

# Mixed-Initiative Visual Exploration of Social Media Text and Events - IEEE VAST Challenge 2021 MC3 Award for Strong Human-in-the-Loop Analysis Methodology

Lei Peng  
Fudan University

Yuheng Zhao  
Fudan University  
Junting Gao  
Data Star Observatory

Yijie Hou  
Fudan University  
Jiaqi Dong  
Fudan University

Qinghong Wang  
Data Star Observatory  
Ziyue Lin  
Fudan University

Siqi Shen  
Fudan University  
Siming Chen\*  
Fudan University

Xingui Lai  
Fudan University

## ABSTRACT

In this paper we present an interactive visualization system designed for solving IEEE VAST Challenge 2021 Mini-Challenge 3. The system enables interactive classification and mining of social media texts, supports user-defined event analysis, and displays risk of public temporally and spatially. It features mixed-initiative visualization design and human-in-the-loop analysis process.

## 1 INTRODUCTION

IEEE VAST Challenge 2021 Mini-Challenge 3 requires participants to retrospectively detect and evaluate public risks in Abila City during the evening of January 23, 2014. The provided data are microblog records and emergency dispatch records from local police and fire departments. We see this challenge as a social media visual analytics task which places emphasize on risk detection and evaluation [2].

Generally, our visual analytics solution has two integrated interactive steps: microblog classification (step 1) and detailed analysis (step 2). The user may first start from an interactive classification and mining process of microblog records to identify normal records, spam and noises (step 1). After observing crucial information, he or she may then interactively explore user-defined risk events in the spatial temporal scenarios (step 2). During the whole analysis process, the user's knowledge of the situation deepens through the iterative exploration.

## 2 VISUALIZATION SYSTEM DESIGN

Our visualization interface is mainly built with D3.js [1] and Vue.js. The system interface has a brushable timeline on the left (Fig. 1.A), a filter panel on the right (Fig. 1.C), and a scrollable central view which can switch between the message view (Fig. 1.B1), the entity graph (Fig. 1.B2), and the general analysis view (Fig. 1.B3).

**Timeline:** The timeline is a crucial part of our system. All the other views are linked to the timeline, and the timeline itself can be rendered based on message type, classification result, user-defined keywords, or sentiments.

**Message View:** In the message view, each category of messages is displayed with its corresponding word cloud. We use a combination of machine learning and human interaction to pre-classify relevant messages, chatter and spams. We first encode messages into tensors with the BERT [3] model and then use a lightGBM [4] model (which is trained with 150 manually labeled messages) as a classifier. Since there might be misclassifications, users can directly relabel a certain message by clicking on the relabel button. The two circles below the relabel button on each message card represent sentiment and objectivity of the message.

\*e-mail: simingchen@fudan.edu.cn

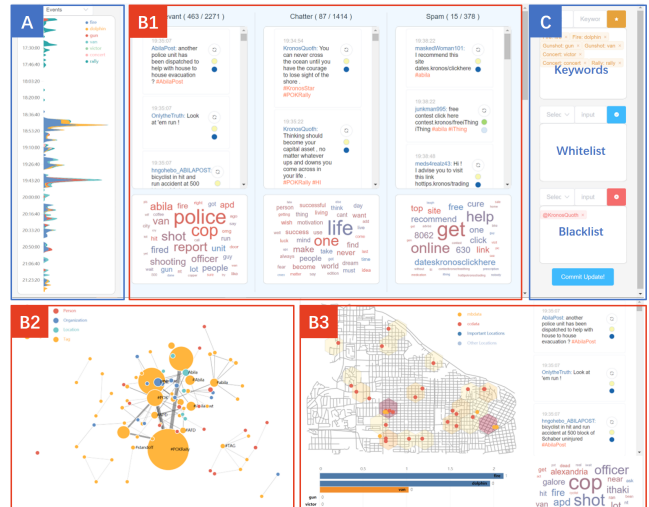


Figure 1: The user interface consisting of: (A) Timeline. (B1) Message View. (B2) Entity Graph. (B3) General Analysis View (C) Filter Panel with the event keyword tool, the white list and the black list.

**Black List and White List:** Users can also do batch operations with the help of the black list and the white list. For example, the user finds that @Clevvah4Evah sent lots of meaningless messages, and @KronosQuoth is a typical chicken soup blogger. Therefore, he or she can add them into the black list to prevent them from appearing in the relevant message set. If the user thinks #AbilaFire must be an important tag directly related to a public risk event, he or she may add it into the white list to avoid potential misclassifications from the model.

**Entity Graph:** From the raw texts, we extract entities such as authors, tags, organizations, locations, and persons to build an entity graph. Two entities are linked together if they both appear in one microblog record. The graph can provide us with event clues and entity relationships.

**Event Keyword Tool:** Having explored with the messages, the word clouds and the entity graph, users may then become curious about several keywords of different events that happen during this evening. Users can now input self-defined events with corresponding keywords into the event keyword tool. The system will render the timeline with the appearance of keywords, generate a bar chart race under in the general analysis view, and highlight every message card that contains a keyword (Fig. 2).

**Map:** We provide a map of Abila City marked with relevant information. The blue rectangles approximately represent the public places that we identify in Abila City. During a certain time period, we mark the locations of microblogs and emergency dispatches

