

Website Advanced Content Analysis Tool

Diplomanden



Ali Al-Kubaisi



Simon Amberg

Einleitung: Das Ziel dieser Bachelorarbeit war die Entwicklung einer Webapplikation `Content Analyzer` zur automatisierten Analyse von Webseiteninhalten. Die Applikation soll Unternehmen dabei unterstützen, den Zustand ihrer Webseiten hinsichtlich Barrierefreiheit in Bezug auf Standards wie den den Web Content Accessibility Guidelines (WCAG) und Schreibstil effizient zu bewerten und zu verbessern.

Vorgehen / Technologien: Im Zentrum der Arbeit stand die Umsetzung eines Web-Scrapers und eines modularen Systems, das open-source Analysetools über einen Plugin-Mechanismus integriert. Der entwickelte Web-Scraper nutzt moderne Technologien, um Inhalte von Webseiten automatisiert zu erfassen und in einem Repository zu speichern. Anhand dieses Repositories wird dann die erfasste Webseitenstruktur mithilfe der eingebundenen Analysetools untersucht. Die initiale Implementierung umfasste insbesondere die Analyse von Barrierefreiheit mithilfe des Tools Axe-Core sowie ergänzende Auswertungen hinsichtlich der WCAG-Richtlinien. Zusätzlich wurde ein Schreibstil-Checker entwickelt, der regelbasierte und KI-gestützte Ansätze für grammatikalische Korrekturen und Textanalysen kombiniert. Darüber hinaus bietet die Applikation Möglichkeiten zur Verwaltung und Annotation von Inhalten und Gruppierung, sowie Verwaltung von erkannten Problemen.

Ergebnis: Besondere Herausforderungen beim Web-Scraping ergaben sich aus der Diversität an Technologien und Architekturen, die für moderne Webseiten eingesetzt werden, sowie der baumartigen Struktur von Webseiten. Diese wurden durch detaillierte Analyse der Problemdomäne, sorgfältige Technologieauswahl, sowie einer iterativen Entwicklung erfolgreich bewältigt.

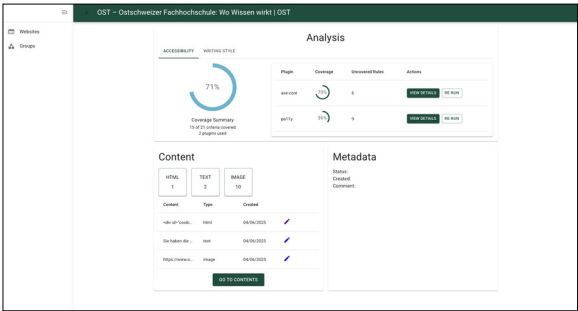
Das Ergebnis der Arbeit ist ein flexibles und wartbares System, das es ermöglicht, Webseiten auf grundlegende Richtlinie zu überprüfen. Über eine anschauliche Darstellung der Resultate können erkannten Problemen verwaltet und behoben werden. In Zukunft soll das System als Grundlage dienen, auf der weitere Analysetools eingebunden werden können, um aufkommende Standards abzudecken und neue Fähigkeiten einzubinden.

Referent
Prof. Dr. Markus Stölze

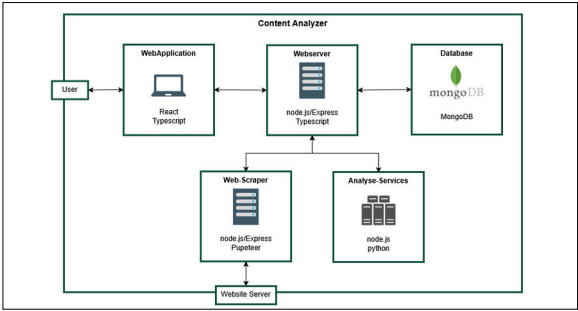
Korreferent
Reto Senn, bitforge AG,
Zürich, ZH

Themengebiet
Software Engineering

Analysis Dashboard: Plugin-Übersicht, Schreibstil, Content-Management, Coverage-Statistiken Eigene Darstellung



Modulare Architektur: Frontend, Webserver, Web-Scraper, Analyse-Services, MongoDB Eigene Darstellung



Accessibility Resultat mit WCAG-Analyse Eigene Darstellung

