Summer 6-27-2016

FEATURE-BASED SENTIMENT ANALYSIS OF CODIFIED

Benjamin Matthies

South Westphalia University, matthies.benjamin@fh-swf.de

FEATURE-BASED SENTIMENT ANALYSIS OF CODIFIED PROJECT KNOWLEDGE: A DICTIONARY APPROACH

Benjamin Matthies, South Westphalia University of Applied Sciences, Hagen, Germany, matthies.benjamin@fh-swf.de

Abstract

Most project-based organizations possess extensive collections of diverse project documents. Exploring the knowledge codified in such project documents is specifically recommended by the common project management guidelines. In practice, however, project managers are faced with the problem of information overload when trying to analyze the extensive document collections. This paper addresses this problem by combining two approaches already established in other disciplines. The first involves the development of a Project Knowledge Dictionary (PKD) for the automated analysis of knowledge contents codified in project documents. The second involves the integration of a sentiment analysis where concrete opinion expressions (positive/negative) are identified in connection with the codified project knowledge. Building on this, three mutually complementary analyses are demonstrated, which provide the following contributions: (1) determining the volume and distribution of five project knowledge types in project documents; (2) determining the general sentiment (positive/negative) in conjunction with the textual description of the project knowledge; (3) classifying project documents by their sentiment. By this means, the proposed solution provides valuable insight into the emotional situation in projects and contributes to the emerging research issue of project sentiment analysis. Furthermore, the solution makes a contribution to overcoming the information overload by assessing and organizing the knowledge content of large document collections.

Keywords: project knowledge, project documentation, sentiment analysis, dictionary.

1   INTRODUCTION

In the course of a project, many different documents are usually produced, such as project appraisals, performance reports, or post-project reviews (Wysocki 2014). Such documents typically report specific experiences gained in projects that may be of great value to a project-based organization (Almeida & Soares 2014; Schindler & Eppler). On the one hand, the analysis of such documents enables the status and development of ongoing projects to be assessed and compared (see Meier 2013), while the analysis of historical project documents allows the lessons learned from past projects to be extracted and used to generate valuable knowledge for the planning of new projects (Koners & Goffin 2007).

In practice, however, project managers are faced with a central problem when trying to analyze available document collections: the information overload (see Caniëls & Bakens 2011; Haksever 2000; Haksever & Fisher 1996; Strait 2006). Document archives in project-based organizations are usually quite extensive. In addition, the documents largely contain unstructured, text-based contents and suffer a lack of contextualization, which means no concrete content references to specific project tasks or problem areas (Caniëls & Bakens 2011). All this results in project managers' scarcely being able to process the document collections to extract relevant project knowledge with any level of efficiency (Haksever 2000; Strait 2006).

One solution for handling these extensive document collections could be the automated text analysis approaches that are currently receiving strong attention (Carrillo et al. 2011; Grobelnik & Mladeni 2005; King 2009). In this regard, established methods from other disciplines are often transferred to the project management context. Choudhary et al. (2009), for example, applied statistical keyword analysis to project documents, a method that has already been used for quite some time within customer management to analyze customer reviews (see Hu & Liu 2004). Given the latest technological innovations, project management is far from exhausting the potential for such interdisciplinary exchange.

This study addresses the problem of information overload in the project environment by combining two approaches already established in other disciplines for handling it. The first involves the development of a *Project Knowledge Dictionary* (PKD) for the automated content analysis of knowledge codified in project documents. The second involves the integration of a sentiment analysis where concrete evaluative comments (positive/negative) are identified in connection with the codified project knowledge. This addresses the emerging research issue of *project sentiment analysis* (see Guzman 2013; Guzman & Bruegge 2013; Prieto 2013).

In summary, the purpose of this research study is: *to develop a Project Knowledge Dictionary (PKD) to analyze the content and associated sentiments in textually codified project knowledge.*

This study follows the design-science research approach (see Hevner et al. 2004) where the PKD is a concrete artifact designed to address an evident problem in project management. The efficacy of the PKD developed here will be demonstrated and evaluated by means of a practical analysis. In this demonstration, the PKD will be used for three mutually complementary analyses: (1) content analysis of the project knowledge contained in the project documents examined; (2) sentiment analysis of the "emotional state" in the respective project descriptions; (3) classification of the project documents by their sentiments.

This paper is structured as follows: Section 2 will first provide an introduction into the relevant foundations for the research, including the nature of codified project knowledge (2.1), the information overload (2.2), as well as the fundamentals of dictionary-based text analysis (2.3), feature-based sentiment analysis (2.4), and document classification (2.5) Section 3 then presents the research approach and the dictionary developed. Section 4 presents the subsequent analyses and their results. Section 5 contains a critical discussion of the study. Section 6 provides a brief summary and a look ahead at future research.

2   RESEARCH BACKGROUND

2.1   The Nature of Codified Knowledge in Project Environments

In the context of knowledge management, the nature of the codification strategy is described as follows: "Knowledge is carefully codified and stored in databases, where it can be accessed and used easily by anyone in the company" (Hansen et al. 1999, p 1). Consistent with this strategy, a large volume of

documents is typically created in the course of a project (e.g., project appraisals, performance reports, or post-project reviews) which codify the collected project knowledge and make it publicly available within a project-based organization (Barclay & Osei-Bryson 2010; Boh 2007; Disterer 2002). This project knowledge can be defined as "key project experiences which have a certain general business relevance for future projects" (Schindler & Eppler 2003, p. 220). In line with this definition, reusing prior key project experiences in project planning can protect future projects from repeating past mistakes or having to solve problems already solved (Koners & Goffin 2007). Knowledge codified in project documents thus represents a significant part of a company's intellectual capital (Almeida & Soares 2014). Multiple research studies already confirmed the advantages in utilizing this accumulated project knowledge in order to increase the efficiency and effectiveness of projects (see Frey et al. 2009; Hong et al. 2008; Kululanga & Kuotcha 2008). This is therefore an essential task of a project knowledge management (Frey et al. 2009).

Several researchers have already addressed the nature of the project knowledge (see, e.g., Chan & Rosemann 2001; Reich et al. 2008). Zhao and Zuo (2011) analyzed relevant studies in this area and provided a summary of five key project knowledge categories (see Table 1): (1) *Business Domain Knowledge*, (2) *Project Product Knowledge*, (3) *Project Engineering (Technical) Knowledge*, (4) *Organization Management Knowledge,* and (5) *Project Management Knowledge*. These five categories can be regarded as a valid classification of knowledge content typically codified in project documents. Therefore, these project knowledge categories will play a central role in the development of the Project Knowledge Dictionary (PKD) by providing the analytical framework.

	Business Domain Knowledge	**Project Product Knowledge**	**Project Engineering (Technical) Knowledge**	**Organization Management Knowledge**	**Project Management Knowledge**
Description	Refers to knowledge about a company´s business context (e.g., business strategy, industry specifics, value creation; culture, employees)	Refers to knowledge on how to select and design a specific project product (e.g., a specific business solution, information system, or software)	Refers to knowledge about the technical characteristics of a business solution development (e.g., website or software development)	Refers to knowledge on how to coordinate the various stakeholders involved in complex project environments (e.g., collaborative work in business networks)	Refers to knowledge on how to conceptualize, plan, coordinate, measure, and manage projects (e.g., project team composition)

Table 1.   Project knowledge categories (cf. Zhao & Zuo 2011, p. 268)

Analyzing the knowledge codified in project documents is specifically recommended by the usual project management guidelines (see PMI 2008; OGC 2009). The literature usually suggests doing this manually (see Schalken et al. 2006), although automated processes are beginning to gain ground (see, e.g., Carrillo et al. 2011; Choudhary et al. 2009). The following sections present three such automated processes: dictionary-based text analysis, feature-based sentiment analysis, and document classification.

2.2   Information Overload in Project Environments

The phenomenon "information overload" can be explained by two different terms (see Eppler & Mengis 2004): *information-processing capability* (i.e., the volume of information an individual can integrate into the decision-making process within a given time interval) and *information-processing requirements* (i.e., the volume of information an individual has to process). Whenever the information-processing requirements exceed the information-processing capabilities, the information overload problem occurs.

Information overload is a frequently observed phenomenon in project environments (see Caniëls & Bakens 2011; Haksever 2000; Haksever & Fisher 1996; Strait 2006). One reason for this is the variety and large volume of documents typically created in the course of a project (Prencipe & Tell 2001). As a result, the extensive documentation stocks available in a project-based organization (i.e., information-processing requirements) exceed the normal information-processing capabilities of project managers. The consequences are that project managers are not able to exploit the full knowledge potential available for the implementation of their projects (Haksever & Fisher 1996). O'Reilly (1980) has even stated that information overload can have a negative impact on project performance.

2.3 Dictionary-Based Text Analysis

Manual analysis of textual content quickly pushes the limits of feasibility when dealing with larger document collections (O'Flaherty & Whalley 2004). Dictionary-based text analyses are one solution for handling larger document collections because they allow a highly automated analysis of textual content (Krippendorff 2013). The dictionary approach can be described as a "bag-of-words" model (Fteimi & Basten 2015), where thematic categories are defined and operationalized using meaningful terminology (i.e. keywords). It is thus a type of conceptual analysis (Indulska et al. 2012), where an a priori analytical construct, i.e. a controlled thesaurus, is predefined and then analyzed on the basis of the underlying documents. Frequency analyses of keywords found in the documents can, when consolidated, then deliver quantitative data about the volume, distribution and ultimately the centrality of the thematic categories investigated (Indulska et al. 2012; Weber 1990).

As with any approach, the dictionary approach has its advantages and disadvantages. A key advantage is the efficiency in analyzing even extensive textual document collections (Beatty & Thomas 2007). Another advantage is the absolute reliability when replicating the analyses (O'Flaherty & Whalley 2004). Furthermore, the quantitative, transparent, and potentially objective results can be seen as an advantage (Boritz et al. 2013). The disadvantages of the dictionary approach lie primarily in the significance of having valid categories and keywords. Thematic categories that are incorrectly defined and keywords that are inappropriately selected can distort the quantitative results (Beatty & Thomas 2007). Another limitation lies in the premise that the simple frequency of keywords (i.e. also their categories) is a direct indicator of relevance (Weber 1990). The actual significance of a keyword within the context of the document contents is largely ignored.

In different disciplines, dictionaries have already been developed for different purposes. Beatty and Thomas (2007), for example, describe the use of dictionaries to analyze intellectual capital disclosures in corporate publications. Fteimi and Basten (2015) developed a dictionary custom-tailored for the domain of knowledge management in order to examine the contents of relevant research publications (abstracts). Vasalou et al. (2011) developed a privacy dictionary for the automated content analysis of privacy aspects in personal communications. Dictionaries for sentiment analysis (see also Section 2.4), which include so-called "sentiment words" (i.e., words that convey positive, neutral, or negative meanings), are another approach frequently used to analyze sentiments, for example, in political texts (see Young & Soroka 2012), financial news (see Loughran & McDonald 2011), or customer reviews (see Qadir, 2009).

A dictionary tailored to the needs of project management has not yet been developed. This study aims to fill this gap by providing a Project Knowledge Dictionary (PKD) for the automated assessment of knowledge content in project documents (see Section 3.3).

2.4 Feature-Based Sentiment Analysis

Subjectivity plays an important role in documents because it normally conveys opinions and emotions (Liu 2010). The analysis of textually codified subjectivity is generally called sentiment analysis: "Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes" (Liu 2012, p. 7). Sentiment analyses are already widespread in several disciplines, being used, for example, by financial management to analyze finance-related news (see Loughran & McDonald 2011) or by customer management to explore customer reviews (see Hu & Liu 2004; Qadir 2009). Sentiment analyses are also becoming increasingly important in project management, too (see Guzman 2013; Guzman & Bruegge 2013; Prieto 2013). The so-called *project sentiment analysis* aims at detecting emotions in textual project communication. The main relevance of such analyses lies in the possibility to automatically identify negative project experiences and, therefore, potential failures and mistakes in the making. Guzman and Bruegge (2013), for example, emphasize the need for such an "emotional awareness" in project management. They suggest using sentiment analysis techniques to analyze project documents to identify the "emotional state" of a project. Prieto (2013) suggests a project sentiment analysis to generate specific performance indicators based on explorations of the emotional states of projects in a diversified project portfolio. By this means, project sentiment analysis can contribute to overcoming the information overload problem in project environments.

This study takes up this fledgling research in project sentiment analysis, with a specific focus on feature-based sentiment analysis, which is based on the investigation of certain objects (such as products) and their features (such as aspects of product quality). The feature-based sentiment analysis takes place mainly at the sentence level, i.e. the sentences in a document are searched for the presence of specific opinions (negative or positive) about an object and its features (Liu 2010). A short example: In the sentence "This notebook's processor is very slow," the feature (<processor>) of an object (<notebook>) is described negatively (<slow>). This sentence can therefore be described as an opinion sentence: the sentence includes one or more features of the object described with one or more opinion words (positive or negative). The analysis of such opinion sentences usually involves three steps (Hu & Liu 2004): (1) identifying the object features in the sentences in one or more documents; (2) identifying opinion words that appear in conjunction with the feature; (3) generating a structured summary of the overall opinion (so-called feature-based summaries of opinions). In the last step where the results are presented, generating a summative frequency analysis is a typical procedure. The simplest form of such a summary could, for example, look like this (see Liu 2010):

OBJECT: *Notebook XYZ*

       FEATURE: *Processor*
            POSITIVE: 241   *< individual opinion sentences >*
            NEGATIVE: 81   *< This notebook's processor is very slow >*
       FEATURE: *Battery*
            POSITIVE: 112   *< individual opinion sentences >*
            NEGATIVE: 321   *< individual opinion sentences >*
       FEATURE: ...

Figure 1.   Feature-based summery of opinions (exemplary)

This paper will take up the concept of a feature-based sentiment analysis and transfer it to the project management context. A project could in this case be regarded as an object and the project knowledge categories could be considered the project features found in the project documents. Consequently, an analysis and summary of the opinions (positive/negative) codified in connection with the project knowledge could provide valuable insight into the emotional situation in projects.

2.5   Document Classification

Automated document classification is an already well-established technique in the context of text analysis (see Feldman & Sanger 2007; Weiss et al. 2010). The task is, in simple terms, to automatically classify a given set of documents according to a pre-defined set of categories (Feldman & Sanger 2007). There are several ways to perform such a document classification (i.e., rule-based classification and super- or unsupervised classification). Rule-based document classification is based on manually pre-defined content categories and corresponding coding rules (e.g., specific keyword searches), and is therefore inspired by dictionary-based text analysis. Supervised classification techniques are based on manually controlled training processes in which a sample of training documents is analyzed in order to automatically formulate classification rules. Unsupervised classification techniques, on the contrary, group documents fully automatically based on their contents.

Document classification techniques have already been implemented in the project environment (see, e.g., Al Qady & Kandil 2013, 2014; Caldas et al. 2002; Ur-Rahman & Harding 2012). Most of the suggested solutions are based on supervised and unsupervised classification techniques. This research, on the other hand, uses a rule-based document classification approach for which categories and coding rules tailored to the specifics of project management were developed (see the following Section 3).

3   DEVELOPMENT OF A PROJECT KNOWLEDGE DICTIONARY

3.1   Research Framework

This study focuses on the development of a *Project Knowledge Dictionary* (PKD) as one way to solve the information overload problem that occurs in the project environment. This study thus fits within the framework of the design-science research paradigm (see Hevner et al. 2004; Peffers et al. 2008) which, at its core, has as its goal the development of workable solutions (called "artifacts") for practical problems.

This paper's research framework is summarized in Figure 1. The research process consists of three steps (shown as input, work process, and output): In the first step, the PKD is developed (see Section 3.3). In the second step, the PKD is used for demonstrative content and sentiment analyses of practical project documents (see Section 3.4). The practicality of the PKD is then evaluated in the third step (see Section 3.5).
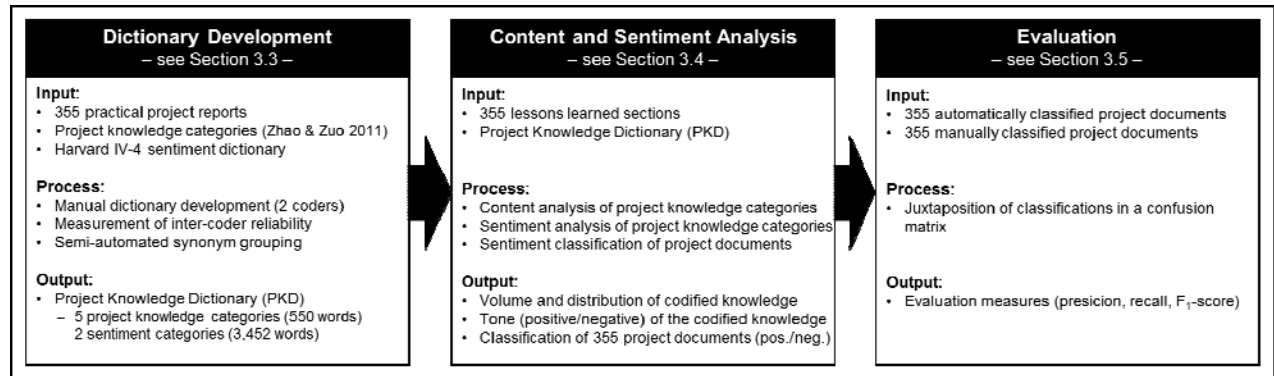

Dictionary Development – see Section 3.3 –	Content and Sentiment Analysis – see Section 3.4 –	Evaluation – see Section 3.5 –

Input:
• 355 practical project reports
• Project knowledge categories (Zhao & Zuo 2011)
• Harvard IV-4 sentiment dictionary
 	
Process:
• Manual dictionary development (2 coders)
• Measurement of inter-coder reliability
• Semi-automated synonym grouping
 	
Output:
• Project Knowledge Dictionary (PKD)
  – 5 project knowledge categories (550 words)
    2 sentiment categories (3,452 words)
	Input:
	• 355 lessons learned sections
	• Project Knowledge Dictionary (PKD)
	
	Process:
	• Content analysis of project knowledge categories
	• Sentiment analysis of project knowledge categories
	• Sentiment classification of project documents
	
	Output:
	• Volume and distribution of codified knowledge
	• Tone (positive/negative) of the codified knowledge
	• Classification of 355 project documents (pos./neg.)
		Input:
		• 355 automatically classified project documents
		• 355 manually classified project documents
		
		Process:
		• Juxtaposition of classifications in a confusion matrix
		
		Output:
		• Evaluation measures (presicion, recall, F_1-score)

Figure 2.   Research framework

3.2   Data Collection

A domain-specific database containing the data potentially to be analyzed is needed to build a domain-specific dictionary (Krippendorff 2013). Given this, collections of practical project documents are particularly well-suited as such, since they are a typical medium for sharing the project knowledge to be analyzed (Schindler & Eppler 2003). Consequently, a total of 355 reports from real-world IT projects were assembled. These documents originate from collections of two research initiatives (*Sectoral e-Business Watch*[1] and *eXperience Online*[2]). The initiatives collected the project reports with a largely uniform format where the objectives of each project were explained, together with the project planning and implementation process, the results, and the lessons learned from the project. The authenticity, quality, and credibility of the reports were checked by the initiatives. The largely textual content of these reports (> 2.000 pages) was then subjected to basic procedures of natural language processing (i.e., word extraction, stop word removal and lemmatization; see Manning & Schuetze 1999), resulting in a textual database (word list) with 8,865 unique words. This database was deemed to represent the typical terminology of IT projects.

The database was further refined and reduced while performing the subsequent content and sentiment analyses by only selecting the concluding lessons learned sections of the reports. This was done because the accumulated project experiences (positive and negative) were particularly interesting. Such concluding project experiences transport the relevant project knowledge to be analyzed. Other contents, such as corporate presentations and technical specifics, are rather irrelevant. This left a total of 355 lessons learned sections (each with 0.25 to 0.75 pages of text) available for subsequent analysis.

3.3   Dictionary Development

The *Project Knowledge Dictionary* (PKD) is designed to identify project-oriented knowledge contained in project documents and the sentiments expressed in conjunction with their codification. The dictionary being developed therefore also consists of two parts: a first part reflecting typical project knowledge categories (see Zhao & Zuo 2011) and a second part reflecting sentiment categories (positive/negative). The dictionary development will be presented in more detail below.

The first part of the PKD was designed to provide a valid and complete reflection of the project knowledge codified in project documents. To ensure the validity, defining appropriate thematic categories, i.e. a valid analytical construct, is essential (Krippendorff 2013). The five project knowledge categories devised by Zhao and Zuo (2011) were therefore used to develop the PKD: (1) *Business Domain Knowledge*, (2) *Project Product Knowledge*, (3) *Project Engineering (Technical) Knowledge*, (4) *Organization*

Management Knowledge, and (5) *Project Management Knowledge*. These five categories can be regarded as a valid classification of knowledge content ordinarily codified in project documents. Next, the categories had to be associated with meaningful keywords that would correctly reflect the categories as comprehensively as possible. This study took a largely manual approach to this task by analyzing the terminology typical of IT project management. The database described in section 3.2 (8,865 unique words) was analyzed by two independent coders (the author and a research assistant), identifying significant keywords and assigning them thematically to appropriate categories. The keyword identification process focused on identifying significant nouns (e.g., *collaboration, industry, PHP*) and compound concepts, i.e. meaningful combinations of nouns (e.g., *project management*), nouns and adjectives (e.g., *organizational coordination*), and nouns and verbs (e.g., *planning tasks*). Such compound concepts allow a generally more accurate and more meaningful delimitation of relevant themes (Fteimi & Basten 2015). To assist this definition process, a keyword-in-context analysis (KWIC) was used, which supported the interpretation of word meanings in individual cases. When defining keywords, wildcard symbols ("*") were also used to address the issue of varying word forms (e.g., *analyz** for *analyze, analyzed,* or *analyzing*). Once the keyword identification process was complete, the word lists generated by the two coders were evaluated for agreement (in the sense of an inter-coder reliability; 76%), with any discrepancies discussed and corrected by consensus. In the next step, these initial word lists were enhanced by a synonym grouping process (see Liu 2012), where this group of reviewed seed words was expanded with synonyms and antonyms using a bootstrapping approach. This development process resulted in a PKD (see Table 2) with five thematic project knowledge categories containing a total of 550 keywords. The PKD was then presented to two experienced IT project managers for a final evaluation.

The second part of the PKD was supposed to reflect typical sentiment classes (positive and negative). Over the years, several solutions have been proposed for implementing such dictionary-based sentiment analyses, including the Regressive Imagery Dictionary (RID) (Martindale 1975), the Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al. 2001), and the Lexicoder Sentiment Dictionary (LSD) (Young & Soroka 2011). An established and already frequently used dictionary is Harvard's General Inquirer, i.e. specifically the Harvard IV-4 sentiment dictionary (see Stone et al. 1968). For example, this dictionary has already been used repeatedly to analyze financial news (see Li et al. 2014). Nevertheless, Loughran and McDonald (2011) expressly recommend that one should pay particular attention to domain-dependent specifics. What may appear to be negative words (e.g., *cost, disposal, duty, rigor,* or *scheme*) can, in specific disciplines such as project management, have a different meaning. Adjustments to the specific linguistic usage and terminology of project management were therefore necessary. For this purpose, the original collections of positive (1,915) and negative words (2,291) were analyzed with 754 words ultimately being removed or corrected. The result of this correction process was a sentiment dictionary with 1,520 positive and 1,932 negative terms tailored to the specifics of project management (see Table 2).

Project Knowledge Dictionary	
Project Knowledge Categories	**Keywords**
Business Domain Knowledge (118 words)	business model; business domain, CEO; competition; corporate strategy; firm size; market share; organizational culture; SME; target group; value chain; …
Project Product Knowledge (88 words)	business application; business solution; CRM; customer satisfaction; patent; project product; prototype, software solution; user requirement; web shop; …
Project Engineering (Technical) Knowledge (112 words)	ASCII; business rule; customization; developer; EDI; engineering; java, PHP; programming, software development; technical development; VBA; XML; …
Organization Management Knowledge (85 words)	alliance; business partner; collaboration; contracting; mediator; lobby; organization management; partner involved; strategic network; third party; …
Project Management Knowledge (147 words)	business project; gantt; launch project; plan project; project complexity; project management; project team; project task; scrum; waterfall; …
Sentiment Categories	**Keywords**
Positive Sentiment (1,520 words)	accomplish; best; compatible, dedicate; efficient; fair; good; improve; loyal; mature; opportunity; prompt, quality; reliable; satisfy; true; valid; well; …
Negative Sentiment (1,932 words)	accuse; blame; collapse, constraint; deceive; delay, error; fail; harm; lack; mistake; obscure; poor; redundant; regret; shortcoming; unable; warning; …

Table 2.   Project Knowledge Dictionary (excerpt)

3.4 Content and Sentiment Analysis

The PKD developed was then applicable for a variety of analytical purposes. Three different analyses were conducted: (1) *content analysis of project knowledge categories*; (2) *sentiment analysis of project knowledge categories*; and (3) *sentiment classification of project documents*. The respective analyses are presented in more detail below. The 355 lessons learned sections in the document collection already described formed the database for these analyses. The analyses were performed using *Provalis Research's* QDA Miner and WordStat 7.1 as well as SPSS Statistics.

Content analysis of project knowledge categories: The purpose of the first analysis was to discover the project knowledge codified in the project documents, or more precisely, its volume and distribution across the document collection. For this purpose, a quantitative content analysis was conducted based on the keywords in the PKD. Since the volume and distribution of the existing project knowledge was primarily of interest and not the ratio of project-specific keywords to the sum of all the words contained in the document, this analysis was based on simple frequency analyses with descriptive statistics.

Sentiment analysis of project knowledge categories: The second analysis involved a feature-based sentiment analysis where the project-specific features (*<project knowledge categories>*) of a given project (*<project no.>*) were evaluated in terms of associated sentiments (*<positive/negative>*). This analysis was based on three steps (see Hu & Liu 2004). In the first step, feature-related keywords were identified in the sentences of the project documents, which was already done with the previous content analysis. In the second step, "opinion expressions" were identified, i.e. sentences in which a feature-related keyword (according to the five project knowledge categories) appears together with an opinion word (according to the sentiment dictionary). For this purpose, this study followed the approach to identify project-related keywords closely associated with positive and negative opinion words, i.e. no more than *n* words apart from each other. Specifically, this means that whenever both word types occur in the same sentence, and are not more than 5 words apart from each other, such a keyword combination will be counted as one codified opinion expression (i.e. one hit). An appropriate algorithm was implemented in the text mining tool WordStat 7.1. In the third step, the identified opinion expressions were aggregated. One way to present the results is to summarize the simple absolute frequencies of identified opinions. It is common, however, to weight the results because this allows a more representative measurement of the overall sentiment (positive or negative) of a document (Pröllochs et al. 2015; Young & Soroka 2012). Both approaches were used in this study. First, the absolute number of positive and negative opinion expressions (hits) was calculated for each project knowledge category to assess the absolute extent of knowledge expressed accordingly. Second, the proportion of opinions was weighted by the total number of sentences in the document. This weighting thus also considers the length of a document. This allows a more representative assessment of the overall tone of the documents, which plays a role in their later classification.

Sentiment classification of project documents: In the third analysis, a document-level sentiment classification was conducted which, in general, served the following purpose: "Given a set of opinionated documents, it determines whether each document expresses a positive or negative opinion (or sentiment) on an object" (Liu 2010, p. 10). There is a wide array of approaches to this task (see Hu & Liu 2004; Liu 2012). In this study, project documents were classified by measuring their predominant tone (positive or negative) with reference to each feature (i.e. project knowledge category) and with reference to each document in summary. The result was a classification scheme that helps a multi-project management to identify the "emotional state" of many projects at once.

3.5 Evaluation

Simply presenting the results of a dictionary-based text analysis is not sufficient. An evaluation was necessary to demonstrate the actual validity of the PKD developed and its associated analyses. In accordance with the design-science principles (see Hevner et al. 2004), this evaluation will be implemented in the framework of a simulation which should represent a case of the real business environment (Zelkowitz & Wallace 1998). To do this, the automated results generated by the PKD were compared with those generated by a human analyst. This is a common approach to evaluating results generated by automated methods (see, e.g., Qadir 2009; Young & Soroka 2012; Ur-Rahman & Harding 2012). If the results of the PKD-based analyses are consistent with the human analyses of the same document content,

this can confirm the validity of the automated approach. In this evaluation, it was of particular interest, whether the automated dictionary approach was actually successful in accurately and fully identifying those project documents with a negative tone, since these are of particular interest to multi-project management.

The sentiment classifications of project documents generated by the PKD were compared to those generated by a human analyst. A common procedure is juxtaposing the results in a confusion matrix which summarizes the proportion of matching classifications in a structured manner (see Walter & Back 2013). Then typical evaluation indicators of the information retrieval discipline could be calculated (see Baeza-Yates & Ribeiro-Neto 2010). In this evaluation, these indicators specifically reflect the degree to which the documents with a negative tone were correctly classified as such:

- *Precision* = true negative classifications / (true negative classifications + false negative classifications)
- *Recall* = true negative classifications / (true negative classification + false positive classifications)
- *F_1-Score* = 2 * (Precision * Recall) / (Precision + Recall)

These metrics, which are customary for evaluations in the design-science research framework (Hevner et al. 2004), reflect different aspects of the efficacy of the classifications. Precision is the proportion of documents correctly classified by the PKD as negative compared to the total number of documents classified as negative (= indicates the degree of wrong classifications). Recall, on the other hand, is the proportion of documents correctly classified by the PKD as negative compared to the total number of existing negative documents (= indicates the degree of completeness). The F_1-score is a harmonic mean of precision and recall.

4   RESULTS

4.1   Content Analysis of Project Knowledge Categories

The content analysis was performed to discover the volume and distribution of various project-specific knowledge content in the document collection ($N = 355$). Table 3 summarizes the results of the frequency analyses. Various findings can be taken from the results. First, the total volume of the project knowledge (4,154 identified keywords) identified had different distributions among the various knowledge categories. The largest amount appears to be discussions of the business domain (1,252 hits; 30%), while the least amount were discussions of project management issues (511 hits; 12%). A look at the average frequency of keyword occurrence in the documents confirms this impression. Here attention must be paid to the partly high standard deviations (S.D.). Second, the codification of project knowledge appears to be common and was present in all documents examined (100% of reports). However, the concrete discussion of the specific project knowledge categories varied remarkably. For example, descriptions of the business domain or the project product can be found in the great majority of the document (more than 79% of the reports). Descriptions of the other three knowledge categories seem to be less common (57-59%).

	Business Domain Knowledge	Project Product Knowledge	Project Engineering Knowledge	Organization Management Knowledge	Project Management Knowledge	Total
Total (%-share)	1,252 (30%)	1,187 (29%)	634 (15%)	570 (14%)	511 (12%)	4,154 (100%)
% of reports	84%	79%	59%	58%	57%	100%
Mean	4.22	4.24	3.02	2.78	2.54	11.70
S.D.	3.60	4.08	2.71	2.50	2.10	8.25
Median	3	3	2	2	2	10
25th perc.	2	1	1	1	1	6
75th perc.	6	6	4	3	3	16

Table 3.   Descriptive statistics of project knowledge keywords per category

4.2   Sentiment Analysis of Project Knowledge Categories

The sentiment analysis was performed to determine the tone (positive/negative) of the codified project experiences. To this end, expressed opinions were identified in the documents, i.e. sentences that described project knowledge keywords in close conjunction with sentiment words. Figure 3 summarizes the opinion

expressions identified (hits, incl. examples) in each project knowledge category. This analysis already allows the interpretation of some initial findings. It becomes clear that the discussions of the project knowledge categories have a mainly positive tone. For example, the discussion of the project product comprises 729 positive hits (80.4%) and only 178 negative hits (19.6%). The discussions of the project management, as another example, are also mostly associated with positive opinions (201 positive hits vs. 139 negative hits), though with less difference between the two extremes (66.1% and 33.9%).

OBJECT: *project #1-355 (total)*

        FEATURE: *Business Domain Knowledge*
                POSITIVE: 438 hits   Example: *<Such standards can benefit the project, particularly for SMEs>*
                NEGATIVE: 213 hits   Example: *<This could be seen as a disadvantage against larger competitors>*
        FEATURE: *Project Product Knowledge*
                POSITIVE: 729 hits   Example: *<These innovations should contribute to higher customer satisfaction>*
                NEGATIVE: 178 hits   Example: *<The management has a very bad perception of the business solution>*
        FEATURE: *Project Engineering (Technical) Knowledge*
                POSITIVE: 285 hits   Example: *<This degree of freedom should boost the technicians' performance>*
                NEGATIVE: 100 hits   Example: *<The expenditure required for customization was underestimated>*
        FEATURE: *Organization Management Knowledge*
                POSITIVE: 272 hits   Example: *<Business partners help firms to overcome lack of technology skills>*
                NEGATIVE: 124 hits   Example: *<Lack of clarity regarding trading partner identification>*
        FEATURE: *Project Management Knowledge*
                POSITIVE: 201 hits   Example: *<Ambitious project goals were achieved through integrating all sites>*
                NEGATIVE: 139 hits   Example: *<This caused interruptions during project execution>*

Figure 3.   *Opinion expressions per project knowledge category*

Looking only at the absolute frequency of opinion expressions found was not enough to assess the general sentiment of a project document. The hits were weighed by the total number of sentences in a document. The resulting measure reflects the general tone (positive vs. negative) of a sentence codified in a project document more precisely. The average values of these weighted calculations are summarized in Table 4 for each category and in total. These results also allow various conclusions. The overall tone in the documents appears to be mostly positive, but, on closer inspection, it is also becomes clear that the discussion of individual project characteristics is more negative in comparison. The descriptions of the project management appear to take a relatively more negative tone (0.0378 positive vs. 0.0300 negative tone) than descriptions of the other project features (such as the project product with 0.1208 positive and 0.0309 negative). The corresponding results can therefore provide concrete evidence of the "emotional state" in a project portfolio. For even more concrete statements, these results were then broken down at the individual project level in the next analysis.

		Business Domain Knowledge	Project Product Knowledge	Project Engineering Knowledge	Organization Management Knowledge	Project Management Knowledge	Total
Positive	Mean	0.0844	0.1208	0.0541	0.0479	0.0378	0.3450
	S.D.	0.1122	0.1366	0.0863	0.0765	0.0682	0.1876
Negative	Mean	0.0380	0.0309	0.0231	0.0234	0.0300	0.1455
	S.D.	0.0649	0.0616	0.0672	0.0538	0.0655	0.1426

| Mean Tone | + | 0.0844 / 0.0380 | 0.1208 / 0.0309 | 0.0541 / 0.0231 | 0.0479 / 0.0234 | 0.0378 / 0.0300 | 0.3450 / 0.1426 |
| | − | | | | | | |

Table 4.   *Sentiment analysis of project knowledge categories*

4.3   Sentiment Classification of Project Documents

The document-based sentiment classification was performed to specify the findings of the general sentiment analysis at a more detailed project level. To this end, the proportion of positive or negative opinion expressions (i.e. tone) in the individual project documents ($N = 355$) were identified. The

individual project documents were then classified according to their predominant tone, i.e. "emotional state" (positive/negative). This was done both at the project knowledge category level and overall. Figure 4 presents the results in an aggregated form (overall summary). The line separates the project documents with a predominantly positive tone (above the line) from the project with a predominantly negative tone (below). It is initially evident that the great majority of the projects have a predominantly positive tone in their project descriptions ($N = 280$; 78.9%). Nevertheless, several projects with a predominantly negative tone in their codified experiences can also be identified ($N = 75$; 21.1%). The identification of such projects is a relevant contribution to multi-project management. Such projects could be examined more closely and eventual grievances headed off early on.
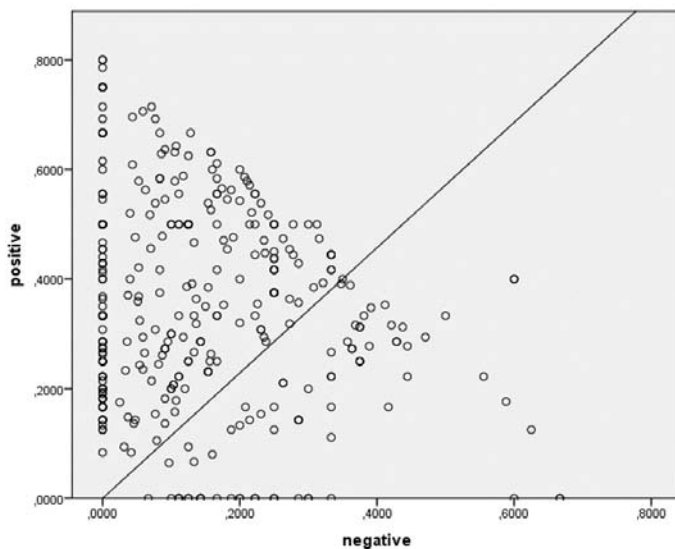

Figure 4.   Document-based sentiment classification of project documents (overall summary)

4.4   Evaluation

The goal of the evaluation was to examine the efficacy of the automated dictionary approach. This evaluation was primarily interested in whether the automated dictionary approach was able to identify the true negative project descriptions. The project document classifications presented above (documents with predominantly negative vs. those with predominantly positive tones) were compared with the classifications manually generated by a human analyst. At the beginning of the study, the analyst read all 355 documents and analyzed for each project knowledge category whether the discussions they contained were primarily positive or negative in tone. In this context, neutral or missing expressions were also considered as positive. The results of the automated and manual approaches were then juxtaposed in a confusion matrix to determine their level of correspondence. Table 5 shows an exemplary evaluation of the classifications for the category of project management knowledge. Here, for example, 56 of the project documents classified as negative by the automated approach were also classified by the analyst as negative. In comparison, 19 of the documents were classified as negative by the automated approach, while the analyst classified these documents as positive. Furthermore, 36 documents were classified as negative by the analyst, while the automated approach classified these documents as positive.

		Human Analyst		
		Positive	Negative	Totals
Dictionary	Positive	244	36	280
Approach	Negative	19	56	75
	Totals	263	92	

Table 5.   Confusion matrix of sentiment classifications (example: project management knowledge)

The confusion matrices were then used to calculate the indicators precision, recall and F_1-score. These indicators are used in such evaluations because they provide individual statements regarding the effectiveness of the automated results (see Table 6). There were comparatively high precision scores across the categories, while recall had generally rather lower scores. This means that the project documents

classified as negative in the automated process were largely also classified by the analyst as negative, although the analyst had usually classified even more documents as negative. In other words: the approach is comparatively precise in the detection of true negative project documents but, however, not fully comprehensive in the detection of all existing negative documents. The F_1-scores varied between 0.5430 and 0.6809 across all categories, with an average macro F_1-score of 0.6240. Here it became clear that the F_1-scores for business domain knowledge and organization management knowledge were comparatively low. This could be due to more complex discussions and thus more difficult contents to be textually analyzed. Therefore, in comparison to the other categories, this indicates some potential need for optimization.

	Business Domain Knowledge	Project Product Knowledge	Project Engineering Knowledge	Organization Management Knowledge	Project Management Knowledge	Average
Precision	66.1%	74.4%	68.4%	59.4%	74.7%	68.6%
Recall	46.1%	62.7%	65.0%	52.8%	60.9%	57.5%
F_1-Score	0.5430	0.6809	0.6667	0.5588	0.6707	0.6240

Table 6.   Precision, recall and F_1-scores of classifications per project knowledge category

5   DISCUSSION

5.1   Contributions

In this study, a Project Knowledge Dictionary (PKD) was developed, demonstrated, and evaluated, that forms the basis for three mutually complementary analyses: dictionary-based text analysis, sentiment analysis, and document classification. In sum, the developed solution makes therefore the following three contributions: (1) determining the volume and distribution of codified descriptions of five project knowledge categories in project documents; (2) determining the general sentiment (positive/negative) in conjunction with the description of the project knowledge; (3) classifying project documents by their sentiment. The implications and limitation of this study and the PKD developed are discussed below.

5.2   Implications for Practice

The PKD proposed offers two potential applications in practical project management environments. First, the approach can be used to track the emotional state of on-going projects within an organization. By discovering which project features (project knowledge categories) are described with certain emotions (positive or negative), multi-project management can identify relevant references to potential failures and mistakes in the making. Second, the PKD provides a practical approach to assessing and organizing the often extensive and unstructured collections of historical project documents. In practice, this means being able to sort, classify, and organize the rich content of document collections in preparation for specific analytical tasks. On the one hand, documents can be organized according to specific knowledge contents; on the other hand, document can be organized according to positive or negative experiences. In this way, this approach also helps overcome the common problem of information overload in project-based organizations.

5.3   Implications for Research

In the framework of the design-science research paradigm, evaluation and (re-)design of artifacts represent an iterative cycle. The PKD therefore provides several avenues for further research. First, there is some need for further evaluation. Although the demonstrated solution was evaluated in the framework of a simulation with practice-oriented data, the evaluation should be extended and transferred to a more complex real-life business environment in the next step. Conducting case studies with project managers and varying data could be an approach here. Second, the dictionary offers potential for further evaluation, correction, and expansion. Although the comparably high precision scores confirm a strong capability in detecting true negative sentiments, the recall scores indicate that the comprehensiveness of such detections could be improved. This applies in particular to discussions of the business domain and organization management. Further research could therefore focus on both to the optimization of the keywords collections used for project knowledge categories and also of the sentiment words used in the sentiment

dictionary. Third, the classification technique demonstrated here can be expanded in further studies. In particular, using computerized text mining processes for document classification (e.g., Naïve Bayes classification) offers great potential for the more accurate and more efficient (automated) classification of project documents according to the project knowledge they contain and the sentiment attached to that knowledge.

5.4   Limitations

This study has some potential limitations. On the one hand, limitations with respect to the dictionary approach must be mentioned. The quantitative content and sentiment analyses were based on a word-frequency count. This technique has proven to be an efficient means of processing large textual data sources as well as a large number of keywords. The technique's advantage, however, also contains its limitations (see also Beattie & Thomson 2007). First, the quantitative dictionary approach does not take the conceptual contexts of the words or word combinations into account. This means, keywords could have a completely different context from the one envisaged by the dictionary. This may have introduced a certain bias into the findings. Second, the selection of the keywords themselves has considerable influence on the results of this analysis. Other keywords or other compositions could certainly generate different results. In order to overcome these concerns, the keyword identification was performed by two independent coders and was tested by means of a reliability check. Furthermore, the dictionary was presented to two experienced IT project managers.

On the other hand, limitations with regard to the database used must be addressed. First, a textual database derived from 355 practical project reports was used for dictionary development. This database should represent the typical terminology used in project documents. However, this collection of project-oriented terms may not be fully complete or representative due to a limited thematic scope. This means in consequence, that special project types (e.g., data migration projects) are not fully covered by this dictionary. Second, the lessons learned sections analyzed revealed a largely positive sentiment, which may be due to the nature of the underlying document collection. Other practical project reports may reveal different results.

Limitations with regard to the sentiment analysis must also be mentioned. This study is based on a polarity approach which differs between positive and negative words (i.e., their semantic orientation). However, natural language is generally complex. The contextual sentiment of a sentence in which a word appears may differ from the word´s assigned polarity. This is a common accuracy problem discussed in the literature (see, e.g., Wilson et al. 2009).

6   CONCLUSION AND OUTLOOK

In this study, a Project Knowledge Dictionary (PKD) was developed to present the project management discipline a way to perform content and sentiment analysis of project-based knowledge content and classify project documents. The approach thus makes a contribution to overcoming the information overload in the project environment. Furthermore, the approach contributes to the emerging research issue of project sentiment analysis. Although an initial evaluation has indicated the general efficacy of the proposed approach, there remain a variety of avenues for its further development and optimization. Future studies will focus on practically testing the PKD and the proposed document-based sentiment classification in order to enable their further development.

References

Al Qady, M. and Kandil, A. (2014). Automatic clustering of construction project documents based on textual similarity. Automation in Construction, 42, 36-49.

Al Qady, M. and Kandil, A. (2013). Automatic Classification of Project Documents on the Basis of Text Content. Journal of Computing in Civil Engineering.

Almeida, M. V. and Soares, A. L. (2014). Knowledge sharing in project-based organizations: Overcoming the informational limbo. International Journal of Information Management, 34 (6), 770-779.

Baeza-Yates R. and Ribeiro-Neto, B. (2010). Modern Information Retrieval: The Concepts and Technology behind Search. Addison-Wesley Professional.

Barclay, C. and Osei-Bryson, K.-M. (2010). An Exploration of Knowledge Management Practices in IT Projects: A Case Study Approach. In Proceedings of the Americas Conference On Information Systems (AMCIS), Paper 452.

Beattie, V. and Thomson, S.J. (2007). Lifting the lid on the use of content analysis to investigate intellectual capital disclosures. Accounting Forum, 31 (2), 129-163.

Boh, W. F. (2007). Mechanisms for sharing knowledge in project-based organizations. Information and Organization, 17 (1), 27-58.

Boritz, J. E., Hayes, L., and Lim, J.-H. (2013). A Content Analysis of Auditors' Reports on IT Internal Control Weaknesses: The Comparative Advantages of an Automated Approach to Control Weakness Identification. International Journal of Accounting Information Systems 14 (2), 138-163.

Caldas, C. H., Soibelman, L., and Han, J. (2002). Automated classification of construction project documents. Journal of Computing in Civil Engineering, 16 (4), 234-243.

Caniëls, M. C. and Bakens, R. J. (2012). The effects of Project Management Information Systems on decision making in a multi project environment. International Journal of Project Management, 30 (2), 162-175.

Carrillo, P., Harding, J., and Choudhary, A. (2011). Knowledge discovery from post-project reviews. Construction Management and Economics, 29 (7), 713-723.

Chan, R. and Rosemann, M.R. (2001). Managing knowledge in enterprise systems. Journal of Systems and Information Technology, 5 (2), 37-54.

Choudhary, A. K., Oluikpe, P. I., Harding, J. A., and Carrillo, P. M. (2009). The needs and benefits of Text Mining applications on Post-Project Reviews. Computers in Industry, 60 (9), 728-740.

Disterer, G. (2002). Management of project knowledge and experiences. Journal of Knowledge Management, 6 (5), 512-520.

Eppler, M. J. and Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. The information society, 20(5), 325-344.

Feldman, R. and Sanger, J. (2007). The text mining handbook: advanced approaches in analyzing unstructured data. University Press, Cambridge.

Frey, P., Lindner, F., Muller, A., and Wald, A. (2009). Project Knowledge Management - Organizational Design and Success Factors - An Empirical Study in Germany. In Proceedings of the 42nd Hawaii International Conference on System Sciences (HICSS), 1-14.

Fteimi, N. and Basten, D. (2015). Impact of Dictionaries on Automated Content Analysis-The Use of Compound Concepts in Analysing Knowledge Management Research. In Proceedings of 23rd European Conference on Information Systems (ECIS), Münster, Germany.

Gamerschlag, R., Möller, K., and Verbeeten, F. (2011). Determinants of voluntary CSR disclosure: empirical evidence from Germany. Review of Managerial Science, 5 (2-3), 233-262.

Grobelnik, M. and Mladeni, D. (2005). Automated knowledge discovery in advanced knowledge management. Journal of Knowledge Management, 9 (5), 132-149.structure in theory and research. Management Science, 34 (5), 583-598.

Guzman, E. (2013). Visualizing emotions in software development projects. In First IEEE Working Conference on Software Visualization (VISSOFT), 1-4.

Guzman, E. and Bruegge, B. (2013). Towards emotional awareness in software development teams. In Proceedings of the 9th Joint Meeting on Foundations of Software Engineering, 671-674.

Haksever, A. M. and Fisher, N. (1996). A method of measuring information overload in construction project management. In Proceedings of the CIB, 21-24 October, Beijing, China.

Haksever, A. M. (2000). A model to predict the occurrence of information overload of project managers. In Proceedings of the CIB, 78.

Hansen, M.T., Nohria, N., and Tierney, T. (1999). What's your Strategy for Managing Knowledge?. Harvard Business Review, 77 (2), 106-116.

Hevner, A.R., March, S.T., Park, J., and Ram, S. (2004). Design science in information systems research. Management Information Systems Quarterly (MISQ), 28 (1), 75-105.

Hong, H. K., Kim, J. S., Kim, T., and Leem, B. H. (2008). The effect of knowledge on system integration project performance. Industrial Management & Data Systems, 108 (3), 385-404.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. 168-177.

Indulska, M., Hovorka, and Recker, J. (2012). Quantitative approaches to content analysis: identifying conceptual drift across publication outlets. European Journal of Information Systems, 21 (1), 49-69.

Kim, S. M. and Hovy, E. (2006). Extracting opinions, opinion holders, and topics expressed in online news media text. In Proceedings of the Workshop on Sentiment and Subjectivity in Text, 1-8.

King, W.R. (2009). Text analytics: boon to knowledge management?. Information Systems Management, 26 (1), 87.

Krippendorff, K. (2013). Content Analysis. An Introduction to its Methodology. 3rd. Edition. SAGE Publications, Thousand Oaks, CA.

Koners, U. and Goffin, K. (2007). Learning from Post-Project Reviews: A Cross-Case Analysis. Journal of Product Innovation Management, 24 (3), 242-258.

Kululanga, G. K. and Kuotcha, W. S. (2008). Measuring organisational learning through project reviews. Engineering, Construction and Architectural Management, 15 (6), 580-595.

Li, X., Xie, H., Chen, L., Wang, J., and Deng, X. (2014). News impact on stock price return via sentiment analysis. Knowledge-Based Systems, (69), 14-23.

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5 (1), 1-167.

Liu, B. (2010). Sentiment Analysis and Subjectivity. Handbook of natural language processing, 2, 627-666.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66 (1), 35-65.

Manning, C. D. and Schuetze, H. (1999). Foundations of statistical natural language processing. MIT press, Cambridge, MA.

Martindale, C. (1975). Romantic progression: The psychology of literary history. Washington, D.C.: Hemisphere.

Meier, C. (2013). Identifying Output Interactions Among Is Projects - A Text Mining Approach. In Proceedings of Twenty-First European Conference on Information Systems (ECIS), Utrecht, NL.

Minanovic, A., Gabelica, H., and Krstic, Z. (2014). Big data and sentiment analysis using KNIME: Online reviews vs. social media. In 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 1464-1468.

O'Flaherty, B. and Whalley, J. (2004). Qualitative Analysis Software Applied to IS Research - Developing a Coding Strategy. In: Proceedings of the European Conference on Information Systems. Turku.

OGC (2009). Managing and Directing Successful Projects with PRINCE2TM. 5th Edition. Office of Government Commerce. The Stationery Office Books, Norwich.

O'Reilly, C. A. (1980). Individuals and information overload in organizations: is more necessarily better?. Academy of management journal, 23(4), 684-696.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of management information systems, 24 (3), 45-77.

PMI (2008). PMBOK guide 4th Edition. Project Management Institute, Inc., Pennsylvania.

Pennebaker, J.W., Francis,M., and Booth, R. (2001). Linguistic Inquiry and Word Count: LIWC 2001. Mahwah, NJ: Erlbaum.

Prencipe, A. and Tell, F. (2001). Inter-project learning: processes and outcomes of knowledge codification in project-based firms. Research policy, 30 (9), 1373-1394.

Prieto, B. (2013). Utilization of Project Sentiment Analysis as a Project Performance Predictor. PM World Journal, 8 (11), 1.

Pröllochs, N., Feuerriegel, S., and Neumann, D. (2015). Generating Domain-Specific Dictionaries Using Bayesian Learning. In Proceedings of Twenty-Third European Conference on Information Systems (ECIS), Münster, Germany.

Qadir, A. (2009). Detecting opinion sentences specific to product features in customer reviews using typed dependency relations. In Proceedings of the Workshop on Events in Emerging Text Types, Association for Computational Linguistics, 38-43.

Reich, B. H., Gemino, A., and Sauer, C. (2008). Modeling the knowledge perspective of IT projects. Project Management Journal, 39 (S1), 4-14.

Schalken, J., Brinkkemper, S., and Van Vliet, H. (2006). A method to draw lessons from project postmortem databases. Software Process: Improvement and Practice, 11 (1), 35-46.

Schindler, M. and Eppler, M. J. (2003). Harvesting project knowledge: a review of project learning methods and success factors. International Journal of Project Management, 21 (3), 219-228.

Stone, P., Dunphy, D. C., Smith, M. S., and Ogilvie, D. M. (1968). The general inquirer: A computer approach to content analysis. Journal of Regional Science, 8 (1), 113-116.

Strait, C. (2006). Too much information. Project Management Institute (PMI). Retrieved December 7, 2015 from http://www.pmi.org/learning/point-information-overload-management-process-2769.

Ur-Rahman, N. and Harding, J. A. (2012). Textual data mining for industrial knowledge management and text classification: A business oriented approach. Expert Systems with Applications, 39 (5), 4729-4739.

Vasalou, A., Gill, A. J., Mazanderani, F., Papoutsi, C., and Joinson, A. (2011). Privacy dictionary: A new resource for the automated content analysis of privacy. Journal of the American Society for Information Science and Technology, 62 (11), 2095-2105.

von Zedtwitz, M. (2002). Organizational Learning Through Post-Project Reviews in R&D. R&D Management, 32 (3), 255-268.

Walter, T. P. and Back, A. (2013). A text mining approach to evaluate submissions to crowdsourcing contests. In 46th Hawaii International Conference on System Sciences (HICSS), 3109-3118.

Weber, R.P. (1990). Basic Content Analysis. 2nd Edition. Sage Publications, Newbury Park, CA.

Weiss, S. M., Indurkhya, N., and Zhang, T. (2010). Fundamentals of predictive text mining. Springer Science & Business Media, New York, NY.

Wilson, T., Wiebe, J., & Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. Computational linguistics, 35(3), 399-433.

Wysocki, R. K. (2014). Effective Project Management: Traditional, Agile, Extreme, Seventh Edition. John Wiley & Sons, Hoboken, NJ.

Young, L. and Soroka, S. (2012). Affective news: The automated coding of sentiment in political texts. Political Communication, 29 (2), 205-231.

Zelkowitz, M. V. and Wallace, D. R. (1998). Experimental models for validating technology. Computer, 31 (5), 23-31.

Zhao, D. and Zuo, M. (2011). Knowledge Transferred across Projects and Its Characteristics in an IT Service Enterprise. In Proceedings of International Conference on Information Technology, Computer Engineering and Management Sciences (ICM), 267-270.