Clickbait Detection
This guy worked all week. You won't believe what he did! WHOA...

Thesis Meeting
26.05.2015
Sebastian Köpsel

Overview

- More Related Work Research

- Build Program to view and categorize Twitter users timeline

- Build corpora of appr. 1k tweets

- Random Stuff

Related Work

- Headline Emotion Classification

- Serendipitous-Entity-Search

- Recognition of Affect

- Emotion Detection

Related Work

- Related Corpora/ Frameworks
    - Word sentiment list
    - WordNet
    - Stop Word Collection
    - SentiWordNet

Program

- Reads a given users timeline
  - Esp. news aggregators
- Console input to classify tweet
  - Clickbait – serious – undefined
- Stores tweets JSON& linked web page
- Demo?

26/05/15

Corpora

- Visited twitter accounts
  - FortuneMagazine
  - bbc
  - TheCut
  - Nytimes
  - HuffingtonPost
  - ViralNova
  - Reuters
  - Slate
  - Washingtonpost
  - Voxdotcom
  - BuzzFeed

Random Stuff

- RSS feeds?
  - They have them
- Twitter streaming API representative?
  - Pretty much[1]
- Not yet searched for only pics + caption :(
  - None in the corpus too

Next Steps

- Bigger Corpus

- Test Corpus

  - Can easily modify the prog so we get unclassified data

- Analyze existing corpora

  - Word frequencies etc

- Keep reading on Machine Learning

Sources

1) Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose

 – Morstatter, Pfeffer, Liu, Carley