SISTEM UNTUK PENGENALPASTIAN TREND BERDASARKAN ANALISIS TWITTER MENGGUNAKAN STUDIO R

HAZIMAH AB HALIM
LAILATUL QADRI ZAKARIA

Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia

ABSTRAK

Twitter adalah salah satu aplikasi media sosial yang semakin mendapat tempat di hati pengguna internet di seluruh dunia. Pengguna Twitter boleh menyebarkan sesuatu informasi dengan pantas melalui *'live update'* dan ia juga membenarkan penggunanya berkongsi maklumat dalam bentuk teks pendek 140 karakter. Kajian ini bertujuan mengenalpasti trend yang berlaku di sekitar kawasan UKM, Bangi. Seperti sedia maklum, maklumat yang disampaikan di Twitter, teks berbentuk pendek dan penggunaan bahasanya yang bercampur. Selain itu, sistem trend yang terdapat di Twitter hanya memfokuskan pada perkataan trend sahaja tanpa menyediakan maklumat tambahan berkaitan dengan trend tersebut. Secara umumnya, pengguna perlu membaca setiap tweet secara manual bagi mendapatkan maklumat yang dikehendaki dan ianya akan mengambil masa yang agak lama. Oleh itu, sebuah sistem analisis trend dibangunkan bagi menganalisis tweet secara automatik dan mengenalpasti trend berdasarkan analisis statistikal perkataan. Pengkaji menggunakan teknik frekuensi perkataan dan TF-IDF serta pengecaman entiti nama untuk mengenalpasti trend di sekitar UKM menggunakan perisian Studio R. Berdasarkan pengujian kajian, teknik TF-IDF tidak sesuai untuk digunakan pada teks yang pendek. Hasil menemukan frekuensi perkataan boleh digunakan untuk mengenalpasti trend, manakala pengecaman entiti nama boleh juga digunakan untuk mendapatkan maklumat tambahan mengenai trend atau data Twitter tersebut.

PENGENALAN

Media sosial merujuk kepada sebuah media untuk manusia berhubung antara satu sama lain yang dilakukan secara online tanpa mengira masa dan tempat dengan hanya memerlukan capaian internet. Ia meliputi dalam pelbagai bentuk aplikasi seperti Facebook, Twitter, Instagram dan juga blog. Pada era yang semakin berkembang teknologinya, kebanyakkan pengguna media sosial bukan sahaja menggunakan media sosial sebagai tempat meluahkan perasaan malah juga ia digunakan untuk tujuan perniagaan. Popularity laman media sosial dan mudahnya data didapati membuatkan platform ini dijadikan sebagai sumber utama untuk penyelidikan sosial (Verzani 2011).

Twitter adalah salah satu media sosial yang semakin mendapat tempat di hati pengguna internet di seluruh dunia. Ini kerana Twitter menawarkan fungsi yang berbeza berbanding dengan media sosial yang lain. Antaranya, Twitter boleh menyebarkan sesuatu informasi dengan lebih pantas atau dengan kata lain *'live update'* dan ia juga membenarkan penggunanya menghantar dan membaca teks dalam 140 karakter (Acar & Deguchi 2013).

Analisis trend boleh diertikan sebagai analisis yang digunakan untuk mengenalpasti corak dalam sesuatu informasi yang diperolehi secara menyeluruh. Analisi trend ini dilakukan berdasarkan maklumat yang dikumpul pada masa lepas bertujuan untuk mengetahui kecenderungan keadaan pada masa akan datang. Antara kepentingan analisis trend ini ialah salah satunya dapat mengkaji ramalan politik di sesuatu kawasan berdasarkan pilihan pengguna di media sosial. Analisis trend ini juga boleh digunakan untuk mengetahui isu-isu panas ataupun sensasi yang hangat diperkatakan di laman sosial.

PERMASALAHAN KAJIAN

Beberapa masalah sudah dikenalpasti. Seperti sedia maklum, maklumat yang disampaikan di Twitter, teks berbentuk pendek dan penggunaan bahasanya yang bercampur.

Jikalau sistem ini tidak dibangunkan, pengguna perlu membaca setiap *tweet* yang dikumpul secara manual ataupun satu persatu. Sistem yang boleh menganalisis *tweet* secara automatik diperlukan untuk menganalisis maklumat yang dikongsi oleh pengguna Twitter di sekitar kawasan UKM untuk mengetahui isu-isu yang sedang trend. Selain itu, sistem trend yang terdapat di Twitter hanya memfokuskan pada perkataan trend sahaja. Tiada maklumat tambahan yang berkaitan dengan trend tersebut.

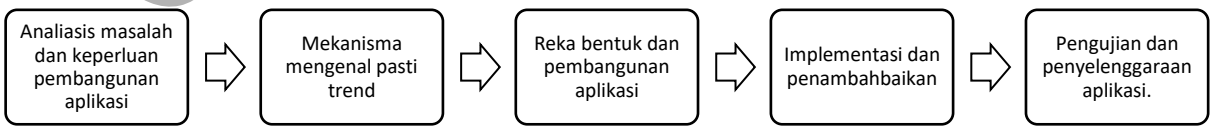
Tambahan pula, pengkaji juga menghadapi masalah dalam melakukan analisis trend itu sendiri. Dalam pelaksanaan analisis trend ini, ada masalah dari sudut pemprosesan bahasa tabii di mana pengkaji perlu menggunakan input data dari *tweets* yang dikumpul. Seperti yang diketahui, pelbagai gaya ataupun cara penulisan yang wujud dalam *tweets* yang dikumpul seperti perkataan singkatan, emotikon. Selain itu, untuk proses pemprosesan ayat di mana pengkaji perlu mengenalpasti kata kunci ataupun perkataan penting yang boleh dijustifikasi ianya sebagai trend ataupun tidak. Manakala masalah dari sudut teknik yang akan digunakan, di mana dalam *opennNLP* banyak teknik yang diperkenalkan seperti frekuenci perkataan, TF-IDF dan juga pengecaman entiti nama. Pengkaji perlu mengkaji teknik-teknik tersebut boleh digunakan untuk mengenalpasti perkara yang sedang trend di sekitar UKM. Di sini, boleh dikatakan agak kompleks untuk melaksanakan sesuatu analisis trend.

OBJEKTIF KAJIAN

Matlamat utama kajian adalah untuk mengenalpasti trend terkini berdasarkan analisis Twitter di kawasan UKM, Bangi dengan mengetahui isu-isu tersebut dapat diberi perhatian oleh sesuatu pihak. Bagi mencapai matlamat kajian ini, pengkaji telah meletakkan beberapa objektif yang perlu dicapai iaitu membangunkan sistem analisis trend menggunakan pemprosesan bahasa tabii dan menguji keberkesanan sistem analisis trend yang dikenal pasti di UKM.

METOD KAJIAN

Metodologi yang digunakan dalam membangunkan sistem pengenalpastian analisis trend ialah Model Air Terjun (Waterfall Model) tetapi ia diolah bersesuaian dengan tajuk kajian. Model ini mempunyai lima fasa yang penting dalam membangunkan sistem ini iaitu analisis masalah dan keperluan pembangunan sistem, mekanisma mengenal pasti trend, reka bentuk dan pembangunan sistem, implementasi dan penambahbaikan dan akhir sekali pengujian dan penyelenggaraan sistem.



Rajah 1: Proses Model Air Terjun

Fasa Analisis Masalah Dan Keperluan Pembangunan Sistem

Fasa analisis masalah dan keperluan pembangunan sistem adalah merupakan fasa pertama dalam membangunkan sistem pengenalpasian trend di Twitter. Dalam fasa ini pemilihan

tajuk telah dijalankan. Setelah tajuk yang dipilih mendapat kelulusan daripada penyelia projek, kajian untuk mengenalpasti permasalahan beserta permintaan terhadap sistem pula akan dijalankan. Selain itu, kajian turut dijalankan untuk mendapatkan latar belakang, objektif yang perlu dicapai, kekangan, skop kajian dan rangka penyelesaian.

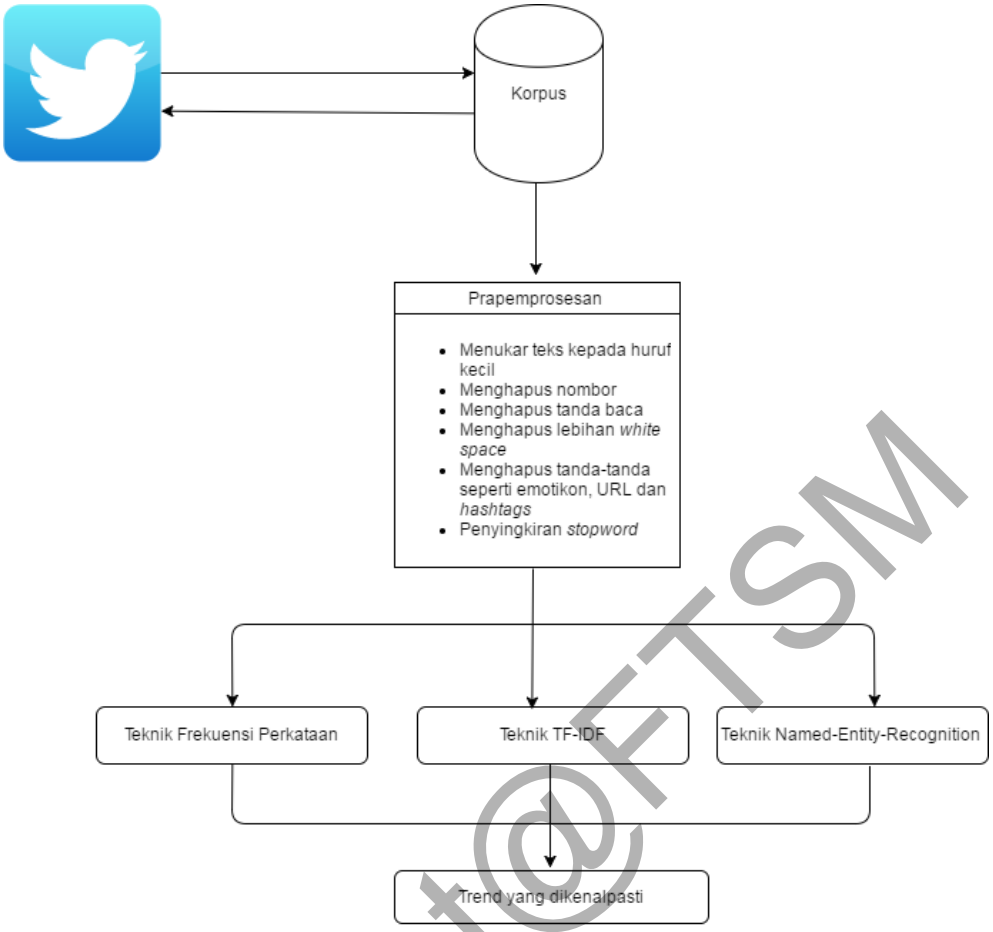
Fasa Mekanisma Mengenal Pasti Trend

Dalam fasa mekanisma mengenal pasti trend, proses ini adalah untuk mengenal pasti bagaimana cara ataupun prosedur untuk mendapatkan trend yang terkini dengan menggunakan studio R. Dalam penjanaan korpus, pengkaji mengumpul data daripada Twitter dengan bantuan Twitter API. Pengkaji telah menggunakan kedua-dua REST API dan Streaming API untuk mendapatkan pelbagai jenis koleksi data.

Carian Twitter API adalah salah satu daripada tiga API iaitu carian, penstriman dan "*firehose".* Carian API ini membolehkan akses kepada subset tweets popular ataupun tweet yang terkini (dalam masa empat hingga enam hari yang lepas). Carian ini membenarkan tweet pertanyaan yang lepas walaupun hanya sebahagian kecil *tweets* yang ketara.

Ini adalah cara yang terbaik untuk mengumpul dan membersihkan dataset *tweets*. Walau bagaimanapun, ia tidak memberi apa-apa utiliti yang sebenar untuk tujuan penyelidikan. Sebahagian kecil daripada *tweets* yang diterima mungkin tidak dapat menggambarkan keseluruhan aliran *tweets*. Dalam proses mengenal pasti konsep yang penting, *tweets* yang dikumpul mungkin akan mengandungi ataupun tidak mengandungi pandangan peribadi pengguna.

Oleh kerana itu, orientasi pengguna telah ditetapkan oleh pengkaji. Ini adalah penting untuk membangunkan pengelasan berdasarkan kandungan ini. Antara teknik yang boleh digunakan ialah penggunaan frekuensi perkataan dan TF-IDF serta pengecaman entiti nama. *Tweets* boleh mempunyai pelbagai bahagian seperti teks, URL, *hashtags, user mention* dan lain-lain.



Rajah 2: Rangakerja Analisis Trend

Fasa Reka Bentuk Dan Pembangunan Sistem

Fasa rekabentuk merupakan rancangan mengenai sistem yang akan dibangunkan berdasarkan hasil analisis. Maklumat-maklumat yang dikumpul dari pelbagai sumber akan dianalisis untuk menentukan bagaimana sistem ini beroperasi. Dalam kajian pengenalpastian trend berdasarkan analisis Twitter, pengkaji meggunakan Studio R. Dalam Studio R ini, pengkaji telah menggunakan pakej *twitteR, ROAuth, plyr, stringr, ggplot2, RColorBrewer, openNLP, wordcloud* dan juga *tm* iaitu rangka untuk perlombongan teks apikasi di Studio R.

Fasa Implementasi Dan Penambahbaikan

Fasa implementasi dan penambahbaikan ini merupakan fasa untuk melaksanakan system yang dibangunkan. Ini adalah bertujuan untuk memastikan tiada masalah yang akan berlaku ketika sistem ini dijalankan dan dapat mencapai objektif. Setiap pandangan mahupun komen akan dititikberatkan bagi penambahbaikan sistem.

Perisian dan perkakasan adalah keperluan yang utama dalam membangunkan sesebuah sistem. Beberapa perisian dan perkakasan yang telah dikenal pasti untuk digunakan dalam proses pengenalpastian trend berdasarkan analisis Twitter. Pembangunan projek ini menggunakan perisian R Studio. R Studio digunakan untuk sebagai program antara muka yang memenuhi keperluan kajian. Selain itu, R Studio digunakan kerana program ini mudah untuk dikemudikan dan mudah untuk difahami. R Studio adalah sumber bebas, percuma dan ia bergerak dengan baik pada Windows. Senarai spesifikasi keperluan perkakasan yang dicadang untuk mengenalpasti trend berdasarkan analisis Twitter adalah seperti dalam jadual.

PERANTI	**ASUS K460 SERIES**
Jenama	ASUS
Sistem Operasi (OS)	Windows 7 (64 bit)
PROCESSOR	
Jenis Processor	Intel Core i5
Kod Processor	3317U
Frekuensi Base Processor	1.70 GHz
MEMORI	
Saiz Memori Capain Rawak (RAM)	6.00GB
SKRIN	
Resolusi Skrin	14”
Keluasan Skrin	1366 x 768

HASIL

Latihan sistem terhadap data Twitter yang dikumpul dijalankan dengan menggunakan perisian Studio R bagi mengenalpasti sesuatu topik itu boleh diklasifikasikan sebagai tepat ataupun tidak secara tidak langsung juga membantu pengkaji menghasilkan satu analisis trend di UKM. Perisian Studio R juga membantu dalam menunjukkan grafik terhadap analisis yang dijalankan. Beberapa modul yang dilakukan untuk mendapatkan analisis yang terbaik. Terdapat tiga modul utama iaitu modul pengumpulan data, modul pembersihan data dan juga modul pengesanan trend.

Dalam modul pengumpulan data, pengkaji perlu mencipta satu permohonan di Twitter untuk melakukan analisis Twitter. Ini membolehkan analisis Twitter dijalankan dengan menghubungkan konsol R dengan Twitter menggunakan Twitter API. Pengkaji perlu memasukkan nama dan menghuraikan aplikasi yang dimohon.

Bagi proses di Studio R, beberapa pakej dan *library* perlu dimuat turun. Kemudian, pengkaji perlu mengakses Twitter API dengan menggunakan *Consumer Key* dan *Consumer Secret* yang telah diberi. Pengkaji telah mengumpul data *tweets* yang berbahasa Inggeris sebanyak 13346 tweets dari tarikh 25 April 2017 sehingga 10 Mei 2017. Data yang di kumpul daripada Twitter dengan bantuan Twitter API adalah bergantung kepada jumlah *tweets* yang diterima pada sesuatu masa.

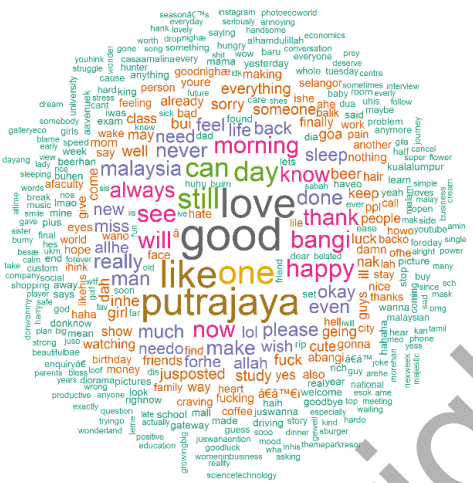
Pengkaji menetapkan lokasi kajian pada titik kedudukan utamanya iaitu kawasan Dewan Canselori Tun Abdul Razak ataupun dikenali sebagai DECTAR sekitar 2 radius. Selepas diberi kelulusan daripada Twitter, pengkaji boleh mendapatkan data *tweets* kemudian data tersebut disimpan dalam fail berformat CSV. Apabila data *tweets* berjaya disimpan, pengkaji perlu menggunakan beberapa fungsi untuk menukarkan data-data tersebut menjadi maklumat yang berguna. Ia dipanggil sebagai proses pembersihan data. Daripada proses tersebut, pengkaji boleh mendapatkan perkataan-perkataan yang boleh digunakan untuk dianalisis.

Untuk modul pembersihan data, teks dalam *tweets* berbeza dengan teks dalam artikel, buku ataupun dalam bahasa pertuturan. Ia termasuk juga teks yang mempunyai emotikon, *Universal Resouce Locator* (URL), RT untuk *retweet, @* untuk *user mention, #* untuk *hashtags* dan juga pengulangan. Jadi adalah perlu untuk melakukan pembersihan data. Terdapat banyak alat NLP yang ada tetapi tidak semua yang sesuai untuk membuat analisis trend di Twitter. Teks di Twitter adalah pendek dan penggunaan bahasanya juga yang tidak rasmi.

Pengkaji memulakan pembersihan data dengan mengambil teks sahaja daripada data *tweet* kepada perkataan. Pengkaji menapis data *tweets* dengan beberapa cara seperti menukarkan teks kepada berhuruf kecil, menghapuskan nombor, menghapuskan tanda baca, menghapuskan lebihan *white space* dan juga membuang tanda-tanda seperti *@-mentions*, emotikon, URL dan *hashtags*. Kemudian disinilah penggunaan *stopword* digunakan.

Stopword ialah perkataan biasa yang banyak digunakan dalam sesuatu ayat dan tidak memberi sebarang maklumat seperti *a, an. be, the, of* dan lain-lain. Pengkaji telah menambah *stopword* yang telah sedia ada dengan *stopword* berbahasa Malaysia. *Stopword* berfungsi sebagai salah satu cara untuk mencantikkan data yang akan dikaji oleh pengkaji. Ini kerana bagi mendapatkan hasil analisis yang terbaik, pengkaji memerlukan data yang berkualiti.

Perkataan yang kerap kali diulang dan tidak memberi sebarang maksud telah dibuang bagi mendapatkan hasil yang cantik dan bermaklumat. Dengan membuang *stopword*, ia akan mengurangkan saiz korpus tanpa menghilangkan maklumat yang penting. Perbezaan pada *wordcloud* sebelum dan selepas menggunakan kaedah *stopword* merujuk pada rajah di bawah.



Rajah 3a: *WordCloud* sebelum analisis kata henti

Rajah 3b: WordCloud selepas analisis kata henti

Manakala untuk modul pengesanan trend, dengan koleksi data yang dikumpul oleh pengkaji dari tarikh 25 April 2017 sehingga 10 Mei 2017, pengkaji berjaya mengumpul sebanyak 13346 *tweets* yang berada di sekitar UKM. Dalam modul pengesanan trend ini, terdapat 3 teknik yang digunakan oleh pengkaji untuk mengetahui trend di kawasan sekitar UKM iaitu frekuensi perkataan, TF-IDF dan pengecaman entiti nama.

Hasil analisis berdasarkan teknik frekuensi perkataan, pengkaji telah menggunakan kaedah frekuensi perkataan kerana dengan mengetahui perkataan yang tinggi bilangannya, ia mungkin memberi maksud ataupun makna yang penting. Dengan itu, ia memberi satu gambaran kepada pengkaji untuk mengetahui perkara apakah yang sering dikongsi di kawasan kajian.

Untuk analisi menggunakan kaedah ini, terdapat 20 perkataan yang mencatatkan frekuensi tertinggi telah diambil kira pada setiap hari bermula pada hari bulan 25 April 2017 sehingga 10 Mei 2017. Daripada 20 perkataan tersebut, pengkaji mendapati terdapat 5 perkataan yang berkaitan dengan perasaan ataupun emosi seperti *love, good, please, happy, feel*. Senarai frekuensi perkataan turut mengandungi satu perkataan lokasi iatu Putrajaya.

Pengguna Twitter juga berkongsi maklumat mengenai ucapan-ucapan sapaan seperti *morning*, Perkataan *birthday* turut dikenalpasti sebagai salah satu frekuensi perkataan yang

direkod tinggi. Ini menunjukkan pengguna Twitter mengucapkan ucapan hari jadi kepada orang yang berkenaan. Jadual di bawah menunjukkan perkataan yang paling tinggi jumlah bilangannya adalah btsbbmas.

Jenis Perkataan	Bilangan Perkataan
Love	421
People	238
Good	255
Never	183
Please	62
Life	163
Happy	252
Btsbbms	451
Putrajaya	73
Feel	222
Source	201
Know	156
Votefor	409
Done	104
Really	211
New	166
Shop	140
Morning	196
birthday	90

Maklumat yang diperolehi berkaitan dengan perasaan ataupun emosi mempunyai data yang stabil. ia tiada peningkatan ataupun penurunan secara mendadak.Perkataan-perkataan tersebut lazimnya hampir setiap hari diucap menyebabkan tiada yang trend untuk dikenalpasti.

Namun, ada dua perkataan yang menarik perhatian pengkaji iatu btsbbma dan *votefor*. Perkataan-perkataan tersebut telah menunjukkan peningkatan yang selari dan meningkat daripada 25 April 2017 sehingga 10 Mei 2017. Ini adalah kerana btsbbmas ini merujuk kepada perkataan gabungan antara nama ahli kumpulan popular di Korea iaitu Bangtan Boys atau dikenali BTS dengan satu syarikat penerbitan majalah yang popular di Amerika Syarikat yang dikenali sebagai Billboard.

Majalah Billboard adalah satu jenama media hiburan Amerika yang dimiliki oleh Hollywood Reporter-Billboard Media Group, sebahagian dari Eldridge Industri. BBMA itu merupakan singkatan bagi *BillBoard Music Award*. Pengguna Twitter menggunakan Twitter untuk mengundi kumpulan tersebut dalam Billboard Music Award 2017 yang akan berlangsung pada 21 May 2017. Diangggarkan data ini akan terus meningkat sehingga pada tarikh tersebut.

Bagi hasil analisis berdasarkan teknik TF-IDF, hasil analisis yang dilakukan menggunakan teknik ini tidak dapat mengeluarkan keputusan yang baik merujuk rajah 4.12. Pengkaji mendapati Teknik TF-IDF tidak sesuai digunakan untuk teks yang pendek untuk mengenalpasti trend dengan menggunakan data *tweets* yang dikumpul. Oleh itu, trend tidak dapat dikenalpasti di kawasan sekitar UKM.

Hasil analisis berdasarkan teknik *pengecaman entiti nama,* pengkaji telah menggunakan teknik ini bagi menguatkan lagi matlumat untuk mengenalpasti trend di sekitar kawasan kajian. Untuk teknik pengecaman entiti nama, terdapat 2 entiti yang telah dikenalpasti iaitu lokasi dan organisasi. Untuk entity lokasi ,pengguna banyak berkongsi

maklumat tentang negara seperti Malaysia, Indonesia, Manchester, China, Real Madrid, Cambodia, India.

Apabila diperhalusi, perhubungan antara lokasi dengan data yang pengkaji ada, didapati pengguna Twitter berkongsi maklumat mengenai isu-isu semasa seperti lokasi Manchester. Lokasi tersebut muncul pada tarikh pada 27 April 2017. Pada masa ini, terdapat perlawanan bola sepak antara Manchester United dengan Manchester City yang mencatatkan keputusan seri (Paul Doyle, 2017). Kawasan sekitar pusat kajian turut dapat dikesan seperti Putrajaya, Bangi dan Kajang. Lokasi seperti Korea dikenalpasti pada tarikh 25 April 2017 kerana pada tarikh tersebut Running Man Korea telah datang ke Malaysia untuk berjumpa dengan peminat-peminat

Manakala untuk entiti organisasi, ia berjaya mengenalpasti organisasi yang berada di sekitar kawasan kajian dengan menggunakan maklumat yang dikaji seperti German Malaysia Institut, KPM Beranang, IOI City Mall, SMK Bandar Baru, UPM, Villa Seafood Restaurant, Gateway Shopping Complex, Bates Motel, Alamanda Shopping Complex dan juga UKM. Berdasarkan pemerhatian pengkaji, *name-entity-recognition* tidak dapat mengenalpasti kesemua organisasi berikutan kekangan pakej openNLP yang digunakan oleh pengkaji.

KESIMPULAN

Sistem pengenalpastian trend berdasarkan analisis Twitter menggunakan perisian Studio R telah berjaya mengenal pasti trend maklumat terkini yang berada disekitar kawasan UKM. Kajian ini telah mendapati teknik TFIDF tidak sesuai digunakan untuk teks pendek seperti teks yang terdapat dalam tweets. Teknik TF dapat mengenal pasti trend dengan mendapatkan senarai perkataan yang mempunyai frekuensi yang tinggi manakala teknik pengecaman entiti nama telah dapat membantu menambah maklumat baharu yang berkaitan dengan trend tersebut.

RUJUKAN

11 Mac 2011 - Bencana Yang Meragut Nyawa Lebih 10,000 Orang. (n.d.). http://pmr.penerangan.gov.my/index.php/antara/9006-11-mac-2011-bencana-yang-meragut-lebih-10000-orang.html

Acar, A. & Deguchi, A. 2013. Culture and social media usage: Analysis of japanese twitter users. *International Journal of Electronic Commerce Studies*, *4*(1), 21–32.

Bhola, A. 2014. Twitter and Polls: Analyzing and estimating political orientation of Twitter users in India General #Elections2014. *arXiv preprint arXiv:1406.5059*,.

Bontcheva, K., Gorrell, G. & Wessels, B. 2013. Social Media and Information Overload: Survey Results. *arXiv preprint arXiv:1306.0813*, 1–31.

Gayo-Avello, D. 2012. No, you cannot predict elections with twitter. *IEEE Internet Computing*, *16*(6), 91–94.

Jacoby, J., Speller, D. E. & Kohn, C. A. 1974. Brand Choice Behavior as a Function of Information Load: Replication and Extension. *Journal of Consumer Research*, *1*(1), 33–42. doi:10.2307/3150994

Kumar, S., Morstatter, F. & Liu, H. 2013. Twitter Data Analytics. *Springer*, 89.

Milstein, S., Chowdhury, A., Hochmuth, G., Lorica, B., Magoulas, R. & Reilly, T. O. 2008. Twitter and the Micro-Messaging Revolution: Communication, Connections, and Immediacy-140 Characters at a time. *Business*, (November), 1–49.

Miyabe, M., Miura, A. & Aramaki, E. 2012. Use trend analysis of twitter after the great east japan earthquake. *... of the ACM 2012 conference on ...*, 175–178.

Peta. (n.d.). http://www.ukm.my/sebumi3/peta.html

Signorini, A., Segre, A. M. & Polgreen, P. M. 2011. The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U . S . during the Influenza A H1N1 Pandemic *6*(5).

doi:10.1371/journal.pone.0019467

Skoric, M., Poor, N., Achananuparp, P., Lim, E. P. & Jiang, J. 2011. Tweets and votes: A study of the 2011 Singapore General Election. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2583–2591.

Tumasjan, A., Sprenger, T., Sandner, P. & Welpe, I. 2010. Predicting elections with Twitter: What 140 characters reveal about political sentiment. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 178–185.

Wang, H., Can, D., Kazemzadeh, A., Bar, F. & Narayanan, S. 2012. A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle. *Jeju, Republic of Korea*, 115–120.