

# Personality Traits Analysis from Forum Dataset

Angeliki Theodorou, Lirije Tahiri, Frederike Gartzky

December 12, 2019

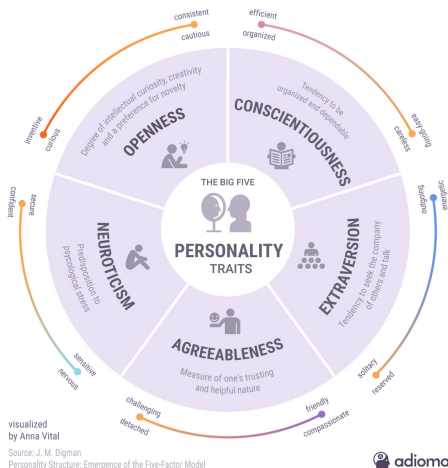
# Project tasks

- 1 Extract text from reddit using BeautifulSoup.
- 2 Identify text related to each user and select users according to some properties.
- 3 Install and run the big-five personality trait implementation in <https://github.com/SenticNet/personality-detection> with sample text.
- 4 Link the output of the webcrawler with the personality trait classification and implement a program that gets a single user as input and outputs their personality trait.
- 5 Compute semantic similarity of messages and predefined keywords associated with the personality traits.
- 6 Create a GUI, enabling the user to filter by subreddit and keyword or subreddit and username.

# Background: 5 personality traits

## OCEAN:

- Openness
- Conscientiousness
- Extroversion
- Agreeableness
- Neuroticism



# What we have done so far

- GUI using Tkinter
- extraction of text from reddit using BeautifulSoup
- used library multiprocessing for threading
- installed and tested personality detection framework

## Reddit Forum and Keyword

Enter the information required:



## Reddit Forum and User

Reddit Forum Subject:

Username:

Submit

## Personality Trait Identification

Reddit Forum and Keyword

Reddit Forum Subject:

starwars


Keyword for Search:

han solo

Submit

Reddit Forum and User

Enter the information required:



# Extracted Text from reddit

```
File Edit View Search Terminal Help
('In': 7, 'u'loves123', 'u'/r/StarWars', 'u'He is sure Han Solo is a looser (u2013 I believe! )')
('In': 23, 'u'SlothsDoinpoga', 'u'/r/StarWars', 'u'The new Han Solo')
('In': 17, 'u'solethprine', 'u'/r/StarWars', 'u'Why didn't u2019t Han Solo come back to help with the Resistance before the events of The Force Awakens?')
('In': 26, 'u'lovedumbcat', 'u'/r/StarWars', 'u'Rewatch, The Empire Strikes Back! This week I made little chocolate Han Solo using a silicone mold and Lili (u2019s Chocolate. They were a hit!')
('In': 8, 'u'SWGeek826', 'u'/r/StarWars', 'u'Found this suspiciously repackaged (u2019Han Solo (u2019 figure while grocery shopping.')
('In': 21, 'u'LowDownnDirt', 'u'/r/StarWars', 'u'Whose the better pilot Han Solo, Lando Calrissian, Atton Rand, Carth Onasi, or Poe Dameron?')
('In': 9, 'u'howlwolffy555', 'u'/r/StarWars', 'u'Who would win in a fight? Han Solo or the Mandalorian??')
('In': 35, 'u'Solidarity64', 'u'/r/StarWars', 'u'How could Han Solo have not believed in the Jedi or the force?')
('In': 17, 'u'Ashklir', 'u'/r/StarWars', 'u'I've only watched Han Solo. Where do I start?')
('In': 33, 'u'awndray97', 'u'/r/StarWars', 'u'The most annoying thing in ALL of Star Wars is that Han Solo says "hell".')
('In': 6, 'u'realgeneral_memeous', 'u'/r/StarWars', 'u'Was Han Solo established as ex-Imperial at any point during the OT/Is there evidence that Lucas and others had this backstory in mind during ANH?')
('In': 19, 'u'tinrodie', 'u'/r/StarWars', 'u'Le Creuset Han Solo roasting pan')
('In': 6, 'u'djcos25', 'u'/r/StarWars', 'u'What Han Solo? Ha, you don't need that guy if I'm around. Custom 6 inch Black Series Dash Ren (u2019s')
('In': 14, 'u'miketheadesome24', 'u'/r/StarWars', 'u'Is anyone not gonna talk about the fact that Finn's ROS outfit is oddly similar to Han Solos?')
('In': 3, 'u'nLaaan515', 'u'/r/StarWars', 'u'At-st, troopers and south park style Han Solo guarding my freezer door at work.')
('In': 11, 'u'rmeendto1a13', 'u'/r/StarWars', 'u'There were concerns with Han Solo's hibernation stickness. #proud papa')
('In': 18, 'u'brownmambak824', 'u'/r/StarWars', 'u'Why did they hire Han Solo?')
('In': 7, 'u'The Zekta', 'u'/r/StarWars', 'u'A Han Solo door I saw at a comic shop')
('In': 5, 'u'trenthescottish', 'u'/r/StarWars', 'u'Han Solo's (u2019s story is amazing')
('In': 3, 'u'KochIsland24', 'u'/r/StarWars', 'u'Han Solo')
('In': 26, 'u'tierfoneyellowaces', 'u'/r/StarWars', 'u'Han Solo donning his golden die on the Falcon (The Force Awakens Documentary)')
('In': 26, 'u'xx_AndyPer_Xx', 'u'/r/StarWars', 'u'Who shot first? Han Solo or Greedo?')
('In': 1, 'u'nattnick1983', 'u'/r/StarWars', 'u'This is my daughter's 2nd Christmas, and I've made it a new tradition to add a Star Wars Ornament every year. The Han Solo Ornament has sound, which my wife is thrilled with...')
('In': 21, 'u'TheDarkTowar', 'u'/r/StarWars', 'u'I find myself really disliking Han Solo for how he treats C3PO')
('In': 17, 'u'sbrlengel74', 'u'/r/StarWars', 'u'After saving the princess and being a hand in the destruction of the Death Star, you would think the rebel alliance would deem Han Solo as someone of great importance. But no, let's just send the princess and his friend to save him from Jabba. They go t this.')
('In': 0, 'u'Hifihedgehog', 'u'/r/StarWars', 'u'The Tragedy of Han Solo')
('In': 0, 'u'LookADrifter', 'u'/r/StarWars', 'u'The Banana Convention - I Wanna Be Han Solo')
('In': 0, 'u'LeMOB', 'u'/r/StarWars', 'u'I found this dly cardboard make of Han Solos blaster online , thought it would belong here.')
('In': 12, 'u'ZikMelster', 'u'/r/StarWars', 'u'Found Han Solo while I was in Slctly')
('In': 0, 'u'Xijustwantahug', 'u'/r/StarWars', 'u'A couple friends and I got dressed up for Halloween at work today. I was Han Solo.')
('In': 5, 'u'Goshawk5', 'u'/r/StarWars', 'u'I started reading the Lost Stars manga and I must say anime Han Solo is a thing of beauty.')
('In': 3, 'u'Darth_Quixote', 'u'/r/StarWars', 'u'Subtle reference to Han Solo's racing league in SW Resistance?')
('In': 9, 'u'KochIsland24', 'u'/r/StarWars', 'u'Han Solo or Anakin Skywalker (Darth Vader)')
```

# How does the given framework work?

## Deep Learning-Based Personality Detection[1]

- Preprocessing: sentence splitting, cleaning
- Document-level feature extraction: Mairesse baseline feature set
- Filtering: ignore sentences that don't carry any personality cues
- Word-level feature extraction: word2vec representation
- Classification: deep CNN



# Personality Trait Classification with given Framework

```
(NLP) lili@grille:~/NLP/personality-detection$ python process_data.py ../GoogleNews-vectors-negative300.bin ./essays.csv ./mairesse.csv
loading data... data loaded!
number of status: 2467
vocab size: 30391
max sentence length: 149
█

(NLP) lili@grille:~/NLP/personality-detection$ python conv_net_train.py -static -word2vec 2
loading data... data loaded!
model architecture: CNN-static
using: word2vec vectors
[('image shape', 153, 300), ('filter shape', [(200, 1, 1, 300), (200, 1, 2, 300), (200, 1, 3, 300)]), ('hidden units', [200, 200, 2]), ('dropout', [0.5, 0.5, 0.5]), ('batch_size', 50), ('non_static', False), ('learn_decay', 0.95), ('conv_non_linear', 'relu'), ('non_static', False), ('sqr_norm_lin', 9), ('shuffle_batch', True)]]
... training
█
```

# What next?

## Todo:

- Link the crawly output with the personality trait classifier.
- Program that takes data of a user and outputs the respective personality trait.
- Personality trait classifier using keywords and semantic similarity, compare results.

## Sharing the work

- divide all tasks evenly
- still collaborate and help each other

# References

- [1] N. Majumder, S. Poria, A. Gelbukh, E. Cambria, *Deep Learning-Based Document Modeling for Personality Detection from Text*, IEEE Intelligent Systems, 2017
- [2] *Personality traits picture* <https://blog.adioma.com/5-personality-traits-infographic/>