Clickbait Detection
10 Frames that will BLOW YOUR MIND

Thesis Meeting
19.05.2015
Sebastian Köpsel

Overview

- General survey
- Related work research
- Started working with Twitter API
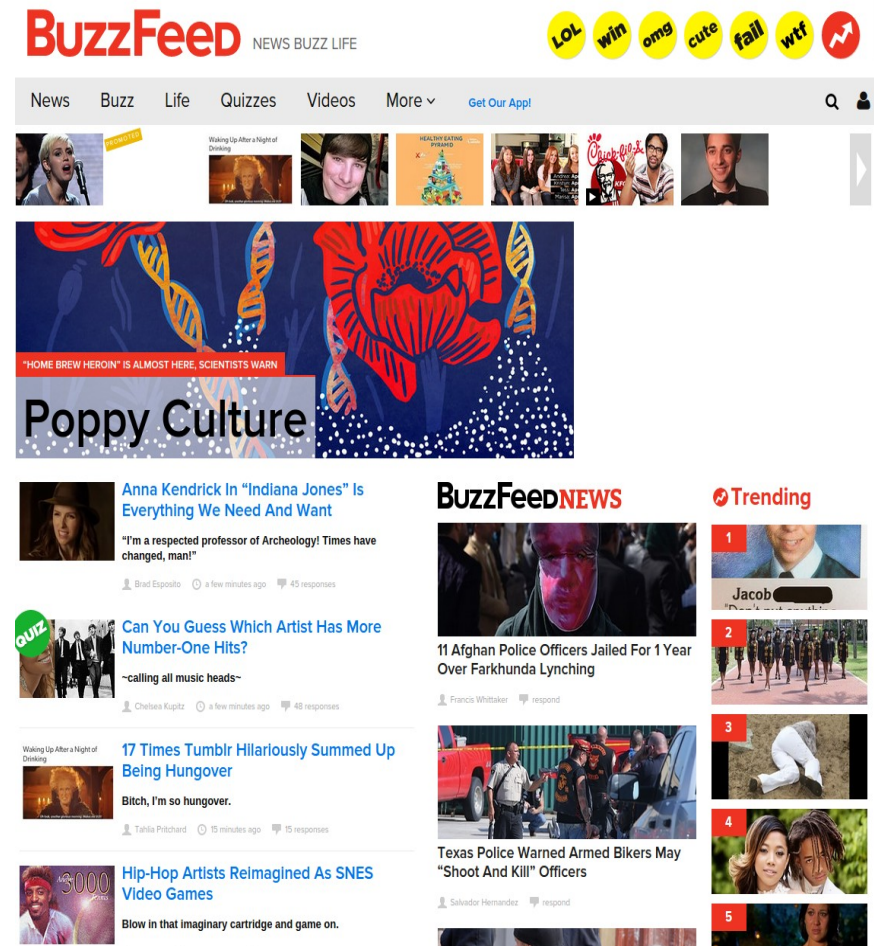- Started Machine Learning learning

General Survey

- Clickbait
  - Clickbait is a pejorative term describing web content that is aimed at generating online advertising revenue, especially at the expense of quality or accuracy, relying on sensationalist headlines to attract click-throughs and to encourage forwarding of the material over online social networks. Clickbait headlines typically aim to exploit the "curiosity gap", providing just enough information to make the reader curious, but not enough to satisfy their curiosity without clicking through to the linked content.

    src: http://en.wikipedia.org/wiki/Clickbait

General Survey

- List of notorious Clickbait pages
  - upworthy.com
  - BuzzFeed
  - heftig.co (deutsch)
  - viralnova
  - likemag
- List of partly clickbaity pages
  - Fox news
  - Huffington post
  - New York Times
- Non Clickbait Corpus
  - Reuters

General Survey

- List of relevant social media websites
  - Facebook
  - Twitter
  - LinkedIn
  - Reddit
  - Pinterest
  - Tumblr
  - Google+

General Survey

- Random Stuff I found
  - Anti Clickbait Twitter Profiles
    - @SavedYouAClick
    - @EmbeddedVideo
    - @HuffPoSpoilers
  - Downworthy
  - Clickbait Headline Generator

Related work research

- I found nothing directly concerning Clickbait detection

- I found a few related fields of research

Related work research

- Text + Genre analysis/ classification

- Buzzword detection

- Spam detection

- Plagiarism detection

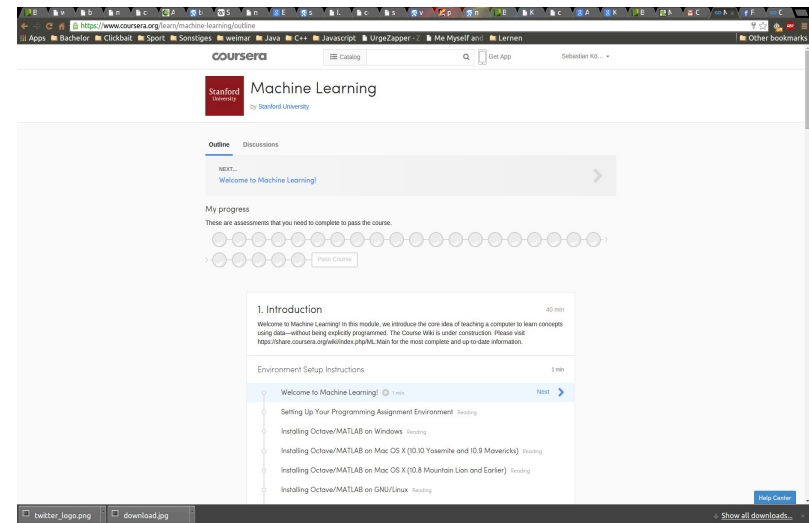- (Advertisement)

Twitter API

- Twitter API application
  - With Java
  - Using twitter4j
  - Can read Stream, filters for URLs   s
- Idea:
  - We read Twitter global stream
  - Search for shared links
  - Store often shared ones to build training corpora

•Machine Learning

- Started reading introduction papers

- Will do coursera course this week

Conclusion

- First ideas to detect Clickbait
  - Common phrases
    - Incomplete sentences with …
    - Quizes-question marks
  - Source pages
  - Existence of social media button
  - Use communities maybe from bloggers like Alison Gianotto(downworthy) to build corpus
  - Plagiarism/ paraphrasing