Visualize my Corpus

Dr Mahmoud El-Haj
Lecturer
SCC, Lancaster University
What is visualisation?

3000 BC

2021

communicate a message
What to visualise?
WHY VISUALISE?

Machine: I can read it!

Human: I can read it too!
Isn’t data enough as it is?
Data in a plot
What can we do with text!

- Tokenization
- Annotation
- Topic Modelling
- Machine Learning
- Visualisation
Or as in the language of memes

tokenization  annotation  topic modelling  machine learning  VISUALISATION
Visualise my corpus.

Text visualisation is awesome!

Visually representing the content of a text document is important.

Today is March 18th and Mahmoud is showing us how to visualising text at Lancaster University, well online!
Visualise [VERB] my [DET] corpus [NOUN].</s>

Text [PROPN] visualisation [PROPN] is [AUX] awesome! [ADJ]


{Today is March 18th}[DATE] and {Mahmoud}[PERSON] is showing us how to visualising text at {Lancaster University}[ORG].
Today is March 18th and Mahmoud is showing us how to visualise text at Lancaster University.

- Text: PROPN
- Visualisation: PROPN
- is: AUX
- awesome!: ADJ

- Visualise: VERB
- my: DET
- corpus.: NOUN

- Visually: ADV
- representing: VERB
- the: DET
- content: NOUN
- of: ADP
- a: DET
- text: NOUN
- document: NOUN
- is: AUX
- important: ADJ

https://universaldependencies.org/docs/en/dep/
What can visualisation tell us?

Things we don’t directly see!
a social news aggregation, web content rating, and discussion website
TRUMP vs OBAMA vs VOTE

https://projects.fivethirtyeight.com/reddit-ngram/?keyword=trump.obama.vote&start=20071015&end=20170731&smoothing=22
How the Internet Talks!

Major events/trends between 2011 - 2015

https://projects.fivethirtyeight.com/reddit-ngram
In May 2014, a study by Matt Daniels found that Aesop Rock's vocabulary in his music surpassed 85 other major hip-hop and rap artists, as well as Shakespeare's works and Herman Melville's Moby Dick, being named the largest vocabulary in Hip Hop. [43][44] Wikipedia
Stanford PhD Theses

https://nlp.stanford.edu/projects/dissertations/brows er.html

Stanford's PhD dissertation abstracts from 1993-2008
Drawing one of my papers*

Try it out:
https://www.lancaster.ac.uk/staff/elhaj/draw/index.html#habibi

*https://www.lancaster.ac.uk/staff/elhaj/docs/habibi.pdf
Text Visualisation Browser
https://textvis.lnu.se/

Data Driven Documents
https://d3js.org/

Jim Vallandingham
http://vallandingham.me
What are we visualising today?

Chaos Communication Congress talks
(https://www.ccc.de/en/)

Tutorial Material:
https://github.com/drelhaj/NLP_ML_Visualization_Tutorial

Conference Data:
https://gitlab.com/maxigas/cc_congresstalks/
Visualisation Roadmap

1. SpaCy Intro
2. Topic Modelling
3. Word Clouds
4. Machine Learning
5. Word Usage
6. Word Embeddings

https://github.com/drelhaj/NLP_ML_Visualization_Tutorial