Extracting tourists movements' information from Twitter data sets

Mickaël Nkunku
Enschede, Netherlands

Thibault Defeyter
Enschede, Netherlands

Jose Ailton Filho
Enschede, Netherlands

Fernando Augusto Machado
Enschede, Netherlands

I. Introduction

The era of the Internet of things we are currently living in is still improving and everyone owning a means to access Internet can express onself. Data published in any way on the network is stored somewhere and can be identified. That identification procedure usually includes retrieving more information than one can see when posting something. Those pieces of information can be anything: IP address, geo-localization, name of the web browser and name of the operating system used are few instances. To put it in a nutshell, from each data set can be extracted a lot of information which anybody who is provided an access to can do analytics on.

The current context enables to use previously unavailable information and create services as to improve users' experience when connected to the network. Information about tourists' movements, for instance, could allow companies and public institutions to do the necessary so that travelers and tourists have a better experience, and even prevent them from doing unnecessary expenses. However, useful information cannot be accessed in a direct way: it must be extracted. Having at disposal an extracting information means which could process huge data sets would be indeed a useful tool in order to analyze such data.

For this research topic, we have chosen to output a table with relevant attributes. Our table will be filled by analyzing huge data sets of tweets which are also used by the *twiqs.nl* website. From each tweet can be extracted useful pieces of information, in particular, geolocalization details. Some papers propose their method to find users' hometowns like the Principal Components Analysis mentionned in paper[1]. As for the technology to be used to approach this research topic, Pig Latin, over the Hadoop cluster, is the most suited. User defined functions will also be implemented in order to increase its efficiency in our favor.

II. Materials and methods

III. Results and discussion

IV. Conclusion

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.