Data Science Project

Social Network Footprinting @ COVID19.4IR.ZA
COMP301

(Adapted from a Project at the University of Fribourg, Switzerland)

- **Level**: 3rd Year B.Sc. Group Project
- **Prerequisites**: Java/Python/Javascript

Description: (A.K.A. Social Data Acquisition and Social Graph Processing)

discover the problem first before doing coding (delay coding as much of possible)

dont reinvent the wheel. use existing APIs and frameworks

Today, social networks are the first choice for marketing campaigns. They promise to serve well targeted, viral, highly customisable advertisements while getting direct customer feedback and engagement. The numbers generated by the Internet companies serving digital advertisements are astronomical: Google $40B, Facebook $32B, etc…

In this context of online marketing through social networks during the COVID pandemic, the tasks of this project are split in three parts, the first and the second parts being more pragmatic (hands-on) and the third more theoretical:

1. Introduction to Web and Graph Processing Frameworks (e.g. Spring Framework) (One Week)

2. Social Data Acquisition (4 Weeks)
   - Crawl a public network (Twitter, etc…)
   - Store the data in a nosql database (neo4j or mongodb)
   - Build a web interface to search through the data (Java, javascript and Cypher/MQL)

3. Social Graph Processing (3 weeks)
   - Starting from one given node in the graph (i.e. a particular company), the social graph will be quantified and analyzed.
   - For instance A always retweet B which always retweet C; A, B and C are part of the same cluster.
   - Analytics: Basic (graph statistics), and advanced (NLP, sentiment analysis, predictive analysis)

can have a mulitlanguage project; a different one for each task

The goal of the project is to quantify and classify the social network footprint of 4IR companies* on social networks before, and during the COVID19 pandemic. You can choose two or three South African companies like Vodacom or MTN, or global companies that have business here.

Working on this project will give you the opportunity to acquire in-depth hands-on experience with the Spring Framework (or similar), state-of-the-art APIs, storage (graph or NoSQL database), graph processing, big data and data science.

The process of software development should follow a Design Thinking approach from beginning to the end. You are required to make use of existing design patterns. Your lecturer will be your client. For the empathy phase you will have to consult him so that he explains the business problem for you to scope and define a clear design challenge. You will be evaluated both individually and as a team on each Design Thinking process for your contribution. There is absolutely no limit to the tools you can use, as long as they will help you come up with an innovative solution. Preferably, make use of open source software for the obvious reason that we do not have a budget. This work is to be done in a group of 2-6 people per group, each with a specific contribution to the sum of all contributions. Your project evaluation includes but is not limited to biweekly progress reports by each team to the lecturer, final solution, technical documentation of the final system, final presentation of the project (zoom or pre-recorded screencast with zoom interview at the time)

You will have to think out of the box and show some creativity to come up with an innovative solution that caters for future functionality. There will be bonus points for going an extra mile (e.g. include fb, Instagram, google, etc. Or advanced analytics like sentiment analysis)

* there is an option of choosing an alternative to marketing analytics – a group could choose any socio/political topic where there is a sizable distinction between one or more communities/factions/groups. Please discuss this with your lecturer before you commit to such a topic.