

The WASABI dataset Cultural,
lyrics and audio analysis
metadata about 2 million
popular commercially
released songs



Series of seminar



→ Michel Buffa
(Professor at Université Côte d'Azur)

→ March 8th, 6pm CET

→ Save the date and join at
the link

→ Abstract

Since 2017, a two-million song database consisting of metadata collected from multiple open data sources and automatically extracted information has been constructed in the context of the WASABI french research project. The goal is to build a knowledge graph linking collected metadata (artists, discography, producers, dates, etc.) with metadata generated by the analysis of both the songs' lyrics (topics, places, emotions, structure, etc.) and audio signal (chords, sound, etc.). It relies on natural language processing and machine learning methods for extraction, and semantic Web frameworks for integration. The dataset describes more than 2 millions commercial songs, 200K albums and 77K artists. It can be exploited by music search engines, music professionals or scientists willing to analyze popular music published since 1950. It is available under an open license² in multiple formats and is accompanied by online applications and open source software including an interactive navigator, a REST API and a SPARQL endpoint.

→ Speaker biography

Since 2017 Michel has been conducting his research within the SPARKS -Scalable and Pervasive softwARe and Knowledge Systems- team of the I3S laboratory. He has been the national coordinator of the WASABI (2017-2020) Web Audio Semantic Aggregated in the Browser for Indexation) research project which includes partners such as Deezer, IRCAM, Radio France, Parisson... He intends to improve music databases on the web and make music even easier to find, but also to play by offering instruments and audio effects usable in his browser! A true music lover, Michel is has been working on this project, which consists in building a metadata database on two million popular music songs (pop, rock, jazz, reggae...), by aggregating and structuring cultural data from the Web of data, audio analysis and natural language lyrics analysis. Interactive web applications, based on the emerging WebAudio W3C standard, exploit this database and are aimed at composers, music schools, sound engineering schools, musicologists, music broadcasters and journalists. His work has won several awards in international scientific conferences and has resulted in the first real-time simulations of tube guitar amplifiers running in a web browser, which are now commercially available.

In parallel, Michel teaches computer engineering at the University of Côte d'Azur. Since 2015, he has set up, in collaboration with the Université Côte d'Azur, the W3C (World Wide Web Consortium), MIT and Harvard MIT and Harvard, several MOOCs (Massive Open Online Course) on HTML5 and web technologies, followed by more than 700,000 students.

Music language models:
A need for symbolic
representations?



Series of seminar



→ Albert Meroño Peñuela
(Lecturer at King's College London)

→ March 8th, 6pm CET

→ Save the date and join at
the link

→ Abstract

In recent times we have seen various breakthroughs in natural language processing, particularly on various architectures learning language models that achieve impressive performances at various tasks. In this talk I present some work that leverages these architectures to learn music language models that can generate believable and novel musical compositions.

An important challenge in order to achieve this is the question of how to symbolically represent music, and what the role of those symbolic representation is in an increasingly machine-learning dominated world. I argue for the use of symbolic music knowledge graphs, and address some of their challenges such as music knowledge graph completion with the novel midi2vec embeddings technique.

→ Speaker biography

Dr. Albert Meroño Peñuela is a Lecturer (Assistant Professor) in Computer Science and Knowledge Engineering at the Department of Informatics of King's College London (United Kingdom). He obtained his PhD at the Vrije Universiteit Amsterdam in 2016, under the supervision of Frank van Harmelen, Stefan Schlobach, and Andrea Scharnhorst; and has done research at the Netherlands Academy of Arts and Sciences and the Autonomous University of Barcelona. His research focuses on Multimodal Knowledge Graphs, Web querying, and Cultural AI. Albert has participated in large Knowledge Graph infrastructure projects in Europe, such as CLARIAH, DARIAH and Polifonia H2020; and has published research in ISWC, ESWC, the Semantic Web Journal and the Journal of Web Semantics.