IoTNegViz: An interactive tool for visualizing negative aspects of IoT

***Abstract*—The "Internet of Things", or IoT, brings to people a brand new approach with a wide range of applications of how we live in this ever changing world. Besides various advantages of IoT that users can benefit from, the downsides of such cyber physical system are often neglected. This paper introduces *IoTNegViz*, an analytic tool for visualizing possible negative aspects from a cyber physical system in IoT. First, the data is extracted from Twitter, then a natural language processing tool is applied for generating and categorizing keywords. The visualization is utilized for exploring and capturing information in a spatial manner. To evaluate the usefulness and efficiency of *IoTNegViz*, we conduct a use-case to gather feedback and experience form users.**

***Keywords*-IoT, Cyber Physical System, WordCluster, Negative aspects, Natural Language Processing**

I. Introduction

As the term "Internet of Things" (hereafter, IoT) becomes more and more popular as a result of various domains it applies to, its potential downsides should be taken into account. IoT can be describe as a giant network in which data are integrated, devices are connected to each other and to the backbone - the Internet. Therefore, the subsequent risks from such cyber physical systems can become severe due to the pervasive influence on many aspects in life. These shortcomings include, but are not limited to, cyber attacks, privacy risks, and safety violations.

It is important to understand the downside of IoT, which in turns can help scientists and domain experts to be prepared, raise awareness and alleviate the existing problems. Therefore, there is a need to have a synthesized method that allows users to capture deluge information with ease in the ocean of data. Response to this need, this paper introduces *IoTNegViz* that enables users to explore and investigate a large number of corpus available on social media. IoT enthusiast are able to quickly detect challenging terms and its related events over time. Our approach is different from the existing techniques in terms of combining both spatial and temporal perspectives. In particular, the contributions of this paper are:

- We propose an approach for exploring and exploiting downsides of IoT, a class of cyber physical systems;
- We develop and implement a visual analytics tool, called *IoTNegViz* based on the approach proposed.
- We evaluate the usefulness and feasibility of *IoTNegViz* with a case study.

The rest of the paper is organized as follows. We present existing approaches in terms of visualization perspective in the following section. Design considerations and design choices are introduced in Sect. III. Then we describe the *IoTNegViz* interface and its components in Sect. IV. User interactions are described in Sect.V. We illustrate the use of *IoTNegViz* on social media data and present the results of our user study in Sect. VI. Finally, we conclude the paper and discuss future work in Sect. VII.

II. Existing Approaches

Latent Dirichlet Allocation (LDA) [1] is used to find potential topics in the text corpora using a flexible generative probabilistic model for data mining based on the words. LDA classifies the topics based on distribution across the words through a three-level hierarchical Bayesian model. In this approach, documents are interpreted as a distribution over topics and topics are formed by a certain number of correlated words.

Topic modeling allows us to discover, organize, re-order, and summarize the topics from the large text corpora in an efficient way and hence many good visualization tools and techniques have been developed for the visualization of topics based on topic modeling. TIARA [2], a topic visualization tool developed by Wei et al. which determines time-sensitive keywords to portray the content evolution of each topic over time using stack graph metaphor. ParallelTopics [3] represent the temporal changes of topics using Parallel Coordinates view. It enables the user to examine large text corpora in a structured way to understand the correlation between the terms. In this paper, we use the disease ontology defined in the Healthcare Hashtag Project [4] in order to quickly narrow down to disease-related tweets discussed in Sect. III.

In text visualization, *Wordle* [5] is widely used because it could locate words into two-dimensional space and could quickly give emphases to important words using font-size and colors. *Wordle* uses a randomized greedy algorithm to place words randomly and relatively close to the center of the allocated position. It is greedy since it prioritizes the words with the larger sizes (more frequent). In addition, *Wordle* is aesthetic and visual appealing [6]. In the past few years, there many efforts to optimize the *Wordle* layout. *ManiWordle* [7] provides more flexible control over how to form the word layout with the interaction from the users. Rolled-out Wordles [8] makes use of Linear Sorting (*RWordle-L*) and Concentric Sorting (*RWordle-C*) to place the words more compactly and still preserve the orthogonal ordering and topology. WordlePlus [9] extends the idea of ManiWordle to provide some further natural interaction supported for pen- and touch-enabled tablets while controlling the overall *Wordle*

978-1-5386-5035-6/18/$31.00 ©2018 IEEE

layout such as resizing, adding, deleting elements. A recent work, called EdWordle [10], allows editing the *Wordle* layout but still reserving the word neighborhoods (related terms should be close to one another). These work mostly focus on extending/optimizing the *Wordle* layout and discard the time element.

III. The *IoTNegViz* Approach

IoTNegViz[1] is developed using JavaScript and in particular the D3.js library [11]. The primary goal of *IoTNegViz* is to create an interactive visual analytic tool that presents cyber physical systems and IoT experts a high level view of negative aspects from an IoT system with spatial and temporal perspectives. The tool enables users to interact and investigate the relationships of individual issues in spacial and chronological manners. To meet this goal, this paper proposes several features that are implemented in *IoTNegViz*:

- **Overview Display (F1)**. Display overview of the issues' distribution.
- **Details-On-Demand (F2)**. Present details on demand of the issues.
- **Spatial Relationship (F3)**. Show the relationship between concerning words in an intuitive manner.
- **Temporal Relationship (F4)**. Visualize trending words and their connections over time.
- **Filter (F5)**. Search or filter out desired IoT-related issues.

A. *The Format of the Processing Dataset*

The dataset for this study was retrieved from Twitter[2] API, based on some filtering keywords such as "internet of things", "iot", "vulnerability", "cyber physical systems". The purpose of these keywords is to extract information related to IOT and cyber physical systems issues. The raw data set contains 2597 tweets. Terms and topics are extracted from tweets by using Natural Language Processing library (i.e., spaCY[3]), resulting in two topics and 26519 terms. We classify topics based on two categories (security and physical crash). Terms are classified into topic based on NLP. The extracted topics and their terms will be fed into the *IoTNegViz* tool. The flow of data visualization process is depicted in Fig. 1.

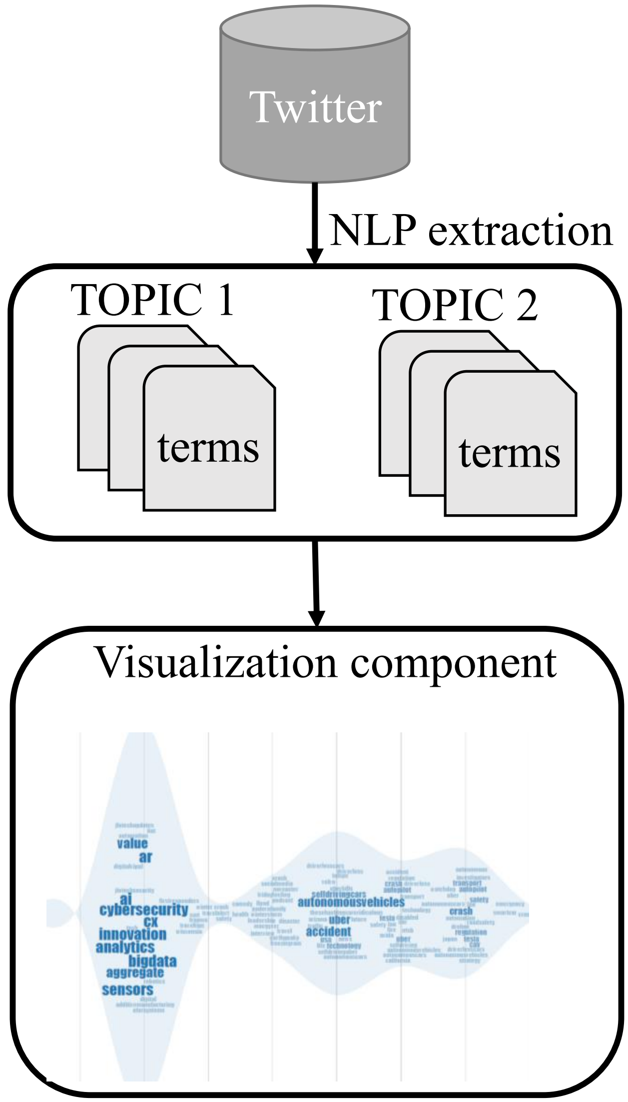

Figure 1. The Schema of the *IoTNegViz*

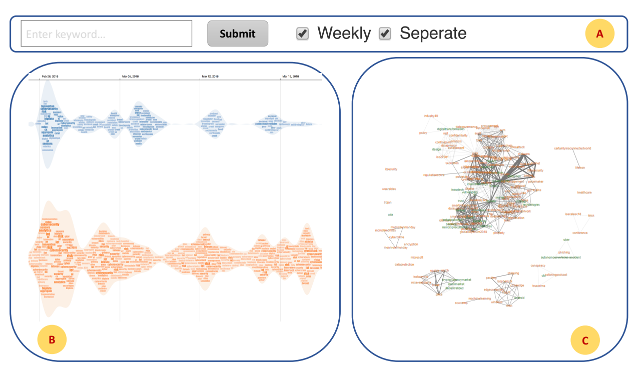

Figure 2. Three main components of the *IoTNegViz*: A) Utility component, B) Topics stream component, C) Dynamic network component

IV. The *IoTNegViz* Architecture

IoTNegViz consists of three main components as depicted in Fig. 2 where 1) Box A contains *Utility* component, 2) Box B displays *Trending Terms* component, and 3) Box C shows the *Dynamic network* component.

Box A - The *Utility* component The *Utility* component allows users to search keywords or terms of interest, layout of the viz tool in Box B can be reorganized in terms of time (daily or weekly by checking/unchecking the "Weekly'

[1]For a short demo of *IoTNegViz*, visit https://goo.gl/UBi8eQ/
[2]https://twitter.com/
[3]https://spacy.io

option) or uniformly (when users want to see each topic in a separated view or a single view by checking/unchecking the "Stack' option)

Box B - The *Trending Terms* component encourages users exploring interesting terms for each topic (security and physical crash) over time. The *Trending Terms* provides an overall distribution of terms during a specified period (the visualization feature **F1** and **F4**). In the upper part of the graph, a list of time steps is given to show the corresponding timestep of the stream below. The shape of the stream directly associates with the stream's thickness, which indicates the total frequency of the inside terms. Terms are organized in such a way that overlaps are avoided and all terms are fit within the outbound shape. The significance of each word is constructed based on its frequency, then emphasized in the visualization by its size and color opacity. From the *Trending Terms*, users can recognize the terms' evolution and changes in its importance over time [12]. Fig. 3 shows an example of the security topic with terms.

Figure 3. An enlarged version of the security topic with terms

V. User Interaction

The *IoTNegViz* supports three types of interactions on components as follows:

Mouse over:

When users mouse over a trending term, a tooltip will be popped up to display all tweets (the visualization feature **F2**) that contain the current terms. More interestingly, related emerging terms are also highlighted within the term's topic and across other topics. Fig. 4 illustrates an example when users mouse over a term "CyberSecurity", related terms in other topic within the same period as well as related terms in the same topic are highlighted.

Mouse click: When users click on a given terms, a list of related terms (colored by topics) in the same period will be displayed below the selected term.

Search box: To narrow down the results shown on the overview, or to explore contents of interest, *IoTNegViz* provides a search box (the visualization feature **F5**), on the top left of the screen, that allows user to input desired

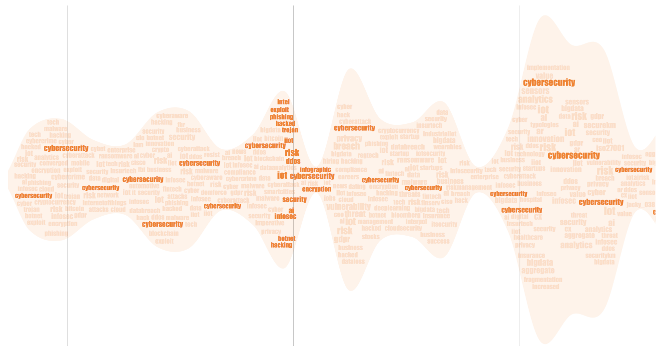

Figure 4. Mouse over a term show its long term impact and related terms in the same period

information. The components update the results according to the search word interactively (as depicted in Fig. 5.

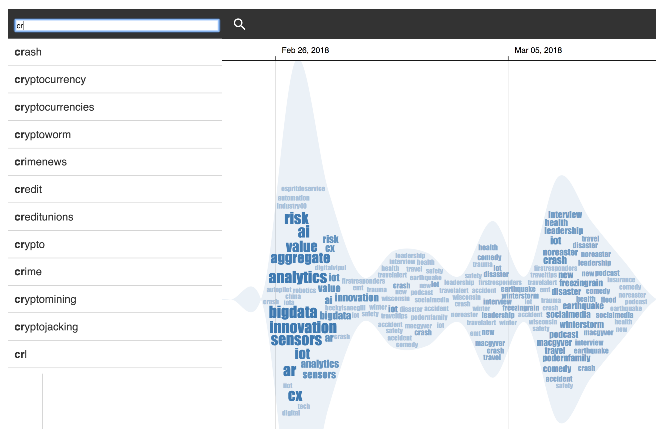

Figure 5. Search box with "auto complete" widget

VI. Case Study

To evaluate the usefulness and effectiveness of the *IoTNegViz*, we conducted a use case with three volunteer users who have experience on utilizing social media as a means to update information. The purpose of this study is to gather users' feedback which can be used as an indicator to improve *IoTNegViz*. Each user study is carried out in approximately 15 minutes. An introduction about the purpose of *IoTNegViz*, the source of data, the structure of the visualization representation is presented to individual before moving on to the actual experiment. After playing with the application, users are asked by a set of proposed questions to measure the level of understanding on the tool.

The questions are as follows:

- (R1) What are the issues that get the most concern? Which time step has the biggest amount of interest?
- (R2) What are the most interesting patterns observed from *IoTNegViz*?

Findings:

For the first question (R1), users are able to find the issues that raise the most concern based on the big font size and bold color, these terms are also emphasized by the repeatedly pattern throughout the timeline. In addition, to identify the time step that acquire the biggest amount of interest, users feel at ease to detect them according to the points that obtain large thickness of the streams.

For the next question (R2), users pay more attention to the specific terms within each stream. Users recognized that *risk* and *crash* are the most popular words that span over a long period of time of the two categories shown. One interesting finding for the term *crash* is that in spite of its popularity along the timeline, it is almost never the word with high frequency within any time step.

Furthermore, users noticed that there are groups of words that often appear together. For example, *cybersecurity*, *iot* (IoT) and *analytics* are tend to be seen together as shown in Fig. 6. This may shows the strong relation between the meaning of these terms.

Regarding to the *cybersecurity* term, this pattern is clearly visible in the dynamic network, i has strong and multiple relations with the terms *datasecurity*, *gdpr*, *bigdata* and *irc*.

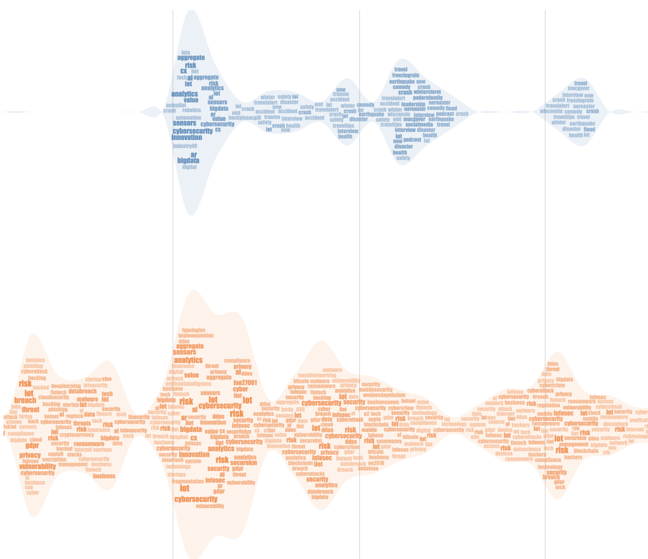

Figure 6. The terms *cybersecurity*, *iot* and *analytics* repeatedly appear together

VII. Conclusion and Future Works

In this paper, we introduce an interactive data analytic to help users to summarize the IoT-related concerning issues. *IoTNegViz* supports of a range of interactive features, such as linking and filtering, allowing users to quickly narrow down events of interest. We demonstrate the usefulness of our tool through different use cases. An important aspects of the paper is to get users to raise awareness, be prepared, and therefore alleviate the existing problems.

References

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[2] 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 *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010, pp. 153–162.

[3] W. Dou, X. Wang, R. Chang, and W. Ribarsky, "Paralleltopics: A probabilistic approach to exploring document collections," in *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*. IEEE, 2011, pp. 231–240.

[4] "Healthcare Hashtag Project, howpublished = https://www.com/healthcare-social-media-research, note = Accessed: 2017-05."

[5] F. B. Viegas, M. Wattenberg, J. Feinberg, H. Dwwhqehuj, D. Q. G. Hlqehuj, P. Dxglhqfh, Z. Dv, R. X. U. Zrun, L. V. Ylhzhg, P. Ri, W. Rq, V. Zrug, F. Ru, Y. Dwlrq, P. Wdnh, L. Q. Pdvv, S. Ri, F. W. Srjudsk, D. Q. G. Frpsrvlwlrq, F. Volghv, D. Q. G. Krxvhv, R. I. Zruvkls, and W. K. H. Kdyh, "Participatory Visualization with Wordle," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 15, no. 6, pp. 1137–1144, 2009.

[6] J. Feinberg, "Wordle," *Beautiful Visualization: Looking at data through the eyes of experts.*, pp. 37–58, 2010.

[7] K. Koh, B. Lee, B. Kim, and J. Seo, "ManiWordle: Providing flexible control over wordle," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1190–1197, 2010.

[8] H. Strobelt, M. Spicker, A. Stoffel, D. Keim, and O. Deussen, "Rolled-out Wordles: A Heuristic Method for Overlap Removal of 2D Data Representatives," *Computer Graphics Forum*, vol. 31, no. 3pt3, pp. 1135–1144, 2012. [Online]. Available: http://doi.wiley.com/10.1111/j.1467-8659.2012.03106.x

[9] H.-c. D. Visualization, "WordlePlus : Expanding Wordle ' s Use through Natural Interaction," *IEEE Computer Graphics and Applications*, 2015.

[10] Y. Wang, X. Chu, C. Bao, L. Zhu, O. Deussen, B. Chen, and M. Sedlmair, "EdWordle: Consistency-Preserving Word Cloud Editing," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 647–656, 2018.

[11] M. Bostock, V. Ogievetsky, and J. Heer, "D^3 data-driven documents," *IEEE Transactions on Visualization & Computer Graphics*, no. 12, pp. 2301–2309, 2011.

[12] T. Dang and V. T. Nguyen, "ComModeler: Topic Modeling Using Community Detection," in *EuroVis Workshop on Visual Analytics (EuroVA)*, C. Tominski and T. von Landesberger, Eds. The Eurographics Association, 2018.