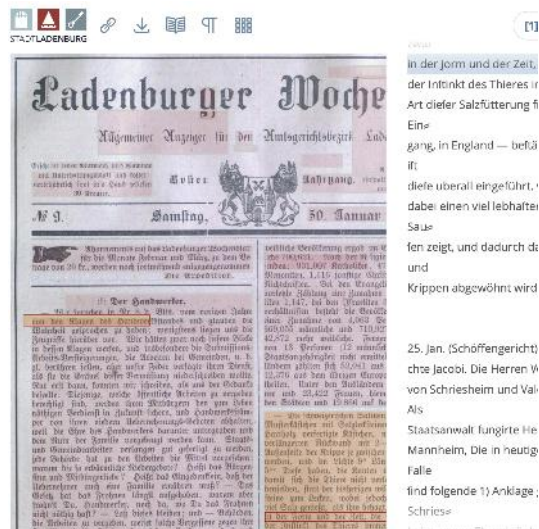


Historische Lokalzeitungen

Kitodo.Presentation

Künstliche Intelligenz



Kitodo Praxistreffen
2025 in Köln

Stefan Weil
UB Mannheim



Programm

- Historische Lokalzeitungen online
- Texterkennung mit KI
- Weitere KI-Anwendung für Kitodo.Presentation

Historische Lokalzeitungen online

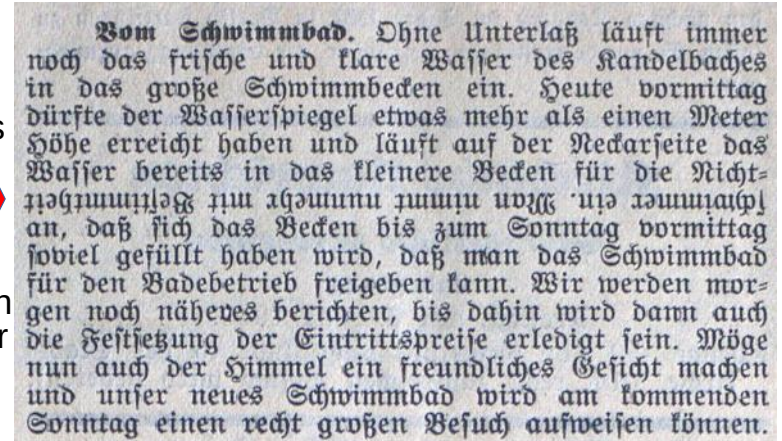
- Ladenburger Wochenblatt (1869–1907), Ladenburger Tageblatt (1907–1919), Neckar-Bergstraß-Post (1919–1941), alle digitalisiert durch Stadtarchiv Ladenburg, 12986 Ausgaben, 65886 Seiten
- Standardkonforme Daten (METS/MODS, ALTO, JPEG)
- PDF nach JPEG mit Python, Metadaten erzeugt mit Python aus einfacher JSON-Beschreibung, Texterkennung mit Tesseract
- Kitodo.Presentation (main branch), DFG-Viewer, Debian trixie, TYPO3 v12.4.39, PHP 8.4
- Verfügbar im Deutschen Zeitungsportal:
<https://www.deutsche-digitale-bibliothek.de/newspaper>



Texterkennung mit LLM

Aufgabe: Transkribiere originalgetreu den kompletten Text.

- **ChatGPT:** Vom Schwimmbad. Ohne Unterlaß läuft immer noch das frische und klare Wasser des Kandelbaches in das große Schwimmbecken ein. Heute vormittag dürfte der Wasserspiegel etwas mehr als einen Meter Höhe erreicht haben und läuft auf der Neckarseite das Wasser bereits in das kleinere Becken für die Nichtschwimmer mit ab. Es ist anzunehmen, daß sich das Becken bis zum Sonntag vormittag soviel gefüllt haben wird, daß man das Schwimmbad für den Badebetrieb freigeben kann. Wir werden morgen noch näheres berichten, bis dahin wird dann auch die Festsetzung der Eintrittspreise erledigt sein. Möge nun auch der Himmel ein freundliches Gesicht machen und unser neues Schwimmbad wird am kommenden Sonntag einen recht großen Besuch aufweisen können.
sehr gutes Ergebnis
- **Claude.AI:** Claude kann auf diese Anfrage nicht antworten, da sie offenbar gegen unsere Nutzungsrichtlinien verstößt. Bitte starten Sie einen neuen Chat.
kein Ergebnis

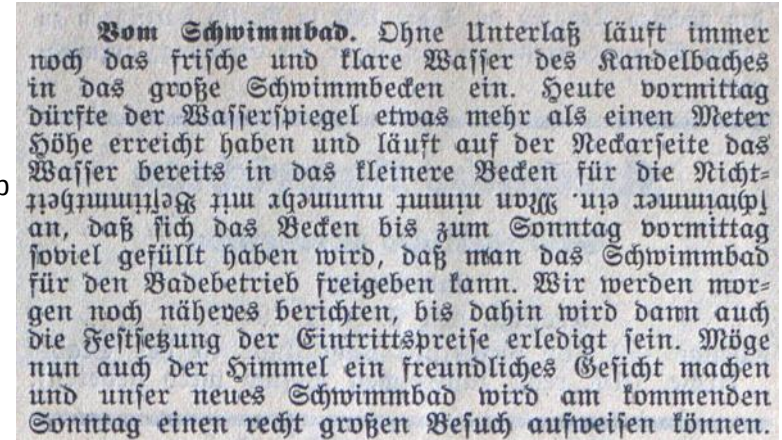


Neckar-Bergstraß-Post Nr. 135
vom 5. Juli 1935 mit
drucktechnischer Besonderheit

Texterkennung mit LLM

Aufgabe: Transkribiere originalgetreu den kompletten Text.

- **Llama4:** Vom Schwimmbad. Ohne Unterlaß läuft immer das frische und klare Wasser des Randlebaches in das große Schwimmbecken ein. Heute vormittag dürfte der Wasserspiegel etwas mehr als einen Meter erreicht haben und läuft auf der rechten Seite das Wasser bereits in das kleinere Becken für die Nichtschwimmer. Bis Sonntag vormittag ist das Becken soweit gefüllt, daß man das Schwimmbad für den Badebetrieb freigeben kann. Wir werden morgen früh näheres berichten, bis dahin wird auch die Festsetzung der Eintrittspreise erledigt sein. Mög der Himmel ein freundliches Gesicht machen und unser neues Schwimmbad am kommenden Sonntag einen recht großen Besuch aufweisen.
mehrere Fehler, fehlende Worte
- **Mistral-small3.2:** Vom Schwimmbad. Ohne Unterlaß läuft immer noch das frische und klare Wasser des Kandelbaches in das große Schwimmbecken ein. Heute vormittag dürfte der Wasserspiegel etwas mehr als einen Meter Höhe erreicht haben und läuft auf der Medarseite das Wasser bereits in das kleinere Becken für die Nichtschwimmer an, daß sich das Becken bis zum Sonntag vormittag soviel gefüllt haben wird, daß man das Schwimmbad für den Badebetrieb freigeben kann. Wir werden morgen noch näheres berichten, bis dahin wird dann auch die Festsetzung der Eintrittspreise erledigt sein. Möge nun auch der Himmel ein freundliches Gesicht machen und unser neues Schwimmbad wird am kommenden Sonntag einen recht großen Besuch aufweisen können.
gutes Ergebnis, wenige Fehler

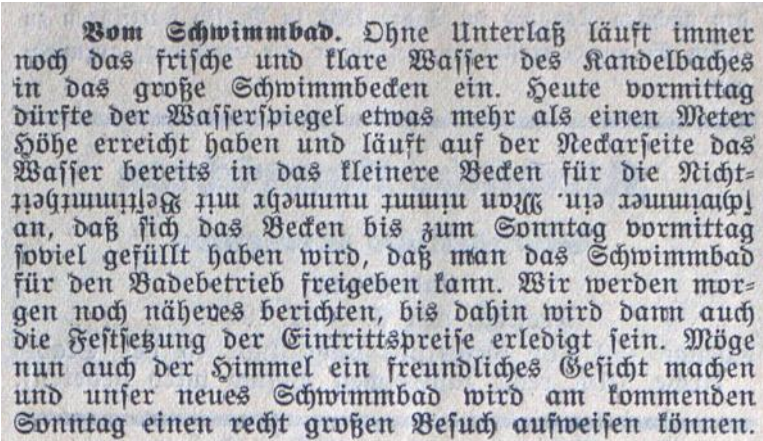


Vom Schwimmbad. Ohne Unterlaß läuft immer noch das frische und klare Wasser des Kandelbaches in das große Schwimmbecken ein. Heute vormittag dürfte der Wasserspiegel etwas mehr als einen Meter Höhe erreicht haben und läuft auf der Medarseite das Wasser bereits in das kleinere Becken für die Nichtschwimmer an, daß sich das Becken bis zum Sonntag vormittag soviel gefüllt haben wird, daß man das Schwimmbad für den Badebetrieb freigeben kann. Wir werden morgen noch näheres berichten, bis dahin wird dann auch die Festsetzung der Eintrittspreise erledigt sein. Möge nun auch der Himmel ein freundliches Gesicht machen und unser neues Schwimmbad wird am kommenden Sonntag einen recht großen Besuch aufweisen können.

Texterkennung mit LLM

Aufgabe: Transkribiere originalgetreu den kompletten Text.

- **qwen3-vl:** Vom Schwimmbad. Ohne Unterlaß läuft immer noch das frische und klare Wasser des **H**andelbaches in das große Schwimmbecken ein. Heute vormittag dürfte der Wasserspiegel etwas mehr als einen Meter Höhe erreicht haben und läuft auf der Neckarseite das Wasser bereits in das kleinere Becken für die Nichtschwimmer hin, so daß dasselbe nunmehr voll 'bis anmaßlich' an, daß sich das Becken bis zum Sonntag vormittag voll gefüllt haben wird, daß man das Schwimmbad für den Badebetrieb freigeben kann. Wir werden morgen noch näheres berichten, bis dahin wird dann auch die Festsetzung der Eintrittspreise erledigt sein. Möge nun auch der Himmel ein freundliches Gesicht machen und unser neues Schwimmbad wird am kommenden Sonntag einen recht großen Besuch aufweisen können.
gutes Ergebnis, wenige Fehler, zeilengetreu



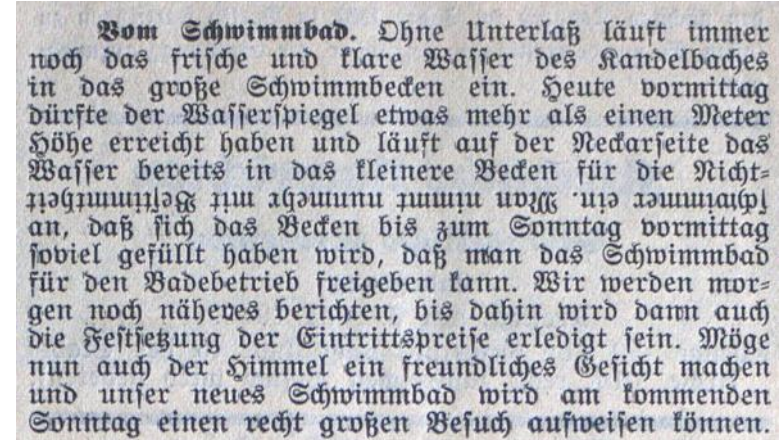
Vom Schwimmbad. Ohne Unterlaß läuft immer noch das frische und klare Wasser des Handelbaches in das große Schwimmbecken ein. Heute vormittag dürfte der Wasserspiegel etwas mehr als einen Meter Höhe erreicht haben und läuft auf der Neckarseite das Wasser bereits in das kleinere Becken für die Nichtschwimmer hin, so daß dasselbe nunmehr voll 'bis anmaßlich' an, daß sich das Becken bis zum Sonntag vormittag voll gefüllt haben wird, daß man das Schwimmbad für den Badebetrieb freigeben kann. Wir werden morgen noch näheres berichten, bis dahin wird dann auch die Festsetzung der Eintrittspreise erledigt sein. Möge nun auch der Himmel ein freundliches Gesicht machen und unser neues Schwimmbad wird am kommenden Sonntag einen recht großen Besuch aufweisen können.

Texterkennung mit Tesseract

Neues Tesseract-Modell german_print2_15

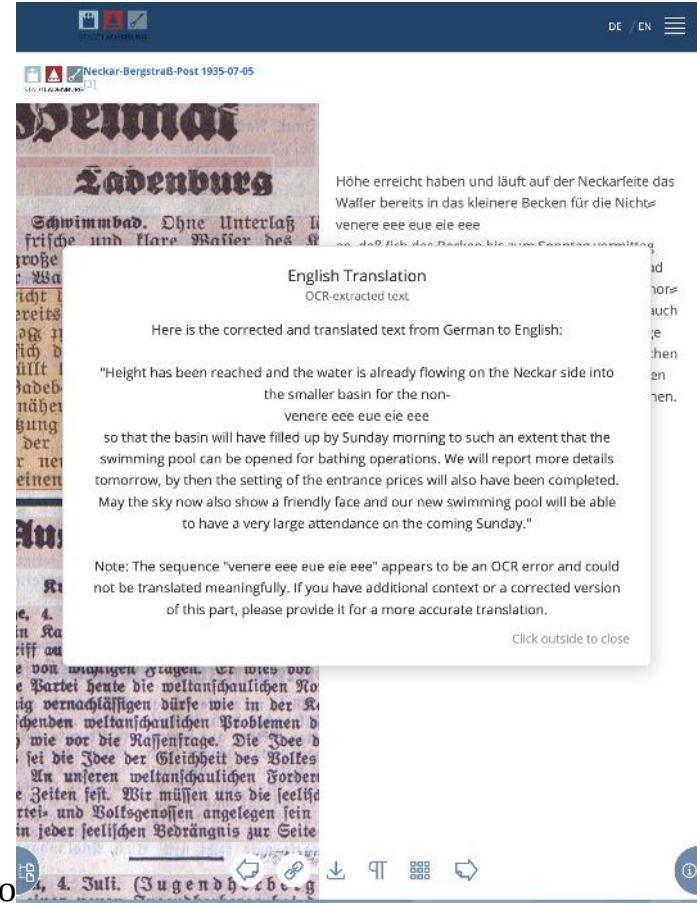
- Vom Schwimmbad. Ohne Unterlaß läuft immer noch das frische und klare Wasser des Kandelbaches in das große Schwimmbecken ein. Heute vormittag dürfte der Wasserspiegel etwas mehr als einen Meter **ö**he erreicht haben und läuft auf der Neckarseite das **a**ler bereits in das kleinere Becken für die Nicht= **oneeeee me ag eene eeeeeee ee eeeeteee** an, daß sich das Becken bis zum Sonntag vormittag soviel gefüllt haben wird, daß man das Schwimmbad für den Badebetrieb freigeben kann. Wir werden mor= **g**en noch näheres berichten, bis dahin wird dann auch die Festsetzung der Eintrittspreise erledigt sein. Möge nun auch der Himmel ein freundliches Gesicht machen und unser neues Schwimmbad wird am kommenden Sonntag einen recht großen Besuch aufweisen können.

**gutes Ergebnis, wenige Fehler, zeilengetreu, unschlagbar
schnell, könnte mit Codekorrektur und Codeerweiterung
perfektes Ergebnis auch für die gedrehte Zeile liefern**



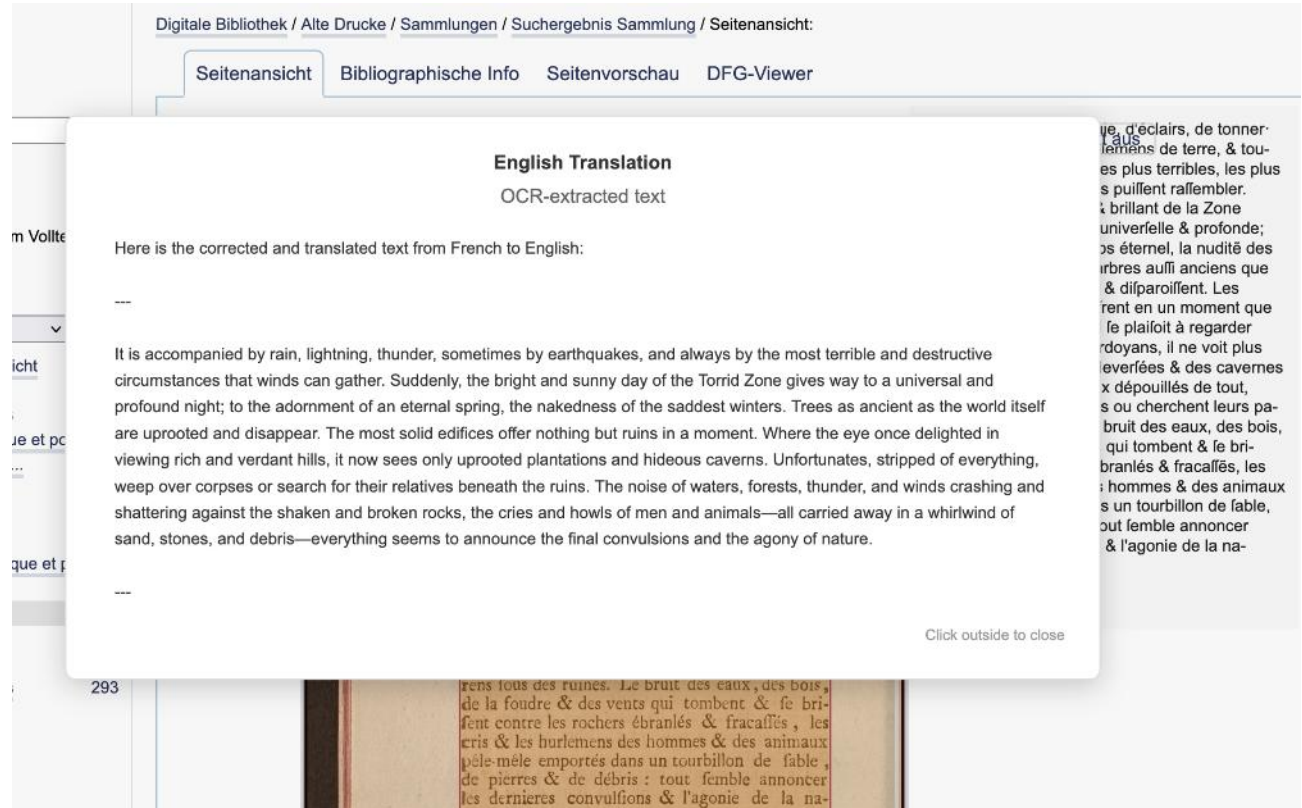
Kitodo.Presentation + KI

- Bookmarklet mit JavaScript für Kitodo-Volltextanzeige
- Serverseitige Übersetzung der OCR-Texte mit LLM
- Funktioniert direkt für jede Installation von Kitodo



Kitodo.Presentation + KI-Übersetzer

- Übersetzt alle gängigen Sprachen (Französisch, Latein, ...) in die Zielsprache
- Beliebige Kitodo-Instanz



Links

- Weil, S. (2024). Digitalisierungsaktivitäten und Kitodo an der UB Mannheim. Kitodo Praxistreffen 2024, Marburg
- Historische Ladenburger Zeitungen
 - Minimalistisch mit DFG-Viewers: <https://digital.ladenburg.de/>
 - Test mit neuestem Presentation: <https://digital.ladenburg.world/>
- Tesseract-Modell `german_print2_15` und weitere:
https://ub-backup.bib.uni-mannheim.de/~stweil/tesstrain/tesseract/german_print/20250915/
- Translation bookmarklet:
<https://github.com/UB-Mannheim/kitodo-presentation/tree/ai-translation/contrib/ai-translation>