Using Text Visualization to Aid Analysis of Machine Maintenance Logs

Senthil Chandrasegaran[1], Xiaoyu Zhang[1], Michael P. Brundage[2], and Kwan-Liu Ma[1]

[1] Department of Computer Science, University of California, Davis, Davis, CA
{schandrasegaran, xybzhang, klma}@ucdavis.edu
[2] National Institute of Standards and Technology, Gaithersburg, MD
michael.brundage@nist.gov

Abstract

Maintenance and error logs for machines in manufacturing organizations are typically written as informal notes by operators or technicians working on the machines. These logs are written using a combination of common language and internally-used abbreviations and jargon. Due to inconsistencies in the terminology used during error logging and in identifying root causes of issues, the data needs to be cleaned before automated analyses can be effectively used. This can require a human to go through and clean/tag the data, disambiguate multiple terms, and sometimes assign additional tags to the data objects to aid automated classification. With some organizations storing over a million records of legacy maintenance report data, this is not entirely feasible. We introduce a visual analytic approach to help analysts sift through such heterogeneous datasets so that the inconsistent data can be tagged and categorized with minimum manual effort. Though such data typically includes metadata such as date, time, severity, machine IDs, etc., in this paper we focus on the manually-entered text descriptions. We use metrics such as word occurrence frequency and information-theoretic metrics to visually highlight common and uncommon issues and fixes that occur in the maintenance logs. We illustrate our approach with data from industry and discuss future research directions to address scalability, metadata, and other approaches for grouping similar logs.

1 Introduction

Machine error diagnosis and prediction is an issue in which the manufacturing industry heavily invests due to its direct effect on machine availability and throughput. Some organizations often maintain cross-functional teams of engineers with expertise covering design, analysis, and manufacturing to help identify and correct such problems quickly. With the growth of Smart Manufacturing and inexpensive and easy to use sensors, the demand for data-driven solutions for machine diagnostics and prognostics has increased. Organizations thus maintain machine maintenance and error logs to help analysts identify patterns and subsequently formalize root-cause analysis of errors. This, in turn, helps organizations plan preventive and predictive maintenance practices.

However, maintenance and error logs for machines can be both human- and sensor-generated. Human-generated logs are typically written as informal notes by technicians, who often use their own jargon when referring to machines, parts, and processes. These terms are frequently not consistent across groups in the organization, making it difficult for analysts to identify similar logs. Sensor-generated logs tend to be highly general and lack relevant and contextual machine-specific information. This data poses the opposite problem: sensors often have similar logs even though the logs are generated from a variety of machines for a variety of problems. Finally, both

DOI:https://doi.org/10.6028/NIST.AMS.100-29#page=86

human- and machine-generated logs are sometimes inconsistent: humans are prone to errors in identification and labeling of symptoms and diagnoses, while sensors can have errors that may result in erroneous/missing logs or corrupted data.

There are ongoing efforts to clean, consolidate, tag, and categorize maintenance and error logs to aid automated diagnostics and prognostics. This cleaning effort requires manual tagging and repairing of data, disambiguation of terms, and assigning of specific terms to aid automated classification of the data. With some organizations storing over a million records of legacy maintenance report data, this is not entirely feasible without aid of tools. Recent semi-automated approaches have used the human in the loop along with natural language processing techniques to aid the above disambiguation and tagging. However, these methods still require making assumptions in the process of cleaning and categorizing data to extract useful and/or actionable information.

To aid human analysts in viewing large datasets, grouping them, and observing patterns and anomalies that aid labeling and categorization of data, we propose the use of visual analytics. The science of visual analytics supports data analysis using computational techniques and interactive visualizations [7]. Specifically, it allows analysts to forage for information, collect evidence, and form schema that leads to hypotheses, a process called the visual sensemaking loop [20]. In this paper, we introduce a visual analytics approach meant for aiding qualitative text analysis and categorization, and apply it to the analysis and categorization of machine log data. We focus on the manually-entered text descriptions and outline requirements that need to be fulfilled to manually analyze and tag such log data for better sensemaking. We describe how the visual analytic approach addresses these requirements, and illustrate the approach with a use-case scenario of maintenance log data from the industry. We close with recommendations for incorporating metadata and approaches for better scalability.

2 Background

With increasing emphasis on smart manufacturing and a push toward eliminating machine downtime, process monitoring, diagnostics, and prognostics have gained prevalence. The complexity and volume of data that needs to be sifted through to achieve this improved maintenance of equipment have prompted the application of visual analytics into product lifecycle management (PLM) [21]. This potential application area was anticipated almost at the inception of visual analytics when Keim et al. [17] suggested that visual analytics may be used in engineering for analyzing complex data that arise from design, production, and feedback from product use. In this section, we will provide a background on the complexity of making sense of machine error and maintenance logs, and the application of visual analytics to address this complexity.

2.1 Processing Human- and Sensor-Generated Logs

System log analysis—analysis of logs automatically generated by the system—is commonly used to track system resilience. It is also used in the case of failure for root cause analysis and in the case of preventive maintenance to identify recurring patterns, such as temporal, systemic, or even seasonal. Typical tools used for such analyses use automated analyses and seldom resort to visualization approaches.

Automatic log analysis tools [8, 13, 14, 16, 19, 30] typically use a range of analyses such as correlation analysis, signal analysis, pattern mining, correlations, resilience analyses at the application level, and spatial/temporal event analysis. For instance, HELO (Hierarchical Event Log Organizer) [14], an event log mining tool, extracts event formats by pattern-mining log files from large-scale supercomputers, using predefined message templates. A model-based approach is used by ELSA (Event Log Signal Analyzer) [13], a toolkit for event prediction. It models the normal flow at a stable event state, and in the event of system failure, tracks the abnormal flow of events using a combination of data mining and signal processing.

There exist visualization-oriented tools for tracking and analyzing machine logs, but these are few, and most of them use relatively basic visualizations. For instance, LogMaster [12] and LogAider [8] use generic visualizations for mining event correlations. LogAider reveals potential correlations that include across-field (through probabilistic analysis of fields), spatial, and temporal correlations. LogMine [16] is a framework for the unsupervised, scalable end-to-end one-pass analysis of large-scale, heterogeneous logs. LogDiver [19] supports lossless data compression, models application failure paths, and cross-validates models and/or results of analyses. More recently, machine learning approaches such as DeepLog [10] have been introduced. Specifically, Deeplog uses a deep neural network model which uses stacked Long Short-Term Memory (LSTM) to detect anomalies, and dynamically updates the models to accommodate for changing log patterns.

The idea of using visualization and visual analytics for monitoring and diagnostics in factories is a relatively new research area. Recent work includes ViDX [29], a visual analytic system for historical analysis and real-time monitoring of factory assembly lines. ViDX uses visualization principles to create outlier-aware aggregate representations of process data and employs user-steerable algorithms for outlier detection. La VALSE [15] is a scalable log visualization tool that uses multiple visualizations for interactive event analysis based on multiple logs. ViBR [5] is a system that visualizes bipartite relationships using a minimum description length principle to aggregate the relationships. The system has been successful in log analyses that include vehicle fault diagnostics by identifying co-occurring faults, comparing faults that co-occur in different vehicle clusters, and comparing faults across vehicles with shared properties.

2.2 Visual Analytics for Text Data

Root cause analysis and preventative action is a crucial area of interest to the manufacturing industry, necessitating logging maintenance and error log data, as discussed earlier. Approaches to parse this data for an automated or even semi-automated solution for diagnosis or prognosis has typically involved knowledge bases [4], manual "tagging" systems assisted by natural-language text parsing support [23][1], and information extraction methods applied to maintenance logs [24, 25]. While our approach also proposes the use of natural language processing (NLP) techniques, we use visual analytics to keep the human in the loop for correcting and tagging the parsed data through the visual representation of and interaction with the data processing results.

Defined as "the science of analytical reasoning facilitated by interactive visual interfaces" [7, p. 4], visual analytics uses visualization support throughout the *process* of analyzing (typically unstructured) datasets. In other words, visual analytics makes "*our way of processing* data and information transparent for an analytic discourse." [17, p. 155]. At the center of all visual analytic systems is the analyst—the human in the loop—who is aided by the system in combining complex datasets, collect evidence, identify correlations, and develop insights. At every stage of this process, the analyst is aided by a combination of visualizations and algorithms.

Visual analytics support for text analysis often focuses on analyzing connections between multiple sources of text, from intelligence reports to news articles to even unstructured social

[1]An open source tool for this process, called Nestor is available here: https://www.nist.gov/services-resources/software/nestor

media texts such as tweets and posts on forums. Some of the earliest text analytic tools were designed for intelligence analysis. Of these, Jigsaw [26] is one of the more prominent and still-used tools. It identifies connections between documents using entities in text data and metadata, highlights these connections to the user, and allows the user to reorganize this information to aid their insight-gathering process. It uses coordinated views such as graphs, calendars, and document overviews, all of which can be filtered and edited by the analyst to identify potential security threats. Other approaches make more use of metadata, such as time-stamps. For instance, Tiara [28], a system for temporal analysis of text documents, is used to analyze data relevant to emails, instant messages, and even patient records. It uses statistical text analysis techniques such as topic modeling to categorize the document collections thematically based on their content, and shows the variation of themes over time. It also allows users to select and examine any theme-based collection in detail, across and at defined time intervals. Other topic-modeling-based text analysis tools include HierarchicalTopics [9], which as the name suggests, uses a hierarchical topic modeling algorithm to identify themes within themes. It combines this with a temporal view showing the evolution of topics over time and allows users to explore and edit topics hierarchically. Other approaches are more suitable for single or very few, but large documents such as historical texts. An example is VariFocal Reader [18], which uses automated annotations and topic modeling to reveal thematic and structural patterns that are useful when analyzing large documents.

In this paper, we adapt our prior work that uses visual analytics with a dominant text visualization component that we developed to aid qualitative analysis of text data [6]. We do this by helping the user identify concepts of interest, categorize associated text, and use their custom categorizations to further analyze the text. We illustrate the suitability of this approach in helping users identify patterns and inconsistencies in any terminology used in machine logs. This will help analysts create useful categorizations of machine logs that will help problem diagnosis, and to subsequently create machine learning models.

3 Design

While individual fields of machine maintenance logs may vary between organizations, they usually have some common features, such as the machine identifier, problem description, the description of the remedial action taken, and the dates on which the problem was reported and closed. While it is possible to "group" these logs by some of the features such as machine ID, it becomes less obvious to group the logs based on the type of problem, the type of solution, or patterns in the dates on which they tend to occur. Such categorization often requires the expertise and insight borne by experience. Our goal is to help such experienced personnel sift through and examine large datasets without needing to examine each record closely.

3.1 Design Rationale

We draw from research in visual analytics—"the science of analytical reasoning facilitated by interactive visual interfaces" [7, p. 421]—to design an appropriate interface for our approach. We identify the following requirements for prognostics of machine maintenance.

- **R1 Identify Common Occurrences:** One of the main requirements in machine maintenance log analysis is to identify recurring problems that—while individually may not cause significant downtime—through their frequency of occurrence cost significant resources in repairing and downtime. These may not always be linked to the same kinds of machines, or even have the same descriptions.

R2 Identify Patterns in Occurrences: Some maintenance issues may manifest as several problems that occur together or in succession to cause a much more significant issue than the individual reports suggest. Other issues may occur only in some kinds of machines, or when some operators are working certain machines, or even certain days of the week, month, or year. Combined with the earlier-identified issue of inconsistency in the descriptions, the need to identify patterns in maintenance logs is only matched by the challenges posed in identifying such patterns.

R3 Identify Anomalies: When taking stock of problems that occur over a long period, there may be a need to identify rare, yet significant problems. These could refer to the problems themselves, or their rare occurrence in a specific machine or part. Such anomalies could be lost to cursory scrutiny when looking for commonly-occurring problems, but if ignored could escalate over time.

R4 Allow Manual Categorization: Identifying patterns, anomalies, and common issues is often not a single-stage process. The relevant analyst or domain expert may need to tag certain groups of problems with a descriptor, add a memo for continued monitoring, or even need verification from a colleague.

R5 Aid Iterative Analyses: Once manual categories are identified, the system should allow the user to filter the existing data using these categories, which will further reveal commonly-occurring keywords.

3.2 Interface Design

Based on the above rationale, we decided on a primarily text-based visualization approach, shown in Fig.1. The visualization is largely extended from our prior work on developing a visual analytic approach to aid qualitative text analysis [6]. Since our focus in this work is on the content (and less so on the metadata), the text component of the data is shown in the central panel. These descriptions are logged by the machine technicians and/or operators and include reports that can describe the problem, the solution, or both. This being a preliminary approach that examines how the existing qualitative analysis system can be used for analyzing patterns in the data, we do not incorporate temporal or other metadata such as machine IDs, operator IDS, severity or cost-related information.

The text shown in the central column follows a "skim formatting" [3] where the font weight for each word corresponds to a predefined criterion. In our case, we use the word information content [22], which is based on the assumption that the less frequently-used a word is in a corpus, the more information it contains. For a more focused application, we can use analyst-defined metrics that give greater weight to keywords associated with rare and severe issues (requirement **R3**) in conjunction with—or instead of—such generic metrics. The information content metric can also be computed on a specific domain, such as existing corpora of operation or repair manuals.

On the right is a word cloud that is automatically computed from the uploaded text. It is scaled proportional to frequency and the words are arranged sequentially in descending order of frequency. The skim formatting described earlier is applied to the word cloud as well. Selecting a set of text in the central column will filter the word cloud to reflect only the selection. The word cloud can highlight commonly-occurring terms in the maintenance logs (requirement **R1**).

A set of checkboxes on the bottom are used to highlight parts of speech or named entities (person, place, names). These can be useful to highlight when the analyst is looking for logs that mention geographic locations, or when the names of operators/repair persons are mentioned.

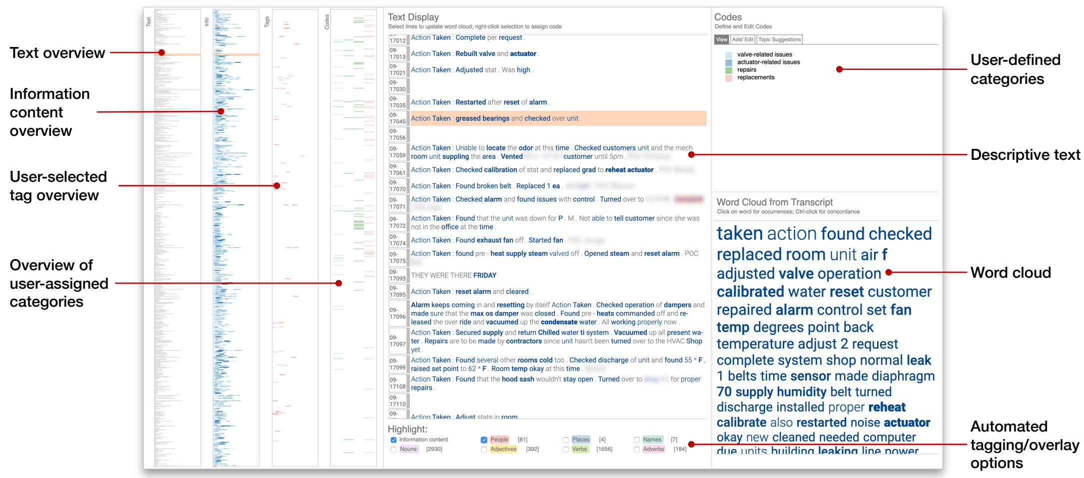

Figure 1: The interface adapted from our earlier work in Chandrasegaran et al. [6], shown here with approximately 600 records of maintenance logs for HVAC (heating, ventilation, and air-conditioning) systems at a specific site.

An overview pane on the left shows several overview visualizations. The first is a "text overview" that simply gives a mini-map view of all the records in the collection. It also shows the position of the current record of interest (as an orange bar) corresponding to the text in the central panel on which the mouse currently hovers. Selecting a word in the word cloud (Fig. 2) shows all its occurrences in the main text as well as in the overview panel (requirement **R2**).

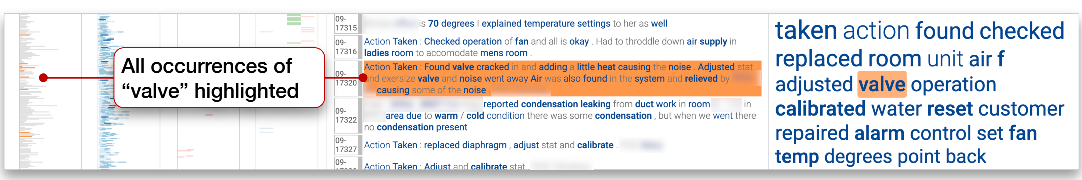

Figure 2: A detail of the interface showing how selecting a word from the word cloud (right) highlights all records where the word occurs in the text overview and detailed view panes.

Additional overviews include an information content heatmap that provides an overview of the skim formatting described earlier, parts-of-speech/named entity tag overviews, and an overview of user-applied categories. These categories are specified in an input field on the top right part of the interface (Fig. 1). Once the categories are specified, they can be assigned to individual fields or groups of sequential fields as shown in Fig. 3 (requirement **R4**). Assigning a category to one or more fields of text updates the overview visualization immediately to the left of the text display. Once the categories of interest have all been assigned, co-occurring problems and—once temporal data can be integrated—temporal and recurring patterns can be visually identified, and these co-occurrences can be further tagged and newer categories assigned to them iteratively (requirement **R5**).

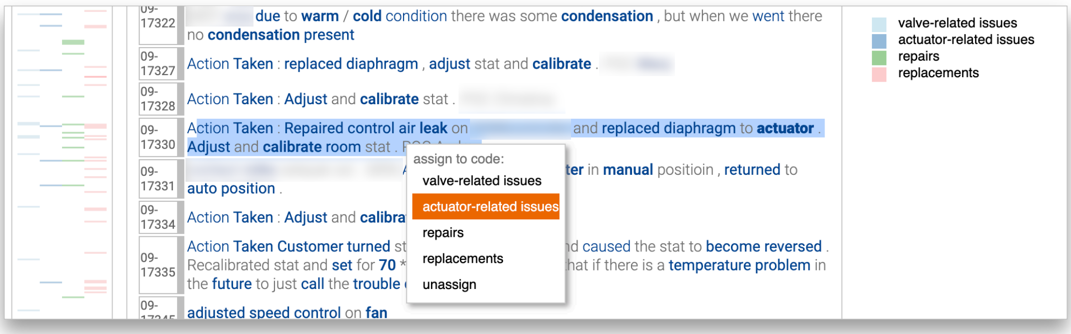

Figure 3: Detailed view of the text and code (category) definition and overview fields showing the categories assigned. A category is assigned by selecting a block of text and assigning a category from a drop-down menu shown above. When a category is assigned to a text, it updates the overview visualization on the left.

4 Implementation

As explained in Section 3, the system presented here is adapted from our earlier work directed at qualitative text analysis [6]. The system is implemented as a web-based application in HTML5 and JavaScript, with a Node.js backend where the data is uploaded and processed to be visualized on the browser. Most of the language processing operations, including tokenization, parts-of-speech tagging, named-entity recognition, and information content measurement are performed at the server end using Python's Natural Language Toolkit [1], and the Stanford POS [27] and NER [11] taggers. At the front end, the interactive visualizations are created using the D3.js [2] JavaScript library. The code is available as open-source[2].

5 Use Case Scenario

To illustrate the system in action, we present a use-case scenario with a dataset of 600 records concerning the maintenance of an HVAC (heating, ventilation, and air-conditioning) system of a set of office buildings. Since our focus is primarily on the text descriptions, we remove all temporal and machine/operator-related metadata before uploading it into our system. Refer back to Fig. 1 for an overview of this dataset when processed and viewed in the system.

We consider an analyst—a maintenance specialist interested in identifying commonly-occurring patterns where repair and/or replacement is required. Once the analyst loads the data, they take a closer look at the word cloud view (Fig. 1) and see that the more commonly-occurring terms seem to be generic terms—mostly verbs—that appear to be concerned with remedial action, such as "taken", "found", "checked" etc. The letter "f" also appears frequently, and upon closer inspection, is revealed to be aggregated from all the mentions of temperatures in Fahrenheit. The first *item*—a component that finds frequent mention—is "valve". Selecting this word in the cloud immediately highlights all its occurrences in the text and the overview (see Fig. 2). The highlights in the overview visualization show that "valve" does indeed appear fairly uniformly across the maintenance records. The analyst is curious if most of the valve-related issues also relate to actuators. They select "actuator" in the word cloud, but realize that they need to see co-occurrence patterns, i.e. cases where valve-related issues co-occur with actuator-related issues.

[2] https://github.com/senthilchandrasegaran/textplorer/

The analyst decides to create a category called "valve-related issues", and another called "actuator-related issues". They manually select every field that shows the occurrence of the word "valve" and assign the category "valve-related issues" to it. They follow a similar process for the actuators (as shown in Fig. 3). They are also curious to see the distribution of valves/actuators repaired and those that are replaced. They create two more categories called "repairs" and "replacements" and through a similar process, continue assigning categories.

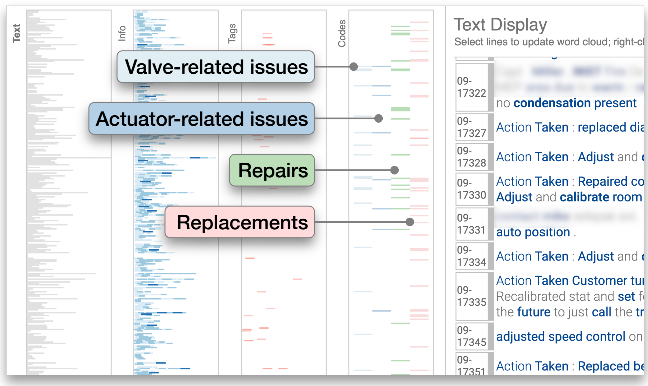

Figure 4: Detail of the categorization overview showing co-occurrences of the manually-created categories.

Once the categories are assigned, they inspect the co-occurrences of these categories closely (see overview in Fig. 1 and detailed view in Fig. 4). A close inspection of the co-occurrences shows the analyst that most valve- and actuator-related issues are repaired (with a few replacements), and that there are few cases where valve-related problems co-occur with actuator-related problems. The analyst continues with more inclusive terms for repair such as "fixed", "removed", "cleaned" and so on, assigning the same category of "repair" to them, to hunt for more patterns.

6 Conclusion

Visual analytics has been shown to be the best solution for sense-making when it comes to semi-structured data such as maintenance logs. In this paper, we illustrate how a visual analytics approach that was designed for qualitative text analysis can be used for analyzing the raw text from machine maintenance logs. Specifically, we identified requirements such as identifying common occurrences, patterns and anomalies, and the need for manual categorization and iterative analyses that an analysis tool should address for use in machine maintenance logs. We made the argument for how a visual analytic approach—which combines automated analysis techniques with human-in-the-loop interfaces—is suitable to address such requirements. We described our interface and with a use-case scenario, illustrated how the system can be used to identify similar maintenance logs, manually assign categories to these entries, and use category co-occurrence to form further insights.

Our current approach was illustrated with a machine log dataset with around 600 records. For our future work, we plan to extend our approach to be more scalable, as machine logs can extend to thousands or even millions of records. While visual representations such as word clouds and information-content maps are scalable, text overview and detail displays need to be redesigned to scale to such large records. One approach we plan to use is to incorporate machine log metadata to separate logs that may be unrelated (and can thus be examined separately). We also plan to use dimensionality-reduction techniques that can make use of metadata to automatically suggest clusters based on similarity metrics, or weights that can be derived through discussions with analysts. We will iteratively refine our approach through longitudinal studies with technicians experienced in machine maintenance for better environmental validity.

NIST Disclaimer

The use of any products described in this paper does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that products are necessarily the best available for the purpose.

References

[1] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc., 2009.

[2] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D^3: Data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 2011.

[3] Richard Brath and Ebad Banissi. Using text in visualizations for micro/macro readings. In *Proceedings of the ACM Intelligent User Interfaces Workshop on Visual Text Analytics*, 2015.

[4] Michael P Brundage, Boonserm Kulvantunyou, Toyosi Ademujimi, and Badarinath Rakshith. Smart manufacturing through a framework for a knowledge-based diagnosis system. In *Proceedings of the ASME International Manufacturing Science and Engineering Conference*, 2017.

[5] Gromit Yeuk-Yin Chan, Panpan Xu, Zeng Dai, and Liu Ren. ViBR: Visualizing bipartite relations at scale with the minimum description length principle. *IEEE transactions on visualization and computer graphics*, 25(1):321–330, 2019.

[6] Senthil Chandrasegaran, Sriram Karthik Badam, Lorraine Kisselburgh, Karthik Ramani, and Niklas Elmqvist. Integrating visual analytics support for grounded theory practice in qualitative text analysis. *Computer Graphics Forum*, 36(3):201–212, 2017.

[7] Kristin A Cook and James J Thomas. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Press, 2005.

[8] Sheng Di, Rinku Gupta, Marc Snir, Eric Pershey, and Franck Cappello. Logaider: A tool for mining potential correlations of hpc log events. In *IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, pages 442–451, 2017.

[9] Wenwen Dou, Li Yu, Xiaoyu Wang, Zhiqiang Ma, and William Ribarsky. Hierarchicaltopics: Visually exploring large text collections using topic hierarchies. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2002–2011, 2013.

[10] Min Du, Feifei Li, Guineng Zheng, and Vivek Srikumar. Deeplog: Anomaly detection and diagnosis from system logs through deep learning. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, pages 1285–1298, 2017.

[11] Jenny Rose Finkel, Trond Grenager, and Christopher Manning. Incorporating non-local information into information extraction systems by Gibbs sampling. In *Proceedings of the Annual Meeting on Association for Computational Linguistics*, pages 363–370, 2005.

[12] Xiaoyu Fu, Rui Ren, Jianfeng Zhan, Wei Zhou, Zhen Jia, and Gang Lu. Logmaster: Mining event correlations in logs of large-scale cluster systems. In *IEEE Symposium on Reliable Distributed Systems*, pages 71–80, 2012.

[13] Ana Gainaru, Franck Cappello, and William Kramer. Taming of the shrew: Modeling the normal and faulty behaviour of large-scale HPC systems. In *IEEE International Parallel and Distributed Processing Symposium*, pages 1168–1179, 2012.

[14] Ana Gainaru, Franck Cappello, Stefan Trausan-Matu, and Bill Kramer. Event log mining tool for large scale hpc systems. In *European Conference on Parallel Processing*, pages 52–64, 2011.

[15] Hanqi Guo, Sheng Di, Rinku Gupta, Tom Peterka, and Franck Cappello. La VALSE: Scalable log visualization for fault characterization in supercomputers. In *EGPGV*, pages 91–100, 2018.

[16] Hossein Hamooni, Biplob Debnath, Jianwu Xu, Hui Zhang, Guofei Jiang, and Abdullah Mueen. Logmine: Fast pattern recognition for log analytics. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 1573–1582, 2016.

[17] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. Visual analytics: Definition, process, and challenges. In *Information visualization*, pages 154–175. Springer, 2008.

[18] Steffen Koch, Markus John, Michael Wörner, Andreas Müller, and Thomas Ertl. Varifocalreader—in-depth visual analysis of large text documents. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1723–1732, 2014.

[19] Catello Di Martino, Saurabh Jha, William Kramer, Zbigniew Kalbarczyk, and Ravishankar K Iyer. Logdiver: A tool for measuring resilience of extreme-scale systems and applications. In *Proceedings of the ACM Workshop on Fault Tolerance for HPC at eXtreme Scale*, pages 11–18, 2015.

[20] Peter Pirolli and Stuart Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, volume 5, pages 2–4, 2005.

[21] Devarajan Ramanujan, William Z Bernstein, Senthil K Chandrasegaran, and Karthik Ramani. Visual analytics tools for sustainable lifecycle design: Current status, challenges, and future opportunities. *Journal of Mechanical Design*, 139(11):111415, 2017.

[22] Philip Resnik. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the International Joint Conferences on Artificial Intelligence*, pages 448–453, 1995.

[23] Thurston Sexton, Michael P Brundage, Michael Hoffman, and Katherine C Morris. Hybrid datafication of maintenance logs from ai-assisted human tags. In *IEEE International Conference on Big Data*, pages 1769–1777, 2017.

[24] Thurston Sexton, Melinda Hodkiewicz, Michael P Brundage, and Thomas Smoker. Benchmarking for keyword extraction methodologies in maintenance work orders. In *PHM society conference*, volume 10, 2018.

[25] Michael Sharp, Thurston Sexton, and Michael P Brundage. Toward semi-autonomous information extraction for unstructured maintenance data in root cause analysis. In *IFIP International Conference on Advances in Production Management Systems*, pages 425–432, 2017.

[26] John Stasko, Carsten Görg, and Zhicheng Liu. Jigsaw: supporting investigative analysis through interactive visualization. *Information visualization*, 7(2):118–132, 2008.

[27] Kristina Toutanova, Dan Klein, Christopher D Manning, and Yoram Singer. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pages 173–180, 2003.

[28] Furu Wei, Shixia Liu, Yangqiu Song, Shimei Pan, Michelle X Zhou, Weihong Qian, Lei Shi, Li Tan, and Qiang Zhang. Tiara: a visual exploratory text analytic system. In *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining*, pages 153–162, 2010.

[29] Panpan Xu, Honghui Mei, Liu Ren, and Wei Chen. Vidx: Visual diagnostics of assembly line performance in smart factories. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):291–300, 2016.

[30] Ziming Zheng, Zhiling Lan, Byung H Park, and Al Geist. System log pre-processing to improve failure prediction. In *IEEE/IFIP International Conference on Dependable Systems & Networks*, pages 572–577, 2009.