Twitter Analytics - London | Infrastructure Summary

SEPTEMBER 2016

Twitter Analytics- London| Infrastructure Summary

Introduction

Transitioning Towards Urban Resilience and Sustainability (TURAS) is an FP7 funded European-wide research and development programme. The "TURAS" project aims to bring together urban communities, researchers, local authorities and SMEs to research, develop, demonstrate and disseminate transition strategies and scenarios to enable European cities and their rural interfaces to build vitally-needed resilience in the face of significant sustainability challenges. As part of this process, the TURAS project has developed a suite of Geo-ICT tools for the project to demonstrate some of the research topic address over the lifespan of the project.

Design Intent

The TURAS London App was designed to be a geospatial analytics dashboard capable of illustrating the social media footprint of disseminated research from UELSRI. The dashboard aims to allow UELSRI assess the engagement, reach and exposure generated by their green infrastructure research, following its dissemination on twitter. In this way, UELSRI can see what research is garnering public attention and can engage with those who have elected to follow UELSRI's twitter for future releases.

Contact Details: James Sweeney, Future Analytics Consulting, Dublin, Ireland @ james.sweeney@futureanalytics.ie

Link to Existing Application: http://www.factest.ie/london/

Principal Elements

The dashboard itself consists of a number of interactive panels which facilitate the user's exploration and analysis of underlying data. There are three main panels, each with interactive elements:

1. A context/info & twitter presence panel
   Showing some brief context on the topic of Green Infrastructure, as well as the most recent tweets, account stats and live tweet facility.

2. A social media engagement panel
   Showing UELSRI's tweets, how they may have been retweeted, favourited and replied to by those on twitter. Also, it looks at those following UELSRI's account and how that community grows over time. An estimate of how many twitter accounts are indirectly exposed to the tweets is also given, as a measure of the potential reach of the research besides that of direct followers.

3. A 'geographic reach' panel
Showing a geographic map view of the location of those who have accessed the research via twitter (and have come to turas-cities.eu). Can be explored based on the particular research.

Tableau

The dashboard was created using Tableau, a proprietary business intelligence development kit (by Tableau) that has at its core the idea that data analysis and reporting should not be isolated activities but should be integrated into a single visual analysis process – one that lets users quickly tailor the information and shift views on the fly.

Tableau's own desktop application is the development and provisioning tool that is used to create the dashboard's multiple information panels and dynamic filtering elements. It also allows for online hosting of the completed dashboards via Tableau's own cloud infrastructure.

Infrastructure

The infrastructure underpinning the dashboard can be considered in three parts:

1. Data collection
Principally, twitter data is sourced by way of queries to the Twitter search API using R. The search API allows for snapshot retrieval of tweets extending back 5-7 days. Therefore, queries are made each week and data is added periodically to the dashboard. This information is supplemented by using Google Web Analytics linked those who are referred to turas-cities.eu by twitter (specifically tracked against the unique url tweeted with the disseminated research).

2. Data retention
This data is stored in structured .csv format until such time as it is integrated into the dashboard. The Tableau development environment accepts input from a wide range of database types, however, the csv files are simply collated (by date) and imported in one exercise. A 'Tableau Data Extract' (TDE) is then created, which is a process by which the data is restructured into a Tableau proprietary format allowing it to be integrated directly into the development files (necessary for online provisioning).

The TDE is a compressed columnar store that splits the data into various tables by dimension and measure. It is compressed using various techniques, such as dictionary compression, run length encoding etc. It is also a memory-mapped file enabling quicker I/O performance, especially as the file grows. We have found this to be very beneficial, as when data is requested from the TDE, it is essentially loaded into memory by the OS. Tableau itself doesn't have to open, process or decompress the TDE to start using it. The mapped columns are loaded first, and data can be read ahead based on the mapping.

3. Data display
The dashboard is segmented into distinct panels which have various interactive elements. Each element is essentially a dynamic filter that allows for multi-filtering across the entire panel simultaneously. Some elements are visualisations that fall under this filtering hierarchy and

allow tailored selections to be set. Each element is individual ordered and arranged to construct the three main panels.
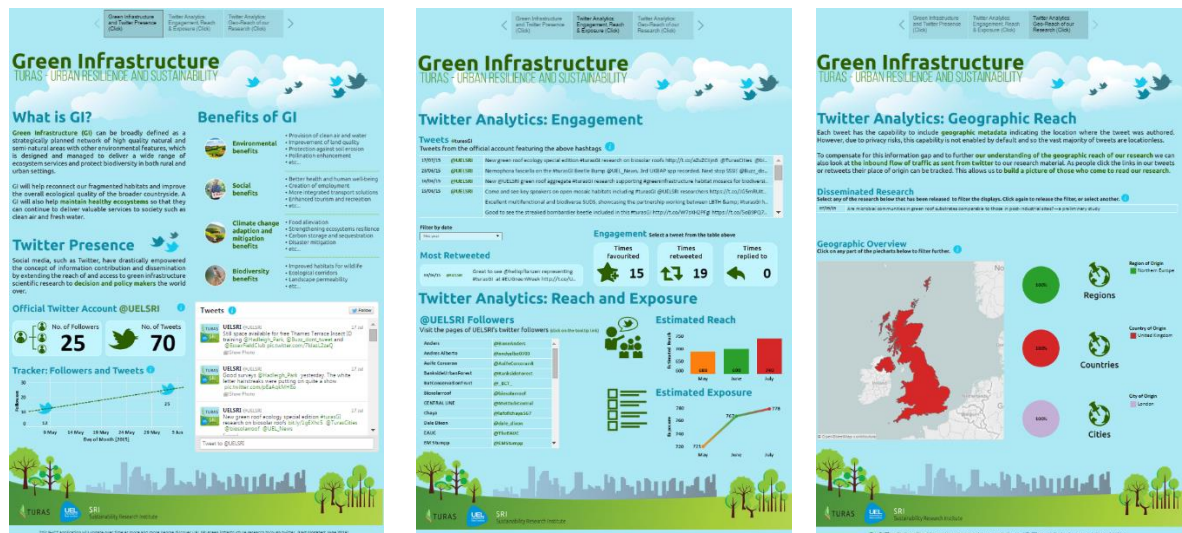
Wider Application Dissemination

In order to create the best environment for re-use, modification and visibility of the TURAS Geo-ICT tools we package all code and documentation for each application and have made them directly available to the public in zip file or available on GitHub @ https://github.com/UCDTURAS.

Compressed Archive File
- To make the development and coding accessible to the public and researchers interested in using, adapting, or further developing the TURAS tools, we have packaged the information (code, development operations and documentation) into a single compressed file which can be downloaded from the final TURAS interface. This package will contain a computer program as well as necessary metadata for its deployment.

GitHub Repository
- The aim of the TURAS project is to bring urban communities and businesses together with local authorities and researchers to collaborate on developing practical new solutions for more sustainable and resilient European cities. Following this, we recognise the importance of having a dedicated modern interface with which to disseminate all the Geo-ICT tools developed as part of the project. TURAS has created a GitHub account to allow end-users, technical developers etc to push/pull data code from the TURAS account.