# Visualization of Heterogeneous Text Data in Collaborative Engineering Projects


**Simon Jones**
Dept. of Computer Science
University of Bath
Bath, United Kingdom

**Stephen Payne**
Dept. of Computer Science
University of Bath
Bath, United Kingdom

**Ben Hicks**
Dept. of Mechanical Engineering
University of Bristol
Bristol, United Kingdom

**Leon Watts**
Dept. of Computer Science
University of Bath
Bath, United Kingdom



### ABSTRACT
In this position paper we describe how visual analysis of heterogeneous text data is likely to benefit large-scale collaborative engineering projects. We describe use cases and tasks performed by project managers, which are potential targets for visual text analysis. We also discuss our objective to develop a structured visualization design process that will enable relevant and insightful heterogeneous data visualizations to be constructed using a systematic approach, rather than being based primarily on the expertise, intuition and creativity of the visualization designer. We suggest that the construction of a text visualization task taxonomy will contribute towards the development of such a visualization design process.



### Author Keywords
Visualization; Text Analysis, Collaborative Engineering; Project Management

### ACM Classification Keywords
H.5.m. Information interfaces and presentation


### INTRODUCTION
Gaining insights from the exploration of large, complex heterogeneous data is the most significant challenge of visual analytics and it is becoming increasingly important in the domain of collaborative engineering. Engineering work is often highly distributed, multi-national and heavily dependent upon electronic communication and digital objects that define the engineered product, the process by which it is designed, and the process by which it will be manufactured. A multitude of electronic communication tools and digital objects are employed; tools include email, instant messaging, video conferencing and social networking and digital objects include, spreadsheets, technical documents, CAD models and specialist simulation models. These communication tools and digital objects have exploded in terms of their prevalence of use, volume of content, variety of type, and overall numbers. By way of examples, a small machine or software project (< £1M) can involve 20+ contributors (engineers from various disciplines, customers, subcontractors, administrators, etc.) generate 20,000+ emails, 3,000+ reports and presentations, hold 500 meetings, generate 1,000+ models (versions) and 40 prototypes [9].

While this explosion has been necessary and beneficial at the detailed application level, it has resulted in overload of information and communication for project managers, with the negative consequence that no individual or management group can be kept continuously up-to-date with all project-related activity. In the context of complex engineering projects, potential issues can be almost impossible to identify early and mitigate; progress monitoring, control and performance measurement are all but impossible; and opportunities to innovate and maximize value are seldom pursued. Thus, effective management and control of collaborative engineering work is highly challenging and problematic.

### TASK-SPECIFIC TEXT VISUALISATION
Information visualization lends itself to the problems of collaborative engineering project management. It aims to amplify cognition by developing effective visual metaphors for mapping data, such that the resulting images are easier and quicker to interpret than the data themselves. While the design of effective data representations is generally supported by insights from visual cognition and perception research, creating an appropriate visualization for large-scale, multi-dimensional data analysis challenges the analyst/designer with an overwhelmingly rich set of choices from a huge design space. Selecting appropriate visual encodings for data is a difficult challenge, and effective solutions require a deep understanding of the available data, the user, their task and the wider context of the analysis problem. Card and Mackinlay [3] describe the structure of the visualization design space with their visualization reference model, but offer little guidance about how to reduce the huge design space based on the characteristics of the analytical task to be completed.

Previous research has attempted to match elementary data types (such as nominal, ordinal, interval, and quantitative data) to the most effective visual mapping techniques (e.g. [12]), however textual data is often particularly challenging from a visualization viewpoint. Text can represent similar concepts by many different means, it can range from being structured to unstructured, it may not contain its complete meaning, and it comprises a large number of heterogeneous dimensions. Although text representation allows a high level of definition it requires a large amount of cognitive effort to interpret and does not draw on the users' inherent ability for pattern recognition and analysis. Graphical representation of text allows the user to reduce their cognitive load and utilize their pattern spotting and visual analysis abilities, but often relies on the number of dimensions to be collapsed, and results in a significant amount of information being lost. Many researchers have shown that a fusion of text and graphical representations is often effective; transforming some qualitative information contained within the text into elementary data types – such as quantitative data, and selecting other components to include as textual objects. For example, a technique such as sentiment tracking is commonly used to assign a positive/negative value to portions of texts, which in turn can be visually encoded, e.g. as a spectrum of typeface colours ranging from green (positive sentiment) to red (negative sentiment).

Although several investigators have offered taxonomies for understanding the structure of task spaces for information visualization in general, e.g. [2] [10], at present the task space for visual text analysis is not well defined. It is not clear how applicable general visualization tasks are to text visualization specifically, or whether there are text-specific tasks that the general taxonomies fail to capture. Shneiderman [10] introduced a taxonomy based on seven data types (temporal, 1D, 2D, 3D, multi-D, Tree, Network, and Workspace) and seven interaction tasks (overview, zoom, filter, details-on-demand, relate. history, extract). Within this taxonomy, text is considered as 1-dimensional linear data accompanied with some metadata. Amar and Stasko [2] separated general analytic tasks into two levels: higher-level tasks and lower-level tasks. Higher-level tasks describe the intents that a visualization aims to support (e.g. "monitor the progress of project tasks"), whereas lower-level tasks specify the visual techniques that could be used to accomplish higher-level tasks (e.g. "display the frequency of a particular term within an e-mail"). Zhou and Feiner [12] introduced a taxonomy that interfaces the high-level presentation intents with low-level visual techniques. In their approach, visual tasks can specify what intents can be achieved and how to achieve them. A similar taxonomy for text visualization tasks remains absent.

While text visualization tasks within a specific domain, such as collaborative engineering, are likely to form a subset of text visualization tasks in general (many may be considered superfluous within the constraints of engineering project management), our work will contribute towards the construction of a complete text visualization task taxonomy.

## ENGINEERING PROJECT MANAGEMENT USE CASES

The proposed visualization methods and approaches are radically different from current project management methodologies. Project management strategies typically rely on frequent progress reviews and status reports to provide information about various dimensions of performance and control, e.g. team cohesion; effectiveness of collaboration and co-creation of digital objects; the control of intellectual property; decision making and rationale capture; uncertainty and problem solving; interface negotiation and concessions; contractual agreements; risk; and costing. Productivity is directly related to the planning and control of engineering projects. According to Coates et al. [4] project outcomes are determined by five success factors: coordination within the project, communication between project partners, reasonable task allocation, effective time management and effective resource planning. In this section we discuss several use cases for engineering project management, whereby heterogeneous text and data visualizations could provide significant benefits towards the issues surrounding performance/control. The first use case relates to monitoring performance and progress on project related tasks. The second use case relates to the management of project members and collaborative teams.

### Use Case 1: Tracking Project Task Progress

*Automatic summarization of progress*

In order to keep a project on track, project managers are required to closely monitor progress. A central way to communicate in a large engineering project is via progress reports - project teams report to project managers; project managers report to important stakeholders, and so on. A project report is fundamentally a summarization of the work that has taken place. It provides situational awareness relating to significant progress, problems, decisions and changes that have occurred since previous progress reviews. We believe that a significant advancement would be to reduce progress reporting as its own task, and rather have progress reports generated as a by-product of the work and communication that is already taking place. For example, documents produced by project teams could be processed using summarization algorithms, either extraction-based: selecting a subset of words and phrases from within the document to form a summary, or abstraction-based: building a semantic representation of the document and then using language generation techniques to create a summary. Summarized representations of text could potentially be visualized in various ways depending on the project manager's task – for example as a visual map of an engineered products status, which can be compared against the intended product specification, or as a timeline of activities and events that have occurred, showing slippages or alignments with the anticipated project schedule.

*Capture of content associated with rationale*

It is important for project managers to be able to capture and view the rationale made during a design process, and the reasons why decisions were made, particularly when a specification has to be re-examined, e.g. for reuse, for validation, or to satisfy stakeholders that a decision is warranted. Rationale capture can involve recording the reasoning behind design decisions, the alternatives considered, the evaluation of trade-offs and the argumentation that led to a particular decision. A major limitation to the capture of rationale information during development is that is can be time consuming and disruptive to project progress.

Several studies have examined Natural Language Processing techniques for rationale capture in decision making processes, for example by analyzing project e-mail repositories or document collections [5]. Others have also investigated ways of visualizing design rationale within industrial projects and argued that a graphical representation is capable of providing a sharper view of rationale than a text does, by showing directly the components of an argument and their connections [1]. Such visualizations are likely to be of great value within engineering projects where teams frequently need to inform collaborators of their rationale.

*Resolving or escalating issues*

Two common techniques used in text mining are to a) identify important entities within text and attempt to show connections among those entities, and b) to identify and extract subjective sentiment information. Such techniques are likely to be useful in the task of resolving or escalating issues within engineering projects. For example, e-mails could be analyzed to help project members identify which product parts are subject to negative sentiment, and how the polarity of sentiment changes over time.

A dashboard visualization which illustrates these changes, would allow project managers to identify areas where issues appear to be arising in relation to the entities involved in the project. Coordinated visualizations of textual and non-textual data could enable effective understanding of these issues, e.g. overlaying data extracted from e-mails on visualizations of the product to which they relate (Figure 1).

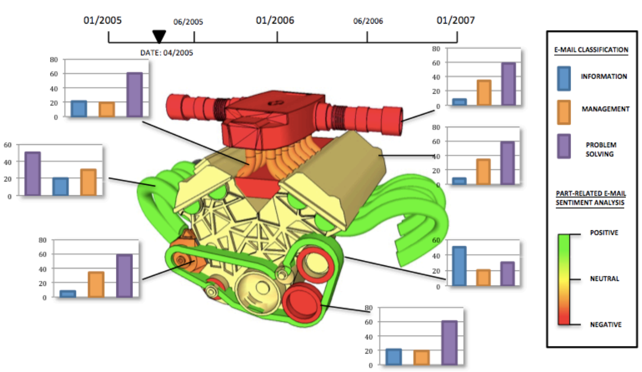


**Figure 1. Example visualization for entity-related sentiment and e-mail classification over time. (Part colours encode average sentiment associated with part entities. Bar charts illustrate frequency of e-mails relating to parts and proportion of e-mails classified as information, management, or problem solving.**

Such a tool could be useful for the prioritization of tasks and activities. An alternative visualization of the same information could be used to illustrate the contents of a project manager's overloaded e-mail inbox, and enable them to respond to messages according to the entities to which they relate and the priority that they are given, rather than the chronological order in which they were received.

By visualizing the connections between entities, e.g. the extent to which two product parts are discussed concurrently, it may be possible to infer dependencies within the project and to predict that issues relating to one entity may affect others. Such visualizations would allow project managers to engage in timely identification and resolution of issues, before they escalate to a level which negatively impacts project progress, as well as provide support for change management and predicting change propagation Similar visualization techniques have been used in [6], where intelligence analysts were able to examine relationships among entities mentioned in transcripts of phone logs.

It may also be useful for project related communications to be automatically classified according to a taxonomy of communication types. For example, e-mail messages could be categorized as being related to information sharing, management activities, or problem solving [7]. The relative frequencies and vectors of change for each communication type could be indicative of significant events e.g. pressure points, work completion or a steady mode of working. A visualization of e-mail classifications over the course of an entire project could be used by managers for the timely identification of events that require some intervention or management coordination.

### Use Case 2: Managing Collaborative Teams

*Opinion and sentiment summarization*

As discussed, sentiment analysis is useful for product managers who need to track the specific aspects of a project. However, sentiment analysis may also be useful to managers wanting to gather information about their own employees. For instance, a review of emails might reveal that among project members, the number of negative words or phrases has increased significantly over a particular period. This might alert managers to examine the project's status and spend more time communicating with key staff and addressing their concerns. By having access to a visualization of employee opinion and sentiment, project managers may identify areas where team members are dissatisfied and implement strategies for enhancing satisfaction and improving productivity.

*View resource capacity and loading on team*

While increased communication between collaborative project members is typically encouraged, it is not only project managers that experience information overload in such situations. It is important for managers to monitor the resource capacity and loading on their teams, in order to ensure that members can perform effectively. Engineering projects often contain key figurehead/expert engineers who are the 'go to people' within the project. These are often referred to in the literature as gatekeepers or information stars as they fill one of two roles; 1) to know 'who knows' and therefore direct engineers to the relevant expert or 2)

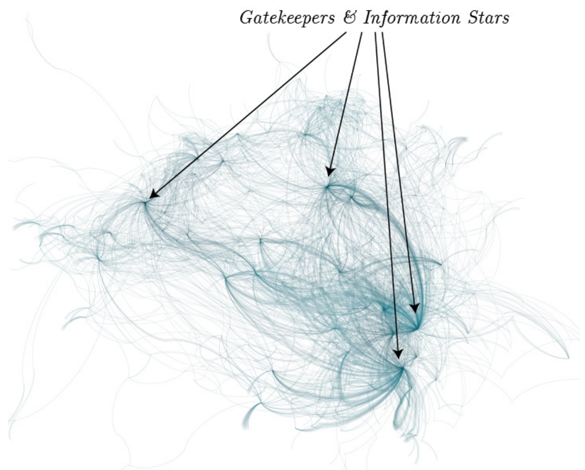


**Figure 2. Visualisation of an Engineering E-Mail Network**

are experts in a particular field themselves. Visualizations of communications within a collaborative engineering project would appear similar to the visualisation in Figure 2, which highlights people who act as information 'gatekeepers'. By increasing the communication activity within the network, the gatekeepers are likely to receive an overwhelming number of e-mails, increasing the issue of information overload. Such visualizations are likely to be useful for managers to identify and address such issues, for example by training staff and reassigning gatekeeper roles to spread the load.

*Assessing linguistic correlates of group cohesiveness and performance*

Increased globalization of collaborative engineering results in a growing number of project members having to interact across geographical and linguistic boundaries. Consequently, the link between communication and group cohesiveness is of great importance to project managers for understanding how effectively teams operate. Language diversity has often been portrayed as a problem for group cohesiveness in multicultural organizations and team performance has been shown to be significantly associated with team members' task-related communications, specifically with the extent to which task-critical information is shared [8]. Visualization of language similarity and other linguistic correlates of team performance and cohesiveness across communication links could enable managers to identify areas where communication breakdowns are occurring, or where project groups may be less cohesive and performance affected.

## CONCLUSION

We posit that analysts rely heavily on cognitive skills, intuition, creativity and experience when designing a visualization that is appropriate for performing a certain data analysis task. We hypothesize that an improved approach would be to develop and employ a structured, automated visualization design process, driven by a framework that captures the interactions between data, task and view. We hope that such a process could be used to guide the creation of data visualizations for collaborative engineering projects, based on formal specification of the data to be analyzed and the task to be performed. In order for the design process to provide guidance on effective visual mappings for text data, we first need to define the task space for text visualization within collaborative engineering project management. This paper presents several examples of important text visualization tasks.

## ACKNOWLEDGMENTS

The research reported in this paper is funded by Engineering and Physical Sciences Research Council (EP/K014196/1).

## REFERENCES

1. Alexander, I. (2008). A Practical Way for Projects to Visualize Design Rationale. In Information Visualisation, 2008. IV'08. 12th International Conference (pp. 553-558). IEEE.
2. Amar, R., Stasko, J. (2004). A knowledge task-based frame- work for design and evaluation of information visualizations. In Proc. of IEEE Symposium on Information Visualization 2004, pp. 143–150.
3. Card, S. K. and Mackinlay, J. (1997). The structure of the information visualization design space. In Proceedings of the 1997 IEEE Symposium on Information Visualization (InfoVis '97) (INFOVIS '97). IEEE Computer Society, Washington, DC, USA, 92.
4. Coates, G. (2004). Engineering management: operational design coordination. Journal of Engineering Design, 15(5), 433-446.
5. De Lucia, A., Fasano, F., Grieco, C., Tortora, G. (2009). Recovering Design Rationale from Email Repositories, Proceedings of 25th IEEE International Conference on Software Maintenance, Edmonton, Canada, IEEE Press, pp. 543-546.
6. Gorg, C., Liu, Z., Parekh, N., Singhal, K. and Stasko. J. Visual Analytics with Jigsaw. IEEE Symposium on Visual Analytics Science and Technology (VAST'07), pp. 201–202.
7. Hicks, B. (2013) The language of collaborative engineering projects. 19th International Conference on Engineering Design (ICED13). Seoul, Korea.
8. Fischer, U., McDonnell, L., & Orasanu, J. (2007). Linguistic correlates of team performance: Toward a tool for monitoring team functioning during space missions. Aviation, Space, and Environmental Medicine, 78, 86-95.
9. Regli, W., (2010). Digital Engineering Archives http://www.digitalpreservation.gov/partners/documents/digarch-regli.pdf.
10. Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy of information visualizations, Proc. IEEE Symposium on Visual Languages '96, IEEE, Los Alamos, CA. pg 336-343.
11. Ware, C. (2000). Information visualization (Vol. 2). San Francisco: Morgan Kaufmann.
12. Zhou, M. X. and Feiner, S. K. (1998). Visual task characterization for automated visual discourse synthesis, Human Factors in Computing Systems (CHI '98), Los Angeles, CA, USA, pp. 392-399