

## **Utilization of Project Sentiment Analysis as a Project Performance Predictor**

By Bob Prieto

The growth in project complexity and scale provides growing challenges for today's project managers.<sup>1</sup> Equally, these challenges provide increased challenges for program and portfolio managers who must look at not only the "sum" of individual project performance but also broader portfolio wide performance patterns<sup>2</sup>. Improvements in traditional project management tools must be coupled with advanced analytics<sup>3,4</sup> and newer tools geared to detection of negative performance precursors. In this paper we examine one possible tool, sentiment analysis, and its application to detection of negative performance precursors.

### **Semantic Analysis**

Early prediction of potential negative trends in project performance is aided by early identification of precursors to sustained negative performance<sup>5,6</sup>. Among the sources of potential precursors that can be utilized is a wide range of project electronic correspondence and reports. These reports may be analyzed in many different ways but one approach is to conduct a semantic analysis of the language utilized in the reports. The appearance of semantically negative terms is an early indicator of potential project issues and often is a prelude to formal identification of an issue with defined impacts in structured project reports. The semantic analysis to be conducted is preferentially (but not exclusively) focused on three primary sources of textual data in order to provide higher computational efficiency and increased confidence in the semantic findings:

- Select pairs of correspondence
- Periodic structured reports
- Periodic narrative management reports

---

<sup>1</sup> The GIGA Factor; Program Management in the Engineering & Construction Industry; CMAA; ISBN 978-1-938014-99-4; 2011

<sup>2</sup> The Focus, Roles & Responsibilities of a Program Management Office; PM World; April 2010

<sup>3</sup> "Generalized Analysis of Value Behavior Over Time as a Project Performance Predictor"; College of Performance Management; Issue 4 2012

<sup>4</sup> Project Categorization and Assessment Utilizing Multivariate Statistical Techniques to Facilitate Project Pattern Recognition, Categorization, Assessment and Pattern Migration Over Time; PM World Journal; June 2013

<sup>5</sup> Resiliency Assessment And Management System; Attny Dkt No. 100325.0531PRO

<sup>6</sup> Strategic Business Objectives Based Program Management Systems and Methods; U.S. patent application 13/709,996 filed on December 10th, 2012; and Attny Dkt No. 100325.0443US2

Let's look at each of these in turn.

### **Select Pairs of Correspondence**

Select pairs of correspondence, including e-mails and texts, between the project manager and identified senior managers and the project manager and his client counterpart would be included in the described semantic analysis of text. The utilization of e-mail and other electronic text messages that are readily captured provides additional richness to the analysis of the following two categories of reports which are more formal in nature.

The decision to conduct a narrower semantic analysis of text versus a broader analysis, such as all incoming and outgoing e-mails, is driven by a decision to select high confidence data sets to improve the so called "signal to noise" ratio by filtering out the significant noise that may exist more broadly in a project's total correspondence. The importance of context sensitive analysis is discussed later in this paper.

Key is the focus on a singular critical node, namely the project manager. In some instances, discrete additional communication pairs may be added but in all instances the intent is to limit such pairs to improve the overall signal to noise ratio. Any added pairs would tend to be focused on the most critical issues and concerns providing a better signal to noise ratio.

For example, the structure and operation of some projects may make it desirable to include correspondence pairs centered on the deputy project manager or executive sponsor but such broadening should be undertaken in such a way as to minimize the addition of large amounts of noise into the semantical analysis that is to be conducted.

### **Periodic Structured Reports**

Textual fields addressing issues and concerns in structured and semi-structured periodic project status reports would be analyzed as part of the sentiment analysis through a similar semantical analysis of structured report text fields.

These reports typically contain identified text fields focused on overall project management assessment; issues and concerns; risks and forecasts which collectively capture a significant portion of negative sentiment while limiting overall text to be semantically analyzed.

The semantical analysis of these text sources may prove to be the weakest signaler of developing problems given their more formal and structured nature but conversely the strongest signal with respect to "maturing" concerns. Even in these reports, domain specific

negative terminology begins to appear before a “problem” has been formally recognized or declared.

### **Periodic Narrative Management Reports**

Periodic, typically weekly or monthly, narrative status reports prepared both for corporate management and importantly for the client are considered in this portion of the analysis. These reports are relatively brief documents and embedded figures and numerical tables can be ignored in favor of the remaining narrative. It is envisioned that together these three data sources will be the primary contextual information to be utilized in a semantic analysis focused on detecting negative “sentiment” as a project precursor.

### **Context Matters**

In looking at broader work on sentiment analysis<sup>7</sup>, there is a growing recognition of the value of broader data sets. This can be seen in the application of sentiment analysis to Twitter feeds for various predictive applications. Concomitantly, there is recognition that broad data sets contain significant amounts of “noise” and even worse uniformed opinions. This is why the inclusion of known, reliable data sources<sup>8</sup> into the analysis has been recognized as important to improving overall predictive capabilities.

These other sentiment analyses are of a broader nature than what we encounter in the project management domain and as we have seen previously there is a need to similarly inform this broad analysis with reliable data sources. The semantical analysis envisioned here has focused on data discretely within the realm of project management as contrasted with a broader project semantical analysis. Why?

The decision to consider only a portion of the text data available within a project context is driven by three considerations:

- The myriad of contexts that are encompassed in a project setting
- Achieving high confidence signal-to-noise ratios
- Computational efficiency

---

<sup>7</sup> Bing Liu. 2008 Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 1–167, Morgan & Claypool Publishers.

<sup>8</sup> How Useful is Social Media-Based Sentiment Analysis to the Buy Side?; Justin Grant; Advanced Trading

## **Project Contexts**

The sum of performance across all project performance aspects provides a deeper view into overall project trends and trajectories. When undertaking semantical analysis it is important that negatively correlating terms be considered within the context to be analyzed.

Let us use an example where we look at one term and its usage across the entirety of a project. For simplicity let us assume that the “strength” of negative sentiment is related to the number of occurrences of a single negative term.

In this case let’s look at the word “variance”. From examination of project management level reports and communications, the preponderance of usage of this term occurs in a negative context. “Variance” from plan or budget is used by a project manager to describe lagging performance while other semantically positive words are used to describe performance that is “ahead of schedule” or “under budget”.

The term “variance” can be used in a semantically neutral way in other project contexts. Examples include presentations of probabilistic data related risk, test and measurement results, and requests for alternative approvals. Thus a broader text analysis for the term “variance” would result in a large set of “noise” diminishing the utility of a semantical search for the term “variance”

Thus, the list of semantical terms with a negative connotation is highly context sensitive. If we desire to expand the contemplated analysis to include a broader array of project level text it will be important to recognize the applicable contexts more narrowly.

Again, let’s use the term “variance” to make the point. If we are to include an analysis of text from a project’s engineering department the same contextual challenge remains. For example, “variance” in boring logs may relate to a degree of statistical confidence. Similarly in testing of project bulk materials or even in related risk analyses we may use the term to categorize the homogeneity of the sample or the confidence level ascribed to our assumptions. At later stages of the engineering effort variance might relate to execution performance and assume a relevancy akin to what we find in analysis of project management related texts.

This suggests two additional considerations:

- The need for refined context sensitive semantical lists that reflect more acutely defined domains such as engineering, supply chain, construction and so forth

- Recognition that in some contexts, semantic relevance may be temporal in nature. This was illustrated in the engineering discussion above where the relevance of the term “variation” changed over the course of the engineering activity.

Let’s explore this notion of temporal variance more explicitly, returning to our project management context.

### **Temporal Variation of Sentiment Strength**

Projects move through various phases, both planned and unplanned. In a well performing, simple project these phases may consist solely of startup, implementation and closeout. The introduction of “disruptions” into project execution patterns whether planned or unplanned changes the project’s performance characteristics. Unplanned disruptions in many instances are a result of decisions to implement “recovery” or “workaround” plans that should be preceded by textual data exhibiting negative sentiment.

Additionally, as a project moves through each of the various project phases it faces, the level of textual data to be considered will wax and wane such that even a well performing project will see various semantic instances increase and later decrease over a project’s normalized lifetime. This temporal variation is important to consider as one conducts semantical analysis in order to establish appropriate threshold levels for detection of the emergence of negative sentiment.

The establishment of temporal patterns and threshold levels is similarly context sensitive such that the levels for an EPC<sup>9</sup> or EPCM<sup>10</sup> project are assumed to be materially different than those for say a FEED<sup>11</sup> project. The establishment of baseline patterns is reliant on a reasonable baseline data set to “tune” the semantic analysis model with. This perhaps represents the greatest challenge faced by project management organizations. As new projects are added to the data set our semantical model is further “tuned”.

### **Risk of Synonyms**

In traditional sentiment analysis the various permutations of a term are considered as well as synonyms for the identified term. In highly targeted domains such as the project management domain described in this paper terminology is very specific and the use of synonyms carries with it the risk of many “false positives”.

---

<sup>9</sup> EPC – Engineer Procure Construct

<sup>10</sup> EPCM – Engineer Procure Construction Management

<sup>11</sup> FEED – Front End Engineering Design

Let's look at the term "variation" again as an example. A listing of synonyms would include words such as:

- Contention
- Difference
- Discrepancy
- Diversity
- Division
- Fluctuation
- Inconsistency
- Separation
- Switch
- Variety

None of these words, in the selected context, would obviously correlate with negative sentiment.

### **Not Just a Word**

Throughout this paper we have used a singular word, "variation", to describe semantical detection of negative sentiment within a project management domain. Effective semantical analysis is not reliant on detection of a singular term but rather a collection of such terms which in their entirety act to convey negative sentiment. Fluor's list of negative sentiment conveying terms within a project management context encompasses over 300 terms. Together these terms<sup>12</sup> foster earlier identification of emerging concerns with the strength of the signal growing in advance of formal declaration of a problem by the project management team.

Equally important, the strength of each term in its correlation to actual negative occurrences varies with some being more statistically relevant than others. This suggests the opportunity to weight the various semantical determinations based on the composition of the terms discovered in looking at threshold and performance index values.

### **Broader Management Relevance**

Large programs or portfolios of projects present many challenges to senior management. Among them is reliance on "summations" of project performance; limited ability to selectively

---

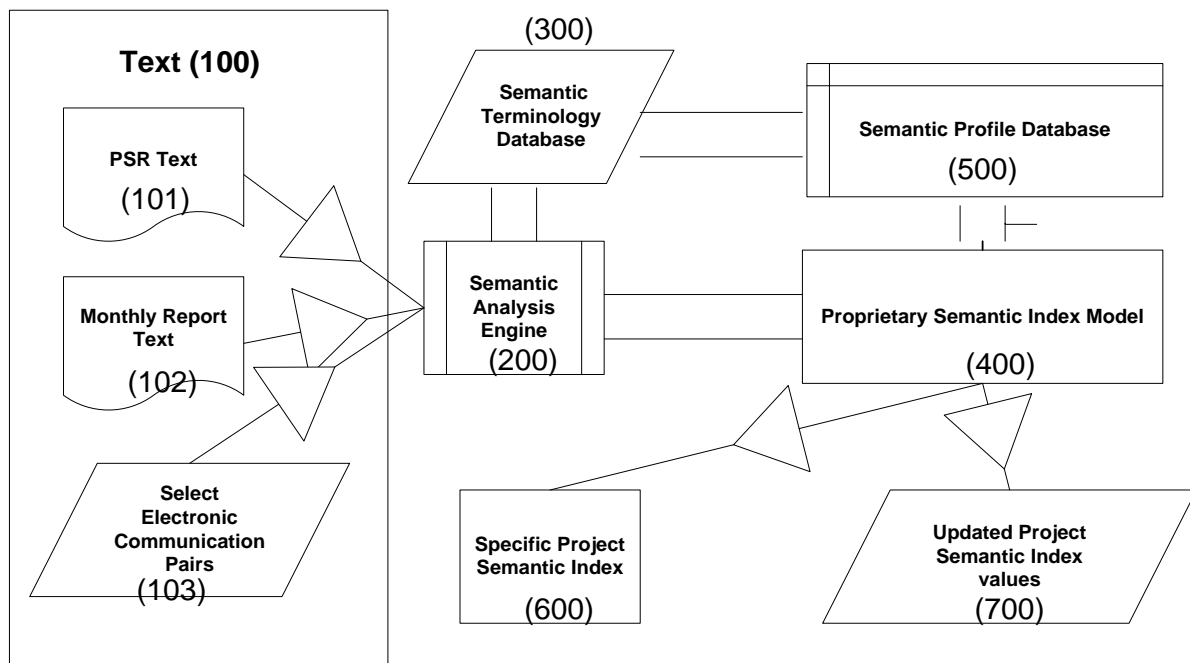
<sup>12</sup> Fluor Trade Secret

drill down into individual project performance, often not conducting such “deep dives” until problems emerge; and difficulty in seeing the effects of more systemic impacts until a full blown crisis has emerged.

The application of sentiment analysis as described in this paper aids in prioritization of focus, better targeting “deep dives” and at an earlier stage as well as detecting the sudden onset of broader performance issues affecting major portions (multiple projects) in a program or project portfolio. This later ability drives the program manager to more quickly seek systemic causes rather than limit initial efforts to a project by project performance examination. Similar benefits may accrue at an enterprise level.

### Words are Not Enough

This paper discusses the potential to use semantical analysis to determine precursors of potential problems through detection of negative sentiment in project level textual data. The discussed approach adds one more tool to the project manager’s toolbox. But just as you can’t build a major project with only one tool, so to must today’s project manager use his full toolbox to meet an expanding array of challenges.



**Figure 1**  
**Project Sentiment Model**

### **Project Management Sentiment Analysis Model**

Figure 1 illustrates the project sentiment model. In the model textual information (100) consists of select text fields from periodic project status reports (PSR) (101); monthly status reports (102) and select electronic communication pairs (103). All the selected textual data flows through servers to/from the project manager to his client counterpart and select corporate managers. These communication flows include e-mails as well as the electronic transmission of PSR (101) and monthly status reports (102).

Textual information (100) of all forms is processed by a semantic analysis engine (200). The semantic analysis engine (200) segregates text by source and pre-processes the text.

The preprocessed list is then compared against a Semantic Terminology Database (300) and for each textual data set (100 comprising 101, 102, 103) the frequency of negative sentiment instances for a given time period calculated.

These sentiment counts by text source are the output of the semantic analysis engine (200) and are compared to a semantic index model (400).

The resulting score for each text stream (101, 102, 103) and a composite score in consideration of all text streams (100) results in a specific project semantic index (600). This index (600) would link to a model of project failure included in the semantic index model (400). As an example a project might obtain a composite score, unadjusted for the index, of "62" at an early stage of the project but score in the lowest quartile of failure probability. That same score at the midpoint of the project might correlate with a failure probability in the highest quartile.

As new periodic data is generated by the semantic analysis engine (200), the semantic profile database (500) is updated after the specific project semantic index (600) has been generated. This allows the index to learn as new data is entered. The new data in the semantic profile database is processed by the semantic index model (400) to produce an updated project semantic index values tabulation (700) for managers to identify potentially comparable projects.



## About the Author



### **Bob Prieto**

Senior Vice President  
Fluor

Princeton, NJ, USA

Bob Prieto is a senior vice president of Fluor, one of the largest, publicly traded engineering and construction companies in the world. He is responsible for strategy for the firm's Industrial & Infrastructure group which focuses on the development and delivery of large, complex projects worldwide. The group encompasses three major business lines including Infrastructure, with an emphasis on Public Private Partnerships; Mining; and Industrial Services. Bob consults with owners of large engineering & construction capital construction programs across all market sectors in the development of programmatic delivery strategies encompassing planning, engineering, procurement, construction and financing. He is author of "Strategic Program Management", "The Giga Factor: Program Management in the Engineering and Construction Industry" and "Application of Life Cycle Analysis in the Capital Assets Industry" published by the Construction Management Association of America (CMAA) and "Topics in Strategic Program Management" as well as over 450 other papers and presentations.

Bob is a member of the ASCE Industry Leaders Council, National Academy of Construction and a Fellow of the Construction Management Association of America. Bob served until 2006 as one of three U.S. presidential appointees to the Asia Pacific Economic Cooperation (APEC) Business Advisory Council (ABAC), working with U.S. and Asia-Pacific business leaders to shape the framework for trade and economic growth and had previously served as both as Chairman of the Engineering and Construction Governors of the World Economic Forum and co-chair of the infrastructure task force formed after September 11th by the New York City Chamber of Commerce. Previously, he served as Chairman at Parsons Brinckerhoff (PB), one of the world's leading engineering companies. Bob Prieto can be contacted at [Bob.Prieto@fluor.com](mailto:Bob.Prieto@fluor.com).