



This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

The Needs And Benefits Of Text Mining Applications On Post Project Reviews.

A. K. Choudhary¹, P. I. Oluikpe², J. A. Harding^{1*} and P. M. Carrillo²

Abstract

Post Project Reviews (PPRs) are a rich source of knowledge and data for organisations - if organisations have the time and resources to analyse them. Too often these reports are stored, unread by many who could benefit from them. PPR reports attempt to document the project experience – both good and bad. If these reports were analysed collectively, they may expose important detail, e.g. recurring problems or examples of good practice, perhaps repeated across a number of projects. However, because most companies do not have the resources to thoroughly examine PPR reports, either individually or collectively, important insights and opportunities to learn from previous projects, are missed. This research explores the application of knowledge discovery techniques and text mining to uncover patterns, associations, and trends from PPR reports. The results might then be used to address problem areas, enhance processes, and improve customer relationships. A case study related to two construction companies is presented in this paper and knowledge discovery techniques are used to analyze 50 PPR reports collected during the last three years. The case study has been examined in six contexts and the results show that Text Mining has a good potential to improve overall knowledge reuse and exploitation.

Keywords: Text mining; knowledge discovery; post project reviews (PPRs); manufacturing and construction.

-
1. Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, Leicestershire, UK, LE11 3TU
 2. Department of Civil and Building Engineering, Loughborough University, Loughborough, Leicestershire, UK, LE11 3TU
- Communicating Author: J.A.Harding@lboro.ac.uk

1. Overview

Post Project Reviews (PPRs) are one of the most important and common approaches for the capture of project knowledge. They provide opportunities for project teams to share, discuss and even explain their experiences through face-to-face, facilitated interactions before a project is closed and the team is dissolved. PPRs therefore allow multi-disciplinary teams to critique a project to determine both positive and negative aspects, potentially capturing tacit knowledge as learning points to improve the planning and execution of future projects. Interactive debates between the project team members during PPRs may lead to greater innovation and better ideas than can be achieved from any individual. This shared communication is crucial as each individual contributor will inevitably have his or her own perspective or viewpoint and will only know part of the whole project story [1]. In the UK, major companies such as BP Amoco, BAA plc, National Grid Transco and construction companies such as Bovis Lend Lease, IPSL, Simons Design and Buro Happold, have adopted PPRs in an effort to learn from experience.

Conducting PPRs is time consuming, manpower intensive and expensive in terms of company overheads. Disterer [2] mentioned that after finishing the project, team members are spread all over the company and project documentation is stored in folders without retaining the essentials for later use. In his assessment of PPRs, Busby [3] highlighted that “PPRs were important learning mechanisms and their value seems to be underestimated by individuals who do not appreciate the need to disseminate insights throughout the organization”. PPRs are often conducted as a part of companies’ quality systems, however a major problem lies in the fact that companies have insufficient resource to act on the outcome of PPRs[4]. Companies often have individuals with responsibility for creating PPRs, but they do not have individuals or teams responsible for subsequently analyzing the PPRs collectively to identify the good or bad practice and areas requiring improvements across a range of projects. In addition, Orange *et al.* [5] and Kamara *et al.* [6] also identified that PPRs have huge potential for much more thorough exploitation. If information and knowledge from PPRs can be extracted and analysed effectively, good and bad practices might be identified so that lessons are learnt from past projects and knowledge is reused or exploited to improve the quality and levels of success in future projects.

Knowledge Discovery in Text (KDT) and Text Mining (TM)[7] are very recent and increasingly interesting areas of research in computer science. KDT and TM are mostly automated techniques that aim to discover high level information from huge amounts of textual data and present it in a useful form to the potential user, who might be an analyst, decision maker, project manager etc.. PPRs are generally recorded in reports or other text-based documentation so KDT or TM techniques might therefore be applied on PPR documents from multiple projects, to potentially identify information relating to good, bad or even best practices. The benefits of mining PPR texts lie in the possibilities of discovering patterns, associations, and linkages of processes, activities and terms occurring in the reports. If these can then be identified as commonly occurring problems, or as good practices, they may highlight key learning points for the organisation. The organisation may then adjust its activities to reflect what is learned from the KDT and TM with the aim of improving processes, optimising profit and improving client retention. In addition, whenever a new project is initiated, it would be very beneficial if lessons learned from similar projects could be quickly identified to reduce the chances of errors being repeated and increase the potential for savings in cost and time. This may also be possible through the identification of patterns or associations using KDT or TM techniques.

2. Current practices in PPRs and Common Problems

A PPR is a process through which an organisation looks at the project process retrospectively with a view to learning from activities carried out, to avoid mistakes in the future and also to learn from successes and failures. A PPR is also defined as “a formal review of the project which examines the lessons which may be learnt and used to the benefit of future projects”[8]. In the remainder of this section, current practices in the application of PPR are discussed through a review of PPR literature. It is clear from this review that whilst many authors acknowledge the benefits of conducting PPRs, several recurring problem areas can also be identified. These will be summarised in section 2.4.

2.1 Benefits of conducting PPR

Benefits gained by organisations from conducting PPRs have been highlighted in Tan *et al.* [9] and Carrillo [4] and include:

- Facilitating collective learning: PPRs provide an opportunity for people involved in the project to come together and examine what went right or wrong during the project. The forum for the PPR should provide an atmosphere for knowledge sharing, exchange of ideas, brainstorming, identifying good and bad practices and contributions which will lead to learning.
- Provide utilisable knowledge: The outcome of a good PPR process should be knowledge that can be utilised for future projects, but it is often tacit knowledge and therefore can be difficult to reuse.
- Benefit client organisations: Review processes should aim to provide greater insight into how assets are managed, and this should help the project organisation to improve its processes and manage its assets better.
- Better project phase management: Reviewing each phase of a project provides opportunities for better project management at the phase level, rather than carrying out a single review at the end. Hence, mistakes might be corrected earlier at the phase level perhaps benefiting the remaining project phases.
- Prevent knowledge loss: When a project team disbands the knowledge carried by the team members can be lost. A PPR process should capture their project related knowledge and make it explicit for others to utilise.

2.2 PPRs in Manufacturing and Construction

Very few research projects have studied PPRs in construction or manufacturing organisations, but the limited published literature that does exist in construction includes Kamara *et al.* [6], Sowards [10], Carrillo [4] and in manufacturing includes Motteram, and Sizer, [11]; Koners and Keith [12]. The importance of PPR is emphasised by its frequent mention in different knowledge management literature including business, information technology, space project management and aerospace^[pmc1] [9], construction [4], manufacturing [12], information technology [2,33], space project management [14], R & D [15], software development [16], environmental studies [17], finance [18] and operations research [19].

The benefits of PPR reports are not fully achieved if the knowledge they contain is not appropriately disseminated and used, and whilst some recommendations to address this issue have been made in the published literature these are mostly vague and unspecific. Roth and Kleiner [20] recommended a six stage process of learning histories which constitute planning, reflective interviews, distillation, writing, validation and dissemination, but greater details are needed in each stage to enrich the process. Also no feedback loops for reuse of learning from previous projects were included and this makes it difficult to translate the results from PPRs into real benefits for the organisation. Busby [3] advocated a process that integrates feedback loops for effective learning but did not elaborate on how this could be done. Schindler and Eppler [21] suggested that a “project knowledge broker” should be responsible for the processes of reviewing a project and transferring lessons learnt within and between project teams. However, this recommendation implies outsourcing the review process and this would have obvious implications for the organisation. Sowards [10] recommended a five stage process which focused on establishing criteria, involving key people, discussing an agenda, documenting key learning points and disseminating these to people who should see them. PPRs can be criticised for only bringing together people considered as “key people” for the project and thus denying others the opportunity to contribute to project learning.

2.3 Review of PPR approaches in literature

Several approaches to conducting PPR were discovered in the reviewed literature as follows:

PPR as KM technique: Tan *et al.*, [9], Robinson *et al.*, [22] and Newell [23] believe that PPRs are necessary for contributing to the knowledge management perspectives of the organisation and relate this to the ability of the organisation to share knowledge across the enterprise. This is always with the aim of consciously exploiting lessons learnt from projects in order to avoid past mistakes and derive innovation.

Process: Sowards [10], Roth and Kleiner [20], Braniš and Christopoulos [17] and von Zedtwitz [15] suggested that PPRs should be done in phases during the lifetime of a project and also at the end of the project. A key learning point of this approach is that

PPR should be adopted as part of the project evaluation mechanisms for defining success criteria and improving project control, quality assurance, and risk assessment.

Systemic Approach: Garon [14] and Terry [18] suggest the use of cognitive mapping and system dynamics in carrying out PPRs. Mapping identifies feedback loops which affect the success of the project. They argue that traditional lessons learnt systems obstruct the examination of deep and messy reasons for project failure or problems.

Collective Learning: Carrillo [4] and Grobelnik and Mladenic [24] indicate the usefulness of PPRs in fostering collective learning.

2.4 Current Limitations and Research Gaps

A survey into research and development (R&D) in automobile manufacturing companies and computer components manufacturers indicates that organisations are aware of the benefits of PPR but do not utilise the full opportunities to learn from them [15]. Newell [23] explored reasons why organisations capture project knowledge but do not seem to subsequently utilise that knowledge in future projects. One of the conclusions of Newell's research is that project teams lack awareness that such critical knowledge exists and could improve their processes. The PPR is sometimes a huge silo of information which rarely gets analysed critically to reveal patterns of information that could help decision making in the project process. The contents of the PPR are therefore not converted into useful knowledge, project teams abandon the report on the shelf and move on, thus ignoring knowledge which might be useful and even critical to future projects.

The recurring themes in much of the published PPR literature are that PPRs are useful and can generate very valuable knowledge, of many types, but this knowledge (and consequent lessons learnt) are not effectively exploited. The major limitations can be summarised as follows:

1. Knowledge and learning from previous PPRs are not routinely transferred to future or ongoing projects and therefore learnt lessons are not properly exploited or reused.

2. Collections of PPRs are not commonly analyzed or explored together to find recurring problems or patterns of behaviour that may exist over a period of time.
3. PPRs may identify good, and bad practices but existing PPR processes do not detail how these should be disseminated in order to improve project processes or performance.
4. PPRs commonly generate large quantities of documentation which may be stored on a company network. However research is still required to determine how to effectively analyse, disseminate and exploit PPRs throughout an organization.

It is clear from the above that major challenges for organisations are how to extract knowledge from (single and multiple) PPRs and then disseminate and exploit this valuable experience and learning appropriately across the organisation to ensure optimum improvements in the organisation's project processes. The remainder of this paper will examine how KDT and TM methods may be applied on PPRs to analyse their contents and extract useful results which will facilitate dissemination and exploitation of PPR learning. An important reason for using KDT and TM techniques is to facilitate the detailed analysis of multiple PPRs simultaneously as this should substantially increase the potential for identifying useful information and patterns of operation. The next section discusses KDT and TM methods in detail.

3. Knowledge Discovery in Text (KDT) and Text Mining (TM).

KDT and TM are multi-disciplinary fields of research that try to resolve the problem of information overload by identifying and retrieving useful knowledge from large stores of text-based data. KDT and TM involve techniques like information retrieval, text analysis, information extraction, clustering categorization, visualization, database technology, machine learning, natural language processing, data mining [29, 42-46] and knowledge management [30]. KDT refers to the overall process of turning unstructured or semi structured textual data into high level information and knowledge. Following the definition of KDD by Fayyad *et al.* [7], Karanikous and Theoudoulidis [30] defined KDT as “the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in unstructured data”. On the other hand, *Text Mining (TM)* is a step in the KDT process consisting of particular

data mining and natural language processing algorithms that under certain computational efficiency and limitations produces a particular enumeration of patterns over a set of unstructured textual data. Text mining uses unstructured or semi structured textual data or information and examines it in an attempt to discover structure and implicit meanings “hidden” within the text.

<<Include figure 1 about here>>

As shown in Figure 1, KDT consists of three main steps.

1. Document Collection: The very first step in the KDT process is to identify and collect the documents which need to be retrieved for example, customer emails, technical reports, patents, project reviews etc.
2. Retrieving and pre-processing documents: When the documents have been retrieved, they will normally need to go through a pre-processing stage to transform them into a form suitable for the TM techniques which are going to be applied. The precise nature of the pre-processing will vary depending on the characteristics of the documents and the types of TM to be used. For example, unwanted text may be removed to reduce the size of text. Transformations may also be done to represent the documents in another form such as XML, Standard Generalized Markup Language (SGML), etc. The resulting documents may then be processed to provide basic linguistic information about the content of each document.
3. Text Mining: As a part of the KDT process, TM uses various algorithms and tools to extract metadata or high-level information and / or to discover patterns and relationship within the extracted information. Results are generated in the form of reports, which, if required can be further processed to extract the high-level knowledge. Therefore, generation of reports and text mining work as closed loop feedback for continuous improvement of mining results.

The results of TM are “knowledge” (in different forms) which can be used by decision makers to take further decisions and make improvements.

To date, very little research has been published in the area of text mining applications but the main application areas identified include analyzing biomedical documents [32-33], patents [34], financial reports [35], customer relations management [36], product development processes [37] and medical records [38]. The vast majority of this research has been done within the last 5 years. This paper is the first of its kind where TM is applied on PPRs from the construction industries.

TM combines techniques from several areas, including natural language processing, data mining, machine learning, information retrieval and information extraction to automatically discover patterns, extract information and generate meta data from large quantities of text. Cunningham [39-41] (2006) described information extraction (IE) as a technology based on analyzing natural language in order to extract snippets of information. This process takes texts as input and produces fixed format, unambiguous information as output. Information extraction is different from Information Retrieval (IR) in the following sense:

- An IR system finds relevant texts and presents them to the user.
- An IE application analyzes text and presents only the specific information that the user is interested in.

Text Mining covers many different types of techniques and it is therefore difficult to identify the “best” tool for a particular job. Some of the functions performed by TM are summarised below:

- Feature extraction: identification and extraction of key features from the text that can be used as the data and dimensions for analysis. The number of times each term appears in a document (Word frequency) should also be indicated.
- Text Based Navigation: enables users to move about in a document collection by relating topics and significant terms. It helps to identify key concepts and some of their relationships .
- Search and retrieval: searching and retrieving for particular text.
- Text Categorization: identification of the main themes of a document. Categorization counts the appearances of words and from the counts, identifies the main topics covered in the document. Categorization often relies on a thesaurus in which topics are predefined and relationships are identified by looking for broad terms, narrower terms, synonyms, and related terms.

- Clustering: automatically groups documents on the basis of some similarity measure, without predefining the category. Documents can also appear in subtopics, ensuring that useful documents are not omitted from the search results.
- Summarization: reduces the length and detail of a document while retaining its main points and overall meaning. Sentence extraction is a widely used strategy to extract important sentences from a given text by statistically weighting all the sentences in the text. Microsoft Word's Auto-Summarize function is a simple example of text summarization.
- Trends Analysis: identifies trends in documents collected over a period of time, e.g. finding how the performance of a particular sub-contractor varies over a period of time.
- Association: identifies relationships between various attributes (features that have been extracted from the documents) such as whether the presence of one pattern implies the presence of another pattern in a given set of documents.
- Information Visualization: visual text mining or information visualization, puts large textual sources in a visual hierarchy or map and provides browsing capabilities, in addition to simple searching.

4. Text Mining applications on PPRs and benefits

In order to address the recurring problems of PPRs identified in section 2, there is a need for better exploitation and reuse of the information and knowledge stored within PPR documents. However, effective and sustainable exploitation and reuse will only be achieved if relevant knowledge is “quickly” and “easily” available without increasing the workload for new project teams

TM has huge potential for addressing the identified problems of PPRs and to analyse multiple PPR reports simultaneously. In particular TM can be used to support the following tasks in the context of PPRs:

- Extracting common patterns of good or bad practices
- Finding correlation or linkages between commonly occurring terms, for example profit vs. regional targets.
- Clustering or grouping of reports based on pre-determined criteria.

- Summarization and generation of statistics to give a brief overview of groups of reports.
- Search and retrieve a particular context oriented PPR report, and
- Dissemination of analyzed reports across the organization using the World Wide Web.

If useful knowledge can be identified by using these TM techniques, the PPR reports could be effectively sorted or grouped so that the most appropriate documents from previous PPRs could be selected and presented to the project manager and his team for consideration at the start of any new project. In this way, relevant lessons from previous projects might be more appropriately disseminated in order to facilitate the faster exploitation of PPR learning..

Although, there are many commercially available tools for TM their functionality varies. Therefore, to achieve useful results from TM, it is necessary to combine two types of expertise:

- Domain expertise; and
- Text Mining expertise

Figure 2 shows the linkages between these two types of expertise and they will now each be discussed in detail. Domain expertise includes knowledge and understanding of the common content, purpose or intent of PPRs whereas TM expertise includes knowledge and understanding of the available text mining tools, methods and associated software systems.

<<Include figure 2 about here>>

Domain Expertise: Table 1 shows the key topics and typical information found in a PPR report. This information was gathered through reviewing the PPR processes of two construction companies.

<<Include table 1 about here>>

Thus, it can be seen that PPRs of the construction industry^[pmc2] are a huge source of textual data with great potential for textual data mining to extract useful knowledge for future project management, process development, helping decision makers for

appropriate decision, avoiding mistakes, optimizing the process, and analysing trends. However, such analysis of textual data is seldom done in engineering disciplines, which may largely be due to a lack of know-how within this community. This may unfortunately result in a huge loss of potentially valuable information.

Text Mining expertise: There are several commercial products available for text mining, as shown in Table 2 which lists various vendors and their commercial products used for the text mining purposes.

<<Include Table 2 about here>>

The functionality of the Mining products marketed by the above vendors varies substantially, particularly as some of the software systems listed focus exclusively on text mining tools, whilst for other, larger vendors (e.g. SPSS, SAS), the TM tools represent only a small portion of the software developed and marketed.

General Architecture for text Engineering (GATE) [39-41] is another free and open software for both commercial and research purposes. It has been successfully used for Information Extraction purposes. IE differs from traditional techniques in that it does not recover from a collection a subset of documents which are hopefully relevant to a query, based on key-word searching (perhaps augmented by a thesaurus). Instead, the goal is to extract from the documents (which may be in a variety of languages) salient facts about prespecified types of events, entities or relationships. These facts are then usually entered automatically into a database, which may then be used to analyse the data for trends, to give a natural language summary, or simply to serve for on-line access. But, its application is limited to information extraction only.

In table 3, a comparative study has been made of all the tools available for text mining, based on the text mining operations mentioned in Section 3. It can be seen from table 3 that none of the software examined provided the full range of necessary functionalities to achieve all the text mining operations.

<<Include table 3 about here>>

Examples of the application of several of these TM operations on PPR documentation will now be given to further illustrate the potential benefits which may be obtained and facilitate better knowledge reuse and exploitation from PPR processes. The TM suite of Polyanalyst^{pmc3} 5.0 and Text Analytics 2.3 [47] have been used on these

examples. In the next section, an illustrative example has been presented showing how Polyanalyst (Which is one of the leading tools shown in Table 3) could be used to address the exploitation and reuse requirements of PPRs highlighted in section 4.

5. An Industrial Case study

5.1 Description of Scenario and Collection of Documents

This example is based on PPR documentation from two UK based construction companies. 50 PPR documents were collected from work carried out during the last three years and additional information was also collected from the companies, including the PPR agendas, explanations of the process and structure of the PPRs and the criteria and expectations of the companies when they carry out PPRs. There are two, quite different types of reports. The first type of report (Company A) contains textual narration and a description of the project, with the review divided under 16 headings plus textual information providing the lessons learned during the project. These reports were quite long, typically covering between 15 and 25 pages. The second type of reports (Company B) did not have any specific headings, to break the report into sub-sections, and also includes several pictorial notations to express the views of the participants in the PPR. Every PPR included key people who had detailed knowledge of the project and the procedures and the format of the PPR process was uniform across all projects for each of the companies. Analysis of the PPR reports was carried out separately for each of the companies.

It should be emphasised here that all the PPR documentation used in this study were existing documents, i.e. the reports had not been prepared for any TM purposes. Our research is also purely exploratory, with the objective of discovering whether text mining might usefully identify knowledge within such PPR reports, and if so, how might it be presented or used to better disseminate and reuse this knowledge. Although there is currently only limited reuse of PPR results, the practise for identifying and reusing knowledge from old PPR reports would be through manual inspection. An iterative methodology was therefore designed for this research:-

- (1) Manual inspection of PPR reports and discussions with domain experts to determine types and examples of knowledge which should be found within the reports.
- (2) TM experiments using the same PPR reports.

- (3) Manual examination and evaluation of the TM results (with reference to domain experts as and when required).
- (4) Further (usually directed) TM experimentation and manual evaluation (repeating steps 2 and 3 as required).

Before any TM experiments were carried out, the reports were therefore studied manually and the format, terminology used, agenda and key issues were considered and discussed. Utilising an iterative process, the key knowledge areas where KDT and TM should concentrate were identified, discussed with collaborators from the construction company and prioritised (using their Domain Expertise). This enabled the KDT process to concentrate on key areas and processes that were most important to the collaborators. The next section discusses the ontology development process that was used to support knowledge discovery.

5.2 Ontology Development

An ontology of important and common terms within the PPR reports was developed to facilitate the processes of knowledge search and knowledge discovery. In practical terms, to satisfy the requirements of simple exploratory TM experiments, it would be sufficient to analyse example PPR reports, find the common terms (which indicate the types of knowledge that are likely to exist in the reports) and then combine and refine these outputs with information provided by our collaborating companies related to key areas of interest. The results from this analysis and refinement could then be used to direct the text mining experiments to identify any useful knowledge that may lie hidden within the reports. However, our aim in this research is to develop a more generic approach for knowledge discovery within PPR reports, so that particular types of knowledge might be targeted for knowledge discovery. With this in mind, an ontology has been developed based on the combination of key terms specified by our collaborating companies and domain experts, based on the useful knowledge manually identified in the example reports.

A set of hierarchies has been developed where each hierarchy has a single root which indicates the main topic areas shown in Figure 3. The Top Level hierarchy shows that the Learning or Outcome from a project will relate to one or more of the following topics, Finance, Quality, Communication, Building, Health and Safety, Labour,

Environment, Time, Security, Materials, Plant or Project Stages of a construction project.

<<Include figure 3 about here>>

Each of these areas is then considered in further detail in separate hierarchies, using key words from items of learning identified in the example PPR reports. For example, figure 4 shows the hierarchy for finance related key words.

<<Include figure 4 about here>>

In order to focus the text mining on particular topics of interest it is likely that two or more hierarchies will be utilised simultaneously. For example, to identify knowledge about delays on particular parts of buildings, both the “Time” hierarchy and the “Buildings” hierarchy would be used.

Another motive for this Ontology based approach is to deal with the semantic issues and issues of multilinguality. Different terms may be used for the same context by different members of a team or the same term might be used for different contexts. The ontology was used in the TM stage of our methodology by exploiting the dictionary feature of the TM software. Any particular term can be added to the dictionary and modified based on the context. The dictionary feature also helps in identifying the meaning of a particular keyword/phrase which has several meaning for different context. Therefore, the combination of modification of dictionary feature and ontology based approach have been used to take care of semantic and multilingual issues at the beginning of the work. However, our future research will focus on enhancing this semi-automatic process to automatic one.

5.3 KDT and Text Mining process on PPRs

Figure 2 shows the steps involved in the KDT process in the present context. These steps are now described in the following subsections.

5.3.1 Transformation and Loading: The very first step in the KDT process for PPRs is the transformation of gathered textual data into a format suitable for TM. Company A reports were semi-structured manually and divided into categories based on the

original headings (issues) within the PPRs. This was done to enable exploration of possible patterns within particular issues as well as across different issues. All the textual data was imported from .doc file to .txt file to ease the computational burden. Analysis for each company was done separately. The whole file was then loaded as a new project into the PolyAnalyst software system.

5.3.2 Pre-Processing: Unlike other data mining applications, in this research pre-processing of the textual data was carried out after loading the data into the PolyAnalyst system. Pre-processing is mainly done to reduce information overload and generate metadata. Textual data pre-processing steps are as follow:

- Removal of “Unwanted” text: In TM, punctuation delimiters such as commas, apostrophes, exclamation marks, as well as alpha numeric text and numerals are referred to as “Unwanted texts”. These delimiters do not help to differentiate between the textual inputs, since they are reasonably well distributed throughout the report. Therefore, they are removed from the text.
- Removal of non informative words: A common technique to improve the accuracy of TM results and to reduce redundancy in the computation is to remove frequently occurring, common words from the text. These words are often defined by a “stop list” which typically consists of 200-300 words, including articles, prepositions, conjunctions and some high frequency words for example *are, the, from, can, may, etc.* In the present context, pre-identified words, which did not contribute to the essential information within the report were removed to reduce the size of document and avoid information overload.
- Stemming: This is a common pre-processing step in textual databases and refers to a simplified form of morphological analysis by simply truncating a word. For example, agree, agrees, agreed, agreeing can all be stemmed to agree. The inbuilt stemmer of the PolyAnalyst software helps to reduce the document space and can provide a more concise document space representation.

5.3.3 Text-Mining of PPRs: After pre-processing, the PPRs are ready for TM, which involves using various tools and techniques to extract patterns, trends and useful knowledge. It also uses techniques of information extraction and

information retrieval to derive useful information. As this paper focuses on the application of technology to derive useful knowledge, less information is provided about each technique but more details are given about how it can be applied. references for more details of each technology. The following subsections will discuss the key TM tasks (listed in section 4) to address current problems commonly identified in PPRs.

- ***Extracting common patterns of good or bad practices:*** Text Analysis (TA) is one of the important techniques used for this purpose. It provides the morphological and semantic analysis of unstructured textual PPR reports. TA extracts and counts the most important words and word combinations from the textual PPR reports, and stores terms-rules for tokenizing PPR records with pattern of encountered terms. The terms or a combination of terms are generated as rules which record the number of times each term or combination of terms exist and where the occurrences are. These rules can be applied on the PPRs to find patterns of events or activities, issues, causes or achievements. However, domain expertise may be needed to determine the relevancy and importance of combinations of terms identified in this manner. For example, if the word “scaffolding” occurred several times in the Health and Safety sections of the PPRs, perhaps in some combination with other words, this may well indicate that there is a recurring problem with scaffolding in a particular context or with the safety of a particular type of scaffolding. Further, these rules can be used as an input to other visualization techniques (e.g. Link Analysis) and classification tools (e.g. Decision Tree, or Decision Forest) in the software system. TA therefore can be used to highlight the commonly used words in various areas of PPRs such as planning, estimation, errors or mistakes, quality, health and safety, defects and many more. Identifying problems, issues and possibly their causes in this way may help managers to avoid them in future projects. TA can also be used to find where a particular word, or its synonyms are used in various reports. Figure 5 shows the screen shot of Text Analysis on PPR reports and generated rules. The left hand column shows the high-level hierarchies, the middle column shows the time hierarchy and the corresponding keywords and phrases automatically extracted by text analysis in the form of rules and the right

column shows the reports. Furthermore, a combination of generated rules can be used to create a new rule based on the OR, AND, XOR_[pmc4] operators. For example, if a decision maker is interested in finding the correlation between “Extension of Time” of project AND “Profit”. A new rule can be created and corresponding report can be found by applying this rule .

<<Include figure 5 about here>>

- ***Finding correlation or linkages between commonly occurring terms:*** The commonly used technique for this purpose is Link Term (LT) analysis. It visually represents the complex patterns of relations between key words in the textual report. Results of the link terms analysis are presented as a graph displaying a cluster of linked objects supporting various object manipulation and drill down operations. This visualization technique of LT facilitates better understanding of hidden structure within the investigated PPR reports based on textual data, and helps interesting patterns to be identified for further investigation. This mechanism provides the quickest way to understand the most prominent semantic characteristics of the explored textual reports. Here it is important to mention that LT analysis does not require any pre-processing of data.

<<Include figure 6 about here>>

Figure 6 shows an example of the application of link term analysis on the PPR reports. Link term analysis has been applied on the ontology developed for companies. The strength of the link shows the correlation between keywords and phrases. In figure 6, “Lead in time” is directly linked to the “Alteration”, with greater support. Clicking on the link line produces the cluster of reports where Alterations/changes have affected the lead in times. Similarly, linkage between “Value engineering” and “Profit” shows the pattern where adoption of value engineering led to profit in the projects. Other key words and phrases can be explored in a similar manner. This analysis provides a way for information retrieval and enables relevant subsets of the original PPR data to be collected for further exploration based on the particular topics (or terms) of interest. In a similar manner, link analysis between each high level hierarchy has been developed to show one to one relationships between ontology groups

as shown in figure 7 where linkages between finance and time are shown, i.e. various attributes of “Time” are linked with attributes of “finance”.

<<Include figure 7 about here>>

- ***Clustering or grouping of reports based on defined criteria:*** Text categorization (TC) can be used to automatically split up the PPR reports in to more homogeneous clusters based on keywords found in the PPR report. This tool therefore enables important clusters of reports (or pieces of text) to be identified as expressing similar ideas, issues, or concepts. TC first uses TA to find the key topics and subtopics and then TC can extract groups of reports (or collections of sections of text) that focus on particular topics or certain categories of interest. TC is an important tool of TM, which automatically builds a hierarchical tree-like taxonomy of topics and subtopics extracted from unstructured textual reports.

Figure 8 shows an example of TC applied on the present case study. The key topics of interest identified include “good profit”, “small profit”, “changes made to project” and “designing”. Here, “Good” shows the cluster of reports where a good profit has been made during the project. Similarly, “small” shows the cluster of reports of those projects where less profits have been made. “Change” shows the cluster of reports where planning has been changed during the operation of projects. “Design” shows the cluster of reports which were mainly related to design of buildings. Therefore, it can be seen that by using TC the whole dataset has been divided into smaller data sets according to their characteristics.

<<Include figure 8 about here>>

- ***Summarization and generation of statistics to give a brief overview of groups of reports:*** Summary statistics and summarization tools of TM are discussed below:

Summary Statistics: Basic statistics can be generated for the PPR text at various stages of the TM to compare its attributes, key words, or generated rules. These include means, standard deviation, frequencies, frequency chart

for each category, strings, yes and no variable, etc. In the present example, these statistics help in finding the frequencies of frequently used words in the PPR reports. In this way, summary statistics are useful in identifying which reports are more important in the context of a particular issue and an example of their use is given in Figure 9. It shows that for company A out of 27 projects a profit has been achieved in 17 of project and 10 suffered through loss or no profit. Similarly, out of 26 projects, additional cost has been incurred in 8 projects.

<<Include figure 9 about here>>

Summarization: PPRs are commonly neglected and under exploited because of their length and the consequent time and effort that subsequent project teams need to apply in order to read through them to identify any knowledge and lessons that are relevant to their new project. Summarization techniques can reduce the PPR reports to a fraction of their original size whilst still retaining the significant content in the summaries. Summarization techniques determine the semantic weight of sentences written in the PPRs and only those sentences whose semantic weight is higher than the threshold are kept. The final summary then lists the most important sentences written in the PPR report. The size of the summary can be changed by changing the semantic threshold. Useful knowledge can be retained within very short summaries, as shown in the following example.

*No profit gain, with extension of time. It is understood that the **Framework Manager** (at Head Office) has made an alternative concession to Company "X". The people giving instructions to "ABCD" were not in a position to explain and "X" middle **managers** had no wish to assist "ABCD", as they did not want to see the work outsourced. The X **Framework Manager** negotiated this Contract and the arrangements were beyond the control of this Business Unit. "ABCD" have produced them, but the "X" **Manager** concerned, ----, continues to object to minor items and will not sign then off.*

In this case an original document of approximately 12 pages has been reduced to a useful set of summarized information within one paragraph.

- **Search and retrieve a particular context oriented PPR report:** The semantic search engine of TM provides one of the most powerful search capabilities for PPR reports. It is very similar to natural language query, where queries are made using natural language, typing a question in conversational English. The result pane will display all the related answers to the query. Results are

displayed as a tree-like structure based on the question. This sub-tree of concepts that are related to the query may help the user in simulating a better answer to the asked query.

For example, during this study, a search was made for reports which are related to an extension of time for the project. The search result pulls all the sentences from the original PPR reports that best respond to the query and places them in the result pane. It shows all the text where the project report mentions the extension of time, or extension of period as shown in figure 10.

<<Include figure 10 about here>>

- ***Dissemination of analyzed reports across the organization:*** A major benefit of the TM systems discussed in this paper is that all the reports generated by the software system for almost all the aforementioned purposes could be exported as “.htm” files, which can be distributed to various project partners and other interested individuals within the company across the globe via the world wide web (providing appropriate ID, password and access rights are set up for these individuals). This should provide an effective dissemination channel for analyzed PPR reports across many organizations. However, as previously mentioned, people are unlikely to reuse even useful and relevant knowledge if it costs them excessive time and effort in finding or making use of it. The TM system therefore addresses this problem area in at least two ways, firstly through convenient dissemination processes using the world wide web and secondly by reducing effort required for knowledge reuse by providing concise (or summarised) knowledge extracted from the PPRs which is linked to the specific full PPRs, if required.

This section therefore has demonstrated that TM has huge potential to benefit the PPR process by addressing known tasks in many ways.

6. Discussion and Conclusion

Post Project Reviews (PPRs) are a rich source of data and information and valuable knowledge can be extracted if the organisations have the time and resources to analyse them. However, there are several types of problem in PPRs that are repeatedly reported in research literature. These problems have been considered in detail in section 2 of this paper, but in general relate to issues of how knowledge from PPRs and the recorded lessons learnt from projects can be quickly and effectively identified and disseminated to relevant people to maximise the benefits gained by the organization and new projects.

KDT and TM are fairly new research areas which address problems of information overload and provides many tools and techniques to help identify useful relevant information and present it to users in a concise or easily searchable form. As discussed in section 3 of this paper, KDT and TM provide mostly automated techniques that aim to discover high level information in huge amounts of textual data. The main focus of this paper has therefore been to determine the potential benefits that TM can offer to PPR processes.

The amount of knowledge generated during PPRs makes them ideal candidates to explore the use of KDT and TM techniques, and this study has helped to show that TM has an enormous potential for use and benefits for future project improvement, avoiding mistakes, improving customer service, and making the organization aware of previously unknown facts and problem areas. The use of KDT and TM does not imply eliminating the human intervention or responsibilities in knowledge dissemination and exploitation but has major advantages in identifying and disseminating new knowledge in concise, easy to manage and use forms. This has been illustrated using a case-study example to examine the important tasks highlighted in section 4) associated with the thorough analysis of PPR documents. This illustrative example, which is discussed in section 5 of this paper, has shown that TM could prove to be a very useful tool to discover patterns, trends and hidden relationships between various issues, topics and keywords used in PPR reports.

The potential for exploitation of this research goes far beyond PPRs since the KDT and text mining could be applicable to a whole range of text based reports and other documents within the construction industry. The next steps include the application of

GATE software for information extraction. The application of several other text mining techniques: question-answering, expert finding, sentiment/opinion detection, recommendation; part-of-speech tagging, (shallow) parsing, semantic role labelling etc. are also being considered as future research scope.

Reference:

- [1] Kerth, N., (2000), The ritual of retrospectives: how to maximise group learning by understanding past projects, *Software Testing & Quality Engineering*, September/October, 53-57.
- [2] Disterer, G., (2002) Management of Project Knowledge and Experiences. *Journal of Knowledge Management*, vol. 5, 512-520
- [3] Busby, J.S., (1999), An Assessment of Post-Project Reviews. *Project Management Journal*, 30(3), 23-29.
- [4] Carrillo, P.M. (2005) Lessons Learned Practices in the Engineering, Procurement and Construction sector. *Journal of Engineering, Construction and Architectural Management*, 12(3), 236-250.
- [5] Orange, G., Burke, A. and Cushman, M., (1999), An approach to support reflection and organisational learning within the UK construction industry, *Paper presented at BITWorld'99: Cape Town, SA, 30 June -2 July* (<http://is.lse.ac.uk/b-hive>).
- [6] Kamara, J. M., Anumba, C. J., Carrillo, P. M. and Bouchlaghem, N., (2003), Conceptual framework for live capture and reuse of project knowledge, in Amor, R. (ed.) *Construction IT: Bridging the Distance. Proceedings of the CIB W78's 20th International Conference on Information Technology for Construction*, New Zealand, 23-25 April, 178-185.
- [7] Fayyad, U. M., Piatetsky-Shapiro. G., Smyth, P., and Uthuruswamy, R., (1996) Eds, *"Advances in Knowledge Discovery and Data Mining"*, Menlo Park, CA: AAAI/MIT Press.
- [8] Lane, K. (2000) Project Management Today www.projectnet.com (accessed 25.03.07).
- [9] Tan, H C, Carrillo, P, Anumba, C, Kamara, JM, Bouchlaghem, D, Udejaja, C (2006). Live capture and reuse of project knowledge in construction organisations, *Knowledge Management Research and Practice*, 4, 149-161.

- [10] Sowards, D., (2005). The value of post-project reviews. *Contractor*, 52(8), 35-36.
- [11] Motteram, G and Sizer, J. (1992). Evaluating and Controlling Investments in Advanced Manufacturing Technology. *Management Accounting*, 70(1), 26-30.
- [12] Koners, U and Keith, G (2007) Managers perception of learning in new product development. *International Journal of Operations and Production Management UK*, 27(1), 49-68.
- [13] Robertson, S. and Terry, W., (2006). Understanding project failure using cognitive mapping in an insurance project. *Project Management Journal*, 37(4), 55-71.
- [14] Garon, S (2006) Space project management lessons learned: a powerful tool for success. *Journal of Knowledge Management*, 10(2), 103-112.
- [15] Von Zedtwitz, M. .(2003), Post-project reviews in R& D. *Research-Technology Management*, 46(5), 43-49.
- [16] Pyra, J., (2002). Risk management post analysis: gauging the success of a simple strategy in a complex project. *Project Management Journal*, 33(2), 41-48
- [17] Braniš, M and Christopoulos, S (2005) Mandated monitoring of post-project impacts in the Czech EIA. *Environmental Impact Assessment Review*, vol. 25(3), 227–238.
- [18] Terry, W. (2004). Identifying the hard lessons from projects easily. *International Journal of Project Management*, 22(4), 273-279.
- [19] Terry, W.(2003). Learning from projects. *The Journal of the Operational Research Society*, 54(5), 443.
- [20] Roth, G and Kleiner, A.(1998). Developing organisational memory through learning histories. *Organisational Dynamics*, Autumn, 27(2), 43-60.
- [21] Schindler, M. and Eppler, MJ. (2003). Harvesting project knowledge: a review of project learning methods and success factors. *International Journal of Project Management*, 2, 219-228.
- [22] Robinson, H. S, Carrillo, P.M., Anumba, C., Al-Ghassani, A. M.,(2005). Knowledge management practices in large construction organisations. *Engineering, Construction and Architectural Management*, 12(5), 431 – 445.

- [23] Newell, S, Bresnen, M, Edelman, L, Scarbrough, H and Swan, J (2006). Sharing knowledge across projects: limits to ICT-led project review practices. *Management Learning*, 37(2),167-185.
- [24] Grobelnik, M and Mladenic, D (2005). Automated knowledge discovery in advanced knowledge management. *Journal of Knowledge Management*, 9(5),132-149.
- [25] Lindkvist, B. (2001), *Kunskapsöverföring mellan product utvecklings projekt* Stockholm: Stockholm School of Economics.
- [26] Pitman, B. (1991). A systems analysis approach to reviewing completed projects .*Journal of Systems Management*, 42(6),6-37.
- [27] Ayas, K. (1996) Design for Learning and Innovation. *Long Range Planning*, 29 (6), 898-906.
- [28] Prahalad, C. K. and Hamel, G.(1990). The core competence of the corporation. *Harvard Business Review*, 68(3), 79-92.
- [29] Harding, J A, Shahbaz, M, Srinivas and Kusiak, A, (2006), "Data Mining in Manufacturing: A Review", *American Society of Mechanical Engineers (ASME): Journal of Manufacturing Science and Engineering*, 128(4), 969-976.
- [30] Karanikas, H and Theodoulidis, B, 'Knowledge discovery in text and text mining software', *Technical report, UMIST - CRIM*, Manchester,
- [31] Fan, W., Wallace, L., Rich, S., and Zhang, Z., (2006), Tapping the power of text mining, *Communications of the ACM*, 49(9), 77-82.
- [32] Cohen, A.M., and Hersh, W. R., (2005), A survey of current work in biomedical text mining, *Briefing in Bioinformatics*, 6(1), 57-70.
- [33] Uramoto, N., Matsuzawa, H., Nagano, T., Murakami, A., Takeuchi, H., and Takeda, K., (2004), A text mining system for knowledge discovery from biomedical documents, *IBM Systems Journal*, 43 (3), 516-533.
- [34] Tseng, Y. H., Lin, C. J., and Lin, Y. I., (2007), Text Mining techniques for patent analysis, *Information processing and management*, 43(5), 1216-1224.
- [35] Kloptchenko, A., Eklund, T., Karlsson, J., Back, B., Vanharanta, H., and Visa, A., (2004), combining data and text mining techniques for analysing financial reports, *Intelligent system in accounting, finance, and management*, 12, 29-41.
- [36] Dorre, J., Gerstl, P., and Seiffert, R., (1999), Text Mining: Finding Nuggets in mountains of textual data, *KDD-99, Sandiego CA, USA*, 398-403.

- [37] Menon, R., Tong, L.H., Sathiyakeerthi, S., Brombacher, A., and Leong, C., (2004), The needs and benefits of applying textual data mining within product development process, *Quality and reliability engineering international*, 20, 1-15.
- [38] Loh, S., Oliveira, J P M, and Gameiro, M. A., (2003), Knowledge discovery in texts for constructing decision support systems, *Applied Intelligence*, 18, 357-366.
- [39] H. Cunningham, D. Maynard, K. Bontcheva, V. Tablan. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*. Philadelphia, July 2002
- [40] K. Bontcheva, V. Tablan, D. Maynard, H. Cunningham. Evolving GATE to Meet New Challenges in Language Engineering. *Natural Language Engineering*. 10 (3/4), pp. 349-373. 2004.
- [41] Y. Li, K. Bontcheva, and H. Cunningham. SVM Based Learning System For Information Extraction. J. Winkler, M. Niranjan and N. Lawrence (Eds.): *Deterministic and Statistical Methods in Machine Learning*, LNAI 3635, Springer Verlag, pp. 319-339. 2005.
- [42] Harding, J A, Shahbaz, M, Srinivas and Kusiak, A, "Data Mining in Manufacturing: A Review, in *American Society of Mechanical Engineers (ASME): Journal of Manufacturing Science and Engineering*, Vol. 128, No. 4, 2006), 969-976.
- [43] Choudhary, A.K., Harding, J.A. and Popplewell, K., "Knowledge Discovery for Moderating Collaborative Projects", *Proceedings of the 4th IEEE International Conference on Industrial Informatics 2006 (INDIN '06)*, Singapore, August 2006, pp 519-524
- [44] Neaga, E. I. and Harding, J. A., "An Enterprise Modelling and Integration Framework Based on Knowledge Discovery and Data Mining," *International Journal of Production Research*, 43(6), pp. 1089-1108, 2005
- [45] Shahbaz, M, Srinivas, Harding, J A and Turner, M, "Product Design and Manufacturing Process Improvement using Association Rules", *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, Vol. 220, pp. 243-254, 2006.

[46] Choudhary, A. K., Harding, J. A., and Tiwari, M. K., ““Data Mining in Manufacturing: A Review Based on the Kind of Knowledge”, *Journal of Intelligent Manufacturing*, July 2008, (Published online),ISBN 1572-8145.

[47] Polyanalyst 5.0, and Text Analyst 2.3 www.megaputer.com

Table 1: Issues and context included in PPR reports

<u>Project Outcome</u> Financial Contract period variation	How did the job go? Profit or Loss? Extension of project period? Did the project finish on time? Were resource effectively allocated?	<u>Quality Assurance</u>	Was quality combined with speed? Was the project delivered on programme to a good standard? Any comment from Quality auditors?
<u>Estimating</u> Rates No. of staff used costs	Any comments for or from the estimators? Any rate completely out of touch? How many staff were used? Cost of temporary services, scaffolding/skips?	<u>Waste/environmental issues</u>	Were there any particular operations wasteful on Could the operation be improved by changing the material or method?
<u>Planning</u> Sequence Duration Events Programme	Any comments for or from the planners? Was the sequence shown on programme correct? Was the duration reasonable? Were any important events omitted? Were any short term programmes used? Who prepared those and did they work?	<u>Health and Safety</u>	Are there any reported accidents or incidents? What are the causes? How did we perform over all? Any health and safety lessons to be learnt from this project?
<u>Method of work</u>	Was any method of work used on site which differed substantially from what was decided before the project?	<u>Interaction with Design team</u>	How did we interact with the design team? What did we do well? What went wrong? Who from the organization will do the post project review with client?
<u>Procurement</u>	Was there any problem in getting subcontractors on time? Any comments for or from the buyers?	<u>Interaction with Client</u>	How did we interact with client? What did we do well? What did we do wrong? What major changes arose from meeting with clients? Was the customer's feedback, positive or negative?
<u>Teamwork</u>	Did the designer and contractor work well together? How did they work well together- Analyse? Was there good liaison between design team and contractors	<u>Subcontractors</u>	Who will carry out the subcontract review?
<u>Lease agreement</u>	Was the document clearly communicated?	<u>Communication</u>	Was there clear communication between management, client and the project team?
<u>Changes</u>	What changes were made in the project and why?	<u>Snagging/Defects</u>	Where do we stand with snagging/defects?
<u>Mistakes/Errors</u>	Were there any notable mistakes made on the site? What was the cause? How can that be corrected?	<u>Innovation</u>	Did any innovative or interesting ideas emerge during whole project?
<u>General comments and suggestions about the whole project?</u>			

Table 2: Text mining products, vendors and their websites

<u>Vendor</u>	<u>Website</u>	<u>Product name</u>
<u>Inxight</u>	<u>www.inxight.com</u>	<u>Smart discovery, Vizserver</u>
<u>SPSS</u>	<u>www.spss.com</u>	<u>Text smart1.0, Clementine</u>
<u>Autonomy</u>	<u>www.autonomy.com</u>	<u>Autonomy incorporated</u>
<u>SAS</u>	<u>www.sas.com</u>	<u>SAS Text miner</u>
<u>Clear forest</u>	<u>www.clearforest.com</u>	<u>ClearForest text-analysis suite</u>
<u>Megaputer</u>	<u>www.megaputer.com</u>	<u>Polyanalyst5.0, Text Analyst2.3</u>
<u>IBM</u>	<u>www.ibm.com</u>	<u>Intelligent miner for text, TAKMI</u>
<u>Convera</u>	<u>www.comvera.com</u>	<u>Retrieval ware</u>

Table 3: comparative study of text mining software in terms of performing the following text mining operations.

<u>Product</u>	<u>FE</u>	<u>TBN</u>	<u>SR</u>	<u>TC</u>	<u>Clus.</u>	<u>Sum.</u>	<u>TrA</u>	<u>Asc</u>	<u>In.Vs</u>
<u>1 Smart discovery</u>	<u>X</u>			<u>X</u>		<u>X</u>			<u>X</u>
<u>2 Text miner 1.0</u>	<u>X</u>			<u>X</u>	<u>X</u>				<u>X</u>
<u>3 Autonomy</u>	<u>X</u>	<u>X</u>		<u>X</u>	<u>X</u>	<u>X</u>			
<u>4 SAS Text miner</u>	<u>X</u>			<u>X</u>	<u>X</u>			<u>X</u>	
<u>5 Clear forest</u>	<u>X</u>	<u>X</u>	<u>X</u>	<u>X</u>				<u>X</u>	<u>X</u>
<u>6 Polyanalyst 5.0/ Text Analytics2.3</u>	<u>X</u>	<u>X</u>	<u>X</u>	<u>X</u>	<u>X</u>	<u>X</u>			
<u>7 Intelligent Miner</u>	<u>X</u>			<u>X</u>	<u>X</u>	<u>X</u>			
<u>8 Retrieval ware</u>	<u>X</u>			<u>X</u>	<u>X</u>	<u>X</u>			

FE: Feature extraction, TBN: Text Based Navigation, SR: Search and Retrieval, TC: Text Categorization, Clus.: Clustering, Sum: Summarization, TrA: Trend Analysis, Asc: Association, InV: Information visualization.

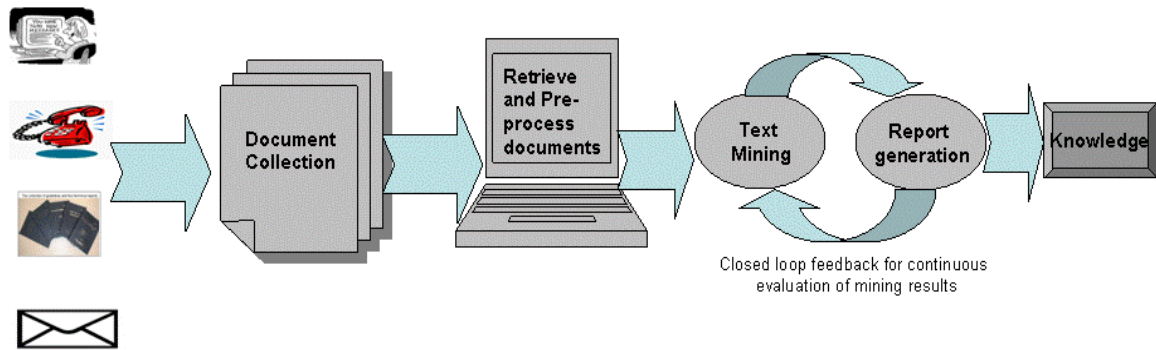


Figure 1. Knowledge discovery in text and text mining process

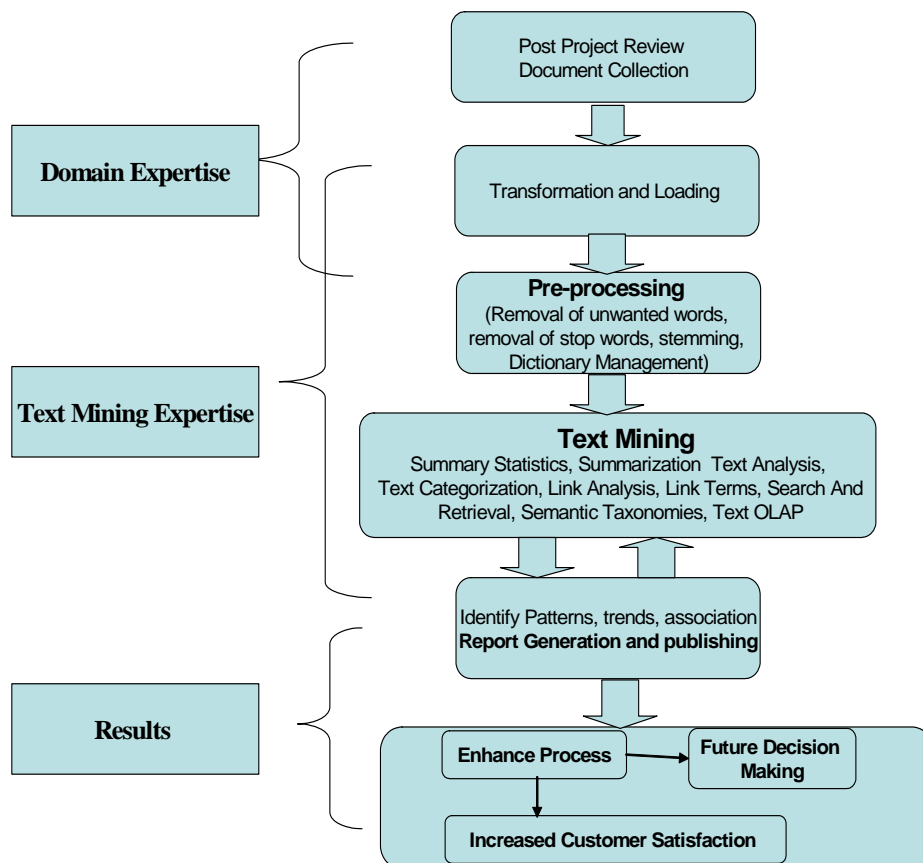


Figure 2: Linkage between Domain expertise and Text mining expertise, and KDT process.

**TOP
LEVEL**

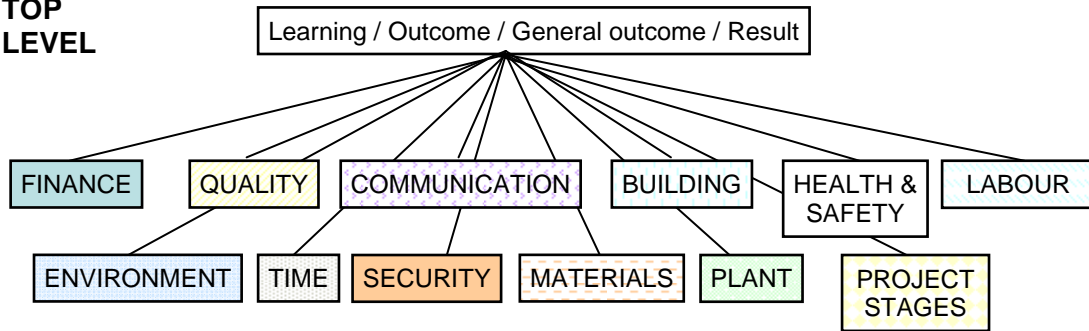


Figure 3. Top Level hierarchy relating to construction PPRs

FINANCE

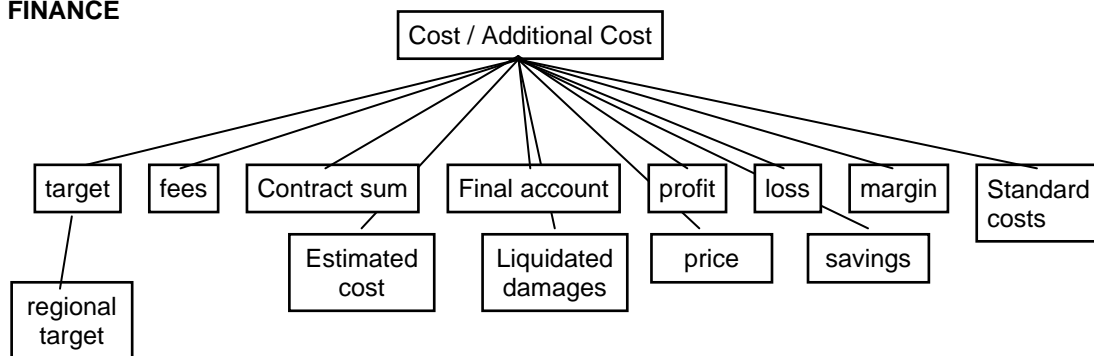


Figure 4. Hierarchy relating to finance from top level hierarchy

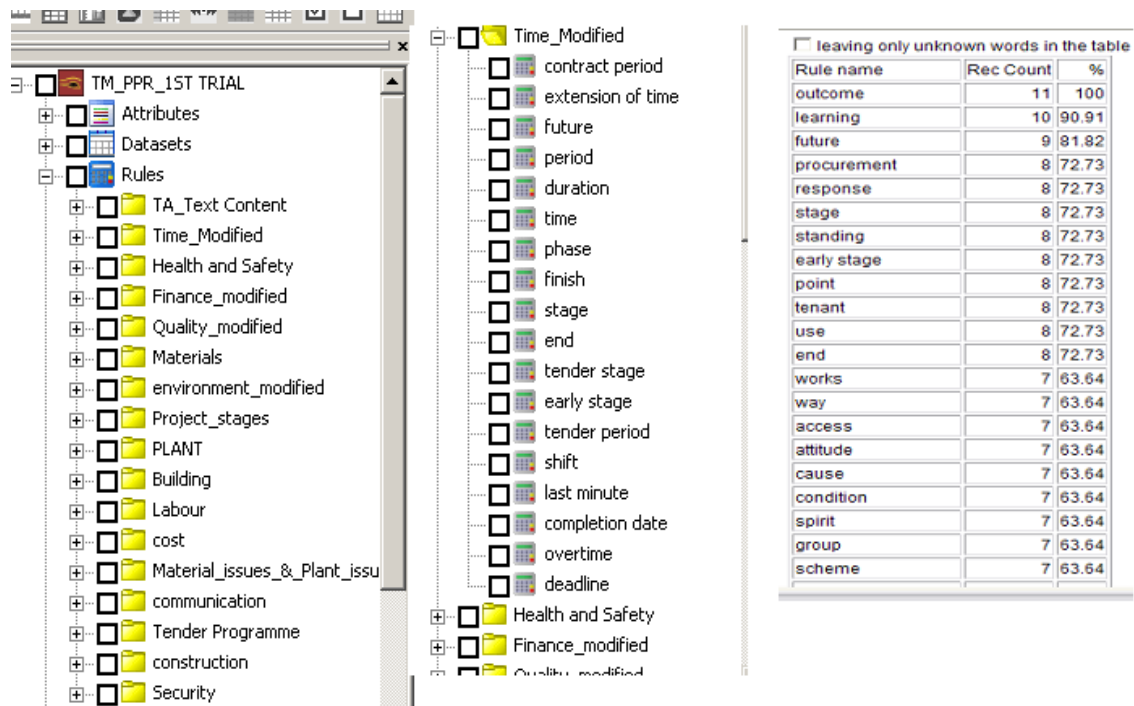


Figure 5. Text Analysis on Post Project reviews and generated rules

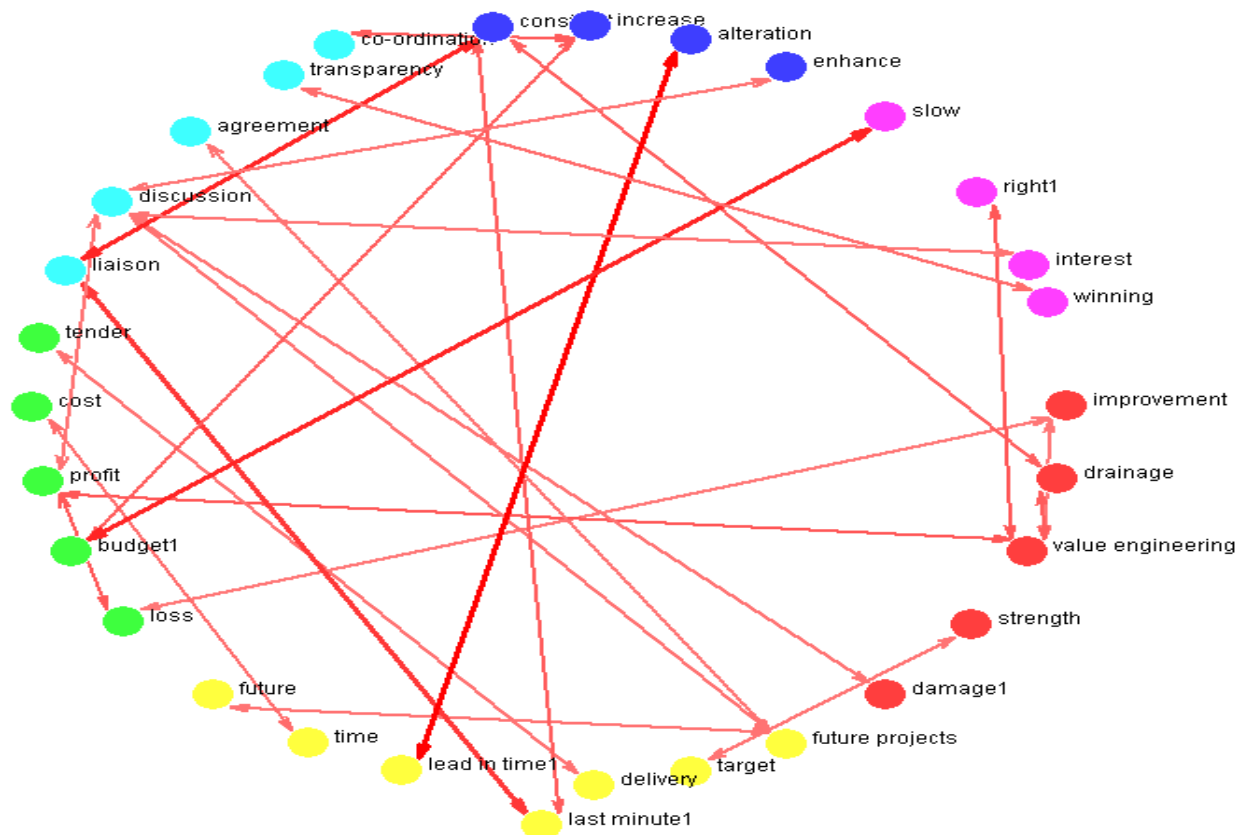


Figure 6. Link analysis between Keywords and their correlation

Time

Finance

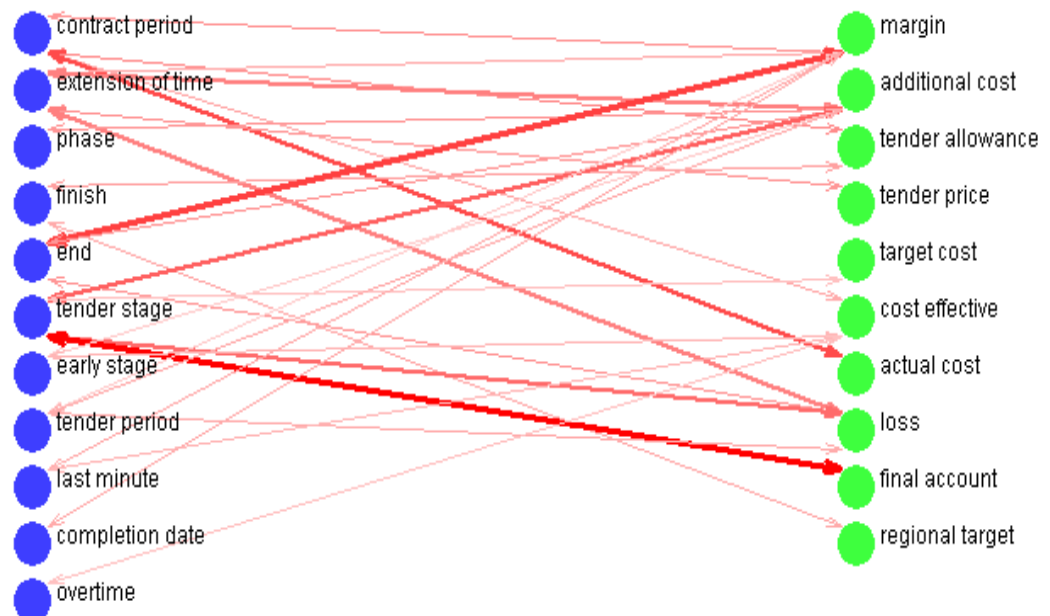


Figure 7. Linkage between high level hierarchies “finance” and “time”.

Categorization Tree

This was a design & build project secured via negotiation with the customer. There was no traditional tender - but there was a considerable amount of negotiation about the content and price of the works. The design consultants were first employed by the customer, prior to the involvement of X, and then transferred to work for X. Cost: The Final Account has been agreed and is marginally below the original, agreed budget - giving a good KPI for Cost. The job made a profit in excess of Regional Targets. Time: The project took 51 weeks and was handed over on programme - giving a good KPI for Time. It was noted that the customer issued very few instructions / variations and simply allowed IPSL to build the project agreed prior to starting work.

Figure 8: Application of Text Categorization on small set of PPR reports

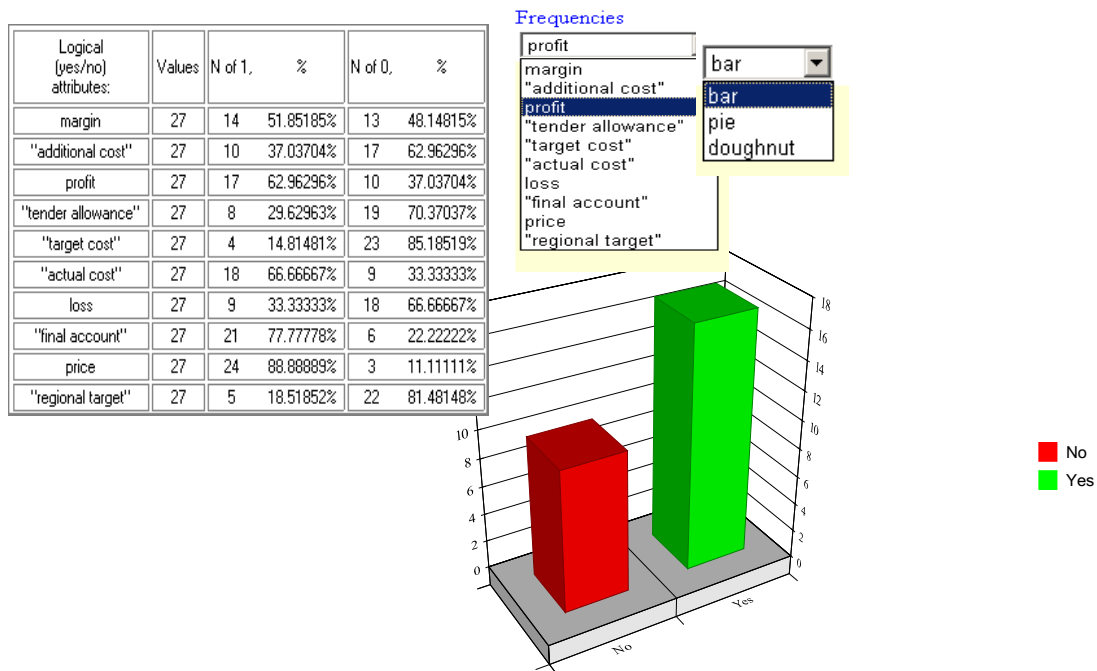


Figure 9: Application of summary statistics on PPRs.

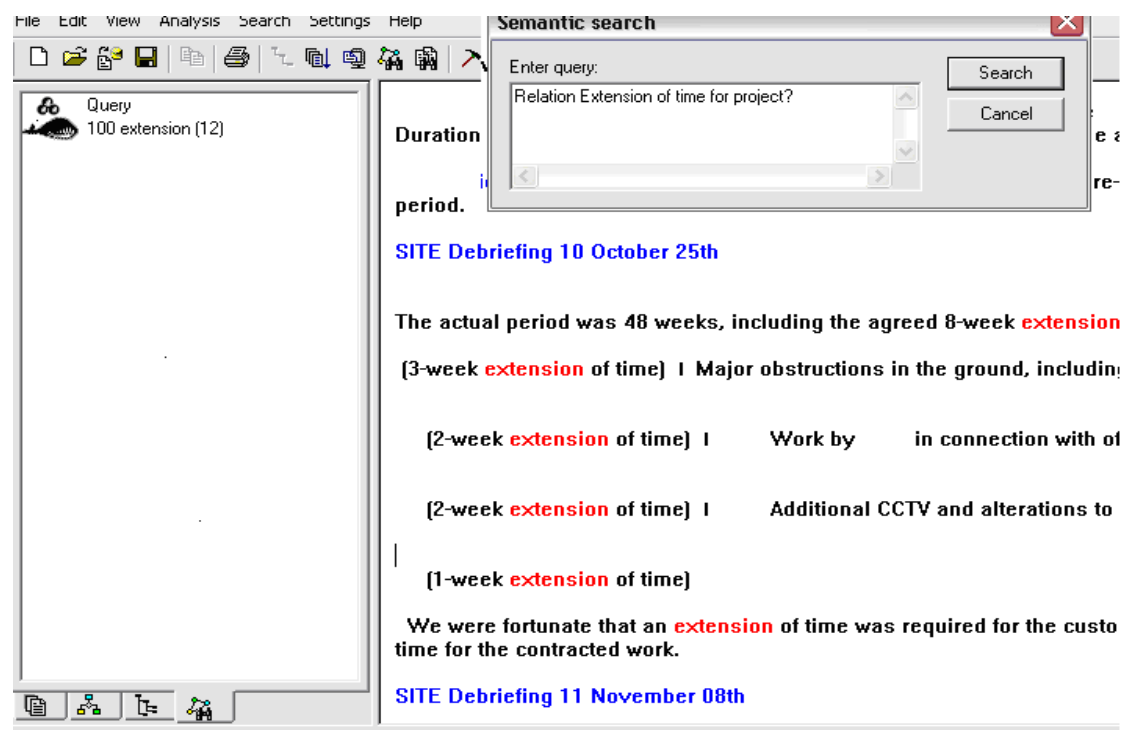


Figure 10: Semantic search on PPR reports
