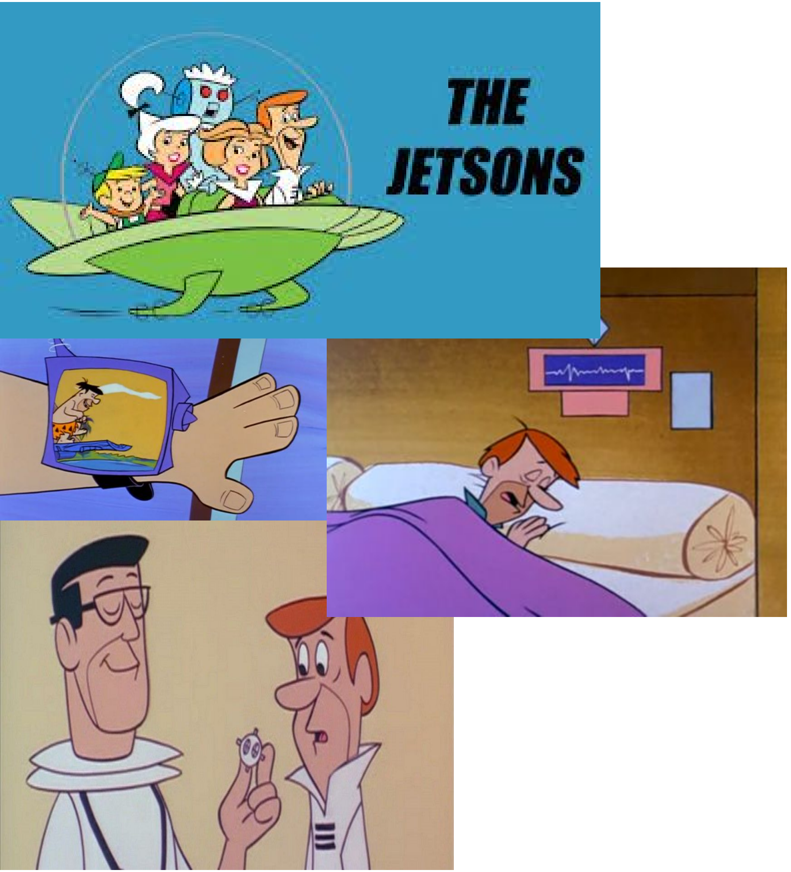
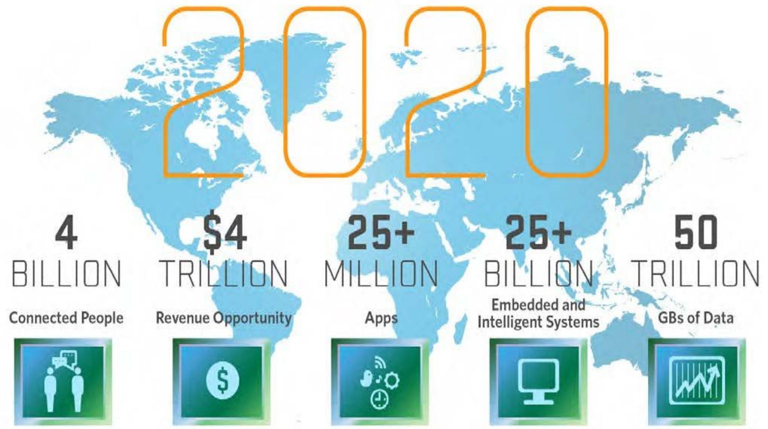
IoT news articles

Project III : WebScraping

Introduction

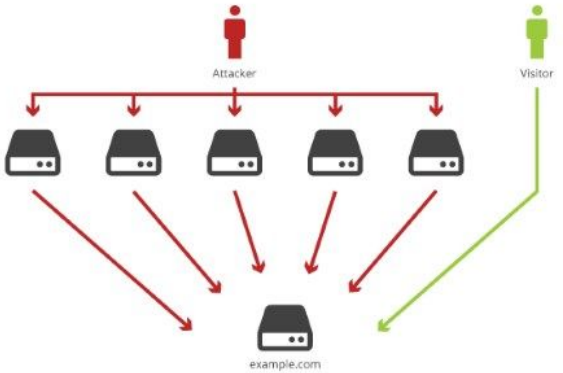




Source: Mario Morales, IDC

Why webscraping IoT news?

October 21 cyber attack came from IoT devices



Sending queries 50 times more queries than the average

The data

Scrapping news headlines from ***Google news***

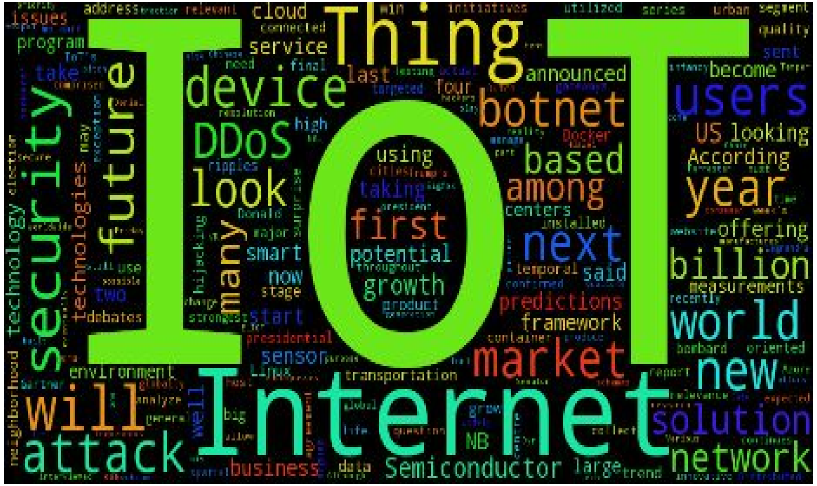
In this work, we focus on :

- Headlines :
  - After 10/21 (from 10/21 to 10/31)
  - Last week(11/1 to 11/5)
  - This week (11/6 to 11/11)

- Compare before and after cyber attack

What happened?

After 10/21 (from 10/21 to 10/31)



→

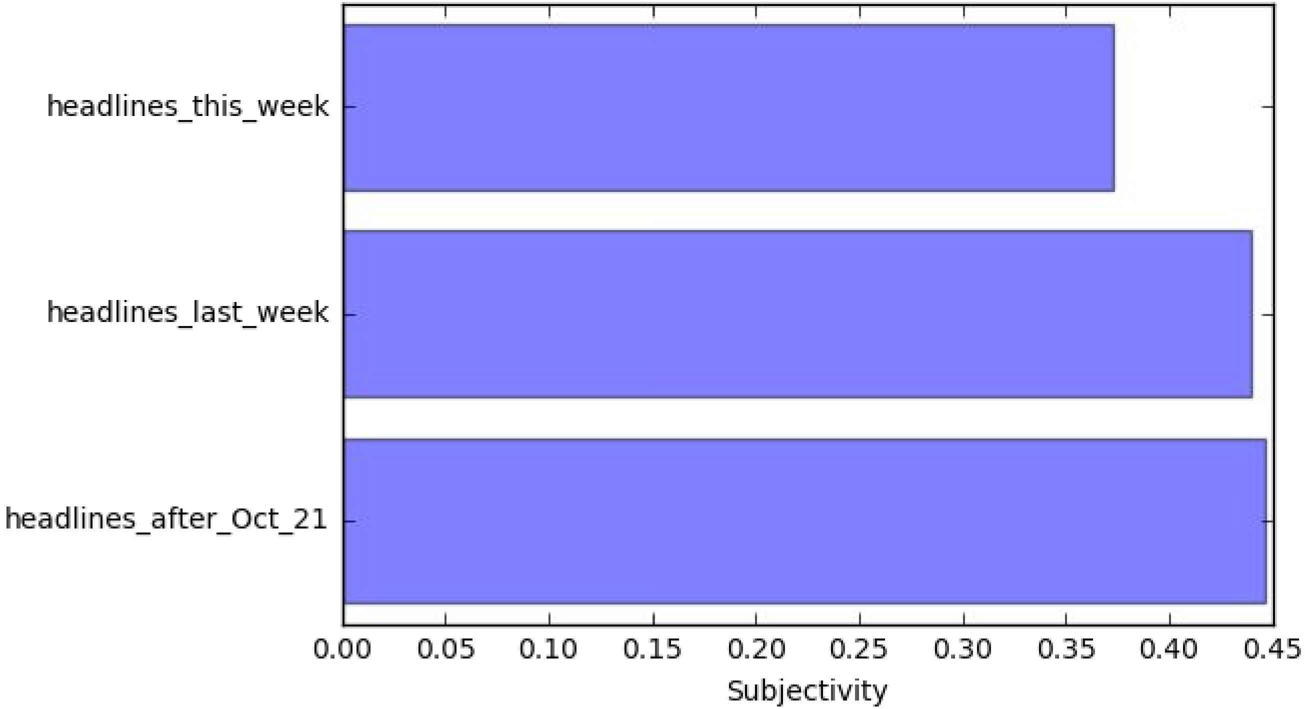
Last week 11/06-11/11

IoT
Bad
Future
Internet
afterthought
Industry's
new
Things
Cities
Overload
Ways
Security
Webcams
Smart
big
Hands
Qualcomm
spread
Version
hack
Ecosystem
change
jobs
bolster
NXP
Review
means
First
create
Ireland's
Recalls
Welcome
success
LoRa
disadvantage
lightbulbs
Problem
Gartner
May
Predictions
Solve
citizen
Cubic
Nationwide
Forrester
Week
DDoS
urban
Auto
India
Bosch
Rs
Operators
Trump
Enable
Adopting
presidency
Semtech
Attack

😟 → 😱

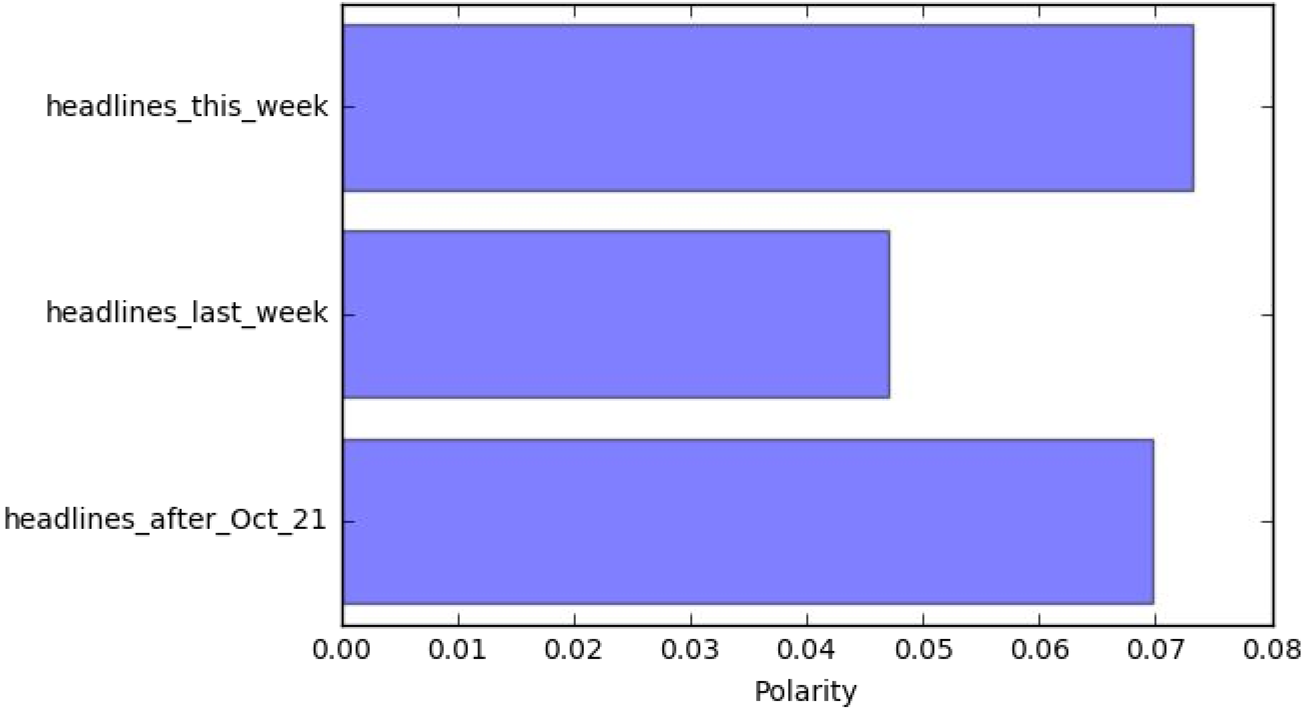
Subjectivity of news headlines?

Measure of subjectivity between 0 (not influenced) and 1 (influenced)

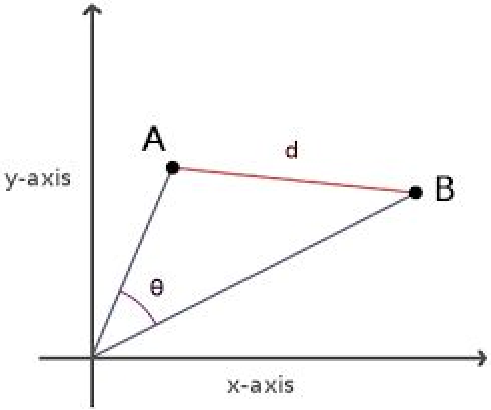


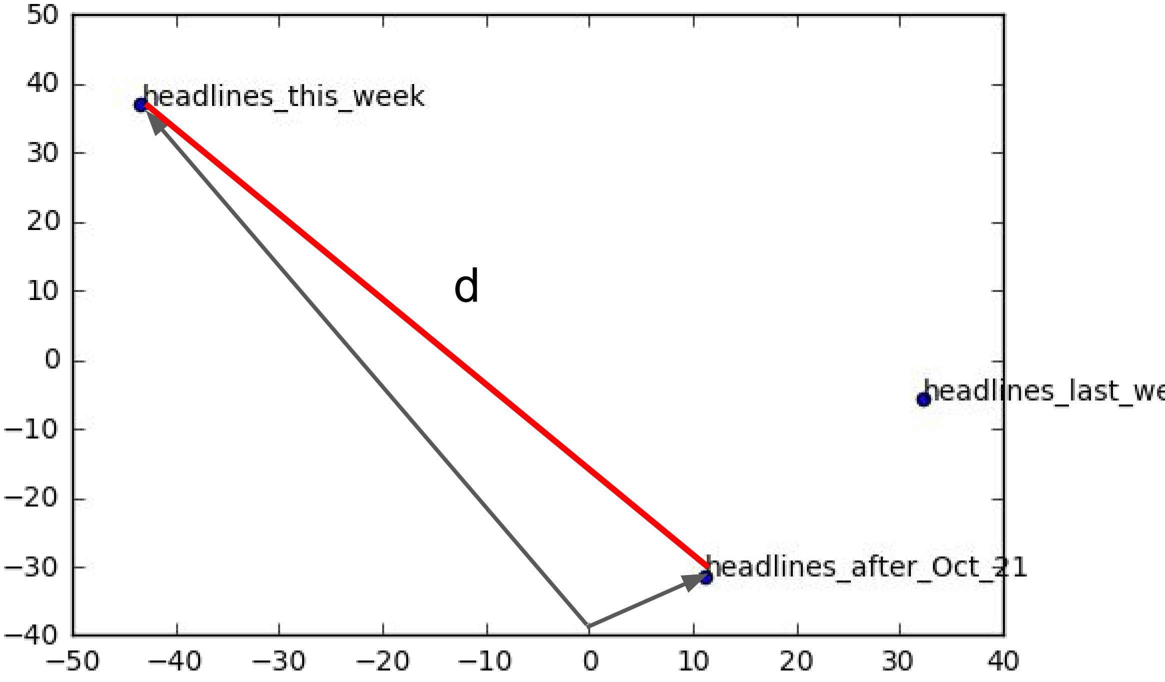
Polarity of news headlines?

Measure of polarity between -1 (negative) , 1 (positive), 0 (neutral)



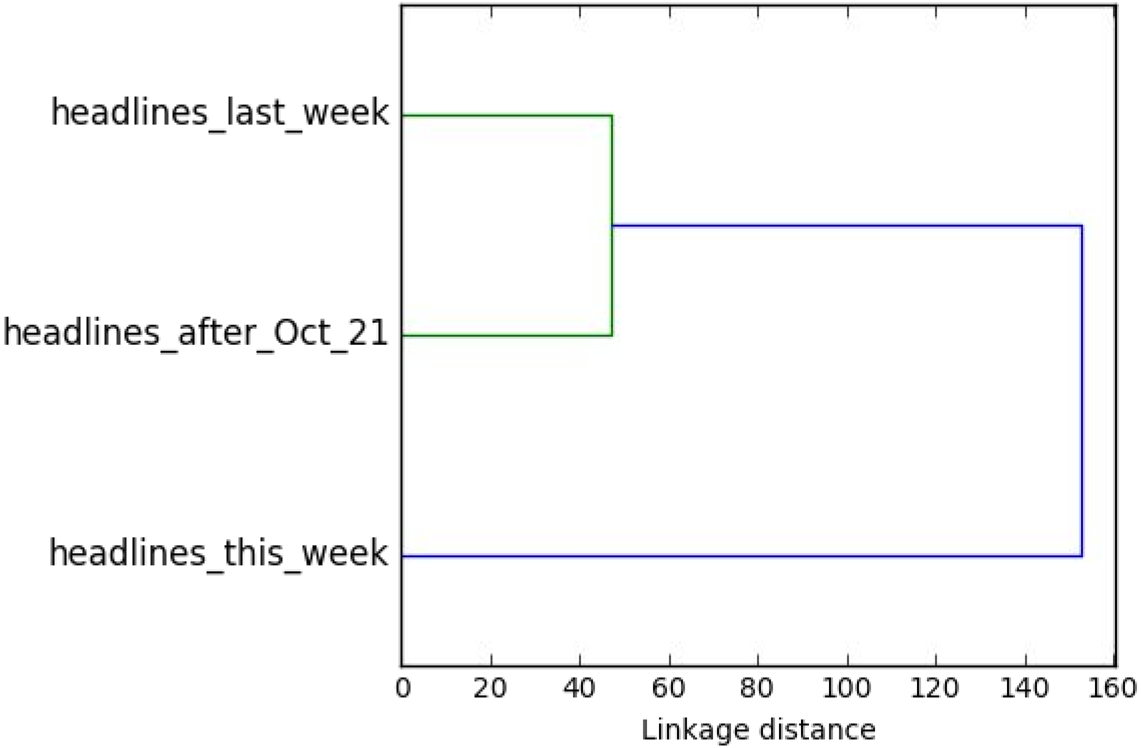
Similarity between headlines





Clustering of headlines

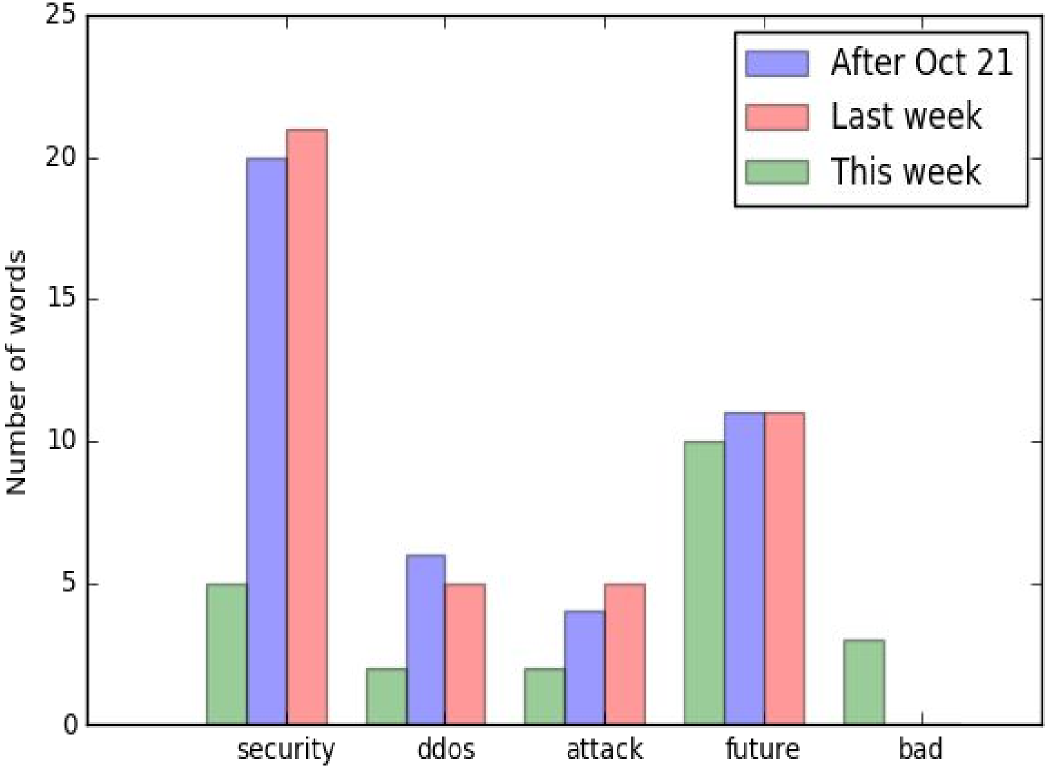
Clustering headlines into groups of similar texts(Ward’s method)



Clustering of topics

Keys words associated with each topic :

Topic 1	Topic 2
iot security market devices users new future attacks world announced 2017 cloud ddos	iot future new bad afterthought smart webcams recalls augmented ddos predictions 2017 information



Conclusion

- Analysis of subjectivity, polarity, headlines similarities and clustering
- Extraction of keywords
- Headline and keyword evolution from a week to another
- IoT is a “neutral” subject for now

Technical Challenges :

- Webscrapping Google News and the risk of being blacklisted
- Working only with text

Perspectives

- Machine learning to detect trending IoT topics