Analytics Practicum Project Details

1. **Please give a brief background of the proposed Analytics Practicum (You can provide links for reference, objective is to get an understanding of the area of work)**

The Green Transformation Lab (GTL) is a joint initiative by SMU and DHL aimed at accelerating the evolution of sustainable logistics across Asia Pacific. Leveraging SMU’s multi-faculty academic excellence and DHL’s sustainability services, expertise and capability in supply chains, the Green Transformation Lab is focused on creating solutions that help companies transform their supply chains, becoming greener, more resource efficient and sustainable.

GTL’s Sustainability Heatmap is a group of heatmaps with global profiling of areas and trends on

1. sustainability-related jobs and
2. sustainability-related topics on Twitter.

Currently these are the only two heatmaps. Both heatmaps has a crawler that crawls the job markets in various parts of the world and Twitter data respectively. Currently, the aim of the application is to administer the crawling, manage keywords and to have an initial visualization of the data. However, little insights have been obtained from the crawled data.

The aim of this analytics project is to analyse the crawled data deeper to gain more insights. Here are some of the areas (not restricted to) that we thought might be interesting:

Heatmap on Sustainability Jobs

* 1. Is there a trend in the type of sustainability jobs in the various regions, e.g., Europe, Asia, America?
	2. We initially observed that the jobs originating from Europe seems to add “sustainable environment” to almost every jobs (related or not related to sustainability jobs). Why? Is it more attractive to potential employees? Or is it policy driven in Europe?
	3. Which is the more important measurement? Number of sustainability jobs per GDP or number of sustainability jobs per population? Is there any other measurements?
	4. …… (we need more insights from analyzing the raw crawled data)…

Heatmap based on Sustainability-related topics from Twitter data

1. How pervasive are the various topics?
2. Which are the topics driven by social network (e.g., Ice-bucket challenge) and what are driven by commercial companies (e.g., Microsoft organic surface pro cover)?
3. ……(we need more insights from analyzing the raw crawled Twitter data)…
4. **Any other specifications based on past work done by your or your colleagues (or any other source) in the area that you would like to quote as a reference in case this is a continuation of a past analysis**

The sustainability heatmap application is hosted at <http://heatmap.greentransformationlab.com/>

More information about general things that GTL does can be found at our website

<http://gtl.smu.edu.sg>

1. **Data Details: Please provide the link or source of the data for the analysis. Data should be readily available and students should not have to spend a lot of time in data collection. Hence, data should be immediately available for download**

We will run the crawler for consecutive days for two weeks in Dec 2014. Data will be read by Jan 2015.

**Data include:**

1. 2 weeks of crawled jobs related to sustainability from 9 different countries
2. 2 weeks of crawled Twitter messages related to sustainability

All data are currently stored on MongoDB. Data and the data dictionary can be exported at start of project.

1. **Please provide your expectation from this analysis and state what type of metrics, trends, predictions and visualizations you think should be done on this data. You can decide the level of detail for this at this point, more guidance might be needed during the duration of the project**

*As described in point #1, last paragraph.*

1. **If you have any specific tools in mind please specify.**

Descriptive and diagnostic can be done using any tool as along as the GTL can get the access of the reports.

Students can use existing heatmap (hosted at <http://heatmap.greentransformationlab.com/>) to help them. Need to know MongoDB to access the data.