

Sara Bauer Ma, MSc

WORKSHOP

Computational Linguistics or "*How your last tweet will be used against you*"

September 28, 2018

Contents

- Introduction
- Social Media
- Background
- Practice and Examples
- Summary

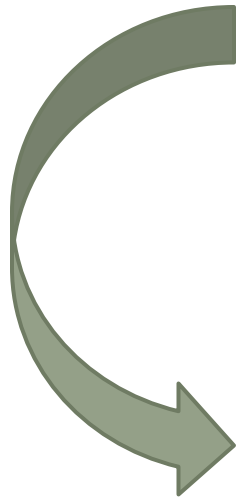
About Me

- MA in English and American Studies
- MSc in Computer Science

About Me

- MA in English and American Studies

MSc in Computer Science



In which **language** shall we proceed?



About Me

- Enough about me....



About YOU

- Which **school** do you go to?
 - Why did you choose THIS **workshop**?
 - What are YOUR **interests**, ideas etc. regarding this topic?



Social Media

- crucial part of digital communications strategies
- **3.196 billion** global social media users in 2018
- globally **1 million new active mobile social users** are added **every day**

Social Media



Commerical Influences – Status '18



Facebook influences 52 % of consumers' online and offline purchases



Twitter: 78% of people who complain to a brand via Twitter expect a response within an hour



Instagram: Top brands on Instagram are seeing a per-follower engagement rate of 4.21% which is 58 times higher than on Facebook and 120 times higher than on Twitter

Social Media - Privacy



Social Media - Privacy

- **Data Scraping.**
 - It involves tracking people's activities online and **harvesting personal data** and conversations from social media, job websites and online forums
 - Usually, research companies are the harvesters, and sell the compiled data to other companies.
 - These, in turn, use these details to design targeted ad campaigns for their products.
- **Facebook apps leaking personal data.**
 - The popular social media site has been plagued by privacy issues over the years.
 - Its highest-profile problem was in October 2010, when Facebook admitted that its top 10 most popular applications including FarmVille and Texas Hold'em shared user data, including **names and addresses**
- **Online social tracking.**
 - Other social websites allow companies to place within ads **cookies** and beacons – pieces of software that can track you and **gather information** about what you are doing on a page
 - Note: these tracking tools are widely used online but mostly on websites dedicated to **kids** and **teens**, which raises a huge children's online privacy concern.

Social Media - Politics

Donald Trump one year on: How the Twitter President changed social media and the country's top office

Mr Trump has weaponised Twitter, using it not just to reach the masses but to control the news agenda through bluster and distraction

Andrew Buncombe New York | @AndrewBuncombe | Wednesday 17 January 2018 21:43 | 6 comments

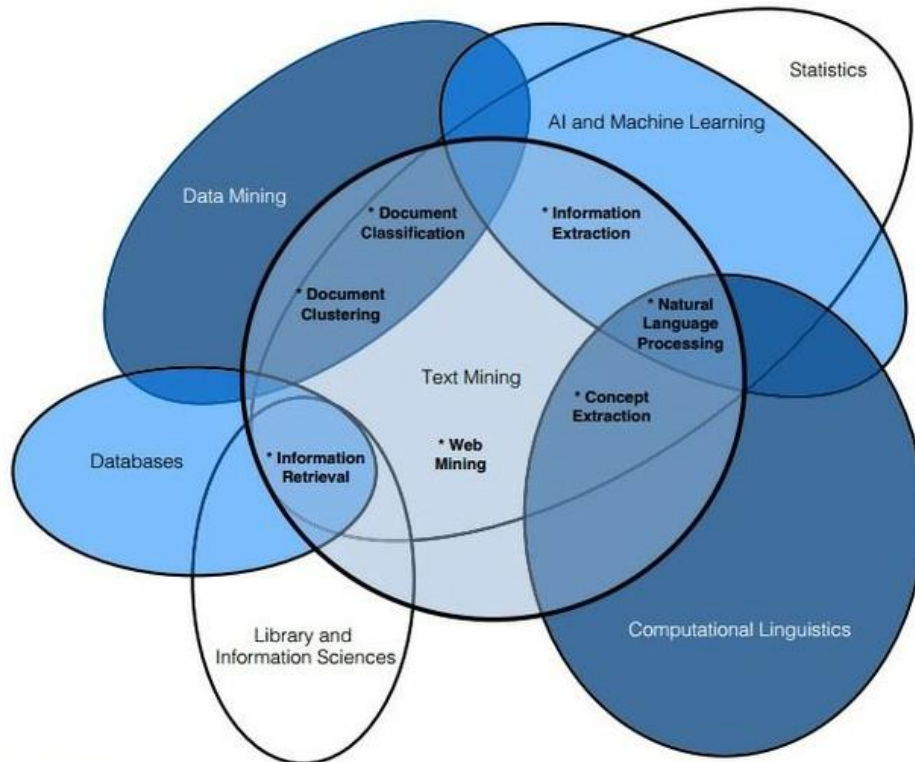
    Click to follow
The Independent US



"Trump uses social media as a weapon to control the news cycle. It works like a charm. His tweets are tactical rather than substantive."

George Lakoff [1]

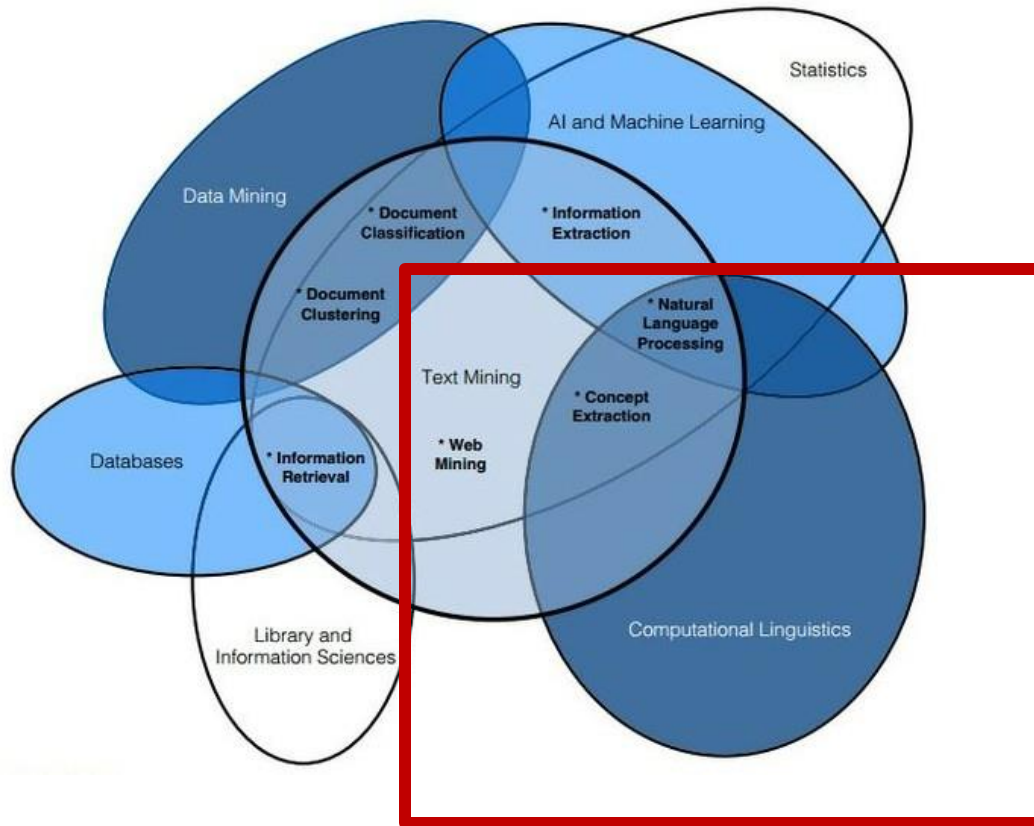
Computational Linguistics



"... is an interdisciplinary field dealing with the statistical and logical modeling of natural language from a computational perspective. "

- Natural Language Processing

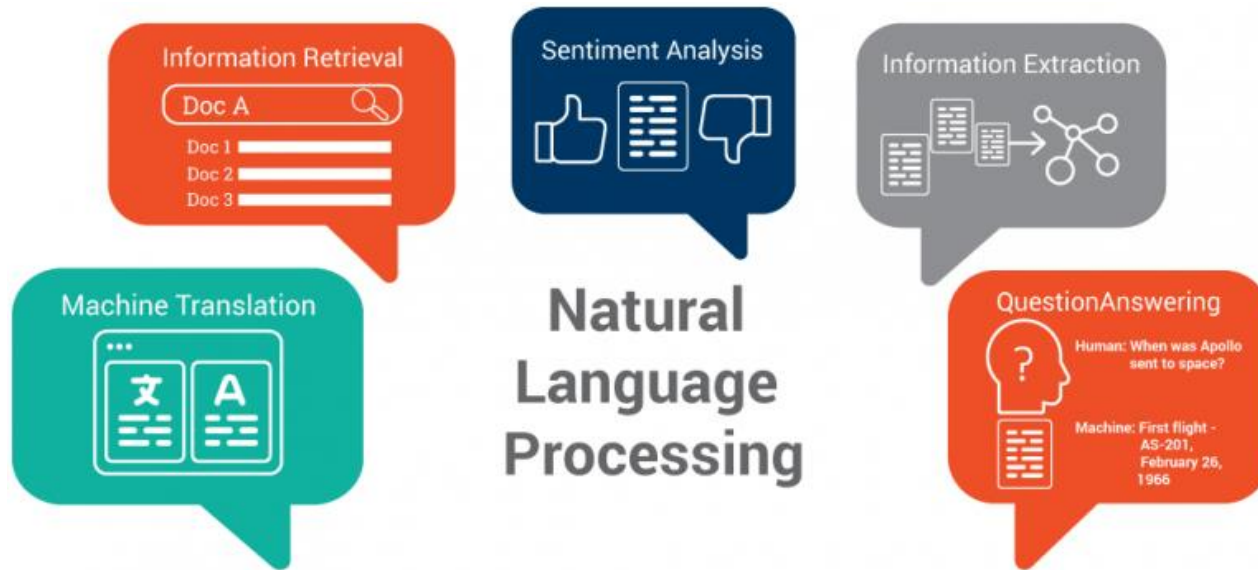
Computational Linguistics



"... is an interdisciplinary field dealing with the statistical and logical modeling of natural language from a computational perspective. "

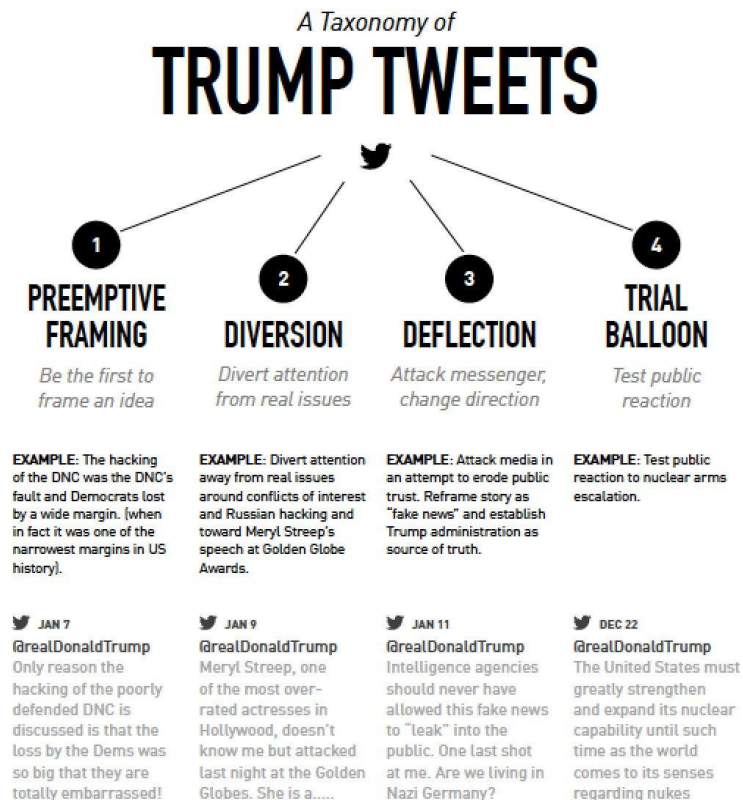
- Natural Language Processing

Natural Language Processing



"...concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data"

Example: George Lakoff on Trump



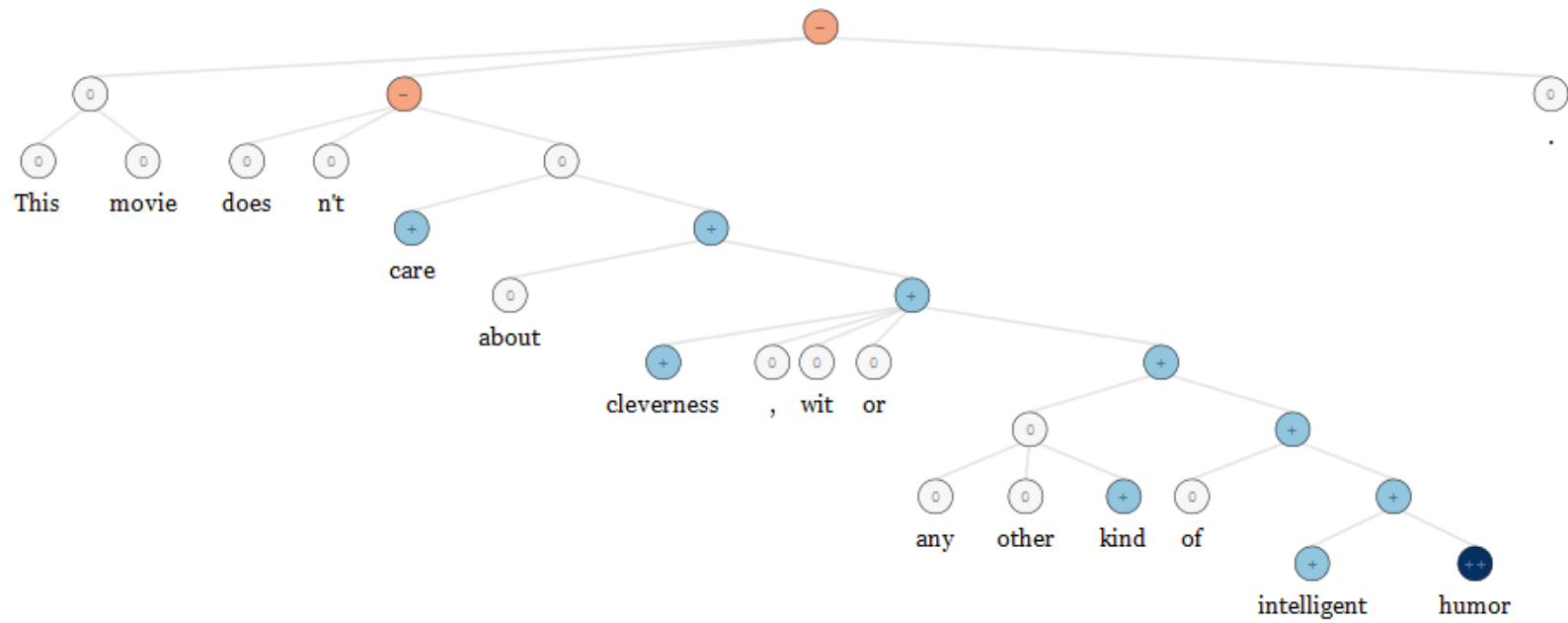
- Linguistic Analysis of Trump's Tweets

"Trump uses social media as a weapon to control the news cycle. It works like a charm. His tweets are tactical rather than substantive."

Code Examples

Sentiment Analysis

You can double-click on each tree figure to see its expanded version with greater details. There are 5 classes of sentiment classification: **very negative**, **negative**, neutral, **positive**, and **very positive**.



All labels are now correct

What can you do?

- Work with **existing github** repositories (don't invent the wheel) to extract data
- Use APIs and **tools** for NLP (or such) interesting fields:
 - Brand analysis
 - Sentiment analysis
 - Word Occurency (Named Entity Recognition)



References

- [1]<https://www.independent.co.uk/news/world/americas/us-politics/the-twitter-president-how-potus-changed-social-media-and-the-presidency-a8164161.html#r3z-addoor>
- Repositories
- **Twitter**
- <https://github.com/Twitter4J/Twitter4J/blob/master/twitter4j-examples/src/main/java/twitter4j/examples/timeline/GetUserTimeline.java>
- **Instagram**
- <https://github.com/postaddictme/instagram-java-scrapers>
- **Facebook**
- <https://github.com/KyleAMathews/facebook-export>

Thank you for your attention!

Questions?