Leveraging Unstructured Data in Higher Education

25th SAAIR Conference 2018

Re-imagining our universities: the role and impact of institutional research in times of disruption

Technology, big data, AI and the future of higher education

Mr Innocent Mamvura

saair
Southern African Association for Institutional Research

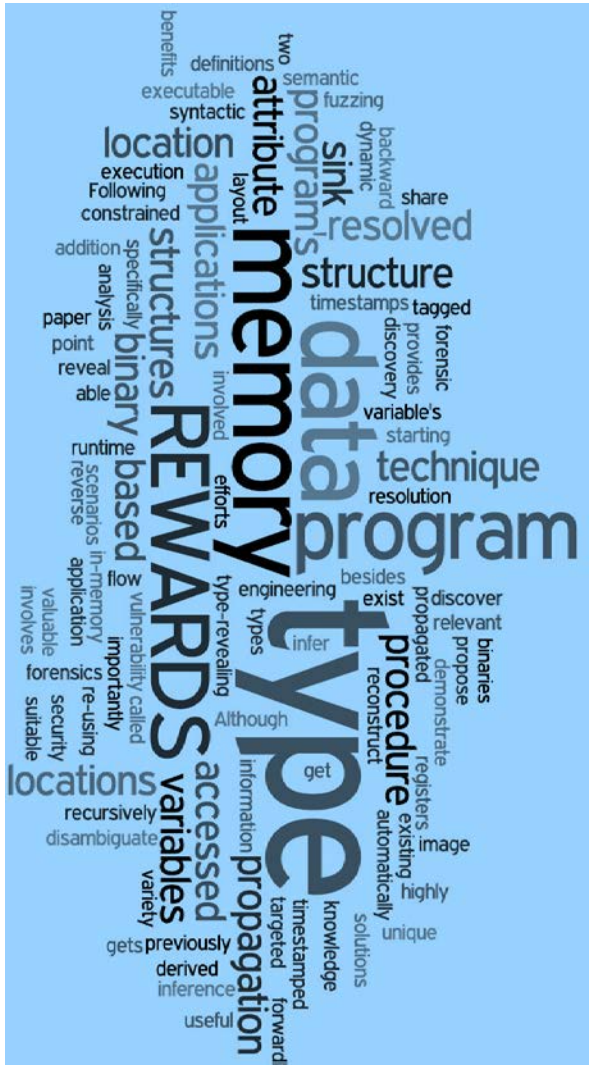
DUT DURBAN UNIVERSITY OF TECHNOLOGY

WITS UNIVERSITY

Abstract

- Data is doubling in size every two years and the total amount of data will reach 44ZB by 2020, with 80% of that unstructured, according to IDC Research. This becomes and even large task when interpreting large sets of unstructured data, which comes in the form of emails, social media, blogs, documents, images and videos, represent a significant source of opportunity for businesses. This paper use machine learning algorithms to analyse social media data. The University of Witwatersrand official twitter accounts were used as data sources to feed into an API connector, the data was then pre-processed to remove irrelevant information such as punctuation marks, full stops and numbers before creating a bag of words. The word cloud technology was then used to generate the words with the highest frequencies and importance. We also demonstrate how institutional researchers can make use of unstructured data to better understand their organisations.

What is unstructured data?



a generic label for describing any corporate information that is not in a database

information that either does not have a pre-defined data model or is not organized in a pre-defined manner

unstructured data : Data that does not reside in fixed locations

any data that has no identifiable structure

Why is unstructured data important?

Unstructured data doubles every three months

7 million web pages are added every day

80% of business is conducted on unstructured information

85% of all data stored is held in an unstructured format

Unstructured data is not organized

- **Text files:** Word processing, spreadsheets, presentations, email, logs.
- **Email:** Email has some internal structure thanks to its metadata, and we sometimes refer to it as semi-structured.
- **Social Media:** Data from Facebook, Twitter, LinkedIn.
- **Website:** YouTube, Instagram, photo sharing sites.
- **Mobile data:** Text messages, locations.
- **Communications:** Chat, IM, phone recordings, collaboration software.
- **Media:** MP3, digital photos, audio and video files.
- **Business applications:** MS Office documents, productivity applications

Real World Applications

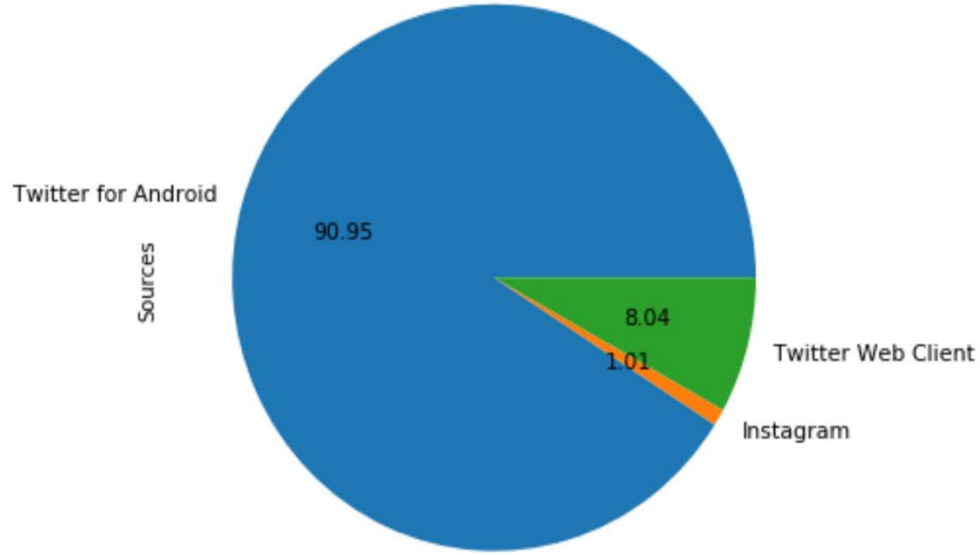
The Wits Inala Food Project

Inala

WITS FOOD SOVEREIGNTY

The tweet with more likes is:
Look at how amazing our Main garden is looking! Love hearing tbe birds sing in the background! #wit
s… https://t.co/AJjZKP4SYW
Number of likes: 18
125 characters.

The tweet with more retweets is:
RT @shawtyarabia: Students are out here in college struggling to afford decent food and it’s dismisse
d as part of the “college experience”…
Number of retweets: 41832
139 characters.



RT @sconnectedness: #supportingstudents https://t.co/SrFhoem6Vn
RT @Courtz_RM: Last 2 weeks of crowdfunding for the Wits Food Sovereignty Center! Food is a human r
ight. Please share link: https://t.co/f…
RT @VishwasSatgar: Climate change aggravates global hunger: UN. Join the dialogue for a Climate Jus
tice Charter. @safoodsov @COPAC_SA …
RT @Courtz_RM: Because the food system in this country is BROKEN. https://t.co/yMyLOhN4WW
RT @synergos: Food security in #Nigeria is impossible without #women - @AuduOgbeh, Minister of @fma
rdng https://t.co/By2b1BlAPE #agriculture
Tell your friends and family about the crowd funding campaign that we are running!

We are raising money to develop… https://t.co/6xTnCDZeLT
RT @Courtz_RM: Food is a human right! Can we come together to restore dignity to hungry students? H
elp us build a Food Sovereignty Centre a…
Today we celebrated the launch of the communial eating space at Wits. Inala is committed to continu
e providing fres… https://t.co/sDk2HRVb5B
RT @VishwasSatgar: Students having a party at the communal eating space launched for them today at
@WitsUniversity . Great music , poetry…
Our new Exec enjoying the entertainment! https://t.co/Myl9JU6Yig

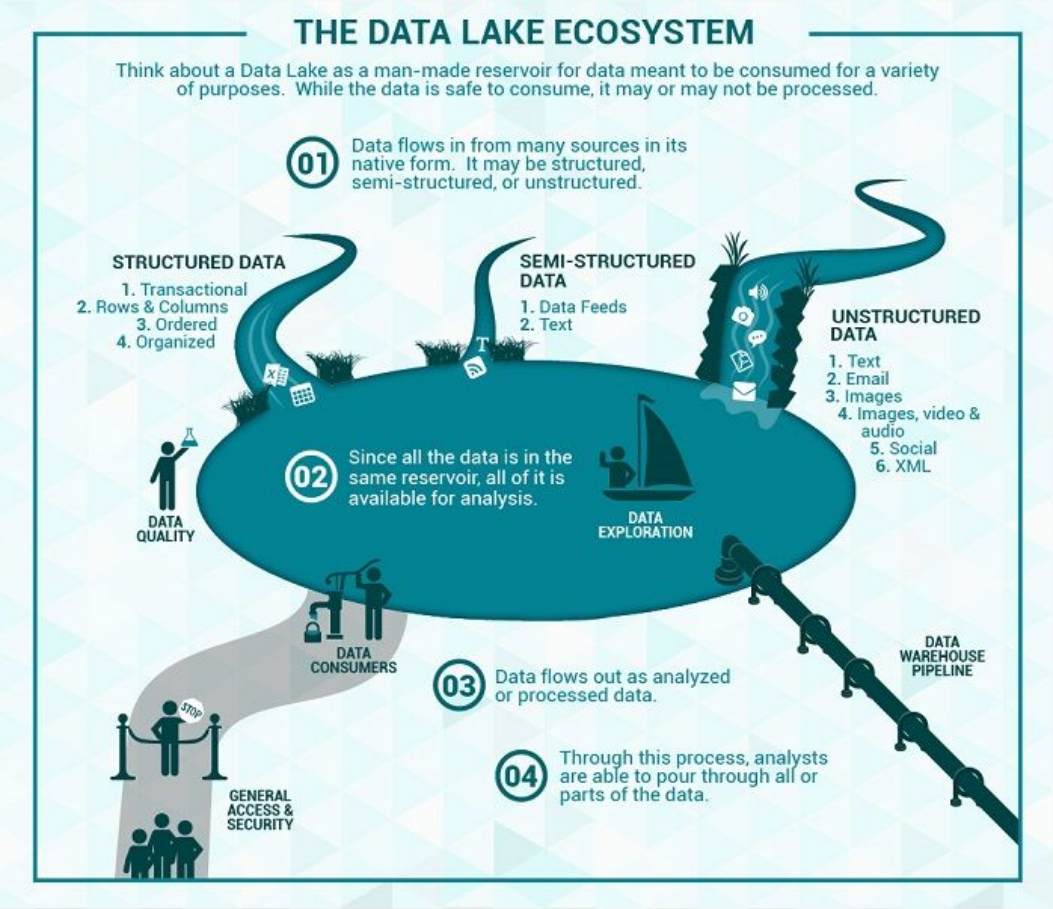
	Tweets	len	ID	Date	Source	Likes	RTs	geo	coordinates
0	RT @sconnectedness: #supportingstudents https:...	63	1056444480772026368	2018-10-28 07:16:06	Twitter for Android	0	2	None	None
1	RT @Courtz_RM: Last 2 weeks of crowdfunding fo...	140	1054714565605371904	2018-10-23 12:42:02	Twitter for Android	0	20	None	None
2	RT @VishwasSatgar: Climate change aggravates g...	140	1052810456103047168	2018-10-18 06:35:47	Twitter for Android	0	4	None	None
3	RT @Courtz_RM: Because the food system in this...	89	1051700680371707907	2018-10-15 05:05:56	Twitter for Android	0	1	None	None
4	RT @synergos: Food security in #Nigeria is imp...	140	1049320930613362688	2018-10-08 15:29:39	Twitter for Android	0	3	None	None
5	Tell your friends and family about the crowd f...	140	1049320250758635520	2018-10-08 15:26:57	Twitter for Android	1	2	None	None
6	RT @Courtz_RM: Food is a human right! Can we c...	140	1047926881281032192	2018-10-04 19:10:12	Twitter for Android	0	6	None	None
7	Today we celebrated the launch of the communia...	140	1045686006752718855	2018-09-28 14:45:46	Twitter for Android	11	6	None	None

The Word Cloud with Natural Language Processing

Sentiment Analysis

- A sentiment analysis was used to score the tweets into positive, neutral and negative. Sentiment analysis can be defined as a process that automates mining of attitudes, opinions, views and emotions from text, speech, tweets and database sources through Natural Language Processing (NLP). Sentiment analysis involves classifying opinions in text into categories like "positive" or "negative" or "neutral". It's also referred as subjectivity analysis, opinion mining, and appraisal extraction. The results of the sentiment analysis shows that 44% percent of the tweets were positive and 44% were neutral tweets whilst 11% of the tweets were negative

How do you integrate your data



THANK YOU

Innocent Mamvura

Tel: 0117171148

Cell: 0834006855

Innocent.Mamvura@wits.ac.za

https://www.linkedin.com/in/innocent-mamvura-673a8915/