

Report P2.2: CaptureBias Prototype and Interface



Capture**Bias**

1 July 2019

Panagiotis Mavridis, Markus de Jong, Xander Wilcke, Alessandro Bozzon, Alec Badenoch, Antoaneta Dimitrova, Honorata Mazepus, Jesse de Vos, Johan Oomen, Lora Aroyo, Tobias Kuhn

Table of Contents

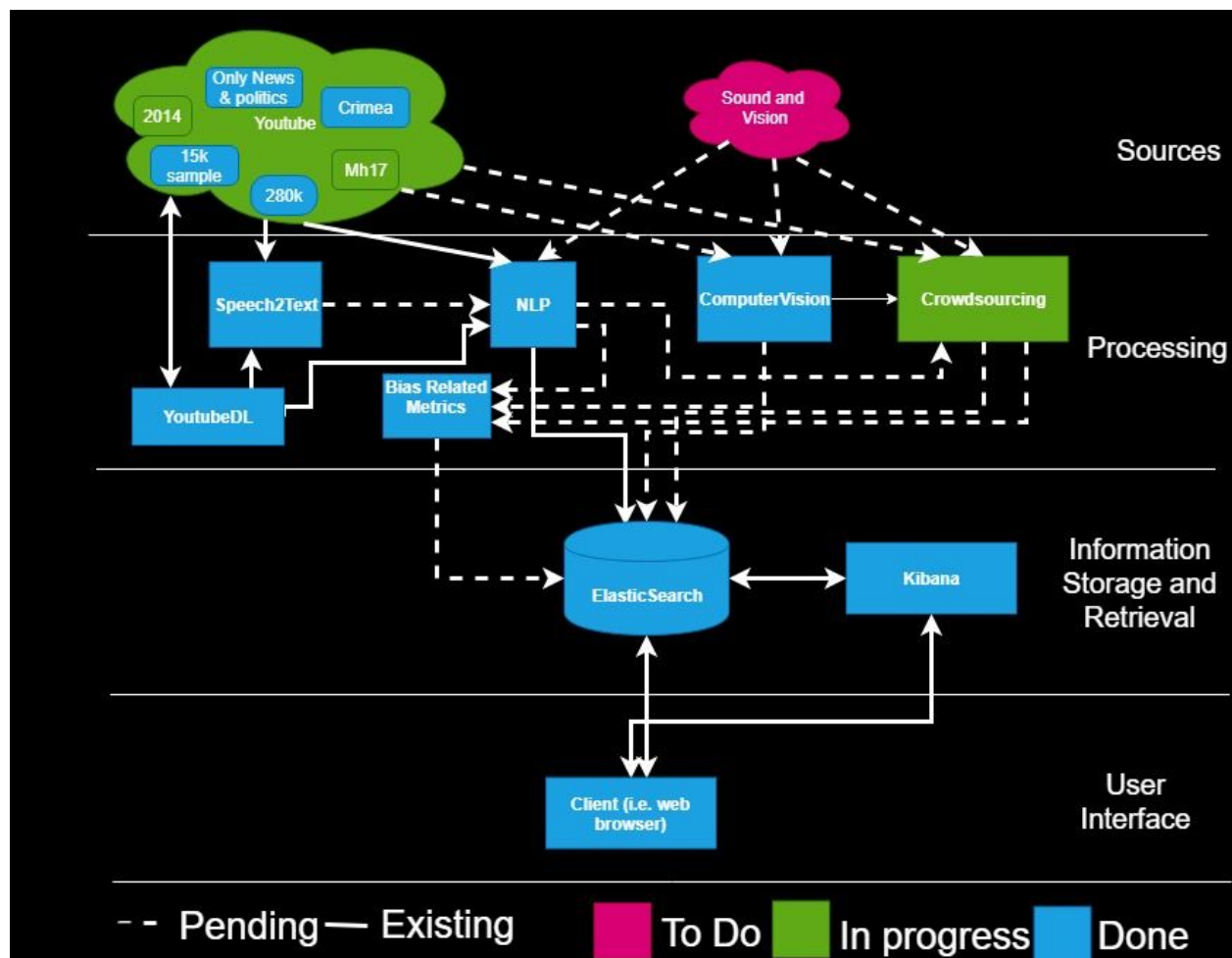
Introduction	3
Pipeline diagrams	3
Conceptual architecture of the overencompassing system	3
NLP pipeline	4
Computer vision pipeline	4
Interface examples	6
Workspace Selection	6
Dashboard example with several bias metrics	7

Introduction

This document illustrates the status at year 1 of the architecture of several data pipelines and proposed user interfaces that show insights on the resulting data for the CaptureBias project. Specifically, we present the conceptual architectural details on the overemcompassing design, as well as the natural language processing (NLP) and Computer Vision. Finally, we present the interface prototype for the interactive presentation of several bias metrics towards our experts via several 'dashboards', and eventually intended for the dissemination towards general public. One can select any of the different workspaces and interact with multiple dashboards, by, for example, selecting different time periods.

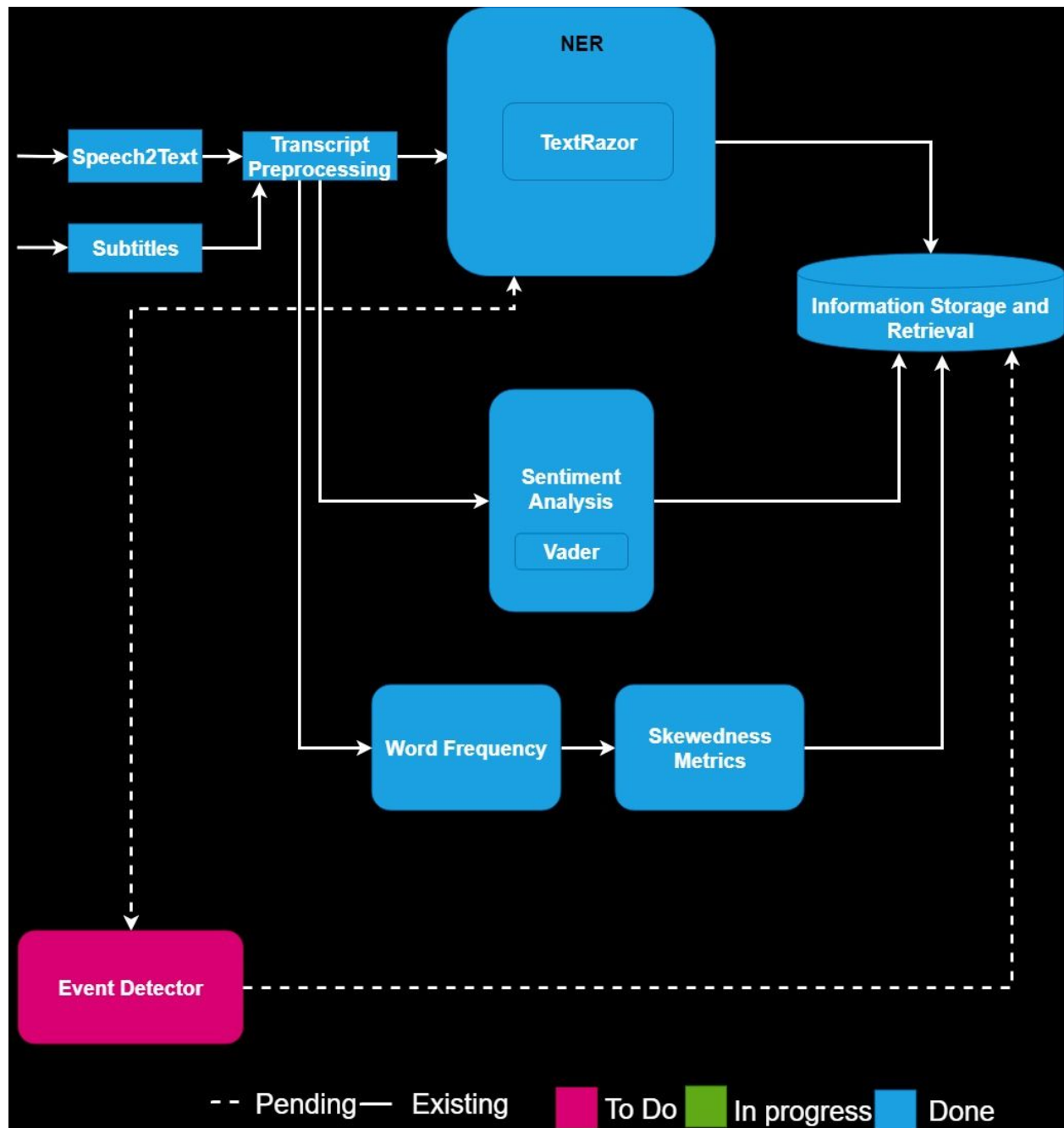
Pipeline diagrams

Conceptual architecture of the overencompassing system



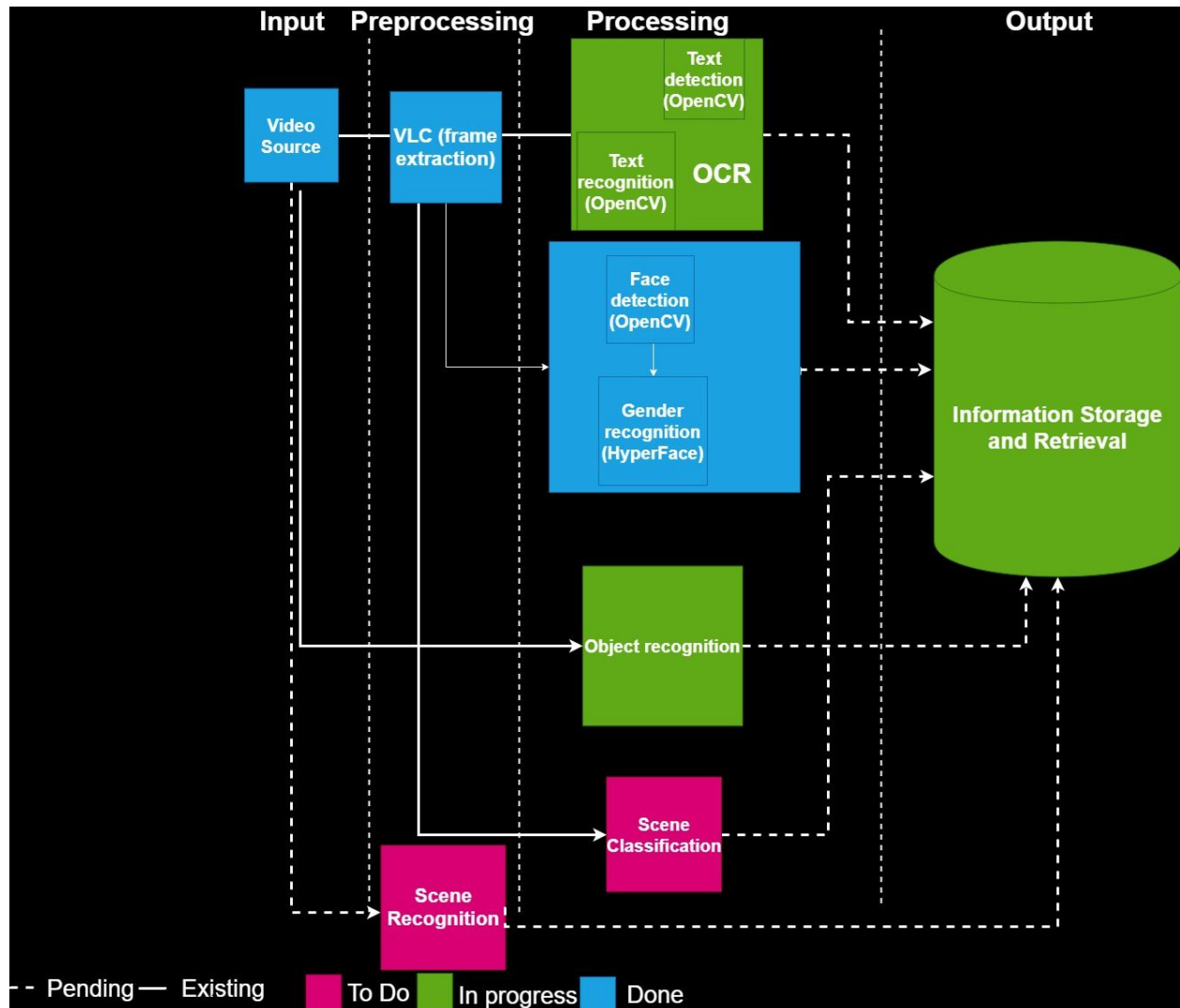
Conceptual architecture of the overencompassing system showing the links between each of the layers: sources, processing, information storage and retrieval and finally the user interface.

NLP pipeline



Conceptual architecture of the NLP pipeline with speech to text analysis (speech2text) to generate transcriptions, NER (Named-entity recognition) to capture the meaning of the words and sentiment analysis to detect positive, neutral or negative sentiments in a text. Word frequency and skewedness metrics are tools we want to use to analyse differences between messages. The event detector is a work in progress which will be used to link messages to specific events (e.g. X declares Y).

Computer vision pipeline



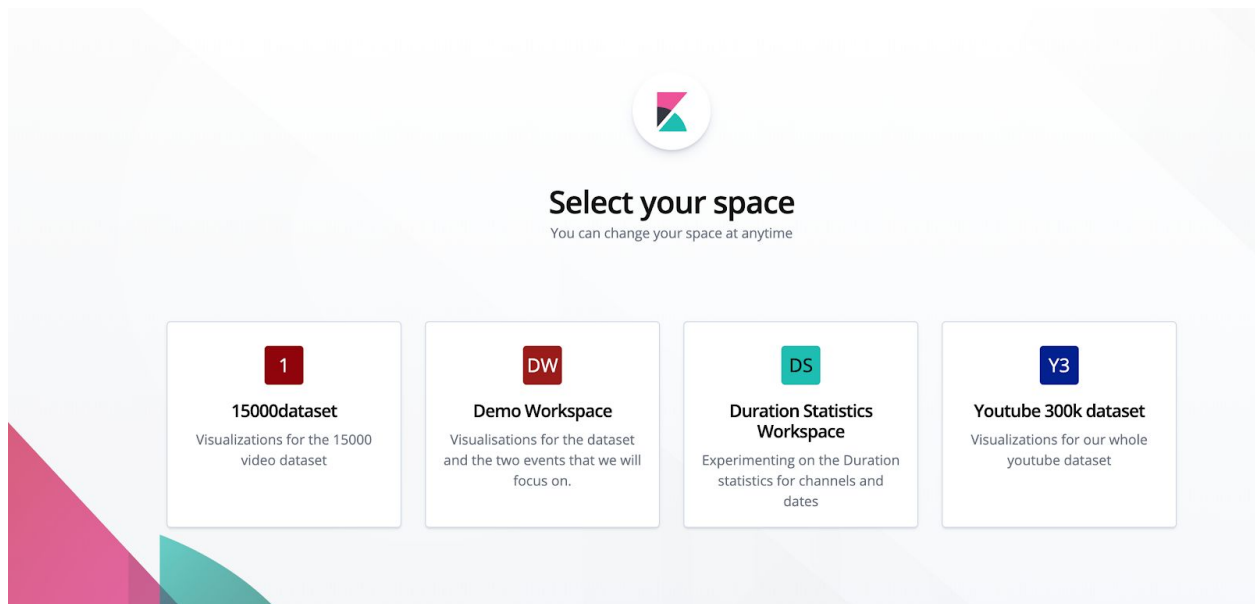
Conceptual architecture of the computer vision pipeline. From the video, frames are extracted for several deep learning analyses: optical character recognition (OCR) to detect text in video, face and gender identification to analyse facial information and object recognition and scene recognition and classification to quantify what is going on in the video (in development).

Interface examples

The following chapter walks through the user interfaces that run on the *Kibana* platform. First, a *workspace* connected to a dataset or dataset sub-sets is selected. This presents the user with several *dashboards* aimed at different sets of prepared graphs based on several bias metrics. Of course, the user will be able to create his or her own graph as well.

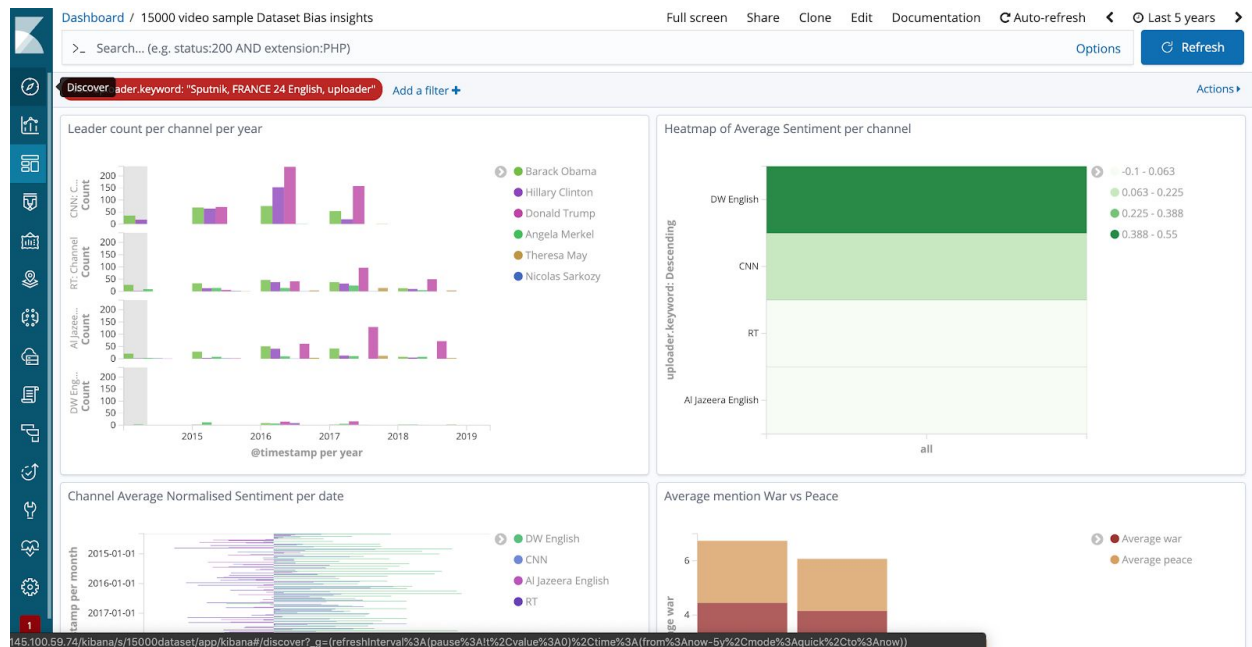
Workspace Selection

The user is presented with a selection of different workspaces. The next dashboards will be based on the 15000 item dataset workspace.

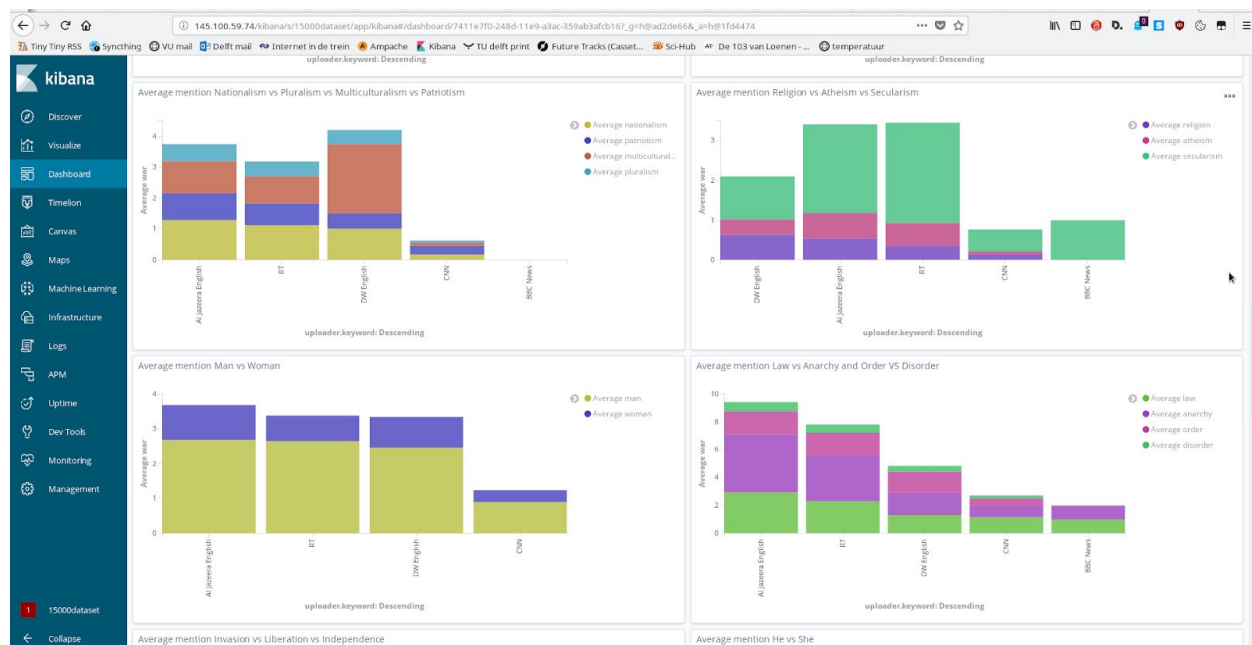


Workspace selection from several available data sets

Dashboard example with several bias metrics



Dashboards illustrating several metrics on word count and sentiment analysis



Dashboards illustrating term usage comparisons across channels