

Meeting minutes

- Team SkyTrek

Date (dd.mm.yyyy)	14.01.2016
Attendee(s)	Jedaiah, Aseem, Huy
Absentee(s)	-

Contents

Item	Discussion point	Remark
1	Retrieving Data set	<ul style="list-style-type: none">- Aseem can pull articles from Google Analytics from 2013 to 2015- Huy will pull attributes of articles<ol style="list-style-type: none">1. Number of words (remove stopwords)2. Number of links3. Images4. Videos- Jedaiah then can pull number of share counts of each article- Huy then can consolidate all the attributes together and settle the final dataset
2	<ul style="list-style-type: none">- Google Trends Timeline- Google Calendar pulling of Public Holidays	<ul style="list-style-type: none">- Huy updated that Google Trends API can only pull data for the previous 30 days, not any arbitrary duration- Huy updated that pulling of Public Holidays from Google Calendar cannot be done on-demand. We will need to pull and save the holidays data in JSON format
3	Proposed Analysis	<ul style="list-style-type: none">- Multi linear regression for identifying significant attributes that determine an article performance- K means clustering for identifying content theme

Actions

Task	Assigned to	Status
Clarify with Prof Kam: Is there a way to dynamically assign new articles to the relevant clusters based on the trained model?	All	Created
Characteristics to scrape from article: <ol style="list-style-type: none"> 1. Number of words (remove stopwords) 2. Number of links 3. Images 4. Videos 5. Number of shares of article (by Jed) 	Huy	Ongoing
Retrieve URL share count	Jedaiah	Ongoing
Put topic modelling and google trends research and implementation into Gantt chart timeline	Jedaiah	Done
Input meeting with client in gantt chart for: <ol style="list-style-type: none"> 1. 25th Jan 2. 2nd Feb 3. 11th Feb 	Jedaiah	Done
Create flow chart documenting process how skyscanner team plans for the next quarter's content postings. To send it to client for verification	Aseem	Done
Test Google API public holiday pulling	Huy	Done
Create a mockup for the following visualization: <ul style="list-style-type: none"> o To plot Unique page views (count) vs Week of article publishing o Clicking this particular bar of interest (week has high count) would drill into the next bar chart (Number of unique page views, vs identifying URL) o Further clicking on the URL would reveal characteristics of the article (e.g. number of images, videos, etc, bounce/exit rate, etc) 	Jedaiah	Done
Create a mockup for the following visualization: <ul style="list-style-type: none"> o Identify key features of popular posts (MLR against unique page views) 	Huy	Done
Create a mockup for the following visualization: <ul style="list-style-type: none"> o Where do they come from (On the whole country news page level, where do the traffic come from) o How long do they spend at the page (average time on page) o Do they visit other pages? (bounce rate vs exit rate) o How effective are the recommended pages? (If GA does not track the unique page views, then this analysis is not going to be implemented) o [Just to prove a logical point] Correlation test between number of shares vs unique page views 	Jedaiah	Done
Deal with proposal segments: <ol style="list-style-type: none"> 1. Sponsor and Background Information 2. Motivation 3. Objectives 	Aseem	Done

4. Data		
5. Scope of Work		
Questions for Prof Kam: 1. Ask Kam about requirement to post sample date. Concern: data confidentiality 2. What is required for scope of work? (different from methodology)	All	Done

The minutes were prepared by: Jedaiah Tan and Nguyen Viet Huy