

BSc Computational and Data Science  
CDS205 Computer Science  
Dozentin: Prof. Dr. Ana Petrus

# Analyse von nachträglichen Artikeländerungen bei Schweizer Online-Zeitschriften

Dario Hollbach<sup>1,\*</sup>

<sup>1</sup>*Fachhochschule Graubünden*

<sup>\*</sup>*E-Mail Adresse: dario.hollbach@stud.fhgr.ch*

9. Dezember 2025

## Zusammenfassung

Im Gegensatz zu Printmedien können Artikel in Online-Zeitschriften auch nach der Veröffentlichung modifiziert werden. Diese Arbeit untersucht die Häufigkeit und Art solcher nachträglichen Artikeländerungen bei den Schweizer Zeitschriften *20 Minuten*, *Watson* und *SRF News*. Hierfür wurden eine Python-Applikation entwickelt, welche auf einem Raspberry-Pi über den Zeitraum von zehn Tagen systematisch Artikel erfasst und Versionen vergleicht. Die in der Datenbank gespeicherten Änderungen wurden mithilfe einer Web-Applikation visualisiert und manuell als Tippfehler oder inhaltliche Anpassung klassifiziert. Die Analyse zeigt, dass rund jeder fünfte Artikel mindestens ein Mal geändert wurde. Die Ergebnisse zeigen, wie Schweizer Online-Zeitschriften Korrekturen und Anpassungen vornehmen.

# 1 Einleitung

Die Präsenz von Online-Zeitschriften wird stetig grösser. Gemäss dem Medienbarometer der Schweiz haben die Online-Zeitschriften im Jahr 2023 eine Meinungsmacht von 27% (BAKOM, 2023). Im Vergleich zu den Printmedien gibt es für die digitalen Newsportale die Möglichkeit, Artikel nach der Veröffentlichung noch zu ändern. Wie häufig Schweizer Online-Zeitschriften davon Gebrauch machen, wird in dieser Arbeit untersucht. Um dies herauszufinden, wurden Medienartikel der ausgewählten Zeitschriften in einem regelmässigen Zeitabstand heruntergeladen und in eine Datenbank gespeichert.

## 1.1 Eingrenzung

Die Auswahl der zu untersuchenden Online-Zeitschriften basiert auf deren unterschiedlichen Platzierungen im Medienqualitätsrating (MQR) (*Medienqualitätsrating MQR-24 | Verein Medienqualität Schweiz (MQS)*, 2024). Folgende Medien wurden ausgewählt: *20 Minuten*, *Watson* und *SRF News*.

# 2 Methodik

Zur Beantwortung der Forschungsfrage wurden mehrere Python-Applikationen entwickelt. Der zugehörige Quellcode ist in einem öffentlichen GitHub-Repository<sup>1</sup> zugänglich.

## 2.1 Web Scraping

Der Prozess der Datenextraktion von Webseiten wird als Web-Scraping bezeichnet und

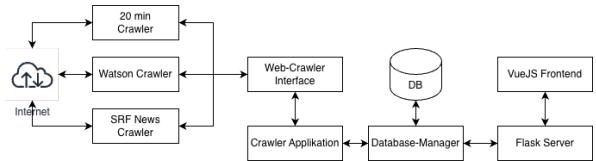


Abbildung 1: Architektur

gewinnt zunehmend an Bedeutung, da immer mehr Informationen online verfügbar sind (Penman, 2015, S. V). Die ausgewählten Medien verfügen alle über einen RSS-Feed. Um diesen auszulesen, wurde das Python Modul `feedparser` verwendet. Von dem RSS Feed können die Links zu den Artikeln herausgelesen werden. Mit dem Python Modul `requests` wurde der Inhalt der Artikel heruntergeladen. Da die Artikel aus dem Web stammen, ist das Format HTML. Für die Datenbank ist nur der Text eines Artikels von Relevanz. Dieser Text-Inhalt wurde mithilfe vom Python Modul `BeautifulSoup` extrahiert. `Beautiful Soup` ist ein beliebtes Modul, das eine Webseite parst und eine Schnittstelle zur Navigation im Inhalt bereitstellt (Penman, 2015, S. 26).

In Python wurde ein Interface für die Crawler der verschiedenen Medien definiert. Jeder Crawler implementiert dieses Interface und ist für das Herunterladen und Parsen der Artikel eines Mediums verantwortlich. Der Hauptteil der Applikation ruft die verschiedenen Crawler auf. Dies ist in Abbildung 1 dargestellt. Um die Applikation in einem regelmässigem Intervall auszuführen, wurde auf dem Raspberry Pi ein Cron Job erstellt. Beim Erstellen des Cron Jobs sind Probleme aufgetreten, welche schwierig nachzuvollziehen waren. Deshalb wurde die Ausgabe der Python Applikation in ein Log-File `swiss_news_change_tracker.log` gespeichert.

<sup>1</sup><https://github.com/dariohollbach/swiss-news-change-tracker>

## 2.2 Datenbank

Für die Speicherung der heruntergeladenen Artikel und deren Änderungen wurde eine SQLite-Datenbank verwendet. SQLite wurde gewählt, da es eine einfache und leichtgewichtige Lösung ist, welche keine zusätzliche Server-Infrastruktur benötigt. Dies ist besonders vorteilhaft, da die Applikation auf einem Raspberry-Pi ausgeführt wird. Die Datenbank besteht aus drei Tabellen: Eine Tabelle für die Zeitschriften, eine für die Artikel und eine für die Änderungen. Das zugehörige Entity-Relationship-Diagramm (ER-Diagramm) der Datenbank ist in Abbildung 4 im Anhang dargestellt. Wie in Abbildung 1 ersichtlich, erfolgt der Zugriff auf die Datenbank durch den Database-Manager mittels SQL-Abfragen.

## 2.3 Visualisierung der Änderungen

Da die reinen Daten der Artikeländerungen mittels SQL-Abfragen nur schwer zu interpretieren sind, wurde eine Web-Applikation zur Visualisierung entwickelt. Das Frontend wurde mit dem JavaScript-Framework `Vue.js` umgesetzt. Für den Datenzugriff wurde ein Flask Server entwickelt, welcher dem Frontend die in der Datenbank gespeicherten Artikel und deren Änderungen zur Verfügung stellt. Diese API greift auf dieselbe Datenbankinfrastruktur zu wie der Web-Scraper (Siehe Abbildung 1). Die Web-Applikation stellt die Artikel der verschiedenen Medien übersichtlich dar. Vorhandene Änderungen eines Artikels können durch Anklicken eines Artikels aufgeklappt werden, was eine visuelle Analyse der Modifikationen ermöglicht.

### 2.3.1 Klassifikation der Änderungen

Eine erste Analyse der Daten zeigte, dass viele Änderungen lediglich Korrekturen von Rechtschreib- oder Tippfehlern umfassen. Um solche geringfügige Anpassungen von inhaltlichen Modifikationen zu unterscheiden, wurde eine Klassifizierung der Änderungen vorgenommen. Aufgrund der überschaubaren Anzahl an Änderungen, wurde dieser Prozess manuell durchgeführt. Dazu wurde die Web-Anwendung um ein Dropdown-Menü erweitert, über welches jede Änderung als Tippfehler oder inhaltliche Anpassung klassifiziert werden kann.

## 2.4 Analyse der Daten

Für die Auswertung der klassifizierten Daten wurde ein Python-Skript mit der Bibliothek Pandas und `matplotlib` entwickelt. Die resultierende Analyse ermöglicht die Beantwortung der Forschungsfrage. Die Relevanz solcher Werkzeuge wird in der Fachliteratur betont: „Im heutigen datengesteuerten Zeitalter stehen Organisationen [...] vor der Herausforderung, aus den riesigen Datenmengen [...] aussagekräftige Informationen zu gewinnen. [...]. Genau hier kommt Pandas ins Spiel.“ (Firdose, 2024, S. 1).

## 3 Resultate

In der Zeit vom 10. November 2025 bis zum 20. November 2025 wurden insgesamt 1767 Artikel von den drei ausgewählten Medien heruntergeladen. Davon wurden 381 Artikel mindestens einmal geändert.

Die Tabelle 1 fasst alle gesammelten Daten zusammen. Obwohl die manuelle Klassifikation

eine gewisse Ungenauigkeit durch subjektive Einschätzungen mit sich bringt, bietet sie dennoch einen Überblick über die Art der vorgenommenen Änderungen.

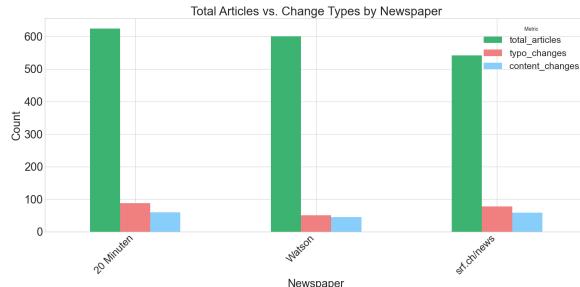


Abbildung 2: Anzahl Änderungen pro Medium

Abbildung 2 zeigt die Anzahl Änderungen pro Medium. Es ist ersichtlich, dass im Durchschnitt ungefähr jeder fünfte Artikel geändert wurde. *20 Minuten* und *SRF News* weisen dabei ähnlich viele Änderungen auf, während *Watson* etwas weniger Anpassungen vornimmt. Bei der Interpretation ist zu beachten, dass bei vielen Artikeln auch nur die Strukturierung von Texten geändert wurde, ohne dass der eigentliche Inhalt angepasst wurde. Dies wurde ebenfalls als Tippfehler klassifiziert.

Zum Resultat der Arbeit gehört auch die Web-Applikation zur Visualisierung der Artikeländerungen. Denn die Analyse der reinen Zahlen gibt keinen Einblick in die inhaltlichen Änderungen. Die Abbildung 3 zeigt einen Screenshot der Web-Applikation.

ID	Date	Content
20	28	Doch aufgefallen ist der ewige «Ricci» selten.
21	21	Ein Bilderbuch-Verteidiger.
22		<b>#Einzig in der Offensive hat der XY-Jährige noch Luft nach oben.</b>
23	23	Mit seiner Technik und Abgeklärtheit trotz des knapp genügenden Schnitts ein enorm wichtiger Wert für Yakin.
24	24	Alle Noten gibt es in der Bildstrecke.
25	25	Sven Forster (fs) arbeitet seit 2018 für 20 Minuten.
26	26	
27	27	Er ist seit Anfang 2024 als Blattmacher im Einsatz.
28	28	Dazu leitet er seit September 2024 das Sportressort.
29	29	

Abbildung 3: Screenshot der Web-Applikation

## 4 Diskussion

Die Auswertung zeigt, dass eine signifikante Anzahl von Artikeln in Schweizer Online-Zeitschriften nach der Publikation geändert wird. Die Unterscheidung zwischen Tippfehlern und inhaltlichen Änderungen gibt Aufschluss über die Art der vorgenommenen Anpassungen, was bei der Interpretation zu beachten ist. Rechtschreibfehler weisen auf journalistische Sorgfalt hin, während inhaltliche Änderungen auf neue Informationen oder Korrekturen hindeuten. Der gesammelte Datensatz bietet die Grundlage für weiterführende Untersuchungen, beispielsweise zur Analyse von inhaltlichen Änderungen bei spezifischen Artikelkategorien.

Tabelle 1: Zusammenfassung der Artikel und Änderungen pro Medium

Zeitung	Total Artikel	Total Änderungen	Tippfehler	Inhaltliche Änd.
20 Minuten	624	148	88	60
Watson	601	96	51	45
SRF News	542	137	78	59
<b>Total</b>	<b>1767</b>	<b>381</b>	<b>217</b>	<b>164</b>

## Literatur

- BAKOM, B. f. K. (2023). *Mediengattungen*. Zugriff am 2025-11-14 auf <https://www.medienmonitor-schweiz.ch/medien/mediengattungen/>
- Firdose, T. (2024). *Ultimate Pandas for Data Manipulation and Visualization: Efficiently Process and Visualize Data with Python's Most Popular Data Manipulation Library (English Edition)*. Orange Education PVT Ltd. Zugriff am 2025-12-04 auf <http://ebookcentral.proquest.com/lib/fh-ostschweiz/detail.action?docID=31466565>
- Medienqualitätsrating mqr-24 | verein medienqualitaet schweiz (mqs). (2024). Zugriff am 2025-11-15 auf <https://medienqualitaet-schweiz.ch/medienqualitaetsrating-mqr-24/>
- Penman, R. (2015). *Web Scraping with Python: Successfully Scrape Data from Any Website with the Power of Python*. Packt Publishing, Limited. Zugriff am 2025-12-04 auf <http://ebookcentral.proquest.com/lib/fh-ostschweiz/detail.action?docID=4191102>

## 5 Anhang

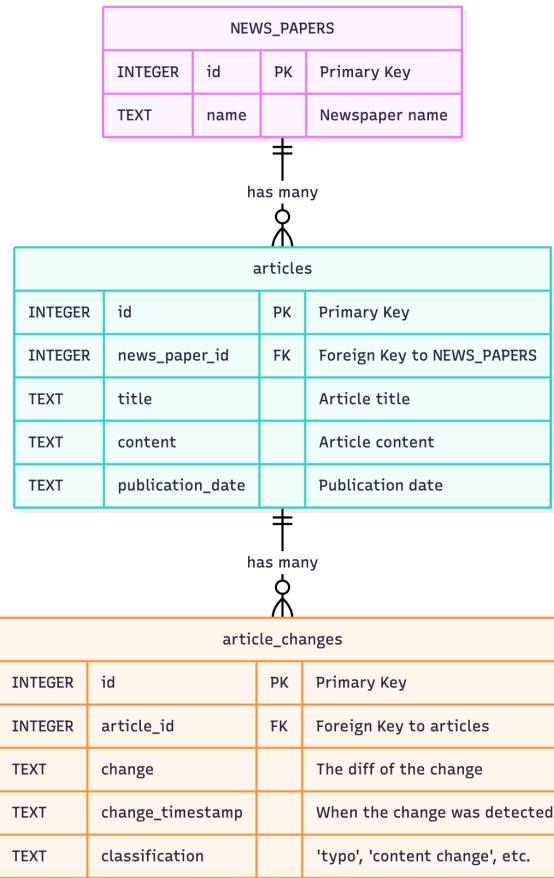


Abbildung 4: Datenbank ER Diagramm