

Social media phenomena: Hate & Conspiracies **+ Ask DS question**

Remaining Reddit data collection

IMT 547 - Social Media Data Mining and Analysis

26-Jan-2021 (Week 4, Day 7)

Today's Topics

- Project Idea generation
- Paper Discussion
- Asking DS questions (breakout rooms)
- Reddit data collection
- PS1 - out

Heilmeir's Questions

(As you start planning for your projects)



The Heilmeier Catechism



DARPA operates on the principle that generating big rewards requires taking big risks. But how does the Agency determine what risks are worth taking?

George H. Heilmeier, a former DARPA director (1975-1977), crafted a set of questions known as the "Heilmeier Catechism" to help Agency officials think through and evaluate proposed research programs.

- What are you trying to do? Articulate your objectives using absolutely no jargon.
- How is it done today, and what are the limits of current practice?
- What is new in your approach and why do you think it will be successful?
- Who cares? If you are successful, what difference will it make?
- What are the risks?

Heilmeir's Questions

(Needed for project pitch)

1. **What are you trying to do?** Articulate your objectives using absolutely no jargon. What is the problem? **Why is it hard?**
2. **How is it done today**, and what are the limits of current practice?
3. **What's new in your approach and why do you think it will be successful?**
4. **Who cares?**
5. **If you're successful, what difference will it make? What impact will success have? How will it be measured?**
6. What are the risks and the payoffs?
7. How much will it cost (in terms of people, resources, etc.)?
8. How long will it take?
9. What are the midterm and final “exams” to check for success? How will progress be measured?

Project Idea Generation

Due this Friday (Grading: You will be graded for completeness.) This step helps in group formation!!

Check Canvas

Project Idea Generation

 Published

 Edit



For your idea generation phase, you should propose 2 to 3 ideas (a minimum of 2 ideas) for your project. You are free to propose more, but your submission should be limited to 2-pages, 12-point font. Each of your ideas should clearly answer two questions and mention the topic of your idea in a few short keywords:

- What you want to do?
- Why should we care?
- Keywords – To mark topic and domain of the idea.

VERY IMP: For at least 1 of your ideas, you should be able to relate the topic of your project to one of the following themes. Both your ideas can be related to any of these themes or your 2nd idea can be a new theme. If it is a new theme say what theme that is.

- Theme 1: News and misinformation
- Theme 2: Hateful and offensive content
- Theme 3: Online communities (community growth, moderation, engagement, deception)
- Some other new theme: say what theme it is.

Papers

Early Public Responses to the Zika-Virus on YouTube: Prevalence of and Differences Between Conspiracy Theory and Informational Videos

Adina Nerghes
KNAW Humanities Cluster
Amsterdam, The Netherlands
adina.nerghes@dh.huc.knaw.nl

Peter Kerkhof
Vrije Universiteit Amsterdam
Amsterdam, The Netherlands
p.kerkhof@vu.nl

Iina Hellsten
University of Amsterdam
Amsterdam, The Netherlands
i.r.hellsten@uva.nl

Automated Hate Speech Detection and the Problem of Offensive Language

Thomas Davidson,¹ Dana Warmley,² Michael Macy,^{1,3} Ingmar Weber⁴

¹Department of Sociology, Cornell University, Ithaca, NY, USA

²Department of Applied Mathematics, Cornell University, Ithaca, NY, USA

³Department of Information Science, Cornell University, Ithaca, NY, USA

⁴Qatar Computing Research Institute, HBKU, Doha, Qatar
{trd54, dw457, mwmacy}@cornell.edu, iweber@hbku.edu.qa

Definition

Paper: What constitutes hate speech and when does it differ from offensive language?

both papers dealt with ambiguous topics that were difficult to both define and analyze.

- Jordan

Is it possible to even come up with a definition of hate speech that the majority of people would be satisfied with for regulatory purposes?

- Stephen

Borderline offensive language like using certain sensitive words is often debatable and that is what makes it sensitive

- Ayushi

Bias of human annotators

“Human coders appear to consider racists or homophobic terms to be hateful but consider words that are sexist and derogatory towards women to be only offensive, consistent prior findings” — Davidson Paper

the demographic of the human coders. Were the human coders predominantly men? Were the human coders randomly selected? What biases occurred because of the demographic of the human coders?

Joshua

Accuracy and reliability of CrowdFlower

Aniruddh

Policies & standards on hate speech

We also have a problem with different standards and policies on various online media. What is considered hate speech in one country might be different.

- Aftab

The complex nature of hate speech can be seen by how frequently websites like Twitter are updating their hate policy.

- Meghana

Using Twitter ▾

Managing your account ▾

Safety and security ▾

Rules and policies ▲

Twitter Rules and policies

General guidelines and policies

Law enforcement guidelines

Research and experiments

General guidelines and policies

About **public-interest** exceptions on Twitter

COVID-19 misleading information policy

Violent threats policy

Abusive profile information

Glorification of violence policy

About **government and state-affiliated** media account labels on Twitter

About **search rules** and restrictions

About **rules and best practices** with account behaviors

Automation rules

About **Twitter’s APIs**

About **Twitter limits**

Guidelines for Promotions on Twitter

Fair use policy

Counterfeit policy

Vine Camera Terms of Service and privacy policy

Updates to our **Terms of Service** and **Privacy Policy**

How to contact Twitter about a **deceased family member's account**

Defending and respecting the rights of people using our service

About specific instances when a **Tweet’s reach may be limited**

Our range of **enforcement options**

Our approach to policy development and **enforcement philosophy**

Notices on Twitter and what they mean

Violent organizations policy

Twitter, our services, and corporate affiliates

Suicide and Self-harm Policy

Additional information about **data processing**

Abusive behavior

Automated copyright claims policy

Distribution of **hacked materials** policy

Platform manipulation and spam policy

Reporting **false information** in France

Civic integrity policy

<https://help.twitter.com/en/rules-and-policies#general-policies>

↑ [Scroll to top](#)

Community drive policies & rules

reddit.com/r/changemyview/

29 minutes ago by [Chrimunn](#)

6 comments share save hide give award report crosspost

CMV: statues of bad people should not be destroyed

20 hours ago by [DrussTheDeathwalker](#)

81 comments share save hide give award report crosspost DELTA(S) FROM OP

CMV: the garbage/recycling industry is highly under-appreciated

14 hours ago by [lucas11119999](#)

18 comments share save hide give award report crosspost

CMV: The constant demonizing of political parties and opinions on social media is the most dangerous assault on democracy to date

21 hours ago by [JackC1126](#)

140 comments share save hide give award report crosspost

CMV: Facebook and Twitter are way more insidious formats than reddit

15 hours ago by * (last edited 6 hours ago) [ArghBlathEh](#)

10 comments share save hide give award report crosspost

CMV: The US Social Security System would be more sustainable if it was tied to the individual.

11 hours ago by * (last edited 11 hours ago) [Evil_Thresh](#) 14Δ

26 comments share save hide give award report crosspost

↑

Made with FDA-approved benzoyl peroxide & probiotics, Go Away Acne Spot Treatment is the acne-fighting formula you need in your routine.

▶

la

promoted save give award report

CMV: Using the pronoun "they/them" is confusing and it should be changed

17 hours ago by * (last edited 17 hours ago) [Jugglamaggot](#)

41 comments share save hide give award report crosspost DELTA(S) FROM OP

CMV: California's (or more specifically, its state government's) relationship with its corporate businesses is killing its economy

9 hours ago by [survivspicymilk](#)

11 comments share save hide give award report crosspost DELTA(S) FROM OP

CMV: it's never okay to strike a puppy (with your hand) if it isn't a threat to you.

13 hours ago by [promise_io](#)

⚙️

Submission Rules

hover over sections for more info

A

Explain the *reasoning behind* your view, not just what that view is (500+ characters required). ▾

B

You must personally hold the view and demonstrate that you are open to it changing. ▾

C

Submission titles must adequately sum up your view and include "CMV:" at the beginning. ▾

D

Posts cannot express a neutral stance, suggest harm against a specific person, be self-promotional, or discuss this subreddit (visit [r/ideasformv](#) instead). ▾

E

Only post if you are willing to have a conversation with those who reply to you, and are available to *start* doing so *within* 3 hours of posting. ▾

⚙️

Comment Rules

hover over sections for more info

1

Direct responses to a CMV post must challenge at least one aspect of OP's stated view (however minor), or ask a clarifying question. ▾

2

Don't be rude or hostile to other users. ▾

3

Refrain from accusing OP or anyone else of being unwilling to change their view, or of arguing in bad faith. ▾

4

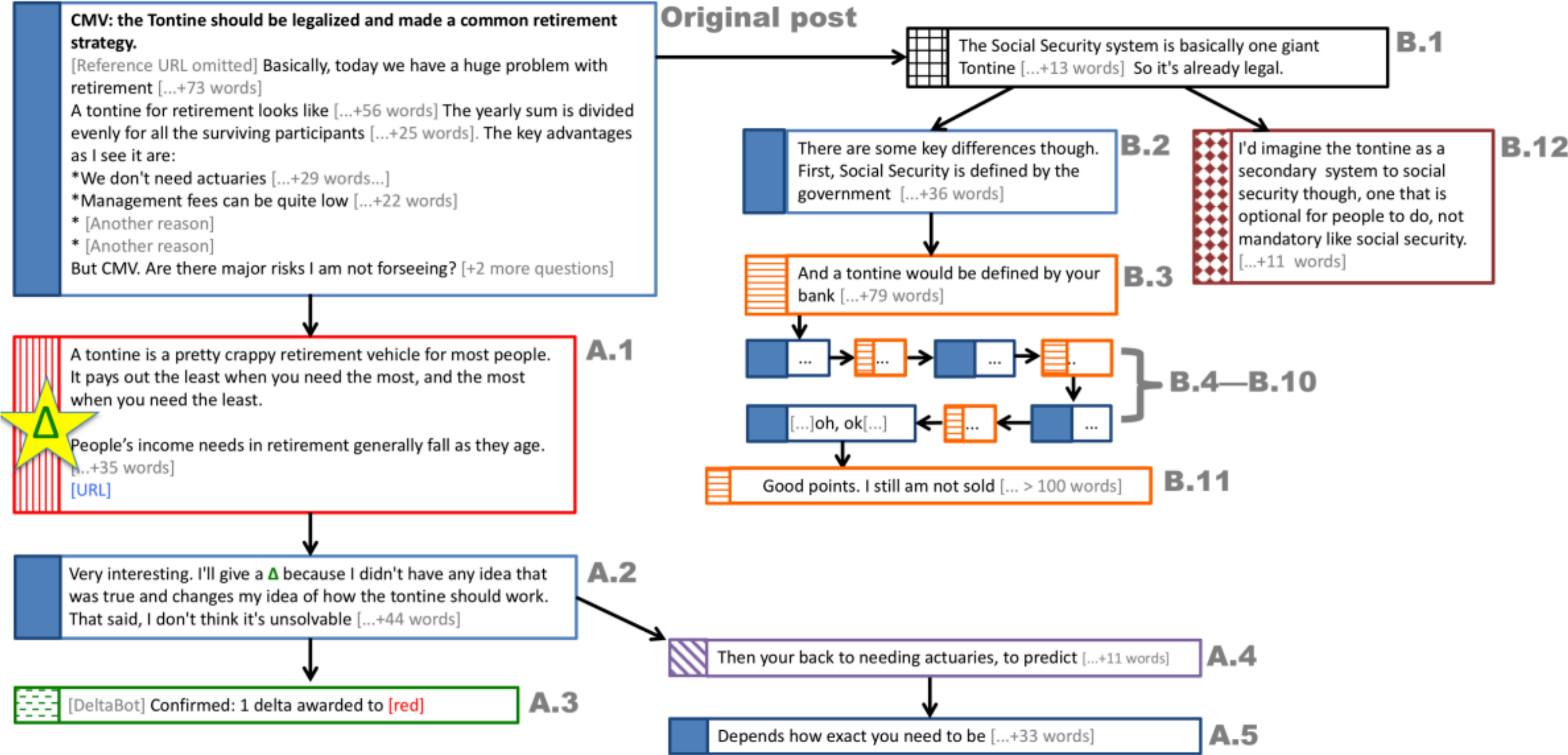
Award a [delta](#) if you've acknowledged a change in your view. Do not use deltas for any other purpose. ▾

5

Comments must contribute meaningfully to the conversation. ▾

Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-faith Online Discussions

Chenhao Tan Vlad Niculae Cristian Danescu-Niculescu-Mizil Lillian Lee
Cornell University
{chenhao|vlad|cristian|llee}@cs.cornell.edu



DISCUSSION TREE

Figure 1: A fragment of a “typical” /r/ChangeMyView *discussion tree*—typical in the sense that the full discussion tree has an average number of replies (54), although we abbreviate or omit many of them for compactness and readability. Colors indicate distinct users. Of the 17 replies shown (in our terminology, every node except the original post is a reply), the OP explicitly acknowledged only one as having changed their view: the starred reply A.1. The explicit signal is the “Δ” character in reply A.2. (The full discussion tree is available at https://www.reddit.com/r/changemyview/comments/3mzc6u/cm��_the_tontine_should_be_legalized_and_made_a/.)

Method issues & suggested improvements

Many suggestions

some hate speech tweets are not in the form of text but some memes like static images and gifs. There might be some hate speech text on the images that cannot be collected by normal lexicon methods.

- Esther

🔗 master ▾ 🔗 1 branch 🏷 0 tags

Go to file

↓ Code ▾

🔗 master ▾ hate-speech-and-offensive-language / lexicons /

🟢 t-davidson Initial commit

📄 readme.md Initial commit

📄 refined_ngram_dict.csv Initial commit

readme.md

Hate Speech Lexicons

This directory contains two lexicons that can be used to i

The file `hatebase_dict.csv` contains the original lexicon high recall it is associated with a high rate of false positive manner (e.g. yellow, oreo, bird).

The file `refined_ngram_dict.csv` contains a refined lexi were contained in our labelled data and for each n-gram (speech by the human coders. We then manually went thr

🟢 t-davidson Updated README 6d15050 on Jun 6, 2019 ⌚ 3 commits

📁 classifier	Initial commit	2 years ago
📁 data	Initial commit	2 years ago
📁 lexicons	Initial commit	2 years ago
📁 src	Initial commit	2 years ago
📄 .gitignore	Initial commit	2 years ago
📄 ICWSM_poster.pdf	Initial commit	2 years ago
📄 LICENSE	Initial commit	2 years ago
📄 README.md	Updated README	2 years ago

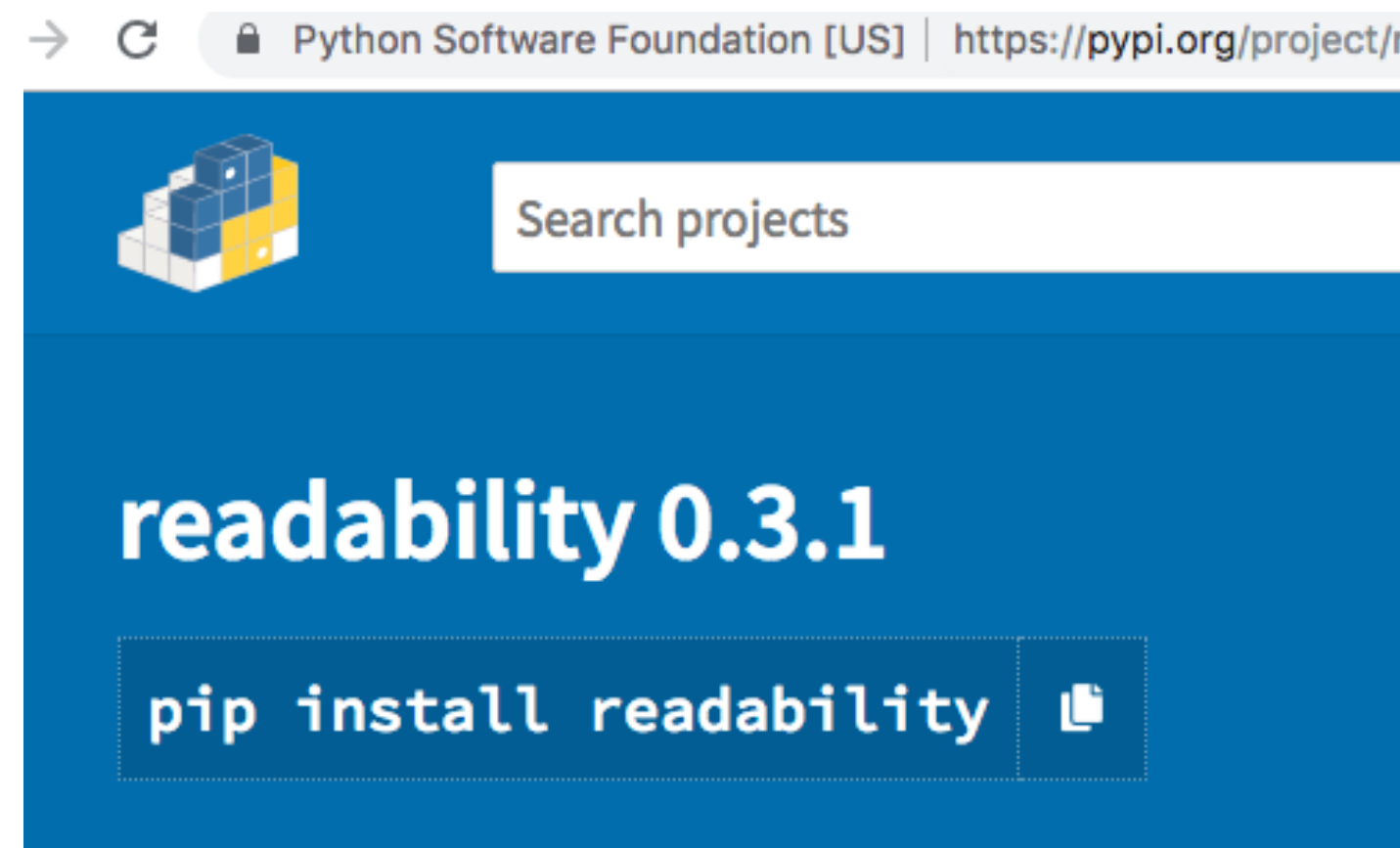
README.md

Automated Hate Speech Detection and the Problem of Offensive Language

Repository for Thomas Davidson, Dana Warmasley, Michael Macy, and Ingmar Weber. 2017. "Automated Hate Speech Detection and the Problem of Offensive Language." ICWSM. You read the paper [here](#).

Recall analytics techniques also used in previous papers

(will do notebook exercise today)



```
$ ucto -L en -n -s '' "CONRAD, Joseph - Lord Jim.txt" | readability
[...]
readability grades:
  Kincaid:                5.44
  ARI:                    6.39
  Coleman-Liau:           6.91
  FleschReadingEase:      85.17
  GunningFogIndex:        9.86
  LIX:                   31.98
  SMOGIndex:              9.39
  RIX:                    2.56
  DaleChallIndex:         8.02
sentence info:
  characters_per_word:     4.17
  syll_per_word:          1.24
  words_per_sentence:     16.35
```



```
>>> textstat.flesch_reading_ease(test_data)
>>> textstat.smog_index(test_data)
>>> textstat.flesch_kincaid_grade(test_data)
>>> textstat.coleman_liau_index(test_data)
>>> textstat.automated_readability_index(test_data)
>>> textstat.dale_chall_readability_score(test_data)
>>> textstat.difficult_words(test_data)
>>> textstat.linsear_write_formula(test_data)
>>> textstat.gunning_fog(test_data)
>>> textstat.text_standard(test_data)
...
```


GET STARTED

- Installation
- Models & Languages
- Facts & Figures
- spaCy 101
- New in v2.0

GUIDES

Linguistic Features

- POS Tagging
- Dependency Parse
- Named Entities
- Tokenization
- Sentence Segmentation
- Rule-based Matching

RUN

TEXT	LEMMA	POS	TAG	DEP	SHAPE	ALPHA	STOP
Apple	apple	PROPN	NNP	nsubj	Xxxxx	True	False
is	be	VERB	VBZ	aux	xx	True	True
looking	look	VERB	VBG	R00T	xxxx	True	False
at	at	ADP	IN	prep	xx	True	True
buying	buy	VERB	VBG	pcomp	xxxx	True	False
U.K.	u.k	PROPN	NNP	compound	X.X.	False	False

NLP techniques

Future labs on spacy and scikit-learn

spaCy

USAGE

GET STARTED

Installation

Models & Languages

Facts & Figures

spaCy 101

New in v2.0

GUIDES

Linguistic Features

● POS Tagging

● Dependency Parse

● Named Entities

● Tokenization

● Sentence Segmentation

● Rule-based Matching

RUN

TEXT	LEMMA	POS	TAG	DEP	SHAPE	ALPHA	STOP
Apple	apple	PROPN	NNP	nsubj	Xxxxx	True	False
is	be	VERB	VBZ	aux	xx	True	True
looking	look	VERB	VBG	ROOT	xxxx	True	False
at	at	ADP	IN	prep	xx	True	True
buying	buy	VERB	VBG	pcomp	xxxx	True	False
ll k	ll k	PROPN	NNP	compound	X.X.X	False	False

Features:
POS, unigrams, bigrams.....

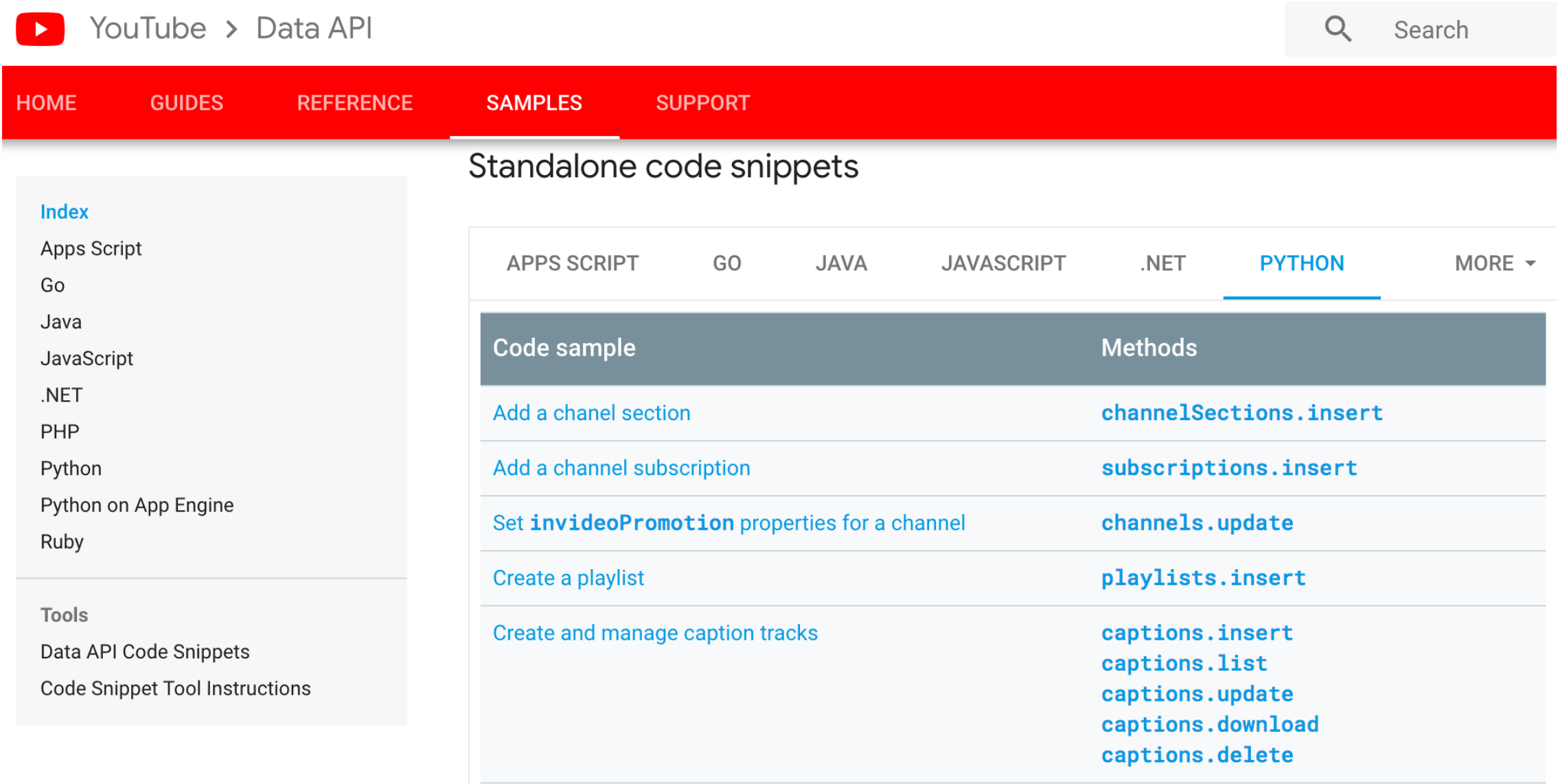
Asking DS questions!

A). Data Sources: You have two data sources at your disposal:

- 1. YouTube Data API. Check all the ways you can get data from YouTube: https://developers.google.com/youtube/v3/code_samples/
- 2. hate speech lexicons from the paper that you read: <https://github.com/t-davidson/hate-speech-and-offensive-language/tree/master/lexicons>

B). Project Themes: Keep the following two project themes

- News and misinformation (*broadly defined*)
- Hateful and offensive content (*broadly defined*)



Q 1). Considering A) and B), what are some key Research Questions you can ask and how you can go about answering them?

Q2). If you had additional datasets at your disposal (Twitter, Reddit, news, Instagram, etc.), how will that help answer your RQs.

BREAK

See you at 9:45am

Leftover Reddit lab

Anonymous Feedback

<https://forms.gle/Yp5suJ8c3qtGAArz9>

Upcoming classes

▼ Week 4 (Jan 25-29): Getting Started with Studying Social Media Phenomena

THU, JAN 28

Asking DS questions with social media

Exploratory Data Analysis

Required Reading

- [Python Plotting for Exploratory Data Analysis](#)

Lab: EDA

Idea generation writeup (due by Fri, 29-Jan)

▼ Week 5 (Feb 1-5): Working with Data

Due: Problem Set I by 5pm, Mon Feb 1

As you think through project ideas, see lecture slides from previous classes for additional pointers