
Collecting and archiving the audiovisual web: Food, cooking and health in a selected corpus of websites and connected YouTube channels in France.

Christian Bonah^{*1}, Solène Lellinger^{*†}, and Caroline Sala^{*‡}

¹Département 'Histoire des Sciences de la Vie et de la Santé (SAGE-DHVS) – université de Strasbourg, CNRS : UMR7363 – Faculté de Médecine Université de Strasbourg 4 rue Kirschleger 67085 Strasbourg Cedex, France

Abstract

The communication will present a joint venture between the research project BodyCapital at the University of Strasbourg and the BNF for a two-step process archiving the audiovisual web related to the subject of food history. Our research question asks how have audiovisuals contributed to shape our eating habits and their connection to concerns about our individual health and healthy eating?

BodyCapital is a research project investigating how audiovisuals have changed body politics and healthy self-perceptions and practices of individuals in market-based societies in twentieth century Europe. Considering audiovisuals as a distinct, interactive performative power of mass media societies we have engaged after film and television analysis a systematic audiovisual web analysis on the subject of food studies and nutrition.

In collaboration with BNF we have organized in a first step a web crawl with Hetrix robot of 158 identified seed URL's compiled on the basis of BNF science and technology lists and URL's identified by the research group. The crawl conducted between 9/3/2021 and 12/3/2021 harvested 1 067 159 URL's at a level n+2. The content analysis with SolRway-back identified 1718 videos in our corpus. After content mapping and identification of links to YouTube videos and complementary research on YouTube channels concerning food studies and their history a second step consisted in a specific collection of 36 YouTube channels collected in June 2022. The crawl harvested 24 427 videos (2,4 TO) that will be analyzed. This second collection includes metadata based on the Youtube API in JSON format for every channel. The YouTube crawl produces MP4 files connected with metadata stored by BNF (Forge) in order to reconstitute YouTube pages for consultation by channel in a guided presentation ("parcours guidé") hosted by BNF.

The presentation will present the result of the two-step crawl in quantitative and qualitative terms and suggest analytical approaches for analysis of the collected corpus. Subsequently, we have elaborated in collaboration with BNF a detailed description of 36 YouTube channels harvested including NGOs, national organizations, individual blogs, food transforming

*Speaker

†Corresponding author: solene.lellinger@u-paris.fr

‡Corresponding author: csala@unistra.fr

professionals and industrial players. Descriptions are based on analysis of channels and individual videos in YouTube channels. Descriptions include data on the channel (name, author, content, goals), interest of the channel for the theme of the research project on food history and finally links to websites collected in step 1 as well as blogs and Instagram links.

Furthermore, the thematic analysis will investigate a comparison of the results of the two-step crawl with audiovisual formats and themes of a television program corpus that has been collected systematically and analyzed in comparison. The trans-media analysis will provide insights into transformations between television and the web concerning specific thematic sub-corpus. A precise example will include cooking recipes audiovisual presentation from TV to the web and their connection to eating as an individual investment in health conceived as a capital.

Keywords: archives du web, images, alimentation, cuisine, Youtube