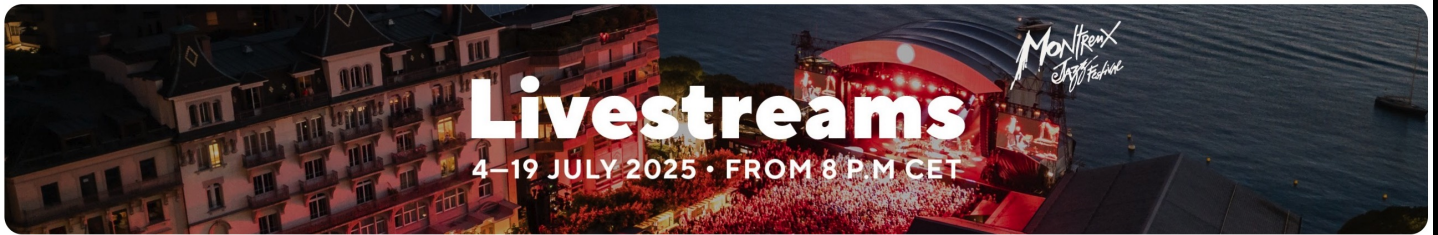


Bachelor Thesis: Data Crawling of Montreux Jazz Festival



University of
Zurich^{UZH}



Introduction

To increase public accessibility to the Montreux Jazz Festival's rich archive, we are developing a database that captures the festival's historical performances. This database will serve as the foundation for an interactive visualization tool that enables users to explore the Montreux Jazz Festival dataset.

A key component of the project involves collecting relevant performance videos from YouTube. To achieve this, we will design and implement a focused web crawler that searches for and extracts video links relevant to Montreux Jazz Festival performances. The system will use existing metadata from the archive, such as performance date, title, and artist names, to construct targeted search queries and filter results for accuracy. To ensure the quality and relevance of the collected content, we will develop a ranking algorithm inspired by YouTube's search and recommendation mechanisms. This algorithm will analyze YouTube video data (e.g., title, description, tags, upload date) and compare it with the festival's existing metadata to identify the most accurate matches. Then, we will apply an audio fingerprinting and recognition algorithm to compare the audio content of the YouTube video candidates with the archive's recording. This will allow us to identify the correct link.

All validated links will be stored in the database and seamlessly integrated with the visualization tool. Users will not only be able to navigate the history of the festival visually, but also directly view and play associated video performances through embedded YouTube links, creating a more immersive and engaging experience.

- **Ranking Algorithm Development:** Design and implement a robust ranking algorithm that leverages available metadata from the archive, such as performance date, artist names, and titles, to identify and prioritize the most relevant YouTube links.
- **Web Crawling:** Build and deploy a targeted web crawler to search YouTube using the constructed queries, and extract video links that are likely relevant to Montreux Jazz Festival performances.
- **Audio Fingerprinting and Recognition:** Apply audio fingerprinting and recognition algorithm to compare the YouTube video candidates with archive recordings, ensuring the selected links correspond to the correct performances.
- **Database and Visualization Integration:** Store the validated links in the existing database and integrate them into a visualization tool, allowing users to view and interact with the videos directly through the interface.

Project Type

Bachelor Thesis

Requirements

Experience in programming with Python.

Supervision

Prof. Dr. Renato Pajarola

Xiao Tan (Assistant)

Workload

- 20% Theory
- 80% Implementation

Contact

Email: xtan@ifi.uzh.ch



Montreux Jazz Festival

@MontreuxJazzFest · 193K subscribers · 815 videos

58th Montreux Jazz Festival | July 5 - 20, 2024 ...more

montreuxjazzfestival.com/en and 5 more links

Subscribe

Montreux Jazz Festival YouTube channel

Assignment

The following tasks will be part of the project:



Professor Dr. Renato Pajarola
Visualization and Multimedia Lab
Department of Informatics
University of Zurich