International Journal of Multimedia and Ubiquitous Engineering
Vol.10, No.7 (2015), pp.351-358
http://dx.doi.org/10.14257/ijmue.2015.10.7.36

# Disasters Tracing with Occurred Events and Unexpected Ones

Yin Lan and Li Shuoming

*School of Mathematics and Computer Science*
*Guizhou Normal University，Guiyang, China, Zip code 550001*
*Computer School, Wuhan University, Wuhan, China, Zip code 430072*
*yl@gznu.edu.cn*
*Department of Electronic Engineering*
*Zhongshan Polytechnic*
*Zhongshan, China, Zip code 528404*
*lishuoming@whu.edu.cn*

***Abstract***

*Disaster comes up always in form of chain of events. These occurrence and evolution are represented dynamically in our daily social media. Retrieving or tracing these user-oriented relevant events around disasters in an aggregated results list through meta-search engine is a useful application. The definition and detection of events has been a hot topic in the study of natural language processing. We give our opinion in this paper, and we gathering the disaster event information via different search interface for an event chain releasing is important. This paper designed an event-triggered model for events detection. As the emergency requirement of government as well as individual activities, disaster events detection is discussed as an example in this paper. A meta-search engine interface model is constructed based on the open source project Carrot2. And a H-T-E (Hazard, Trigger and Event) query expansion approach is applied in our event triggered model. The experimental results show that the proposed method can bring about increased accuracy and extra clues not supplied by commercial search engines. The solution we applied can be useful for other information supervision tasks and some issues like crisis ontology construction etc. also triggered by the project for further study.*

***Keywords:*** *Meta search; Event-triggered model; H-T-E ; Query expansion*

## 1. Introduction

### 1.1 About Events and Event-Triggered Model

In human society, we are facing disasters everyday and the disasters are always without borders and coming up in the form of event chains. Internet is a good mirror for our physical world, and event supervision work in the Internet will be a good clue for human behavioral instruction. But Event is an ambiguous concept for representation and computing, as Charles Lamb said “Nothing puzzles me more than time and space; and yet nothing troubles me less.” Obviously time and space are two important dimensions for event. Generally, events are things that happen. “The things” is diversified like ceremony, wedding, earthquake and so on. In fact, as the term of “ontology” in information engineer is arguing, “event” is also rooted in philosophy and linguistic research, so how to visual event in the semantic space and detect it effectively is discussed a lot and many efforts of rule-based methods or statistical machine learning solutions are applied in digital world, such as template matching or maximum entropy *etc*. statistic methods.

Many natural language processing tasks deal with Verbs as event triggers then find arguments for relation discovery and event detection.[1]Then event templates

ISSN: 1975-0080 IJMUE

are defined for different tasks. Furthermore, the more automatic methods which have no manual intervention are proposed for event detection. [2]With the booming of social network, real time event detection becomes a hot topic, and Spatio-temporal data mining are concerned a lot. [3,4] But the event what we discussed in this paper is not so concrete or detailed in an aspect of linguistic analysis.

In natural language processing, most of researchers achieve a consensus and attempts to provide annotation of event in a limited specific scenarios or domains, for it is still a challenge to deal with domain in open domain for the arbitrary of language.

In this paper, we define event as event chain from hazard to flowing events sequence. Such as "Earthquake- Landslide- Building collapse" represents in Table 1.Which is classified by the American Disaster and Emergency Standard. [5] In our opinion, what we called or concerned event always be triggered by a sequential terms then broadcasting in a large area with interval. Triggered the topic terms then detect the event chain becomes a key issue. Then we can apply the word effectively for disaster tracing. Such as "Ebola" can be defined as a kind of biological hazards, and what we called disasters in this paper are always swarming ones in a period and which at least focus by the main stream media and public views. So the time and place property limited relatively or can be instructed by some metadata clues, what we concentrated in the event-triggered model is the lexical chains or lexical structures for the disaster describing.

**Table 1. Classification of Hazards**

| | |
|---|---|
| Naturally occurring hazards | (a) Geological hazards: earthquake, tsunami, volcano, landslide, mudslide, subsidence, glacier, iceberg |
| | (b) Meteorological hazards: flood, tidal surge, drought, fire (forest, range, urban, wildland, urban interface),snow, ice, hail, sleet, avalanche, windstorm, tropical cyclone, hurricane, tornado, water spout, dust storm or sandstorm, extreme temperatures, lightning strikes, geomagnetic storm |
| | (c) Biological hazards: emerging diseases that impact humans or animals, such as plague, smallpox, anthrax, west Nile virus, foot and mouth disease, severe acute respiratory syndrome, pandemic disease, Animal or insect infestation or damage |
| Human-caused accidental hazards | Hazardous material (explosive, flammable liquid, flammable gas, flammable solid, oxidizer, poison, radiological, corrosive) spill or release, explosion/fire, transportation accident, building/structure collapse, energy/power/utility failure, fuel/resource shortage, air/water pollution, contamination, water control structure/dam/levee failure |
| Human-caused intentional hazards | Terrorism (explosive, chemical, biological, radiological, nuclear, cyber), sabotage, civil disturbance, public unrest, mass hysteria, riot, enemy attack, war, insurrection, strike or labor dispute, disinformation, criminal activity (vandalism, arson, theft, fraud, embezzlement, data theft), electromagnetic pulse, physical or information security breach, workplace/school/university violence, product defect or contamination, harassment, discrimination |

Considering the revolution of language, including the static method for hazard definition, some dynamic methods for dataset construction are also be done by researchers, such as data at http://crisislex.org/, a set for 26 events during year from 2012 to 2013 spawned significant activity on Twitter are provided. These good resources also give us good reference for concept filtering.

### 1.2 Meta Search Interface

We have to commit that there are bias in different search engines. The fact is there are lots of search engines for information retrieval, a meta-search engine deals with the user's query simultaneously by several individual search engines (such as Google, Yahoo, Bing, Baidu *etc*.) and aggregate results into a single list in a user oriented re-rank. The typical architecture of meta-search engine as figure 1 shows. The key issue of models for meta search engine is optimizing the performance of the combination.[6] Ranking aggregation is discussed a lot by researchers.[7-8] The barrier of retrieval question is still the gap of limitation of short and natural language based query and the enormous on-going information. Knowledge bottle is still the barrier in these tasks.

In this paper, we set a data pre-processing module with natural language processing method between the query and response in the Meta Search, the open source Meta search engine frame named Carrots2 [9] are applied for the secondary development and an event-triggered model is applied for the retrieval control.

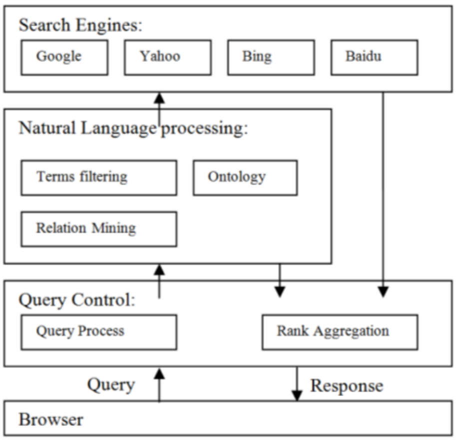


**Figure 1. The Architecture of Meta Search Interface**

### 1.3 About this Paper

In this paper a meta-search engine platform is constructed with the support of open sources software Carrots2 as Figure 2 shows. An Event-triggered Model is proposed and we apply it for query expansion and disaster information supervision, we apply an H-T-E lexical structure for the event tracing and we control the query and query expansion by the model. The results show that the solution what we practice is effective and the project shows that the solution can help us for more user defined tasks for meta-search.

A Java and TOMCAT based frame is realized for this project. Google, and Yahoo, Baidu and Bing search engines web service API for programmers is called.

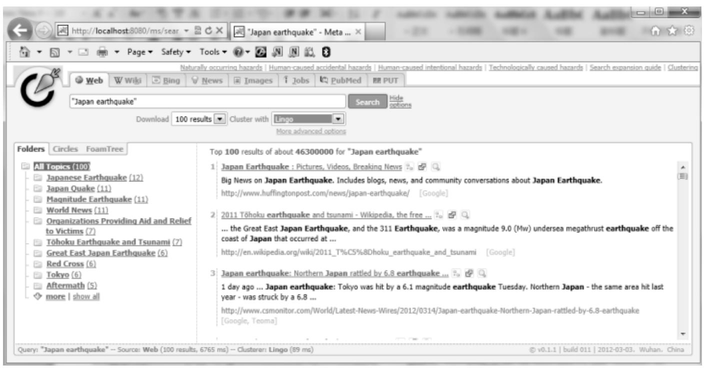

**Figure 2. Carrots2 Meta-Search Interface**

# 2. The Principles and Methodology

## 2.1 The Event-Trigger Model

As we defined that the event are describe in two level of concept clustering for sequential reason. One is Hazards set H {H1, H2, H3… Hi} and the other is Events set E {E1, E2, E3, Ei}. Some concept word in Hazards set can be regarded as trigger word. Figure 3 is the H-T-E model.

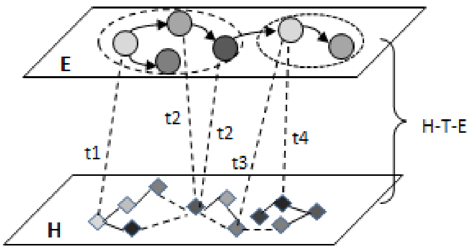


**Figure 3. The Principle of H-T-E Keyword Expansion**

**Table 2. Relationship of Hazards as an Example**

| Primary hazard | Associated hazard | Secondary hazard |
|---|---|---|
| Earthquake | Landslide | Building collapse, Tsunami, Explosion |
| Hurricanes | Cyclone, Flood | Building collapse |
| Snow | Traffic accidents | Fire accidents caused by heating, Avalanche |
| Volcano eruption | Earthquake | Forest Fire, flood, Sparry flood |

Note: Associated hazard relates to those hazards that go along with the primary hazards and usually happens at the same time. Secondary hazards are the hazards that follow as a result of other hazard events. From the view of time, the event can be looked as two levels of concept clustering.

## 2.2 Topic Model for Text Processing

Topic model is applied for discovering the abstract "topics" that occur n a collection of documents, LDA (Latent Dirichlet Allocation) as a successful probabilistic model for unknown document topic detection[10], It learns topics as discrete distributions over the event patterns. We applied the traditional topic model for concept clustering, then we can get some aid for triggered concepts detection and event terms finding. The part of speech such Verbs or Nouns is not concerned in our work, we filter trigger words just by the high frequency co-occurrence with hazards. Then an event lexical ontology for Event-triggered model is made half done manually. We believe the term from the original documents will describe the event well. It can help us adjust the retrieval results by term query strategy Especially clustering concepts with the same topic on primary hazards, associated hazards and secondary hazards is benefic for the term set filtering. And the different concepts set are dynamic selecting for the query.

## 2.3 Algorithm and Questions Discussed in the Project

The following algorithm is the process we performed in our Meta search engine project.
Input: Queries (Topic theme)
Output: Retrievals after query expansion

 

Step1: Boolean D=Compare (Q, H)
//Compare the Query and hazard Similarity
If D=0, conduct a normal search;
Else go to step 2;
Step2: H=Hi, and Hi∈{H1,H2…Hk}
//Present associate and secondary hazards for the alternative expansion
//Hi is inter-connected hazards set which provide relevant hazards for user to confirm
Step3: T=Ti, and Ti∈{T1,T2…Tk}
//T is our word set which is triggered by Hi.
Step3: Refined search result by switch Hi +Ti for event detection
// the word combination for event detection are filtering by the amount of feedbacks in a individual search engine
Step4: Result clustering or integrating by Carrot2
//Up to top 200 feedbacks are processed in clustering with multiple clustering algorithms

In our project, the following two questions will be discussed:

1) The embedded question of source engines dependency.

A. The decoding query forms to different search engine will cost more time and sometime need to form transformation such as language translation.
B. Some vulnerability to search spam.
C. Lack method for comparing relevance scores.

Even so, the advantage is also obviously, and with the booming of big data, these commercial search engines can give a good support for personal utility.

2) The Semantic calculation between concept terms.

In the algorithm, the similarity of concepts pair is a fuzzy question, we combine the mathematical calculation and user interaction in the object.

Comparing work is easy for a numerical data, but it is not so in case of a symbolic data, especially for semantic measurement. It can be looked as a fuzzy set issue. Similarity is determined not only the information content but also the structure，and individual cognition, even it will be a dynamic work based on the current state of knowledge as embedded in the ontology. Two concepts which can be measured semantically similarity discussed a lot, the typical methods including:

D. Numeric the terms for mathematical computation, such as Salton's cosine correlation method. [11]
E. Structure-based methods, such as their literal values *etc*.[12]
F. Information-based methods, typical work as Lin applied information theory for measurement.[13]
G. Corpus-based methods, such as apply Wordnet, Hownet, Wikipedia *etc*. as knowledge for measurement.[14-16]
H. Hybrid method, such as combing local context and Wordnet similarity.[17]

In this paper, we apply it based on the Event-triggered model which can be looked as a semi-automatic ontology for concept detection. In further research, the more practice on similarity measurement and adaptive ontology learning method will be attempted.

## 3. Implementation and Evaluation

The proposed Meta search interface feedback the top 200 results and finish clustering within 3 seconds on a laptop (CPU i5-2520M @ 2.5GHz, RAM 4 GB).

Figure 4 shows the results in clustering of an expanded search. In order to evaluate our experiment, we use a precise of top-N methods, we select the top 100 results from each engine, and define the precision is assessed by human judges. Traditional trigger methods get a lot of redundant result and we just count for a class of relevant results. And after rank aggregation and lexical ontology applied for searching, the results show that we can develop the results well. Furthermore, we can catch back some relevant events we concerned.

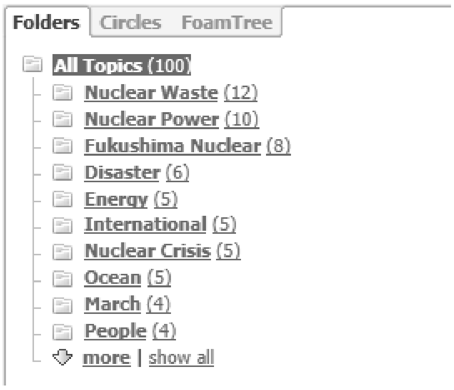


**Figure 4. The Principle of H-T-E Keyword Expansion**

Table 3 is a clustering comparison between the original query and the refined query. In figure 5, R1, R2 and R3 are 3 query results from the individual known search engines, and AR the aggregation result we applied by the open source meta-search engine combine with event-triggered strategy.

In brief, the key words is real the key for searching, but the "key" is limited by human cognition, sometime we have to say these words are should be objective and came from relevant documents by data-driven method which will be better for event described well, meta-search engine have the superiority than individual ones, and our search habits are limited us for one search interface, which make the bias of search engine obviously.

**Table 3. Clustering Comparison of the top 100 Feedbacks**

| *Original query results* | *Alternative query results* |
|---|---|
| *Japan Earthquake(12)* | *Nuclear Waste (12)* |
| *Japan Quake(11)* | *Nuclear Waste (10)* |
| *Magnitude Earthquake(11)* | *Fukushima Nuclear (8)* |
| *Aid and relief (11)* | *Disaster(6)* |
| *Tohoku Earthquake(7)* | *Energy(5)* |
| *Great East Japan Earthquake(6)* | *International(5)* |
| *Red cross(6)* | *Nuclear Crisis(5)* |
| *Tokyo (6)* | *Ocean(5)* |
| *Aftermath (5)* | *March(4)* |
| *Others (25)* | *people(4) ; Others (36)* |

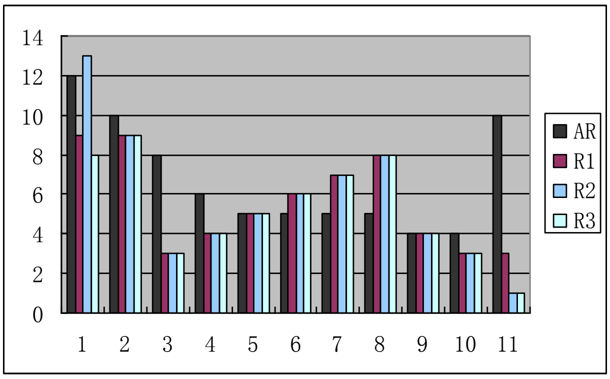


**Figure 5. Results between the Aggregated Results(AR) and Other 3 Indivadual Results (R1,R2,R3)**

## 4. Conclusion

As the lexical resources on crisislex.org, an adaptive lexical ontology for disaster tracing is very important. We constructed an application by open source meta-search engine. And we proposed a search expansion method in our experiment.

In summary, the Meta search interface and the event-triggered model for expansion we proposed facilitated response personnel and decision maker to gather useful information and enhance situation awareness. In future work, a more automatic domain event template construction work will be discussed and applied in our Meta search engine work. For the describing the event, the time consumer is enduring in this paper.

But for further study, in order to trigger the event is still a tough work for the property words like place name, relative organization *etc*. can't be predicted. When a disaster occurs, time is very limited, so we need to act as quickly as possible and with as much knowledge of the situation as possible. Recent study shows collecting and filtering lexicons through social network for the disaster supervision has become a direct way [18-19]. How to gather information about a crisis from Weibo or Twitter *etc*. social media in time for the crisis alarming is very essential. Meanwhile a fuzzy hierarchical decision method for concept selecting will be discussed in the view of computation. However, the millions of Twitter messages ("tweets") broadcast at any given time can be overwhelming and confusing, finding the relevant trust ones timely still be our further study.

## Acknowledgments

This work is supported by the united fund of Guizhou science and technology department, Guizhou Normal University No. LKS [2012] 37. We would like to thank Carrot2 project for offering a wonderful open programming source which enabled us to create the Meta search interface.

This paper is a revised and expanded version of a paper entitled "An Event-Triggered Model for Disaster Retrial Via Meta Search Interface" presented at the 8th International Conference on Future Generation Communication and Networking (FGCN 2014) to be held 20-23 December 2014 in Haikou, Hainan, China.

## References

[1] Qin B., "Event type recognition based on trigger expansion," Tsinghua Science & Technology, vol. 15 no. 3, **(2010)**, pp. 251-258.
[2] Chambers N. and D. Jurafsky, "Template-Based Information Extraction without the Templates," in ACL, Portland, USA, **(2011).**
[3] Lauw H. W., "Social network discovery by mining spatio-temporal events," Computational & Mathematical Organization Theory, vol. 11 no. 2, **(2005)**, pp. 97-118.

[4] Zheng Y., "GeoLife2. 0: a location-based social networking service. in Mobile Data Management: Systems, Services and Middleware, 2009," MDM'09. Tenth International Conference on, Taipei, Taiwan, China, **(2009)**.
[5] Management, N.F.P.A.T.C.o.D., NFPA 1600 standard on disaster/emergency management and business continuity programs. National Fire Protection Association, **(2004)**.
[6] Aslam J. A. and M. Montague, "Models for metasearch," in Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, New Orleans, LA, USA, **(2001)**.
[7] Dwork C., "Rank aggregation methods for the web," in Proceedings of the 10th international conference on World Wide Web, Hong Kong, China, **(2001)**.
[8] Renda M. E. and U. Straccia, "Web metasearch: rank vs. score based rank aggregation methods," in Proceedings of the 2003 ACM symposium on Applied computing, **(2003)**.
[9] Osiński S. and D. Weiss, "Carrot2: Design of a flexible and efficient web information retrieval framework," in Advances in Web Intelligence. Springer. **(2005)**, pp. 439-444.
[10] Blei D. M., A. Y. Ng and M. I. Jordan, "Latent dirichlet allocation," The Journal of Machine Learning Research, vol. 3, **(2003)**, pp. 993-1022.
[11] Salton G. and C. Buckley, "Term-weighting approaches in automatic text retrieval," Information processing & management, vol. 24 no. 5, **(1988)**, pp. 513-523.
[12] Levenshtein V. I., "Binary codes capable of correcting deletions, insertions and reversals," in Soviet physics doklady, **(1966)**.
[13] Lin D., "An information-theoretic definition of similarity," in ICML, Madison, Wisconsin, USA, **(1998)**.
[14] Pedersen T., S. Patwardhan and J. Michelizzi, "WordNet: Similarity: measuring the relatedness of concepts," in Demonstration Papers at HLT-NAACL 2004, **(2004)**.
[15] Gabrilovich E. and S. Markovitch, "Computing Semantic Relatedness Using Wikipedia-based Explicit Semantic Analysis," in IJCAI, **(2007)**.
[16] Zhu Y. L., "Semantic orientation computing based on HowNet," Journal of Chinese Information Processing, vol. 20 no. 1, **(2006)**, pp. 14-20.
[17] Leacock C. and M. Chodorow, "Combining local context and WordNet similarity for word sense identification," WordNet: An electronic lexical database, vol. 49 no. 2, **(1998)**, pp. 265-283.
[18] A. Olteanu, S. Vieweg and C. Castillo, "2015.What to Expect When the Unexpected Happens: Social Media Communications Across Crises," In Proceedings of the ACM 2015 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '15), Vancouver, BC, Canada, **(2015)**.
[19] A. Olteanu, C. Castillo, F. Diaz and S. Vieweg, "2014. CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises," In Proceedings of the AAAI Conference on Weblogs and Social Media (ICWSM'14), Ann Arbor, MI, USA, **(2014)**.

# Authors

**Yin Lan**, (1979) she is a teacher in the School of Mathematics and Computer Science, Guizhou Normal University, Guiyang, China. She is a Ph.D. candidate in Wuhan University. Her research Interests include Natural Language Processing, Knowledge representation and discovery, Information Retrieval.

**Li Shuoming**, (1981) he is a teacher in the Department of Electronic Engineering, Zhongshan Polytechnic, Zhongshan, China. He is a Ph.D. candidate in Wuhan University. His research interests include Data mining, Machine learning, Natural language processing and Information retrieval.

 