

Hot Topics Detection by Using 2-Layers Keywords Extraction

I-Hsien Ting
Department of Information
Management
National University of
Kaohsiung, Kaohsiung, Taiwan
iting@nuk.edu.tw

Su-Chen Yang
Department of Information
Management
National University of
Kaohsiung, Kaohsiung, Taiwan
henryyang@nuk.edu.tw

Chia-Sung Yen
Department of Cultural &
Creative Industries
National Chin-Yi University of
Technology, Taichung, Taiwan
jsn1003@gmail.com

Tsung-Hsing Tsai
Department of Information
Management
National University of
Kaohsiung, Kaohsiung, Taiwan
hotgm99@gmail.com

Abstract—Hot topics analysis is one of the important task for users to organize information from WWW and especially for social networking websites. Therefore, how to design an efficient approach for users to extract those hot topics is very essential. Thus, we proposed a so-called 2-layers keywords extraction approach and three empirical analyses are then be applied to validate the usability and performance of the proposed approach.

Keywords—Keywords extraction, hot topics, social network

I. INTRODUCTION

With the rapid growth of contents of World Wide Web, it is an information exploration era and we all facing information overloading problem. It is not an easy task for us to find useful information and therefore an efficient approach for extracting important information is essential. Search engine is a powerful tool for us to search information in the Internet, however, it is not an efficient tool for us to organization the information well.

Recently, there are many services provide services for users to extract and organize information. The major technique is keyword extraction to extract keyword from the content of internet, it is also called text mining, opinion mining. Some research also devoting on understand the positive and negative sentiment of the keyword, which is called sentimental analysis. These tools and techniques are very helpful for us to extract and analyze information automatically.

In this paper, we intent to develop a keyword extraction and analysis approach to detect hot topics in the world wide web based on the techniques that mentioned above. The data sources are from Facebook, discussion forum or bulletin board system. The approach is so called 2-layers keywords extraction, the first layer to selection related posts based on keyword and ontology of the specific hot topic and the second layer are mainly focusing on extract keywords from the posts as well as measure the frequency of the keyword.

The organization of the rest part of the paper is as follow, in section 2 we proposed the approach of the 2-layers keywords extraction and explain the technical details. We then show three interesting hot topics analysis results based on the approach in section 3. In section 4, the paper is concluded and future research suggestions will be provided.

IEEE/ACM ASONAM 2020, December 7-10, 2020
978-1-7281-1056-1/20/\$31.00 © 2020 IEEE

II. THE 2-LAYERS KEYWORDS EXTRACTION APPROACH

In order to achieve the objectives that discussed above, we therefore developed a so called 2-layers keywords extraction approach. In Figure 1, it shows the mentioned process.

In the first step of the approach, we collected data from three difference sources, which are the most popular discussion social networking websites and discussion forums. Including Facebook (<https://www.facebook.com>), PTT (<https://term.ptt.cc/>) and Dcard (www.dcard.tw). The data collection period is from 1/1/2020 to 1/10/2020 which cover the period of the designed analyses.

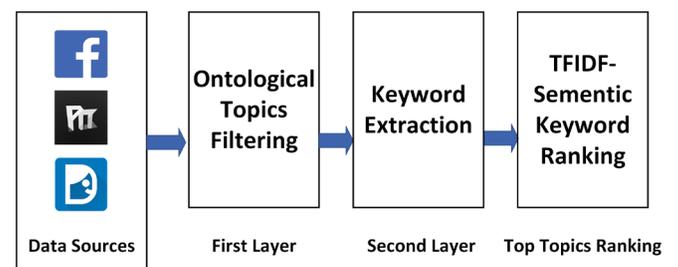


Fig. 1. The 2-layers Keywords Extraction Approach

In the second step of the approach, we then use ontological approach for topics filtering, to filter out related topics from the collected data. In the third step of the approach, we then extract all the keywords from the filtered topics, common natural language processing techniques are used, such as keyword segmentation, etc. In the last step of the approach, we use semantic TF-IDF techniques to rank the frequency and importance of the keyword. After the implementation of the approach, we can then get the ranking of the keywords which can be treated as the top-N hot topics.

III. ANALYSIS RESULTS OF HOT TOPICS ANALYSIS

In this section, we use the proposed 2-layers keyword extraction approach to analyze three different topics to test the approach. The first topics is travel preference during COVID-19 pandemic period and the second topic is about the hot discussion

topics in Chinese Valentine’s day(7 July of Lunar Calendar). The third topic is about the most popular cheat dating places.

Table 1 is the analysis result of the top 10 travel preference city during COVID-19 period, so that we collected data from March of 2020 to October 2020 and sources are from Ptt (the most popular BBS in Taiwan), Mobile 01 (The most discussion forum in Taiwan) and Facebook (public fans page only). The 2-layers keywords extraction approach is then used to analyze the collected data. From the results in table 1, the top 10 cities are Taipei city, New Taipei city, Yi-Lan City, Peng Hu County, Taina City, Kaohsiung City, Hsin-Chu City, Ma-Zuo County, Nan-Tou County and Hua-Lian County.

TABLE I. TOP 10 TRAVEL PREFERENCE DURING COVID-19 PERIOD

No.	Place
1	Taipei City
2	New Taipei City
3	Yi-Lan City
4	Peng-Hu County
5	Tainan City
6	Kaohsiung City
7	Hsin-Chu City
8	Ma-Zuo County
9	Nan-Tou County
1	Hua-Lian County

Table 2 is the top 10 discussion topics in Chinese Valentine’s day, so that we collected data also from the three platforms that mentioned above from two weeks before 7 July, 2020 (Lunar Calendar). The top 10 results are A-A Dating Model, Interaction between Lovers, Sex, Marriage, Income, Outlook/Dressing, Cross-culture Romance, Unfaithful, Confess and Homosexual issues.

TABLE II. TOP 10 DISCUSSION TOPICS IN CHINESE VALENTINE’S DAY

No.	Place
1	A-A Dating Model
2	Interaction Between Lovers
3	Sex
4	Marriage
5	Income
6	Outlook/Dressing
7	Cross-Culture Romance
8	Unfaithful

No.	Place
9	Confess
1	Homosexual

In table 3, the analysis is about the top 10 discussed cheat dating places. The reason to do this analysis is due to a very famous politician cheating news (the place is in the office of city hall) and it is widely discussed national wide. Therefore, we selected data one month after the exact dates of the event happens. The result is shown in table 3, which are, Motel, Office, Hospital, Home, Restroom, Car, Park, MRT, Bar and public area.

TABLE III. TOP 10 CHEAT DATING PLACES

No.	Place
1	Motel
2	Office
3	Hospital
4	Home
5	Restroom
6	Car
7	Park
8	MRT
9	Bar
1	Public

IV. CONCLUSION

In this paper, we proposed a so called 2-layers hot topics extraction approach. The approach is then empirically validated by applying three different analyses. From the empirical studies, it shows the hot topics can be easily extracted when applying the proposed approach. In the future, we expect to develop a system based on this approach for users to identify keywords and extract keywords easier and more accurate.

REFERENCES

- [1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955. (*references*)
- [2] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
- [4] K. Elissa, “Title of paper if known,” unpublished.
- [5] R. Nicole, “Title of paper with only first word capitalized,” *J. Name Stand. Abbrev.*, in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetism Japan, p. 301, 1982].

M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.