere Informationen zu finden. Bestimmte Textpassagen erlauben eine direkte Zuordnung zu Bildern, da sie detaillierte Beschreibungen zu architektonischen Elementen enthalten und zusätzliche Angaben zur Position wie nördlich, rechts oder unten. Mit Hilfe von textlichen Quellen lassen sich meist weitere Schlagwörter oder Quellen finden.

Der Vergleich ist eine unglaublich wichtige Methode in der kunsthistorischen Forschung. Ein Ziel kann sein, die Entwicklung eines Künstlers anhand seiner Werke zu untersuchen oder die Inspiration für ein bestimmtes Werk zu erforschen. In vielen Fällen liefert die Biographie oder das Werksverzeichnis eines Künstlers oder Architekten die notwendigen Namen und Orte, welche in den Texten im 4D Browser annotiert sind und somit schnell erschlossen und zeitlich eingeordnet werden können. Eine Recherche nach relevanten Informationen zu einer Person kommt mit dem 4D Browser einer ganzheitlichen Erschließung der Thematik nahe, da die Quellen miteinander verknüpft sind und man sich von Information zu Information weiterbewegt. Orte und Objekte regen zu weiterer Erkundung an oder bieten neue Schlüsselwörter. Texte enthalten Beschreibungen, die über Annotationen mit Bildern verknüpft sind, welche wiederum visuell verglichen werden können. Da multimodale Daten im 4D-Browser verknüpft sind, wäre die Recherche schneller und umfassender im Sinne einer ganzheitlichen Erschließung.

Fazit

Der 4D Browser wird seit 2016 stetig weiterentwickelt. Die Verknüpfung der verschiedenen Datentypen durch Annotationen eröffnet neue Forschungs- und Analyseansätze und ist ein großer Gewinn für die Forschungsplattform. Um einen richtigen Mehrwert für Wissenschaftler darzustellen, bedarf es aber eines sehr großen Datensatzes mit automatisch annotierten Daten von hoher Qualität und Zuverlässigkeit. Dann können neue Filter- und Analysewerkzeuge helfen, den Daten weitere Erkenntnisse zu entlocken. Dieser Werkstattbericht bietet dafür einen Einblick in die aktuellen Möglichkeiten und Entwicklungen.

Bibliographie

Baca, Murtha, and Melissa Gill. 2015. "Encoding multilingual knowledge systems in the digital age: the getty vocabularies." *NASKO*:41-63.

Beaudoin, Joan E., and Jessica Evans Brady. 2011. "Finding Visual Information: A Study of Image Resources Used by Archaeologists, Architects, Art Historians, and Artists." *Art Documentation: Journal of the Art Libraries Society of North America* 30 (2):24-36.

Bruschke, J, C Kröber, F Maiwald, R Utescher, and A Pattee. 2023. "Introducing a Multimodal Dataset for The Research of Architectural Elements." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 48:325-331.

Dewitz, Leyla, Cindy Kröber, Heike Messemer, Ferdinand Maiwald, Sander Münster, Jonas Bruschke, and Florian Niebling. 2019. "Historical Photos and Visualizations: Potential for Research." CIPA 2019, Ávila.

Kröber, Cindy, Katharina Hammel, Cornelia Schade, Nicole Filz, and Leyla Dewitz. 2021. "User Involvement for Application Development: Methods, Opportunities and Experiences from Three Different Academic Projects." Workshop on Research and Education in Urban History in the Age of Digital Libraries.

Maiwald, Ferdinand. 2022. "A window to the past through modern urban environments." Phd, Technische Universität Dresden.

Maiwald, Ferdinand, Jonas Bruschke, Danilo Schneider, Markus Wacker, and Florian Niebling. 2023. "Giving Historical Photographs a New Perspective: Introducing Camera Orientation Parameters as New Metadata in a Large-Scale 4D Application." *Remote Sensing* 15 (7):1879.

Schmidt, Sophie C, Florian Thiery, and Martina Trognitz. 2022. "Practices of linked open data in archaeology and their realisation in Wikidata." *Digital* 2 (3):333-364.

Utescher, Ronja, Aaron Patee, Ferdinand Maiwald, Jonas Bruschke, Stephan Hoppe, Sander Münster, Florian Niebling, and Sina Zarriß. 2022. "Exploring Naming Inventories for Architectural Elements for Use in Multi-modal Machine Learning Applications." Workshop on Computational Methods in the Humanities 2022.