Meeting minutes
- Team SkyTrek

Date (dd.mm.yyyy)	29.01.2016
Attendee(s)	Jedaiah, Huy, Prof Kam
Absentee(s)	Aseem

Contents

Item	Discussion point	Remark
1	Project Focus	- We explained to Prof Kam that the focus of the project is identifying what attributes influent content view - Prof Kam recommended reading more literature research on how to analyze the data for the focus

2	- Data Analysis	- Example:
		Gather data for
		   - titles [classification]
		   - media (photo, video)
		   - main text [text mining]
		   - ancillary
		Build regression model to model page view against the designated attributes
		
		- Details:
		Title: come up with classification, manually classify each title
		
		Text mining: clustering
		• distance value of members within cluster to cluster central
		• detect core words how many time it appear
		• core words → go through and eliminate unimportant terms → combine terms into themes → rerun text mining and determine the cluster again → use the new distance value to plug into regression
		• get content expert to filter the list of terms to help elimination of unimportant terms
		
		- Prof Kam recommended look at page view distribution based on date.
		- Example: 90% of page views come from first 5 days after publish → cut off point of 5 days for gathering data

3	Midterm	- Focus on reporting progress of work until week 7 - Rubrics will be released in week 5 or 6
4	Questions for sponsor	- What is their main objective for this project? Is it to study what are the key content factors that drive high unique page views? - Do they have the data? For each URL, breakdown of page views across the days following the date of publishing - Who writes their articles (crowd sourced to freelancers or they have a few editors)? Can we sit down with them and find out what are the key phrases they tend to use when writing foe each of the 7 content themes?

Actions

Task	Assigned to	Status
Come up with list of possible title categories to clear with sponsor	All	Created
De-aggregate shares count for each URL	Jedaiah & Huy	Created
Build regression model based on current dataset	Aseem	Created
Clarify with Prof Kam: Is there a way to dynamically assign new articles to the relevant clusters based on the trained model?	All	Done

Characteristics to scrape from article:
  1. Number of words (remove stopwords)
  2. Number of links
  3. Images
  4. Videos
  5. Number of shares of article (by Jed)
	Huy	Done

Retrieve URL share count	Jedaiah	Done
Put topic modelling and google trends research and implementation into Gantt chart timeline	Jedaiah	Done

Input meeting with client in gantt chart for:
  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: o To plot Unique page views 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:	Huy	Done

o Identify key features of popular posts (MLR against unique page views)		
Create a mockup for the following visualization: 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:
  1. Sponsor and Background Information
  2. Motivation
  3. Objectives
  4. Data
5. Scope of Work
	Aseem	Done

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