**1 Introduction**

* short description of code
* this corpus will help with deep learning
* list of features (code tool to add to corpus, collected a sizeable amount of data, data is relatively unique (very few corpus built on youtube data, and no corpus (to our knowledge) built on youtube video metadata)

## 2 Video metadata

* few datasets built on youtube related data
* no datasets built on romanian youtube channels
* no datasets built on video metadata

## 3 Data Collection

* list of channels chosen with rationale of choosing
* Antena 1 for TV entertainment type stuff
* Stirile Pro TV for formal news
* Radio Zu + Kiss fm for radio type content
* Selly + Daniel Pirvan + Dorian Popa for vlogger type content

## 4 Code

## -vedem

## 5 References

* <https://aclanthology.org/D18-1394/>
* <https://aclanthology.org/W18-3817/>
* <https://aclanthology.org/2020.peoples-1.2.pdf>
* <https://aclanthology.org/2020.sltu-1.28/>
* https://blog.devgenius.io/creating-a-youtube-summariser-bd9f41109d40