NEI:A Framework for Dynamic News Event Exploration and Visualization

Xiaofei Guo, Juanzi Li, Ruibing Yang, Xiaoli Ma
Knowledge Engineering Group
Department of Computer Science and Technology
Tsinghua University, China, 100084
{guoxiafei,ljz,yangruibing,maxiaoli}@keg.cs.tsinghua.edu.cn

ABSTRACT
Nowadays, there are many events reported by News Media everyday, which contains a massive number of news. People are getting more and more interested in understanding how an event evolves after it happens. News related to the same event or similar events usually has more common entities and stronger topic correlations, which is a new perspective to study news event. Due to the complexity of event evolving process, event visualization has been a big challenge for a long time.

In this paper, we design a novel four-phase framework NEI(News Event Insight) that focuses on visualizing a news event properly and clearly, namely (1)Entity Topic Modeling. We extract topics and entities through timeline. (2)Temporal Topic Correlation Analysis. Based on the topic modeling result, we design two methods to select hot topics and build links for them. (3)Keyword Extraction. Specially, we combine string frequency with syntax features and use language models to acquire candidate keywords for representing topics. (4)Visualization. Visualization demonstrates the quantifying properties of topics related to a certain event. A case study shows our framework achieves promising results on both single event and similar events.

Author Keywords
visualization, keyword extraction, entity, topic similarity

1. INTRODUCTION

With the rapid development of Web 2.0 techniques, there are overwhelming news and user-generated contents on the web. People are getting more and more interested in understanding how an event evolves after an event happens. For example, most people care about the event *American Presidential Election*(APE) which contains almost five-thousand documents in the whole time in Sina specific news portal[1]. However, it is unrealistic for people to read all of the articles. Indeed, what people want to know is not just a simple concept of news event, but also the inner relationships within an event without heavy reading task. Thus, a novel method to model the evolution and development of a news event is required.

News event visualization is challenging. Different from other text streams, a news event has more named entities (e.g. Noun. Person, Organization or Location name) and strong topic correlations. For an event, people are often interested in the whole picture of what a news event evolves along timeline[4]. To measure and visualize news event, many problems should be tackled. Topic representation is the main problem. Instead of using top words to display a topic in word sense, it is necessary to employ advanced method to represent the topic in semantic level. In addition, temporal topic correlation analysis may also be the key point to decide which topic will be shown. Finally, the third goal of this paper is to incorporate topic extraction technologies and temporal topic correlation analytic algorithms for event visualization.

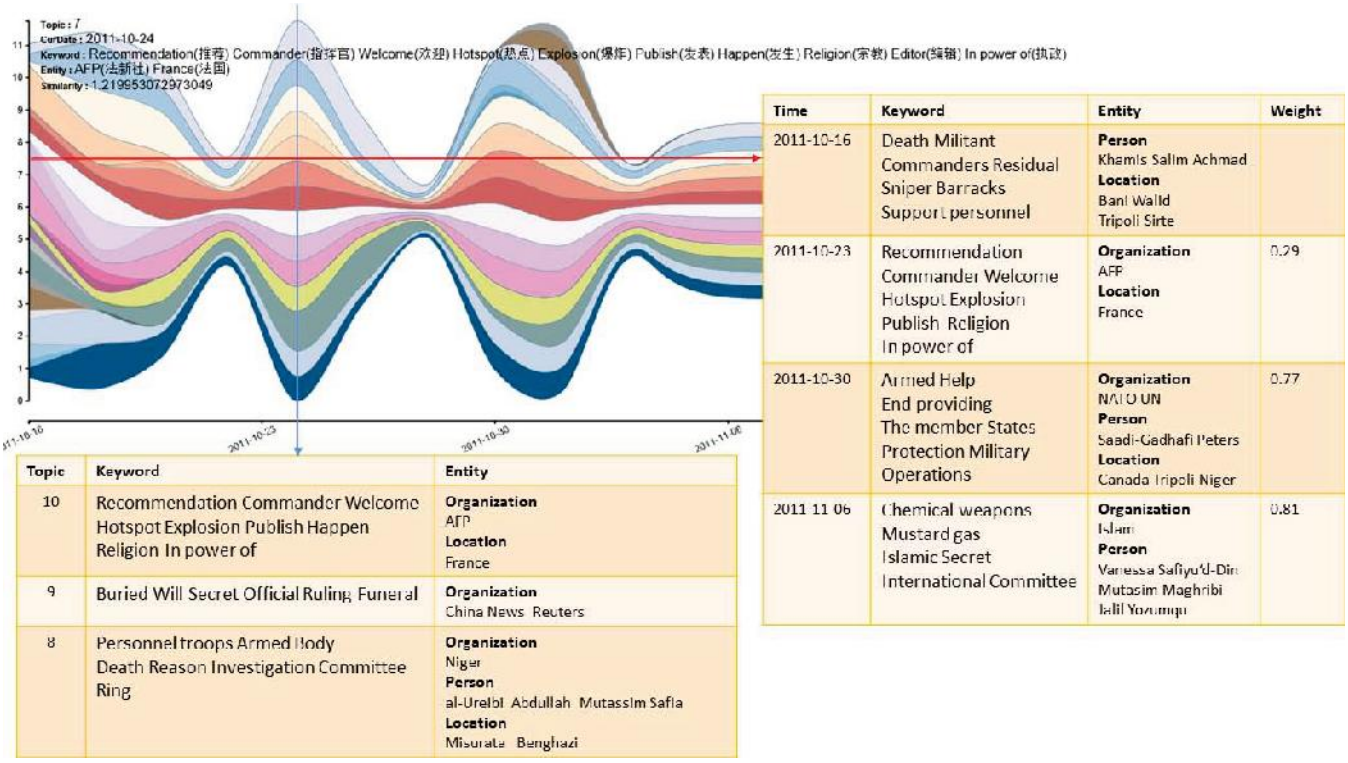
Fig. 1 shows the evolution of event *Moammar Gadhafi's Death*(MGD), where x-axis represents time and y-axis represents the number of topic streams. We can see clearly that the event has two significant time points , one is about Oct. 23, when people confirmed Moammar Gadhafi's Death, one is about Nov. 3, which may be a big shock after the war. What's more, detail information about each topic will be shown clearly in the upleft corner of the figure. On the one hand, each layer in the picture represents a topic stream, while you move the focus along the layer, it shows the details of the topic evolution. On the other hand, if you move the focus cross different layers, you'll know what happens at the same time. As a result, we see after topics' evolving, the event stream becomes stable, as no more topics born or die.

In this paper, we offer a method for people to show an news event as Fig. 1. For an event, we design algorithms to select hot topics and compute temporal topic

[1]`http://www.news.sina.com.cn/zt/`.

VINCI '14, August 05 - 08 2014, Sydney, NSW, Australia

http://dx.doi.org/10.1145/2636240.2636845



Topic	Keyword	Entity
10	Recommendation Commander Welcome Hotspot Explosion Publish Happen Religion In power of	**Organization** AFP **Location** France
9	Buried Will Secret Official Ruling Funeral	**Organization** China News Reuters
8	Personnel troops Armed Body Death Reason Investigation Committee Ring	**Organization** Niger **Person** al-Urelbi Abdullah Mutassim Safia **Location** Misurata Benghazi

Time	Keyword	Entity	Weight
2011-10-16	Death Militant Commanders Residual Sniper Barracks Support personnel	**Person** Khamis Salim Achmad **Location** Bani Walid Tripoli Sirte	
2011-10-23	Recommendation Commander Welcome Hotspot Explosion Publish Religion In power of	**Organization** AFP **Location** France	0.29
2011-10-30	Armed Help End providing The member States Protection Military Operations	**Organization** NATO UN **Person** Saadi-Gadhafi Peters **Location** Canada Tripoli Niger	0.77
2011 11 06	Chemical weapons Mustard gas Islamic Secret International Committee	**Organization** Islam **Person** Vanessa Safiyu'd-Din Mutasim Maghribi Ialil Yozumpu	0.81

Figure 1. Libya News Event Visualization

correlations. For a topic, we design a method to discover phrases and entities, so that people could visually see the event effectively and efficiently. Besides, the information could be employed as the prior knowledge for analyzing similar events in the future.

To summarize, our contributions are listed as follows:

- We develop a unified framework for news event visualization, which exploits the popular events efficiently and effectively. Experiments show that our approach achieves promising results on both single event and similar events.
- We design an algorithm to select hot topics and compute similarities for them. In the meantime in order to represent a topic more effectively, we use key phrases, which integrate more semantic information.
- We use the entity topic model to track a topic with its useful entities.

The rest of the paper is organized as follows. We start by summarizing the previous work. After that we formalize the problem of NEI and explain the architecture of it. Then we describes our novel method including the four main components. Two case studies are presented to evaluate our framework. Finally, we wrap up the experiment results with the conclusion and future work.

2. RELATED WORK

Previous methods have tackled the problem at different research orientations. Mining evolution patterns of critical events has three important branches.

(1)**Topic Representation**. It is quite often to discover key phrases to represent topics[11] for event tracking. [9] has predicted and analyzed the busty of event. However, there might be some key phrases and entities which are actually related and representative in different topics of event. [15] discovered latent themes for describing themes. What's more, [21] used named entity as important additional information for news event detection. [18] tried to extract burst words to help readers find useful information on Twitter.

(2)**Event Analyzing**. For this issue, most work to date focused on topic modeling. [1] [8][10] introduced their methods based on topic models to discover latent structures in document collections. [3] regarded event as discrete temporal sequences instead of topic models. For the content, [2][8][17] were focused on how to display text streams such as news articles, while [14] attempted to discover the relationships between different events. Furthermore, [12] developed a method for tracking units of information as they spread over the Web. They were successful on analyzing text streams or events relationship by using time sequence and content similarity between two component events.

(3)**Visualization**. After analyzing, we need visualization methods to show the result intuitively with time-

line. Recently, various methods have been proposed to rank topics, relationships etc[13]. [20] presented a graph-based method for news summarization by extracting useful sentences. [5][19] proposed an approach to tracking and connecting clusters in text data. That is, they get the result of news evolution by displaying key word. In contrast to previous works, [4] defined three relationships combined into a uniform relationship to show a temporal event map. More similar to our NEI framework in visual is the system named "ThemeRiver"[6][7]. The ThemeRiver system, depicted thematic variations in a collection of visualization patents from one company over several years, which is quite different from our work in intent.

To the best of our knowledge, the three branches of analyzing and measuring topics mentioned above are usually independent in event visualization. If combined into one, it could strengthen the understanding of news event for the reader. This is why we propose a unified model that takes topic representation, relationship analyzing and event visualization into consideration at the same time. It is worth mentioning that in this work, we use entities and N-gram phrases to represent a topic, which contains much more useful information.

In next section, we present a novel framework to achieve the task of measuring and visualizing news event.

3. NEWS EVENT INSIGHT

In this section, we present required definitions and formulate the problem of event visualization.

3.1 Formalization

First of all, a news *event* is something that happens at special time and place. Usually, a news event often represents a process of one thing which is usually reported by a set of related documents, such as news articles on the Web.

Definition 1 (Event). Let e be a news event, which contains a set of related documents, denoted by $\boldsymbol{D}_{(e)} = \{(d_i, t_i)|i = 1 \dots n_d\}$, where d_i is the news article reporting e published at time t_i. Each document d_i is a set of words, denoted as $d_i = \{w_{i,j}|j = 1 \dots n_w,\ w_{i,j} \in V\}$, where V is the word vocabulary.

There are two assumptions we used in analyzing event.

- When an event happens, news articles report the event from different aspects, which we define them as topics.
- However, the topics of event change with the evolution of event. That is , there are dependence relationships between earlier topics and later ones.

Based on these assumptions, we model event as a set of topics, denoted by $\Im_{(e)} = \{z_i|i = 1 \dots n_z\}$, where z_i is the topic of event e.

Definition 2 (Topic). A topic z is defined as a triple (τ, EN_z, KP_z), where τ is the time window of topic z, and EN_z, KP_z are the entities and keyword phrases in topic z. In fact, for an event e we divided it into a set of time windows $\{\tau_i|\tau_i \in T, i = 1 \dots n_t\}$, where n_t is the number of time windows and $T = \tau_1 \cup \tau_2 \cup \dots \cup \tau_{n_t}$ is the whole time of event e, while τ_i, τ_{i+1} are two successive time slices. The set of EN_z of topic z is a set of named entities denoted by $EN_z = \{en_{ij}|i = 1 \dots n_z, j = 1 \dots n_{en}\}$, where en_{ij} is a series of terms in topic z_i which represent the names of location, people or organization. Similarly, we use kp_{z_i} to represent a set of key phrases to describe topic z_i, while $KP_z = \{kp_{im}|i = 1 \dots n_z, m = 1 \dots n_{kp}\}$ represents the whole key phrases in event e. Specifically, we consider for a topic, named entities and key phrases can keep enough information in both word and semantic meaning.

Definition 3 (Relationship). We represent the evolution of topic by $\Re_{z_i z_j}$, where z_i, z_j are two topics which z_i is in the time window τ_i, z_j is in the window $\tau_{i+1(e)}$. $\Re_{z_i z_j}$ is the value which denotes the relationship between z_i and z_j.

From above definitions, the modeling of event e can be described as the procedures to obtain $\Im_{(e)}$ and $\Re_{(e)}$ from $\boldsymbol{D}_{(e)}$ for the target event e. Hence, we could formulate the problem of measuring and visualizing temporal topics as function Φ :

$$\Phi : \boldsymbol{D}_{(e)} \ \rightarrow \ (\ \Im_{(e)}\ ,\ \Re_{(e)}\ )$$

News Event Insight. Formally, we want to visualize the various elements of event e. Besides topics, an event may consists of many entities and key phrases, which could be shown in a whole picture. We propose a sequential framework namely News Event Insight(NEI) for visualizing news event evolution.

3.2 Architecture

Fig. 2 illustrates the architecture of the NEI framework. It is described as two parallel work flows. As shown in the right side, given a set of documents reporting one particular event, NEI generates a visualization of the event after five main processes shown at the left side.

(1)**Preprocessing**. It includes two main sub-component work: Data Cleaning and Preprocessing. With data cleaning, we set some important parameters through observation[2]. For Chinese text, there are three things that we have to do with preprocessing: tokenization, stop word filtering and articles splitting by timespan.

(2)**Entity Topic Modeling**. Here we get a document list over time ready for topic modeling, which is the basic component of NEI. Then we use entity topic model to train the data, which will be introduced later.

[2]We set $\tau = 7$ days in event APE, while $\tau = 2$ days in event MGD, where $n_t = \lceil \frac{T}{\tau} \rceil$ is the number of time windows.

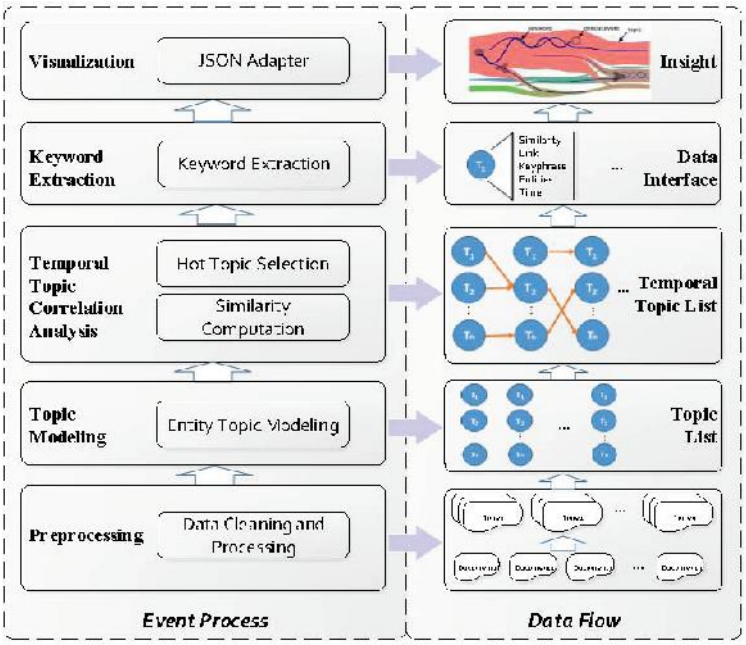


Figure 2. Architecture of News Event Insight

(3)**Temporal Topic Correlation Analysis**. Based on the extracted result with entity topic model, the temporal topic correlation analysis aims at finding the correlations between different topics in different time. It has two main functions: hot topic selection to decide which topic is more significant and similarity computation to show the dynamic continuity of topics. Furthermore, we present these specific approaches in temporal topic analyzing.

(4)**Keyword Extraction**. Subsequently, we propose an N-gram based keyword extraction method to provide the semantic information of a topic, which is shown keyword extraction section.

(5)**Visualization**. After all of the processes above, we define an interface to change the serialized results into JSON object, which are then used to produce the final visual summary of news event.

4. OUR APPROACH

In this section, we explain the three main components of NEI's mechanism: Entity Topic Modeling, Temporal Topic Correlation Analysis and Keyword Extraction in details. After that, our data interface used to transport data from the text module to visualization module will be introduced.

4.1 Entity Topic Modeling

After preprocessing, we get a document list by timespan, denoted by $\boldsymbol{D}_{(e)} = \{(D_i, \tau_i)|i = 1 \dots n_t\}$ where D_i represents the news articles in time window τ_i. Different from previous work, we build the connections between topics and entities through the inherent relations of words using entity topic model(ETM)[16][3], which can better track topic and predict entity. For each time window, we get a list of topics $\Im_{(e)} = \{Z_i|i = 1 \dots n_t\}$, where Z_i is a collection of topics for D_i at τ_i. More detailed, for each topic z in Z_i, we get a triple (θ, ϕ, EN_z), where θ is the matrix of the document-topic distribution $p(z|d)$ and ϕ is the matrix of topic-word distribution $p(w|z)$, while EN_z is the entities related to topic z.

4.2 Temporal Topic Correlation Analysis

Different from other text streams, a news event has more entities and strong topic correlations, which means topics will be generated or continued or disappeared through timeline. It is a dynamic process like people's mind which changes along time.

Hot Topic Selection. The result of ETM may not be intuitionistic understanding, especially when the topic number is large. Here we offer a strategy that is based on the document-topic distribution matrix θ to decide which topic should be given preference to visualize. To be specific, we consider the more portions a topic has, the more important it will be. What's more, we think each document has more than one important topic, instead, it has N important topics[4]. As a result, we update the topics of event with hot topics, denoted as $\Im_{(e)} = \{Z_{h_i}|i = 1 \dots n_t\}$. The detailed hot topic selecting algorithm is presented in Algorithm 1.

Topic Similarity Computation. After topic modeling, each topic is constructed with a vector of words.

[3]We set $K = 20$, $\alpha = \gamma = \frac{50}{K}$, and $\beta_1 = \beta_2 = 0.1$.
[4]We set $N = 10$ to select 10 hot topics for each time window.

Algorithm 1 Hot topic selecting algorithm.

Require: document-topic distribution matrix θ, timespan τ, hot topic number N,
Ensure: hot topic list Z_h.
1: **for** $d_i = 1 \ldots n$ **do**
2: $Z_i = \max_N\{p(z_j|di)\}$, where z_j is one topic of Z_τ and $p(z_j|di)$ is the document-topic distribution,
3: **end for**
4: $Z_h = \max_N \bigcup_{n_d}\{Z_i\}$.
5: **return** hot topic list T_h.

Here we introduce a Bayesian-based method to compute the topic similarity using the topic-word distribution matrix ϕ. A topic $z_{i,j}$ denoted as
$z_{i,j} =< w_{ij,1}, w_{ij,2} \ldots, w_{ij,m} >$
represents the jth topic of Z_{h_i} in τ_i.
Equally, we define topic
$z_{i+1,k} =< w_{(i+1)k,1}, w_{(i+1)k,2} \ldots, w_{(i+1)k,n} >$
as the kth topic of $Z_{h_{i+1}}$ in τ_{i+1}.
Algorithm 2 shows the method to compute the similarity $\Re(z_{i+1,k}|z_{i,j})$.

Algorithm 2 Similarity computing algorithm.

Require: topic-word distribution matrix ϕ, $z_{i,j}$, $z_{i+1,k}$,
Ensure: the topic similarity $\Re(z_{i,j}, z_{i+1,k})$.
1: $\Re(z_{i,j}, z_{i+1,k}) = 0.0$,
2: **for** each word w_i in $z_{i+1,k}$ **do**
3: calculate $p(w_{(i+1)k,l}|z_{i,j})$,
4: $\Re(z_{i,j}, z_{i+1,k}) += p(w_{(i+1)k,l}|z_{i,j})$,
5: **end for**
6: **return** $\Re(T_{i,j}, T_{i+1,k})$.

Thus, the final similarity function is as following:

$$\Re(z_{i+1,k}, z_{i,j}) = \sum_{l=1}^{l=n} p(w_{(i+1)k,l}|z_{i,j}) \quad (1)$$

where $p(w_{(i+1)k,l}|z_{i,j})$ means the word in topic $z_{i+1,k}$ has how much similarity with topic $T_{i,j}$, which is defined as following:

$$p(w_{(i+1)k,l}|z_{i,j}) = \begin{cases} p(w_{(i+1)k,l}|z_{(i+1),k}), & w \in z_{i,j}, \\ 0, & w \notin z_{i,j}. \end{cases} \quad (2)$$

4.3 Keyword Extraction

To better help readers understand hot topics' result $\Im_{(e)} = \{Z_{h_i}|i = 1 \ldots n_t\}$, NEI framework automatically extracts keywords denoted as KP_z with semantic information for each topic z of Z_i. Given the fact that the number of times a word appears reflects the importance of notional word, based on the string-frequency method, we develop an approach to extract keywords through adding syntax rules and features.

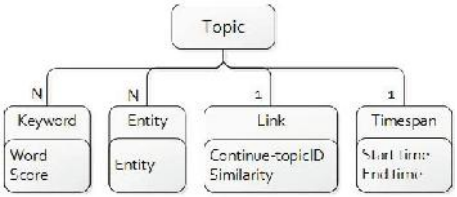


Figure 3. data interface of NEI

First we conduct the sentence split, word segmentation, stop-word filtering and POS tagging to get the candidate keywords of uni-grams. Then we employ bi-gram and tri-gram to create combined candidate words/phrases which will undertake a filtering process. We only keep those keywords whose frequencies are above a certain predefined threshold determined by experimental value.

Considering word frequency, POS, position, and morphology, we define eight features in Table 1. To calculate the final score for every keyword, we define the score as following equation:

$$score(w) = (w.tf)^{t_1}(1 + \sum_{fi \in F} w.fi \cdot t_{fi}) \ln \frac{termSum^{t_2}}{w.ctf^{t_3}} \quad (3)$$

where F = inTitle, quo, inFirst, sign is a set of feature values; t^{fi} is corresponding weight; t_1, t_2, and t_3[5] are the weights of $w.tf$, $termSum$ and $w.ctf$.

4.4 Data Interface

We define a standard interface to find a way of describing topic effectively. It is used as an API to transport data from its analytic result to its visual tool. As previously mentioned, a topic z is defined as a triple (τ, EN_z, KP_z), which represents keyword phrases, entities and timespan. In addiction, We add the temporal correlated information to the topic as Fig. 3 shows, while the words on the boxes denote the number and the boxes represent the details.

5. EXPERIMENT

In this section, we describe the design of our experiments and analyze the effectiveness of NEI's visualization.

5.1 Data Set

Since there is no standard corpus of news event visualization, we collect three datasets from Sina specific news portal. For single event, we crawl news of event MGD which is the central point till today. For similar events, we used the datasets of APE08 and APE12, which is one of the hottest news events in the world every four years. The statistics of the three events are shown in Table 2.

[5]We set $t_1 = 0.99$, $t_2 = t_3 = 2.3$, and $\beta_1 = \beta_2 = 0.1$.

Table 1. Definitions of Features

Features		Definition
word frequency	*w.tf*	noting the word frequency
	w.ctf	noting the total word frequency in the set of documents
	termSum	noting the total frequency of uni-grams, bi-grams, and tri-grams according to the grammar
position	*w.inTitle*	noting whether the word occuring in the title, if yes, value 1, else, value 0
	w.inFirst	noting whether the word occuring in the first sentence, if yes, value 1, else, value 0
POS	*w.POS*	the POS of the word
morphology	*w.quo*	noting whether the word is quoted, if yes, value 1, else, value 0
	w.sgin	noting the amount of information contained in the word according to its length (l), its value is set to: if l = 1, 0; if l is bigger than 2 and smaller than 8, value $\log l$; if l > 8, 3.

Table 2. Statistics of news event

Dataset	Number	Start	End	Days
Libya	466	2011/10/16	2011/11/22	51
08APE	4403	2008/06/28	2008/12/06	161
12APE	779	2012/07/05	2012/11/29	147

5.2 Visualization

In order to present experimental results of our framework, we first use the worldwide spot news event (i.e. MGD) as shown in Fig. 1 in introduction, which shows both the evolution of topics extracted from the MGD news dataset and their inner-relationships through a timeline. The MGD news dataset covers a 51-day period event crawled from Sina news. In this figure, each colored layer represents a topic stream generalized with the ETM model. Moreover, the topic of each timestamp are associated with its keywords and entities. All of these summarize the content of topic and its evolution in visualization.

Stream Layer. We use different colors to represent different topics to make the visualization more intuitive. A continued topic stream is colored in one all the time, which only has width changes to represent the strength or popularity of the topic at that time.

Topic Annotation. For a news event, people not only want to track the bursts, but also know how the stream evolves over time in details. We create a structure to tackle this problem by showing topic keywords and entities as well as their similarities at the same time. The annotation is shown at the upleft corner of the picture. What's more we translate the result into English to facilitate reading. In Fig. 4 we use a green star to mark the position of the current topic.

5.3 Similar Event Analysis

Quite often, what people interested in is not just the single event's evolution, but also the comparison result of events related to it. Indeed, they are pay attention to the differences between similar events. This calls for similar event analysis.

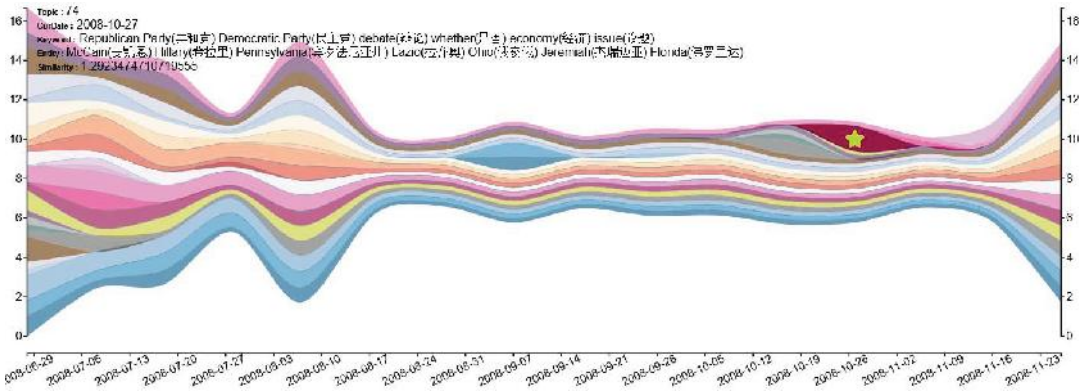
As aforementioned, everybody cares about the event *American Presidential Election* (APE). We present our experiments with the two APE real world data sets: 161-day news and 147-day news with the same theme to show similar events evolution. For simplicity and clarity, we fix the number of topics, each time span we use the algorithms introduced before to choose hot topics. The comparison result is shown as in Fig. 4.

Topic Distribution Profile. As the expected result shown in Fig. 4, the topics are sequentially ordered by timeline. But the same event have different bursts. While Fig. 4(a) has one big burst on Aug.5 and two little waves in Sep. 7 and Oct. 20, Fig. 4(b) has only one big burst on about Sep. 20. That is, the same event may has different skeletons. What's more, there is an obviously difference that the topic streams in 08 is more complicated, that means the APE in 08 is much more fierce.

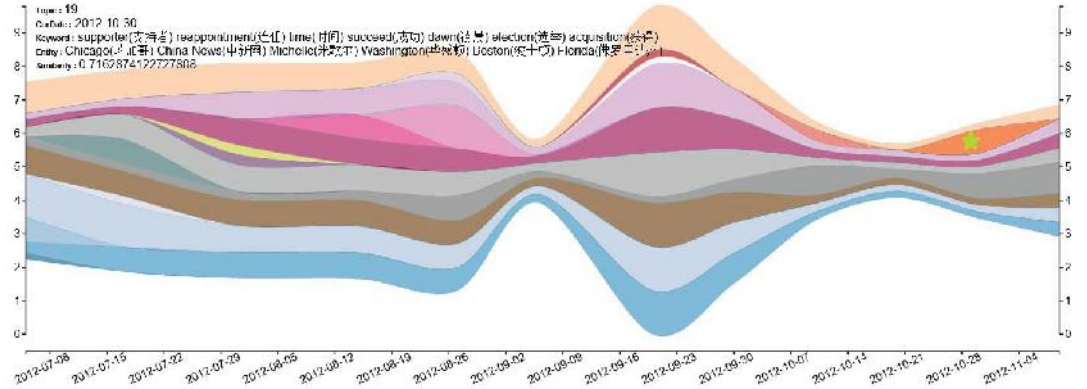
Topic Content Draft. Because it is a dynamic figure displayed with the browser, we could not see all of the topic annotations drafting. From the same topic in 08, we can easily guess that in Oct. 27, there may be a debate about the economy issue between Republican Party and Democratic Party. In addition, some people like McCain and Hillary are probably join it. And three places may have relationships with this debate. Another sample shows another topic annotation in Fig. 4(b), which may mainly reported the reappointment of Obama.

6. CONCLUSIONS

In this paper, we gave detailed explanations of event and topic, and built a novel framework which could show people the dynamic exploration of news event. We also formalized the three changing situations of a topic:



(a) APE08



(b) APE12

Figure 4. American President Election in 08 and 12

generation, continuation and disappearance. While applying our framework on both single event and similar events, we found that the result was significative for event retrieval in the future.

There are some future directions to extend our work. We will consider to use non-parameter method to generate topics of the event freely. That is, topics are no more dependent on our experiments, it only have the effect on how the event happened. We can expand this framework to display news streams which contains much more events in a whole picture.

7. ACKNOWLEDGMENTS

The work is supported by 973 Program(No. 2014CB340504), NSFC (No. 61035004), NSFC-ANR(No. 61261130588), FP7-288342, Tsinghua University Initiative Scientific Research Program (20131089256) MCM20130321 and THU-NUS NExT Co-Lab.

8. REFERENCES

1. A. Ahmed and E. P. Xing. Timeline: A dynamic hierarchical dirichlet process model for recovering birth/death and evolution of topics in text stream. *CoRR*, abs/1203.3463, 2012.

2. J. Alsakran, Y. Chen, D. Luo, Y. Zhao, J. Yang, W. Dou, and S. Liu. Real-time visualization of streaming text with a force-based dynamic system. *IEEE Computer Graphics and Applications*, 32(1):34–45, 2012.

3. L. Araujo, J. A. Cuesta, and J. J. M. Guervós. Genetic algorithm for burst detection and activity tracking in event streams. In *PPSN*, pages 302–311, 2006.

4. Y. Cai, Q. Li, H. Xie, T. Wang, and H. Min. Event relationship analysis for temporal event search. In *DASFAA (2)*, pages 179–193, 2013.

5. Z. Gao, Y. Song, S. Liu, H. Wang, H. Wei, Y. Chen, and W. Cui. Tracking and connecting topics via incremental hierarchical dirichlet processes. In *ICDM*, pages 1056–1061, 2011.

6. S. Havre, B. Hetzler, and L. Nowell. Themeriver: Visualizing theme changes over time. In *Information Visualization, 2000. InfoVis 2000. IEEE Symposium on*, pages 115–123. IEEE, 2000.

7. S. Havre, E. Hetzler, P. Whitney, and L. Nowell. Themeriver: Visualizing thematic changes in large document collections. *Visualization and Computer Graphics, IEEE Transactions on*, 8(1):9–20, 2002.

8. L. Hong, B. Dom, S. Gurumurthy, and K. Tsioutsiouliklis. A time-dependent topic model for multiple text streams. In *KDD*, pages 832–840, 2011.

9. M. Hu, S. Liu, F. Wei, Y. Wu, J. T. Stasko, and K.-L. Ma. Breaking news on twitter. In *CHI*, pages 2751–2754, 2012.

10. Y. Hu, A. John, F. Wang, and S. Kambhampati. Et-lda: Joint topic modeling for aligning events and their twitter feedback. In *AAAI*, 2012.

11. L. Kong, R. Yan, H. Jiang, Y. Zhang, Y. Gao, and L. Fu. Mining event temporal boundaries from news corpora through evolution phase discovery. In *WAIM*, pages 554–565, 2011.

12. J. Leskovec, L. Backstrom, and J. M. Kleinberg. Meme-tracking and the dynamics of the news cycle. In *KDD*, pages 497–506, 2009.

13. J. Li, J. Li, and J. Tang. A flexible topic-driven framework for news exploration. In *Proceedings of KDD*, volume 2007, 2007.

14. C. X. Lin, B. Zhao, Q. Mei, and J. Han. Pet: a statistical model for popular events tracking in social communities. In *KDD*, pages 929–938, 2010.

15. Q. Mei and C. Zhai. Discovering evolutionary theme patterns from text: an exploration of temporal text mining. In *KDD*, pages 198–207, 2005.

16. D. Newman, C. Chemudugunta, and P. Smyth. Statistical entity-topic models. In *KDD*, pages 680–686, 2006.

17. X. Wang, C. Zhai, X. Hu, and R. Sproat. Mining correlated bursty topic patterns from coordinated text streams. In *KDD*, pages 784–793, 2007.

18. X. Wang, F. Zhu, J. Jiang, and S. Li. Real time event detection in twitter. In *WAIM*, pages 502–513, 2013.

19. F. Wei, S. Liu, Y. Song, S. Pan, M. X. Zhou, W. Qian, L. Shi, L. Tan, and Q. Zhang. Tiara: a visual exploratory text analytic system. In *KDD*, pages 153–162, 2010.

20. R. Yan, X. Wan, M. Lapata, W. X. Zhao, P.-J. Cheng, and X. Li. Visualizing timelines: evolutionary summarization via iterative reinforcement between text and image streams. In *CIKM*, pages 275–284, 2012.

21. K. Zhang, J. Zi, and L. G. Wu. New event detection based on indexing-tree and named entity. In *SIGIR*, pages 215–222, 2007.