An Algorithm for the Disaster Information Extraction using English Translation API and SNS

SungJe Kim', Hyun Suk Hwang[2], Chang Soo Kim[3]

'Pukyong Unversity, Interdisciplinary Program of Information Security,
[2]Pukyong National University, Interdisciplinary Program of Information Systems,
[3]Pukyong National University, Dept. Of IT Covergence and Application
Engineering, 45, Yongso-ro, Nam-Gu, 608-737, Busan, South Korea
{comgosul, hhs[2], cskim[3]} @pknu.ac.kr

Abstract. Information from SNS posts is a rapid media for sharing disaster information. However, it takes a long time to find desired information in huge amounts of SNS posts. Specifically, it is important to receive related information rapidly when searching for disasters. Some research has progressed to extract disaster information using the analytical method of related keywords. Some search results can be irrelevant because Korean words are combined of more than one morpheme. Therefore, this paper proposes an algorithm to remove the unrelated posts by translating search words containing a keyword into English and comparing them with a translated keyword.

Keywords: Social Network Services, Twitter, Search Engines, Disaster Information Extraction

1   Introduction

Social Network Services have been the most attractive services for people to share information since 2009, such as their location, situations, interests or real-life stories. For example, in Seoul in July of 2011, there was damage caused by torrential rains which occurred simultaneously in different areas of the city. Specifically, some Twitter users shared photos or videos as well as text messages about the urgent situation which caused damage in Gangnam-gu Seoul, where roads were flooded. At that time, the SNS was one of the quickest media. However, when performing a general search, it takes a long time to find desired information in large amounts of SNS posts.

The search engines have been researched based on social network structures [2][3][4] and some common interests[5][6]. In the case of disasters, it is the most important to rapidly attain accurate information from a great deal of the posts. Seo [1] proposed a model to extract related words disasters by categorizing posts into disaster codes. Some results of the searching can be irrelevant because Korean words are combined of more than one morpheme [7].

---

[3] Corresponding Author

In filtering search results, the posts with unrelated disaster words can be deleted by defining exception rules. However, this method needs to periodically survey lots of exception rules because Twitter users may express them as different words. Therefore, in this paper, we propose an algorithm to remove unrelated posts by translating search words containing a specific keyword into English and comparing them with the translated keyword.

2 SNS Search Engines

In general, a social search is defined as a method to support an advanced search by using social information, such as social networks, data mining and collaborated intelligence. The social searches can be categorized into two types: common interests [2][3][4] and users social network structures [5][6]. The PeopleRank [3] and SaND [4] method defined the level of closeness using the PageRank algorithm on Google. The Aardvark method [2] identified the level of closeness using a routing engine for calculating the distance between users.

The FolkRank [5], SNDocRank [6], and Social SimRank [7] methods are search engines to find web pages or people who have similar interests based on words of common interests. Seo [1] proposed a model to extract disaster-related words by analyzing SNS posts. This model can provide better search results than methods by general categorized codes of disasters because Twitter users usually use simple and conversational words. The semantic matching algorithms are utilized to study keyword searches considering similarity by means of word matching, abbreviation expansion, synonym and association between words [10].

3 An Extraction Algorithm

We propose an algorithm model to extract the disaster information with different meanings from SNS posts. Some compound words for disasters are shown in Table 1. Some search results can be irrelevant because Korean words are combined of more than one morpheme.

Table 1. Compound words containing the disaster keyword

Disaster keyword	Compound word
II ^{2}r$^-$ (Flooding) Il=r1- (The rainy season)	°} p^1 –"F-"d (Morning classes), $^\circ$Ili ----,- °O (Morning swim) Y_IP1$^-$4(Stalls), 14 ,$^{-1-1\cdot1}$1^{-0}1 (Heart failure)
all °(Tsunami)	1^{7}2_4 V (Johaeil), ^{1}44 V (Park Haeil)
^{1}1,$^{1\cdot}$4 (Disaster Prevention)	1.$^-$1$^-$1 el TA q(Foundation for Youth Violence Prevention), $^{11}_0$$^{1\cdot}$4 1 (Fiscal)
AN tfi (Disaster)	AN till (Reinterpretation), AI Al Al ot(To be published)

First, this algorithm uses the search method of our previous work [1] based on categorized codes of disasters. Next, the algorithm filters unrelated posts by translating the words containing the keyword into English and comparing them with a translated keyword. The Google translating API will be used to filter the composite morphemes, which are different in meaning with the keyword.

The code below shows an algorithm to filter unrelated posts from searched posts. First of all, the SNS posts will be displayed by executing the Search(keyword) function with an input keyword based on the categorized disasters. Next, the input keyword and words containing the keyword will be stored into extraction storage for each post. Some posts, which do not have the same meaning with the keyword, will be deleted after both words are translated into English by the Trans(words) function. Finally, filtered results will be displayed by the Display(rePost) function.

```
Algorithm ExtractionMethod(Input keyword, Output rePost)
    i   0, j   0, k   0, 1   0, m   0
    Post[i,j]    0
    Extract[k]    0
    Postᵣ, Search(keyword) : Search results of n
    DO WHILE i < Post.max size
       DO WHILE j < Post[i].max size
          If contain(keyword, Post[i,j]) Then
             Extract[l]   Post[i,j]
             1   1 + 1
          End If
       EndDo j
       DO WHILE k < Extract.max size
          If Trans (keyword) = Trans (Extract [k] ) Then
             rePost[m]    Insert(Post[i])
             m m + 1
             Exit
          Endlf
       EndDo k
    EndDo i
    Display (rePost)
  END.
```

4 Conclusion

SNS users want to share their information and rapidly extract related information. Social engines have actively been researched to effectively filter desired information from amounts of posts. Specifically, in the case of disasters, it is important to share real-time information via SNS in order to prevent and respond to disasters.

In this paper, we propose an algorithm to filter disaster information from SNS posts. This model executes a search based on categorized codes of disasters. Unrelated posts are filtered by translating the words containing a keyword into English and comparing them with a translated keyword. We are going to implement the proposed algorithm after collecting and categorizing the SNS posts with regard to disasters. The algorithm is required to verify accuracy and effectiveness of search results as well.

References

1. Tae Woong Seo : A Study of Real-Time Disaster Information Extraction and Display using the Mash-up based on SNS — using the Twitter API and Google map API -: Master's Thesis, Pukyong National University, Busan, Korea (2012)
2. Damon Horowitz, SepandarD.Kamvar :The Anatomy of a LargeScale Social Search Engine : International World Wide Web Conference, pp.431--440 (2010)
3. B Carterette, R Jones :Evaluating Web Search Engines Using Clickthrough Data : SIGIR (2007)
4. David Carmel et al :Personalized Social Search Based on the User's Social Network : Conference on Information and Knowledge Management, Proceeding of the 18th ACM conference on Information and knowledge management, pp. 1227--1236 (2009)
5. ShenghuaBao et al :Optimizing Web Search Using Social Annotations : International World Wide Web Conference, Proceedings of the 16th international conference on World Wide Web, pp. 501--510 (2007)
6. Liang Gou, Hung-Hsuan Chen, Jung-Hyun Kim, Xiaolong Zhang, C.Lee Giles : SNDocRank: a social network-based video search ranking framework : MIR '10 Proceedings of the international conference on Multimedia information retrieval, pp. 367--376 (2010)
7. Seung Sik Kang : Morpheme Analysis and Information Search in Korean :hongrung Publishing Company (2003)
8. Dong-Kyun Lee, JoonHee Kwon : Social Search Algorithm considering Recent Interests of User : Journal of Korean Institute of Information Technology, Vol. 9, No. 4, pp. 187--194 (2011)
9. Korea Institute of Geoscience and Mineral Resources : Constriction of the cyber technical information integration center for natural disasters : NEMA (2008)
10. Anne H.H. Ngu, Michael P. Carlson, Quan Z. Sheng, Member, IEEE, and Hye-young Paik : Semantic-Based Mashup of Composite Applications : IEEE Transactions on services computing, vol. 3, no.1, pp.2--15 (2010)