International Journal of
Information Systems and Project Management
ISSN (print):2182-7796, ISSN (online):2182-7788, ISSN (cd-rom):2182-780X
Available online at www.sciencesphere.org/ijispm

Perspectives on reusing codified project knowledge: a structured literature review

André Coners
South Westphalia University of Applied Sciences
Haldener Strasse 182, 58095 Hagen
Germany
www.shortbio.org/coners.andre@fh-swf.de

Benjamin Matthies
South Westphalia University of Applied Sciences
Haldener Strasse 182, 58095 Hagen
Germany
www.shortbio.org/matthies.benjamin@fh-swf.de

Abstract:
Project documentation represents a valuable source of knowledge in project-based organizations. The practical reality is, however, that the knowledge codified in project documents is hardly re-used in future projects. A central problem in this context is the extensive amount of usually textual material. As a consequence, computer-assisted processes are indispensable in order to analytically manage the constantly growing and evolving databases of available project documents. The goal of this study is to summarize the current research focusing on the computer-assisted reuse of textually codified project knowledge and to define the corresponding state-of-the-art in this this specific field of information systems research. As a result of a literature review, this study structures the body of research contributions and outlines what kinds of computer-assisted techniques are incorporated, what practical application areas these solutions address, and in what business domains they are applied. In particular, this should point out research opportunities and thereby make a contribution to the further development of knowledge management in project environments.

Keywords:
project knowledge management; knowledge management; project documentation; codified knowledge; text mining.

DOI: 10.12821/ijispm060202

Manuscript received: 27 September 2017
Manuscript accepted: 12 November 2017

1. Introduction

In general, there are two strategies for managing and transferring knowledge in an organization [1]: codification strategy (i.e., transferring codified knowledge via documents) and personalization strategy (i.e., sharing tacit knowledge directly from person to person). In regard to the codification strategy, project-based organizations usually invest a lot of time and effort in the preparation of diverse project documents [2-4] such as project appraisals, requirement specifications, or post-project reviews [5]. The knowledge codified in such documentations fundamentally represents a valuable source of lessons learned, concepts, and interdependencies that can be relevant for future projects [6-9]. Consequently, in the sense of a continuous learning process, the PMBOK® guide and PRINCE2 framework specifically recommend archiving all project documentation produced in publicly accessible organizational knowledge databases as a rich trove of expertise [10, 11]. The reuse of this constantly growing base of knowledge allows a project-based organization to develop continuously by disseminating the lessons learned and best practices from completed projects and thus avoiding the repetition of errors and extra work in future projects [12-14]. In this context, a consequent project knowledge management plays a vital role in creating, archiving, sharing, and finally utilizing such knowledge in project-based organizations [13].

The practical reality is, however, that the knowledge codified in project documents is hardly reused in project-based organizations [2, 15-17], which are characterized by the fact that business functions are primarily performed in interdisciplinary, temporal, and constantly evolving projects rather than in functional organizations [18]. A central problem in this context is the extensive amount of usually textual material [15, 19-21]. As a consequence of this, the main challenges in reusing such textually codified knowledge are: (1) the effective identification and efficient retrieval of relevant knowledge content and (2) the combined, i.e. summarizing analysis and presentation of various knowledge sources. The first challenge results from the circumstance that the extensive and constantly evolving knowledge repositories lead to an information overload in project-based organizations [2, 17, 22, 23]. The databases of material are too large for project managers to handle manually – in a reasonable amount of time –, meaning for them to read and interpret all potentially relevant document contents. Moreover, project managers find the search for relevant knowledge frequently to be a "waste of time" and are not very motivated to search electronic document databases for relevant pieces of knowledge [24]. The second challenge results from the fact that the isolated analysis of individual cases often does not supply any representative insight into the standard risks, success factors and their relationships [19, 20]. Project managers therefore need to combine and synthesize the knowledge described in several contextually related project documentations in order to ultimately achieve a profound information basis for decision-making [25]. As a consequence, these described challenges usually lead to the fact that the potential for reusing codified knowledge sources often remains untapped. Specifically, this means that the available project knowledge is lost in the evolving environment of project organizations – the phenomenon of so-called "project amnesia" – and that mistakes already made are repeated in the worst case or that already solved problems must be solved again.

As a consequence of the practical problems described, research has increasingly focused on effective reuse of textually codified project knowledge. In particular, the implementation of computer-assisted text analyses has gained importance and established itself as a particular field of research in the meantime [26, 27], which involves the systematic classification, structuring, retrieval, summarizing and visualization of textual knowledge content (see e.g., [20, 25, 28]). Since there has been an increasing amount of recommendations for such computer-assisted reuse of knowledge in recent years, this study follows the call of Almeida and Soares [2], who recommend the structured summary of research efforts in the field project knowledge management. Although several forms of knowledge transfer exist, this study specifically focusses on the technical perspectives on reusing codified project knowledge in this context. The goal is to bring to light what research has contributed to solving the above-mentioned problems and what concrete concepts and approaches have been incorporated. For this purpose, a structured literature review is used to provide a summary of related research contributions, providing an overview of the technical approaches, practical application cases and business areas of use.

The following three research questions (RQ) are asked for this:

RQ1: *What computer-assisted technologies are employed for reusing textually codified project knowledge?*

RQ2: *What practical areas of application are recommended for the implementation of appropriate computer-assisted solutions in project management?*

RQ3: *In what business domains are appropriate computer-assisted solutions used?*

Answering these questions should allow for a representation of the state-of-the-art in the computer-assisted reuse of textually codified project knowledge. In general, this creates awareness of textual project documents as valuable knowledge carriers and describes the practical options of such knowledge reuse scenarios. In particular, this structured overview should also serve as a reference for defining future research opportunities. Finally, this should make a contribution towards the further development of knowledge management in project environments (see [29-31]).

This paper is structured as follows: Section 2 introduces the theoretical background, explaining the role of project documentation as a source of knowledge and discussing the fundamentals of computer-assisted text analyses. Section 3 describes the design of the literature review. Section 4 summarizes the results of the literature review. Section 5 contains a critical discussion of the findings and their implications for the project management. Section 6 provides the final conclusions and outlook.

2. Research background

2.1 Project documentation as a source of codified knowledge

According to the knowledge codification strategy (see [1]), project documentation is intended for the purpose of explicitly saving experience implicitly gained by the project employees, transferring it to organizational knowledge databases and thereby making such experiences available for subsequent projects as a valuable source of knowledge [16]. A special role is played by so-called post-project reviews [6, 9, 12, 14, 32], which can be defined as "the final formal review in the course of a project that examines any lessons that may be learned and used to the benefit of future projects" [32, p. 256]. As this definition already suggests, the "lessons learned" from the projects are the focus of interest in such reports. Lessons learned can be described as "key project experiences which have a certain general business relevance for future projects. They have been validated by a project team and represent a consensus on a key insight that should be considered in future projects" [14, p. 220]. Lessons learned are experience reports codified textually and describing not only the questions of "what?", "where?" or "how many?", but rather also the questions of "how?", "why?" and "with what effect?" [14]. Such codified project knowledge usually contains innovative knowledge that can be useful in various ways for future projects [5]:

- It is a starting point for conceptualizations and for feasibility studies of future projects.
- It is a reference to the planning of work packages, project milestones and project costs.
- It provides instructions for executing project management functions.
- It is a training resource for new project managers or inexperienced team members.
- It allows for a collective analysis of the causes of failed or problematic projects.

The established project management guidelines explicitly recommend that the knowledge codified in project documents be secured and systematically reused [10, 11]. The PMBoK® guide refers to such project knowledge as "organizational process assets" that should act as a fundamental basis of knowledge for projects [11]. Likewise, the PRINCE2 framework recommends the reuse and analysis of codified project knowledge in order to summarize the relevant lessons learned in a so-called "lessons log" during the project planning phase [10]. Multiple empirical studies in this context have already been able to prove that the systematic reusing of available project knowledge can have a significant positive impact on project performance (see, e.g., [33-36]). However, the true potential of such project-based learning is often not fully exploited. In this context, the term "project amnesia" stands for the untapped potential of profiting from past (and documented) successes and failures [14]. The reasons are, among others, insufficient willingness for learning

from documented experiences, difficulties in retrieving relevant knowledge content, or lacking enforcement of the project knowledge management procedures [14].

2.2 Knowledge discovery in textual databases

In order to analytically manage extensive textual databases, computer-assisted processes are indispensable [15, 20, 25]. Computer-assisted, i.e. (partially) automated analyses of textual databases have made great progress over the last decade and are often referred to by the term text mining, although there are also very different descriptions in this context (see, e.g., [37-40]). Text mining, however, is less an individual technique than a generic term for multiple complementary techniques from various sub-disciplines such as classical data mining, (computer)-linguistics or information retrieval [39]. Text mining combines such competencies and can be described collectively as "[t]he semiautomated process of extracting patterns (useful information and knowledge) from large amounts of unstructured data sources" [39, p. 1023].

For the purpose of this study, a differentiation shall be made between six fundamental text mining areas (according to [39, pp. 29-41]):

(I) Search & information retrieval: The computer-assisted naming, searching and retrieval of content in extensive, textual databases.

(II) Document classification: The partially automated (supervised) classification of text content or entire documents in predefined thematic categories based on trained classification models.

(III) Document clustering: The automated (unsupervised) grouping of text content or entire documents according to their semantic similarity based on clustering algorithms.

(IV) Information extraction: The targeted extraction of specific facts from unstructured (textual) databases in structured, relational data formats (e.g., data tables).

(V) Concept extraction: The statistical extraction and summarization of interpretable content structures and relationship patterns (e.g., semantically connected words) from textual databases.

(VI) Natural language processing: The automated processing and analysis of linguistic facets in texts.

The use of such text mining techniques has been proposed multiple times in the context of project management. Choudhary et al. [20] stress here that the lack of efficient computer-assisted tools can make the reuse of textually codified project knowledge a time- and cost-intensive process. Therefore, computer-assisted processes are indispensable in order to analytically manage the constantly growing and evolving databases of available project documents.

3. Literature review design

A literature review is a suitable method for providing a structured summary of the current research efforts in a specific discipline [41]. Since there may be differences between the design and execution of such a review, a detailed documentation of the review process is required [42]. To understand the idea of the following review design, therefore, it is necessary to address the six characteristics proposed by Cooper [43] in the following: (1) *goal*, (2) *focus*, (3) *perspective*, (4) *coverage*, (5) *organization*, and (6) *audience*.

(1) The *goal* of this literature review is to summarize and structure the research to date focusing on the computer-assisted reuse of codified knowledge in project management. Specifically, this involves answering the question of what concrete text analysis techniques are used (RQ1), what practical application areas are conceivable in project management (RQ2) and in what business domains such solutions are used (RQ3).

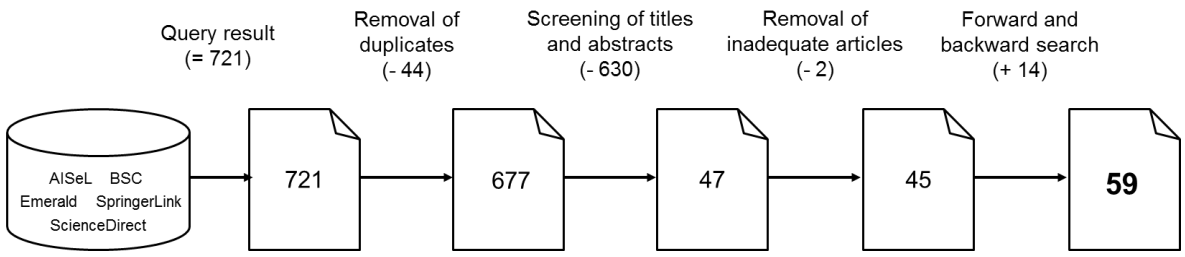
(2) The *focus* lies on scientific journals and conference proceedings reviewed by professionals (peer-reviewed), which propose practical solutions for the computer-assisted reuse of textual documentation in project

environments. In general, the resulting literature mainly derived from the fields of information systems and management (e.g., *Automation in Construction*, *Computers in Industry*, *Decision Support Systems*, *Expert Systems with Applications*), project management (e.g., *International Journal of Project Management*, *International Journal of Managing Projects in Business*), and knowledge management (e.g., *Journal of Knowledge Management*, *Knowledge Management Research & Practice*).

(3) A neutral *perspective* is adopted, i.e. the research papers are solely structured and summarized in terms of content, without taking a critical position in regard to their proposed solutions.

(4) The *coverage* of the review includes a systematic keyword search in five literature databases (see Table 1 for details). The corresponding retrieval combined two content dimensions by making typical terms for project documentation (cited from [12] and [5]) and typical terms in data mining (based on [39]) appear closely together (search fields: title, abstract or keywords). The corresponding search led to a comparatively large number of articles being discovered (721 with duplicates; 677 without duplicates). Potentially relevant articles were then selected on the basis of a screening of the titles and the abstracts (total: 47) and then analyzed in full. In this process, two articles were removed from the database on account of a lack of relevance and scientific suitability. Additionally, a reference search (forward- und backward search; as recommended by [44]) was conducted on the basis of the remaining 45 relevant articles. By this means, it was possible to identify another 14 relevant reference articles so that a collection with a total of 59 articles was ultimately available for the review.

Table 1. Literature review design: retrieval and determination of relevant research contributions



Notes:

Databases:	AIS electronic Library, EBSCO BSP, Emerald, ScieneDirect, SpringerLink
Keyword search:	("project review" OR "post-project" OR "post project" OR "postmortem review" OR "project appraisal" OR "project reflection" OR "project evaluation" OR "project debriefing" OR "project document*" OR "project knowledge" OR "project experience*") AND ("content analys*" OR "text analys*" OR "discovery" OR "textual analys*" OR "mining" OR "detection" OR "modeling" OR "mapping" OR "extraction" OR "classification" OR "clustering" OR "grouping" OR "synthe*" OR "retriev*" OR "reusing" OR "reuse" OR "coding")
Search fields:	title, abstract, keywords
Search restrictions:	peer-reviewed research (scientific journals and conference proceedings)

(5) The *organization* of the review is based on a primarily methodological procedure, with the articles being assessed and categorized independently by two experts according to their content. The creation of these categories is fundamentally based on a qualitative content analysis (i.e., reading, interpreting, and grouping findings into meaningful thematic categories; see [45]). By comparing the respective groupings (three grouped categories according to RQs 1-3), it was possible to evaluate the inter-coder agreements (between 78 and 91% in the course of three groupings), which were on an acceptable level. Then the differences were discussed and specific articles categorized by consensus. Initially, the respective papers were categorized according to which of the six typical text mining techniques were used (see Section 2.2). This categorization offers insight into the

dissemination of technical procedures and methods. Then the articles were grouped on the basis of the respective practical areas of application (such as the assessment of project risks or the evaluation of project documentation quality). Here, the underlying question was: What is the practical purpose of the proposed solution? The last groupings were based on areas of use (business domains) for the proposed solutions (such as, e.g., construction or industrial engineering).

(6) The *audiences* of interest in this review are researchers and practitioners of the project management and knowledge management disciplines in the broader sense and experts in project knowledge management in the narrower sense.

4. Findings

The 59 proposed solutions identified in the literature review were initially organized according to the six typical text mining areas (see Figures 1), whereby the solutions can also use multiple techniques at the same time. As a result of this categorization, a qualitative overview of the most widespread techniques in project management was created (RQ1). This shows that the solutions proposed in the context of project management are largely focused on techniques in the area of *search & information retrieval* (43 papers), which support the systematic finding and efficient retrieval of specific knowledge content. *Document classification* (21) and *document clustering* (16) are also quite common for thematic classification or semantic grouping of textual content. A total of 21 papers used techniques related to *concept extraction* in order to create new and enriched content concepts from a combined and summarized corpus of project documents. Relatively less widespread are the techniques of *information extraction* (4) for transferring facts from unstructured information in structured formats and the *natural language processing* (2) for the analysis of linguistic facets in documents.

Figure 1. Quantitative results of the literature review: text mining areas

Text mining areas

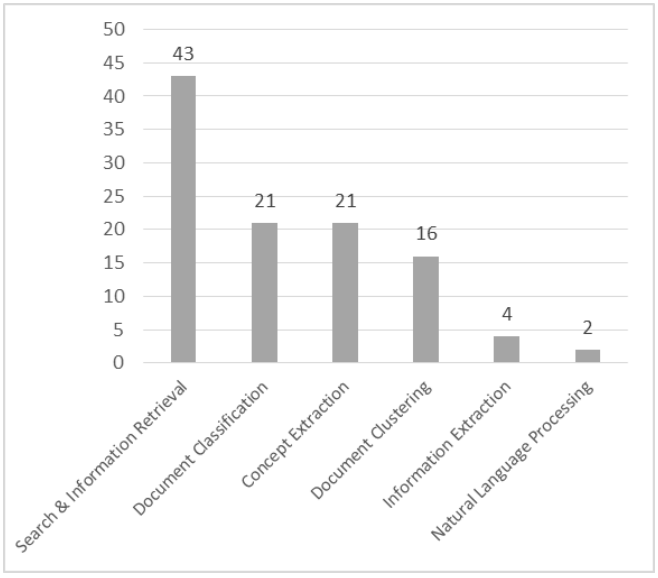


Figure 2. Quantitative results of the literature review: application areas

Application areas

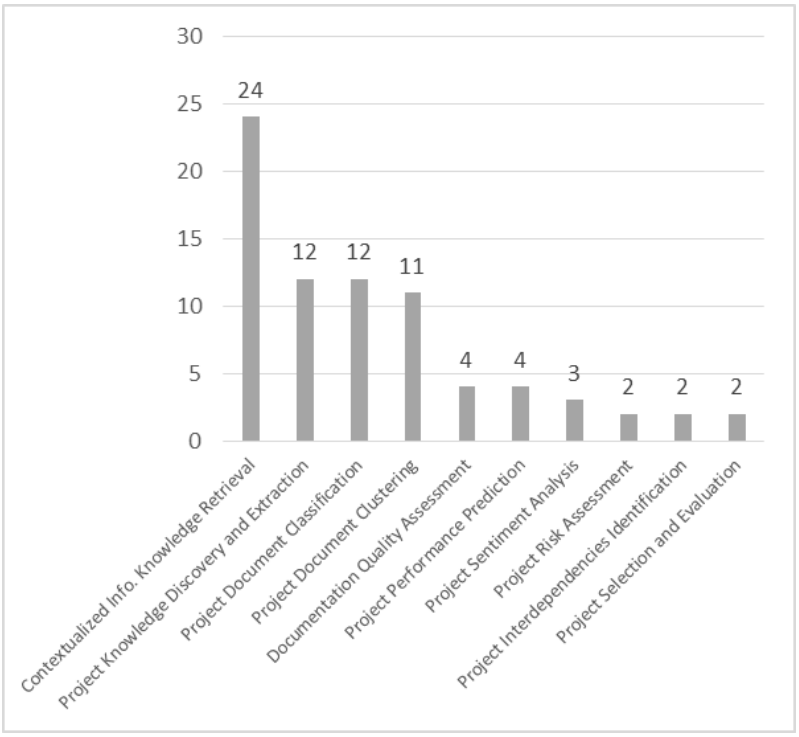
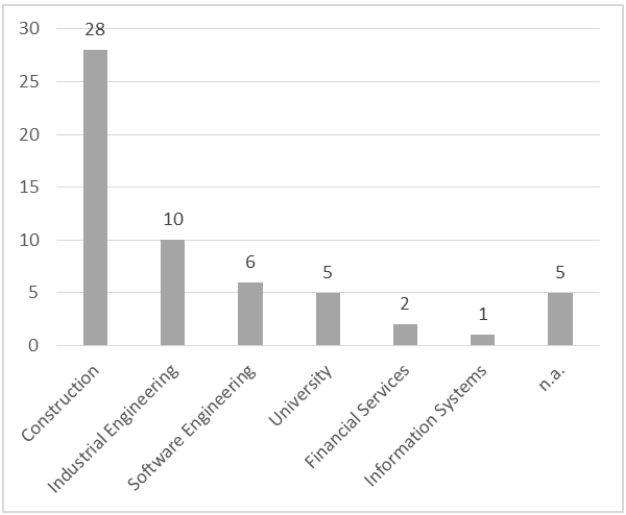


Figure 3. Quantitative results of the literature review: business areas

Business areas



The next step involved a qualitative content analysis, in which the proposed solutions were categorized in practical areas of application for which the corresponding tools were used (RQ2). The proposed solutions here can also address multiple cases of applications simultaneously. A total of 10 practical application areas could be identified, which stand for different problems or tasks as part of the reuse of the codified project knowledge (see Figure 2 for quantitative results and Table 2 for a complete description of the areas with references). In the following these application areas (frequencies in brackets) shall be described in more detail.

Contextualized information and knowledge retrieval (24): Multiple authors have stressed the need for conceptualized knowledge structures, i.e. an organization of the documents and individual knowledge content according to thematic focal points such as project phases or processes, products or specific situations (see, e.g., [46-49]). As demonstrated by Demain and Balatsoukas [46]), for example, such contextualized as-needed information can be the facts related to the project tasks or problems (i.e., a specific piece of information) or a large number of references for decision-making (e.g., for identifying standard procedures or finding the best option among alternatives). The proposed solutions in the application area of contextualized information and knowledge retrieval aim to give project employees appropriately targeted access to contextualized and thus task-specific knowledge content. On the basis of such contextualized databases, more user-oriented search strategies can be developed for project managers, with these strategies allowing for systematic keyword searches for relevant documents and document content, for example (see [50]). In this context, multiple papers describe the underlying conception of domain-specific taxonomies or, even more specifically, technical ontologies that can be referred to as "a body of formally represented knowledge" [51, p. 199]. In ontological databases, domain-specific knowledge areas and technical terms are defined and their relationships are mapped [52]. Thus, they act as a content framework and simultaneously as a technical basis in order to save project documents and content in databases in a contextualized way (e.g., according to specific project themes, tasks or problems). To give an example, Kamara et al. [53] use XML-based structures here in order to implement a so-called "project-in-context" model according to project-specific task stages and thus to contextualize the codified project knowledge in databases accordingly (see also [48] and [49]). Ma et al. [54], to cite another example, integrated knowledge context parameters into project management tools such as PERT (using a so-called building-block approach) in order to record and reuse project task knowledge according to the knowledge context. As another functionality, frequency analyses of subject-specific keywords can also be carried out in order to identify important concepts or trends, for example (see [20]). To this end, Ozorhon et al. [55] developed, for example, a framework with project-specific term dictionaries that can systematically design keyword searches and summaries.

Project knowledge discovery and extraction (12): In regard to the reuse of codified project knowledge, Oluikpe et al. [56] stress that the isolated consideration of individual cases does not offer any representative insight into typical project success factors or errors. Rather, the combined, i.e. summarizing analysis of multiple documented experiences allows for the extraction of new and more enriching knowledge such as thematic patterns and typical connections hidden in a larger corpus of project reports. In particular, the techniques of concept extraction can be used for this, as they combine larger corpora of documentation in an inductive process and summarize the codified knowledge in interpretable concepts [39]. Such concepts extracted can be described as semantic patterns (e.g., a bundle of highly correlated words, i.e. words frequently used together) that can be interpreted as widespread thematic concepts included in the examined corpus of project documents. For example, Carrillo et al. [19] used link analyses where correlations between a set of keywords and phrases were revealed in order to formulate the rules for connections in project business. For example, such analyses highlighted the importance of maintaining a good relationship with construction parties, which leads to substantial discounts and cheap rates. Information extraction techniques can also be used in order to transfer certain facts (e.g., names, techniques or dates) from unstructured databases of texts and transfer them to structured formats (e.g., relational data tables). This is helpful in the discovery of sources and networks of expertise, for example. Liu and Hsu [57] used such techniques to extract relevant facts from a body of relevant documentation, such as project members, activities, tools and costs, and to define relationship rules beyond this (e.g., “john“ has skills in “java“ and assisted in “Marketing, CRM“). Carrillo et al. [19] also describe the possibility of extracting information such as deadlines or supplier names and using it in subsequent steps for correlation and trend analyses.

Project document classification (12): In order not to lose the overview of the large databases of heterogeneous project documents, Frey et al. [33] stress that a clear organization of documents is initially of central importance. In order to organize orderly access to relevant project documents, document classification techniques are useful for classifying heterogeneous project documents in predefined thematic categories. Document classification is based on a semi-automatic process in which the analyst initially predefines thematic categories that are then learned algorithmically on the basis of a sample set of documents preorganized by content so that following documents can be automatically classified in appropriate categories according to their content context (see [39]). Such categories can represent many subjects in the daily project work, such as specific project products or attributes. Multiple papers proposed corresponding procedures for such an organization of project documents (see, e.g., [28, 58, 59]). Ur-Rahman and Harding [60] demonstrated, for example, an approach that analyzes content statement patterns and makes it possible to differentiate between project documents that report on project success or failure. Jiang et al. [61] also describe an automatic classification of content according to domain-specific standards for projects in the construction industry (BIM; Building Information Models).

Project document clustering (11): Multiple solutions incorporate clustering methods in order to group heterogeneous project documents in a completely automatic way on the basis of their semantic similarity and thus to ultimately facilitate organized document management. Such techniques are suited in particular for overcoming the limits of document classification techniques, which only function when sufficiently suitable documents are available for training (algorithmic learning) of the thematic categories. Al Qady and Kandil [62] used document clustering techniques accordingly, for example, in order to organize an unorganized body of project documents in the construction sector according to textual similarity. Document clustering techniques are also used as a preparatory work step in order to arrange heterogeneous document databases for additional analyses in homogeneous project document groups. In this context Liu and Hsu [57] stress the structuring of large heterogeneous document databases in conceptually homogeneous groups that maximize the content similarity of the textual data and thus minimize distortions in the following statistical analyses as a result of unsuitable document content (statistical "noise"). Wu and Lin [48, p. 131] used clustering techniques accordingly in order to group a large database of project documents in "conceptually highly homogeneous documents" so that the individual content concepts of these appropriately separated documents groups can be analyzed in isolation in the next step and summarized (among others, with the techniques of concept extraction).

Documentation quality assessment (4): Reviewing the content of project documentation pursues the goal of ensuring the quality, completeness, and richness of documented content. The content completeness of documentation can be defined on the basis of specific taxonomies and subsequently verified. To this end, Mena-Nieto et al. [63] developed, for example, the eXtensible Project Documentation Reporting Language (XPDRL) which maps specific documentation standards in the construction sector and thus facilitates the effective exchange of documents and also the subsequent verification of the completeness of documents. Arthur and Stevens [64] pursued another approach by calculating the "goodness" of software documentation on the basis of so-called "documentation quality indicators" (i.e., accuracy, completeness, usability, and expandability). To record the content richness, Campbell et al. [65] used topic analyses in order to examine documented project descriptions in terms of whether certain relevant topics in documentation are inadequately covered. Deficient documentation was identified in software developments by comparing extracted and potentially relevant topics in the stack overflow with the topics actually treated in the documents.

Project performance prediction (4): The forecast of the developments in projects is a field in which negative issues in current project documentation are identified in particular and used in order to counter their potentially negative impact in due time. Williams and Gong [66] as well as Singla and Kakkar [67] have focused here on the prediction of cost overruns by using semantic analyses to identify typical indicators for cost problems and then classifying projects with potential deficient developments. Prieto [68] also used comparable semantic analyses in order to identify so-called "negative performance precursors" and to draw attention early on to problems and potential project failure.

Project sentiment analysis (3): Project sentiment analyses are intended for extracting project-specific sentiment patterns from the textual project communications and thus promoting "emotional awareness" in project management. The emotions expressed in connection with a specific subject (e.g., project changes) can be evaluated and support in

particular the reactions to negative developments. Prieto [68] as well as Guzman and Bruegge [69] use lexical sentiment analyses, i.e. evaluations of specific emotional terms (positive/negative) in order to examine the emotional state in project communication. Prieto [68] used the corresponding analyses for the calculation of specific "project sentiment indices," while Guzmann and Bruegge [69] primarily used visualizations of emotions in projects.

Project risk assessment (2): The organizational lessons learned databases can be incorporated for the assessment of potential project risks. The evaluation, linking and modeling of experiences gained in past projects allows the project manager to gain an overall picture of the sources of problems, causal connections, interdependencies and vulnerability factors for the planning of future projects. For this purpose Dikmen et al. [70] developed a risk-related information taxonomy in lessons learned databases that allows the orderly saving and modeling of project-specific risks in knowledge databases and, building on this, facilitates systematic risk assessment throughout the project life cycle. Yildiz et al. [71], as another example, recommend a comparable risk mapping tool that makes it possible for the user to search systematically for relevant sources of risks, past problematic cases and potential vulnerability factors and to also identify relationships.

Project interdependencies identification (2): The discovery of patterns and similarities in project documents makes it possible to identify interactions and interdependencies between projects for the most part automatically. For this purpose, Lin et al. [72] used, for example, information retrieval and concept extraction techniques in order to visualize project team coordination patterns (e.g., task categories, execution sequences, connections and durations) for the support of a multi-project management. Meier [73], as another example, recommended a document clustering approach in order to evaluate overlapping project descriptions and thus to identify project output interactions in project portfolios.

Project evaluation and selection (2): The selection of projects can also be supported by having project proposals presorted automatically, e.g., by grouping them according to content similarity or in predefined classes (e.g., research disciplines or consistency of proposal descriptions), or on the basis of content indices. Such proposals were made by Ma et al. [74] and Zhu et al. [75].

In a final categorization, the articles were arranged according to their business domains (RQ3), for example, by construction, industrial engineering, or software development. As a result (see Figure 3), it can be stated that a large share of the research papers (28) comes from the area of construction or address such projects specifically. The second-largest share of papers comes from the area of industrial engineering (10), followed by domains with less frequent papers (software engineering, university, financial services and information systems development).

Table 2. Qualitative results of the literature review

Application Area	Description	References
Contextualized information and knowledge retrieval	Computer-assisted search and retrieval of documents or specific knowledge content relevant for specific project contexts (e.g., project phases or tasks).	[24], [25], [46], [47], [48], [49], [53], [55], [57], [61], [72], [76], [77], [78], [79], [80], [81], [82], [83-85], [86], [87], [88], [89], [103]
Project knowledge discovery and extraction	Extraction of new knowledge by revealing semantic patters contained in combined project documents for the modeling of content relationships and extraction of thematic concepts.	[19], [20], [48], [50], [57], [56], [59], [76], [82], [90], [91], [92]
Project document classification	Semi-automated classification of project documents and document contents according to manually predefined and trained thematic categories.	[28], [58], [59], [60], [61], [82], [90], [93], [94], [95-97], [98]
Project document clustering	Automated organization of project documents or document contents in hierarchical groups and structures according to content similarity.	[19], [20], [48], [50], [56], [57], [59], [62], [74], [76], [92]

Application Area	Description	References
Documentation quality assessment	The evaluation of content quality, richness and completeness of textual project documentation.	[63], [64], [65], [99], [100]
Project performance prediction	The forecast and evaluation of future project performance on the basis of performance indicators codified in project documents.	[66], [67], [68], [101]
Project sentiment analysis	The discovery, evaluation and visualization of sentiment patterns in project communication.	[68], [69], [102]
Project risk assessment	Support of risk analyses through the discovery, organization and modeling of risk factors codified in project documents.	[70], [71]
Project interdependencies identification	The identification of project interactions and interdependencies on the basis of similarities and overlaps in content across project documents.	[72], [73]
Project evaluation and selection	Support of the evaluation and selection of projects through thematic analysis and content-related grouping of project descriptions.	[74], [75]

5. Discussion

The findings brought to light through the literature review can feed into the discussion of future paths of development. Fundamentally, the examination of the reuse of project knowledge is an interesting and promising field of research (see, e.g., [24, 29]). In their work on project knowledge reuse, Schacht et al. [24] stress that the usual knowledge management literature to date has primarily focused on the creation, storing and the transfer of knowledge. Knowledge reuse is often viewed implicitly as a natural result of available, saved and transferred knowledge and is underrepresented in the literature. Having this in mind, the results of the literature review should be summarized, research questions answered, and research prospects discussed below.

RQ1: First, it became apparent that a large part of the research focusses on the development of techniques to identify and organize (contextualized) knowledge fragments, which can then be retrieved by project managers according to their specific information needs. In contrast, techniques regarding the discovery and prediction of previously unknown and potentially useful issues from the textual data are underrepresented in the reviewed body of literature. Here, recent advances in predictive text mining (see [40]) hold interesting and untapped potentials for the project management discipline. Several contributions, such as Williams and Gong [66] or Prieto [68], exemplified the potentials of such techniques for predicting project cost overruns or discovering hidden sentiment patterns in project communication.

RQ2: Second, it became clear that a majority of the research papers concentrate on individual application areas as well as text mining techniques. Only a few papers cover a range of multiple problem and application areas. Nonetheless, it could be of great practical interest if research contributions could offer complete solution packages that give project management a more comprehensive range of functions. Such packages could combine multiple application areas as part of an electronic document management (EDM) in project environments and provide solutions that, for example, not only group or classify the available project documents automatically and provide efficient access to individual contextualized knowledge content, but also simultaneously place new knowledge concepts, i.e. combined and therefore enriched knowledge, corresponding to the problem context, at the disposal of the project managers.

RQ3: Third, it could also be discovered that the construction sector in particular supports intensive efforts to reuse codified databases of knowledge. This could be explained, among others, by the fact that construction projects are particularly complex and knowledge-intensive in which historical documentation is also very extensive (see [58]). In

light of this finding, a cross-disciplinary exchange between the business domains (such as construction and software engineering) could be recommended and promise interesting learning effects.

Finally, the reviewed solutions rarely proposed specific procedures or guidelines for integrating the identified knowledge into typical planning and decision-making processes for projects. This is necessary, however, in order to ensure systematic, standardized and thus consistent learning in project-based organizations. In this context, Barclays and Oseri-Bryson [16] criticize the circumstance that even if relevant lessons learned from past projects are at the disposal of project managers, this knowledge often flows into decision-making and learning processes only implicitly, meaning at the discretion of the individual project managers. Systematic, non-intuitive and standardized procedures could therefore result in additional added value for the primarily technically-oriented approaches. Furthermore, the acceptance of such advanced business intelligence processes (e.g., perceived usefulness and ease of implementation) needs to be kept in mind in order to ensure a consequent usage of the solutions by end users [104].

6. Conclusion

In the literature there is widespread agreement that the lack of efficient (i.e. computer-assisted) tools makes the reuse of knowledge codified textually in project documents a time- and cost-intensive process. The goal of this study was to structure the previously proposed research contributions and define the corresponding state-of-the-art in this special field of project knowledge management. The study outlined what computer-assisted techniques are used (RQ1), what application areas these solutions address (RQ2) and in what business domains they are applied (RQ3). The subsequent discussion also provided paths for the use, dissemination and targeted development of these solutions.

The implications for practice are that a summarizing overview of possible solutions for the reuse of knowledge codified in organizational databases is placed at the disposal of practical project management – a task for which no specific procedures have been recommended in the standard project management guidelines. In general, this should create awareness of textual project documents as valuable and also exploitable source of knowledge. In particular, this should provide impulses for the widespread use of such solutions and ultimately also for the further development of project-based learning in organizations. In this regard, a cross-domain exchange between different areas of use (e.g., construction and software engineering) could also supply interesting ideas.

There are implications for research, particularly in the further development of appropriate solutions. First, techniques regarding the discovery and prediction of previously unknown and potentially useful issues (e.g., project cost overruns or sentiment patterns) are underrepresented in the analyzed body of literature. However, several contributions, such as Williams and Gong [66] or Prieto [68], exemplified the potentials of such techniques in the project management context. Second, the combination of application areas has promising potential. For example, the identification of contextualized knowledge with the simultaneous synthesizing of multiple historical sources of knowledge could support a project manager by allowing him or her to receive not only the desired particles of knowledge, but rather also combined, i.e. new and enriched knowledge in a certain context. Third, the development of procedures for the structured implementation of the extracted knowledge in the project management's decision-making processes offers potential hardly tapped to date.

The limitations of this study include, above all, the typical limits of a literature review. In general, a structured literature review cannot identify all potentially relevant research contributions. For example, comparable articles from other domains such as product development or marketing could be of interest as well. Nonetheless, this review focused specially on solutions discussed in the project management context, which were gathered on the basis of relatively broad-based search on databases. In order to address the limitations of a literature review and to gain additional insight into the subject discussed here, it would also be advisable to pursue more related studies, e.g., in the form of expert interviews or surveys conducted with project managers. In this context, it could also be of interest to further study the relationships to other forms of knowledge transfer (e.g., the person-to-person transfer of tacit knowledge), which were out of scope of this study.

References

[1] M.T. Hansen, N. Nohria, and T. Tierney, "What's your Strategy for Managing Knowledge?," *Harvard Business Review*, vol. *77*, no. *2*, pp. *106-116*, 1999.

[2] M. V. Almeida, A. L. Soares, "Knowledge sharing in project-based organizations: Overcoming the informational limbo," *International Journal of Information Management*, vol. *34*, no. *6*, pp. *770-779*, 2014.

[3] N. Potter and M. Sakry, "The Documentation Diet," *Journal Of The Quality Assurance Institute*, vol. *18*, no. *1*, pp. *8-12*, 2004.

[4] A. Prencipe and F. Tell, "Inter-project learning: processes and outcomes of knowledge codification in project-based firms," *Research policy*, vol. *30*, no. *9*, pp. *1373-1394*, 2001.

[5] R. K. Wysocki, *Effective Project Management: Traditional, Agile, Extreme*, seventh ed., Hoboken, NJ: John Wiley & Sons, 2014.

[6] F. T. Anbari, E. G. Carayannis, and R. J. Voetsch, "Post-project reviews as a key project management competence," *Technovation*, vol. *28*, no. *10*, pp. *633-643*, 2008.

[7] W. F. Boh, "Mechanisms for sharing knowledge in project-based organizations," *Information and Organization*, vol. *17*, no. *1*, pp. *27-58*, 2007.

[8] K. Goffin, U. Koners, D. Baxter, and C. Van der Hoven, "Managing lessons learned and tacit knowledge in new product development," *Research-Technology Management*, vol. *53*, no. *4*, pp. *39-51*, 2010.

[9] U. Koners and K. Goffin, "Learning from Postproject Reviews: A Cross-Case Analysis," *Journal of Product Innovation Management*, vol. *24*, no. *3*, pp. *242-258*, 2007.

[10] OGC, *Managing and Directing Successful Projects with PRINCE2™*, 5th ed., Norwich: Office of Government Commerce. The Stationery Office Books, 2009.

[11] PMI, *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)*, Project Management Institute, Incorporated, 2013.

[12] G. Disterer, "Management of project knowledge and experiences," *Journal of Knowledge Management*, vol. *6*, no. *5*, pp. *512-520*, 2002.

[13] B. Hanisch, F. Lindner, A. Mueller, and A. Wald, "Knowledge management in project environments," *Journal of Knowledge Management*, vol. *13*, no. *4*, pp. *148-160*, 2009.

[14] M. Schindler and M. J. Eppler, "Harvesting project knowledge: a review of project learning methods and success factors," *International Journal of Project Management*, vol. *21*, no. *3*, pp. *219-228*, 2003.

[15] M. Ajmal, P. Helo and T. Kekäle, "Critical factors for knowledge management in project business," *Journal of Knowledge Management*, vol. *14*, no. *1*, pp. *156-168*, 2010.

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

[17] S. Newell, M. Bresnen, L. Edelman, H. Scarbrough, and J. Swan, "Sharing Knowledge Across Projects - Limits to ICT-led Project Review Practices," *Management Learning*, vol. *37*, no. *2*, pp. *167-185*, 2006.

[18] C. J. Middleton, "How to set up a project organization," *Harvard Business Review*, vol. *45*, no. *2*, pp. *73-82*, 1967.

[19] P. Carrillo, J. Harding, and A. Choudhary, "Knowledge discovery from post-project reviews," *Construction Management and Economics*, vol. *29*, no. *7*, pp. *713-723*, 2011.

[20] A. K. Choudhary, P. I. Oluikpe, J. A. Harding, and P. M. Carrillo, "The needs and benefits of Text Mining applications on Post-Project Reviews," *Computers in Industry*, vol. *60*, no. *9*, pp. *728-740*, 2009.

[21] H. T. Loh, R. Menon, C. K. Leong, "Mining Of Text In The Product Development Process," *Innovation in Manufacturing Systems and Technology (IMST)*, vol. *01*, 2002.

[22] B. Dave and L. Koskela, "Collaborative knowledge management - A construction case study," *Automation in Construction*, vol. *18*, no. *7*, pp. *894-902*, 2009.

[23] A. M. Haksever, "A model to predict the occurrence of information overload of project managers," In *Proceedings of the CIB*, vol. *78*, 2000.

[24] S. Schacht, S. Morana, and A. Maedche, "The Evolution of Design Principles Enabling Knowledge Reuse for Projects: An Action Design Research Project," *Journal of Information Technology Theory and Application*, vol. *16*, no. *3*, pp. *5-36*, 2015.

[25] R. Menon, L. H. Tong, and S. Sathiyakeerthi, "Analyzing textual databases using data mining to enable fast product development processes," *Reliability Engineering & System Safety*, vol. *88*, no. *2*, pp. *171-180*, 2005.

[26] V. Hausmann and S. P. Williams, "Issues for the longterm management of Social Business Documents," *International Journal of Information Systems and Project Management*, vol. *4*, no. *3*, pp. *45-61*, 2016.

[27] W.R. King, "Text analytics: boon to knowledge management?" *Information Systems Management*, vol. *26*, no. *1*, p. *87*, 2009.

[28] M. Al Qady and A. Kandil, "Automatic classification of project documents on the basis of text content," *Journal of Computing in Civil Engineering*, vol. *29*, no. *3*, 2013.

[29] P. Davidson and J. Rowe, "Systematising knowledge management in projects," *International Journal of Managing Projects in Business*, vol. *2*, no. *4*, pp. *561-576*, 2009.

[30] S. Gasik, "A model of project knowledge management", *Project Management Journal*, vol. *42*, no. *3*, pp. *23-44*, 2011.

[31] B. Hanisch, F. Lindner, A. Mueller, and A. Wald, "Knowledge management in project environments," *Journal of knowledge management*, vol.*13*, no. *4*, pp. *148-160*, 2009.

[32] M. von Zedtwitz, "Organizational Learning Through Post-Project Reviews in R&D," *R&D Management*, vol. *32*, no. *3*, pp. *255-268*, 2002.

[33] P. Frey, F. Lindner, A. Muller, and A. Wald, "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)*, 2009.

[34] H. K. Hong, J. S. Kim, T. Kim, and B. H. Leem, "The effect of knowledge on system integration project performance," *Industrial Management & Data Systems*, vol. *108*, no. *3*, pp. *385-404*, 2008.

[35] T. Kotnour, "Organizational learning practices in the project management environment," *International Journal of Quality & Reliability Management*, vol. *17*, no. *4/5*, pp. *393-406*, 2000.

[36] G. K. Kululanga and W. S. Kuotcha, "Measuring organisational learning through project reviews," *Engineering, Construction and Architectural Management*, vol. *15*, no. *6*, pp. *580-595*, 2008.

[37] W. Fan, L. Wallace, S. Rich, and Z. Zhang, "Tapping the power of text mining," *Communications of the ACM*, vol. *49*, no. *9*, pp. *76-82*, 2006.

[38] R. Feldman and J. Sanger, "The text mining handbook: advanced approaches in analyzing unstructured data," Cambridge: University Press, 2007.

[39] G. Miner, J. Elder, T. Hill, R. Nisbet, D. Delen, and A. Fast, *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*, first ed., Cambridge, MA: Academic Press., 2012.

[40] S. M. Weiss, N. Indurkhya, and T. Zhang, *Fundamentals of predictive text mining*, New York, NY: Springer Science & Business Media, 2010.

[41] J. Webster and R.T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS Quarterly*, vol. *26*, no. *2*, pp. *13-23*, 2002.

[42] J. vom Brocke, A. Simons, B. Niehaves, K. Riemer, R. Plattfaut, and A. Cleven, "Reconstructing the giant: On the importance of rigour in documenting the literature search process," In *Proceedings of the European Conference in Information Systems*, 2009, pp. *2206-2217*.

[43] H. M. Cooper, "Organizing knowledge syntheses: A taxonomy of literature reviews," *Knowledge in Society*, vol. *1*, pp. *104-126*, 1988.

[44] Y. Levy and T.J. Ellis, "A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research," *Informing Science Journal*, vol. *9*, pp. *181-212*, 2006.

[45] P. Mayring, "Qualitative content analysis," *Forum: Qualitative Social Research*, vol. *1*, no. *2*, retrieved December 7, from http://nbn-resolving.de/urn:nbn:de:0114-fqs0002204, 2000.

[46] P. Demian and P. Balatsoukas, "Information retrieval from civil engineering repositories: importance of context and granularity," *Journal of Computing in Civil Engineering*, vol. *26*, no. *6*, pp. *727-740*, 2012.

[47] H. Fan, F. Xue, and H. Li, "Project-Based As-Needed Information Retrieval from Unstructured AEC Documents," *Journal of Management in Engineering*, vol. *31*, no. *1*, 2014.

[48] I. C. Wu and S. W. Lin, "A Platform for Supporting Knowledge Mining and Reuse Based on Context Information of a Project," *Journal of Library and Information Studies*, vol. *8*, no. *1*, pp. *97-136*, 2010.

[49] I. Wu and W. Wang, "An Exploratory Study of Capturing and Reusing Project Knowledge from a Stage-based Perspective," In *Proceedings of the Americas Conference on Information Systems (AMCIS)*, Paper *336*, 2007.

[50] P. Carrillo, A. Choudary, J. Harding, and P. Oluikpe, "Extracting Knowledge from Post Project Review Reports," In *W102-Special Track of the 18th CIB World Building Congress*, 2010.

[51] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowledge acquisition*, vol. *5*, no. *2*, pp. *199-220*, 1993.

[52] R. Studer, V. R. Benjamins, and D. Fensel, "Knowledge engineering: principles and methods," *Data & knowledge engineering*, vol. *25*, no. *1*, pp. *161-197*, 1998.

[53] J. M. Kamara, C. J. Anumba, P. M. Carrillo, and N. Bouchlaghem, "Conceptual framework for live capture and reuse of project knowledge," *CIB REPORT*, vol. *284*, paper *178*, 2003.

[54] C.-H. Ma, D.-B. Luh, and C.-Y. Huang, "Applying a Systematic Building Block to Project Knowledge Management," In *Proceedings of the 4th International Conference on Next Generation Web Services Practices*, pp. *172-178*, 2008.

[55] B. Ozorhon, C. G. Karatas, and S. Demirkesen, "A Web-based Database System for Managing Construction Project Knowledge," *Procedia-Social and Behavioral Sciences*, vol. *119*, pp. *377-386*, 2014.

[56] P. Oluikpe, P. M. Carrillo, J. A. Harding, and A. K. Choudhary, "Text Mining of Post Project Reviews," In *Proceedings of the Conference on Performance and Knowledge Management (CIB W102)*, 2008 pp. *70-81*.

[57] D. R. Liu and C. Hsu, "Project-based knowledge maps: combining project mining and XML-enabled topic maps," *Internet research*, vol.*14*, no. *3*, pp. *254-266*, 2004.

[58] M. Al Qady and A. Kandil, "Document management in construction: practices and opinions," *Journal of Construction Engineering and Management*, vol. *139*, no. *10*, 2013.

[59] A. Riel and P. Boonyasopon, "A Knowledge Mining Approach to Document Classification," *The Asian International Journal of Science and Technology in Production and Manufacturing*, vol. *2*, no. *3*, pp. *1-10*, 2009.

[60] N. Ur-Rahman and J. A. Harding, "Textual data mining for industrial knowledge management and text classification: A business oriented approach," *Expert Systems with Applications*, vol. *39*, no. *5*, pp. *4729-4739*, 2012.

[61] S. Jiang, H. Zhang, and J. Zhang, "Research on BIM-based Construction Domain Text Information Management," *Journal of Networks*, vol. *8*, no. *6*, pp. *1455-1464*, 2013.

[62] M. Al Qady and A. Kandil, "Automatic clustering of construction project documents based on textual similarity," *Automation in Construction*, vol. *42*, pp. *36-49*, 2014.

[63] A. Mena-Nieto, F. J. Moral, and F. López-Rodríguez, "Digital Standard (XBRL) to Evaluate Documentation in Industrial Engineering Projects," In *Proceedings of the 41st International Conference on Computers & Industrial Engineering*, pp. *283-288*, 2011.

[64] J. D. Arthur and K. T. Stevens, "Assessing the adequacy of documentation through document quality indicators," In *Proceedings of the IEEE Conference on Software Maintenance*, 1989, pp. *40-49*,

[65] J. C. Campbell, C. Zhang, Z. Xu, A. Hindle, and J. Miller, "Deficient documentation detection: a methodology to locate deficient project documentation using topic analysis," In *Proceedings of the 10th Working Conference on Mining Software Repositories*, 2013, pp. *57-60*.

[66] T. P. Williams and J. Gong, "Predicting construction cost overruns using text mining, numerical data and ensemble classifiers," *Automation in Construction*, vol. *43*, pp. *23-29*, 2014.

[67] I. Singla and M. Kakkar, "Construction Cost Overturn Prediction Using Artificial Neural Network on Text and Numerical Data," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. *3*, no. *9*, pp. *8075-8081*, 2015.

[68] B. Prieto, "Utilization of Project Sentiment Analysis as a Project Performance Predictor," *PM World Journal*, vol. *8*, no. *11*, p. *1*, 2013.

[69] E. Guzman and B. Bruegge, "Towards emotional awareness in software development teams," In *Proceedings of the 9th Joint Meeting on Foundations of Software Engineering*, 2013, pp. *671-674*.

[70] I. Dikmen, M. T. Birgonul, C. Anac, J. H. M. Tah, and G. Aouad, "Learning from risks: A tool for post-project risk assessment," *Automation in construction*, vol. *18*, no. *1*, pp. *42-50*, 2008.

[71] A. E. Yildiz, I. Dikmen, M. T. Birgonul, K. Ercoskun, and S. Alten, "A risk mapping tool for construction projects," In *Proceedings of the 14th International Conference on Computing in Civil and Building Engineering*, Moscow, Russia, 2012.

[72] F. R. Lin, K. J. Huang, and N. S. Chen, "Integrating information retrieval and data mining to discover project team coordination patterns," *Decision support systems*, vol. *42*, no. *2*, pp. *745-758*, 2006.

[73] C. Meier, "Identifying Output Interactions Among Is Projects-A Text Mining Approach," In *Proceedings of the European Conference on Information Systems (ECIS)*, 2013.

[74] J. Ma, W. Xu, Y. H. Sun, E. Turban, S. Wang, and O. Liu, "An ontology-based text-mining method to cluster proposals for research project selection," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. *42*, no. *3*, pp. *784-790*, 2012.

[75] J. J. Zhu, H. H. Wang, C. Ye, and Q. Lang, "Project evaluation method using non-formatted text information based on multi-granular linguistic labels," *Information Fusion*, vol. *24*, pp. *93-107*, 2015.

[76] S. C. Brandt, M. Schlüter, M. Raddatz, and M. Jarke, "Management of Experience Knowledge in Continuous Production Processes," *IFAC Proceedings*, vol. *42*, no. *4*, pp. *2173-2178*, 2009.

[77] C. S. Chen and Y. C. Lin, "Enhancing Knowledge Management for Engineers Using Mind Mapping in Construction," *New Research on Knowledge Management Technology*, pp. *978-953*, 2012.

[78] M. Chia-Hsiang, L. Ding-Bang, and H. Cheng-Yong, "Applying a Systematic Building Block to Project Knowledge Management," In *Proceedings of the 4th International Conference on Next Generation Web Services Practices*, pp. *172-178*, 2008.

[79] T. E. El-Diraby and J. Zhang, "A semantic framework to support corporate memory management in building construction," *Automation in Construction*, vol. *15*, no. *4*, pp. *504-521*, 2006.

[80] Y. C. Lin and H. Y. Lee, "Developing project communities of practice-based knowledge management system in construction," *Automation in Construction*, vol. *22*, pp. *422-432*, 2012.

[81] N. Matta, G. Ducellier, Y. Charlot, M. R. Beldjoudi, F. Tribouillois, and E. Hibon, "Traceability of design project knowledge using PLM," In *Proceedings of the International Conference on Collaboration Technologies and Systems (CTS)*, pp. *233-240*, 2011.

[82] L. Soibelman, J.F. Wu, C. Caldas, I. Brilakis, and K.J. Lin, "Management and analysis of unstructured construction data types," *Advanced Engineering Informatics*, vol. *22*, pp. *15-27*, 2007.

[83] H. C. Tan, P. M. Carrillo, and C. J. Anumba, "Case study of knowledge management implementation in a medium-sized construction sector firm," *Journal of Management in Engineering*, vol. *28*, no. *3*, pp. *338-347*, 2011.

[84] H. C. Tan, P. M. Carrillo, C. J. Anumba, N. Bouchlaghem, J. M. Kamara, and C. E. Udeaja, "Development of a methodology for live capture and reuse of project knowledge in construction," *Journal of management in engineering*, vol. *23*, no. *1*, pp. *18-26*, 2007.

[85] H. C. Tan, P. Carrillo, C. Anumba, J. M. Kamara, D. Bouchlaghem, and C. Udeaja, "Live capture and reuse of project knowledge in construction organisations," *Knowledge Management Research & Practice*, vol. *4*, no. *2*, pp. *149-161*, 2006.

[86] C. E. Udeaja, J. M. Kamara, P. M. Carrillo, C. J. Anumba, N. D. Bouchlaghem, H. C. Tan, "A web-based prototype for live capture and reuse of construction project knowledge," *Automation in Construction*, vol. *17*, no. *7*, pp. *839-851*, 2008.

[87] T. Williams, "Identifying the hard lessons from projects – easily," *International Journal of Project Management*, vol. *22*, no. *4*, pp. *273-279*, 2004.

[88] I. C. Wu, "Toward supporting information-seeking and retrieval activities based on evolving topic-needs," *Journal of Documentation*, vol. *67*, no. *3*, pp. *525-561*, 2011.

[89] M. C. Yang, I. Wood, and M. R. Cutkosky, "Design information retrieval: a thesauri-based approach for reuse of informal design information," *Engineering with computers*, vol. *21*, no. *2*, pp. *177-192*, 2005.

[90] G. Hébrail and J. Marsais, "Experiments of Textual Data Analysis at Electricité de France," *New Approaches in Classification and Data Analysis Studies in Classification, Data Analysis, and Knowledge Organization*, pp. 569-576, 1994.

[91] P. A. Ménard and S. Ratté, "Concept extraction from business documents for software engineering projects," *Automated Software Engineering*, pp. *1-38*, 2015.

[92] B. Yoon, S. Lee, and G. Lee, "Development and application of a keyword-based knowledge map for effective R&D planning," *Scientometrics*, vol. *85*, no. *3*, pp. *803-820*, 2010.

[93] M. Al Qady and A. Kandil, "Document discourse for managing construction project documents," *Journal of Computing in Civil Engineering*, vol. *27*, no. *5*, pp. *466-475*, 2012.

[94] O. Alam, B. Adams, and A. E. Hassan, "Preserving knowledge in software projects," *Journal of Systems and Software*, vol. *85*, no. *10*, pp. *2318-2330*, 2012.

[95] C. H. Caldas, L. Soibelman, and J. Han, "Automated classification of construction project documents," *Journal of Computing in Civil Engineering*, vol., *16*, no. *4*, pp. *234-243*, 2002.

[96] C. H. Caldas and L. Soibelman, "Implementing automated methods for document classification in construction management information systems," *Computing in Civil Engineering*, pp. *194-210*, 2002.

[97] C. H. Caldas and L. Soibelman, "Automating hierarchical document classification for construction management information systems," *Automation in Construction*, vol. *12*, no. *4*, pp. *395-406*, 2003.

[98] J. Wasiak, B. Hicks, L. Newnes, A. Dong, and L. Burrow, "Understanding engineering email: the development of a taxonomy for identifying and classifying engineering work," *Research in Engineering Design*, vol. *21*, no. *1*, pp. *43-64*, 2010.

[99] C. Y. Chen and W. L. Tsai, "Toward Single-sourcing of Software Project Documented Contents: A Preliminary Study," *International Journal of Electronic Business Management*, vol. *9*, no. *1*, pp. *37-44*, 2011.

[100] Á. Mena, F. López, J. M. Framiñan, F. Flores, and J. M. Gallego, "XPDRL project: Improving the project documentation quality in the Spanish architectural, engineering and construction sector," *Automation in Construction*, vol. *19*, no. *2*, 2010.

[101] I. Singla and M. Kakkar, "Construction Cost Overturn Prediction Using Porter on Text and Numerical Data," *International Journal of Advanced Research in Computer Science & Technology*, vol. *3*. no. *3*, pp. *51-52*, 2015.

[102] E. Guzman, "Visualizing emotions in software development projects," In *Proceedings of the First IEEE Working Conference on Software Visualization (VISSOFT)*, 2013.

[103] Y. C. Lin, L. C. Wang, and H. P. Tserng, "Enhancing knowledge exchange through web map-based knowledge management system in construction: Lessons learned in Taiwan," *Automation in Construction*, vol. *15*, no. *6*, pp. *693-705*, 2006.

[104] M. P. Bach, J. Zoroja, and A. Čeljo, "An extension of the technology acceptance model for business intelligence systems: project management maturity perspective," *International Journal of Information Systems and Project Management*, vol. *5*, no. *2*, pp. *5-21*, 2017.

Biographical notes

André Coners
André Coners is a Professor at South Westphalia University of Applied Sciences, Germany. Before joining the South Westphalia University of Applied Sciences he was Principal at Horváth and Partners Management Consultants and responsible for the business segment Business Process Management. He has more than 10 years experience as a Management Consultant regarding Strategic Management, Business Process Management and Managerial Accounting. He published books and articles about strategic management, process mining, activity-based costing, cost management, and process management.

www.shortbio.org/coners.andre@fh-swf.de

Benjamin Matthies
Benjamin Matthies is a research assistant at the South Westphalia University of Applied Sciences, Germany. Before joining the team, he worked as business analyst for a multinational retail company and was involved in a variety of IT und business projects. His research interests include project and knowledge management, management accounting, and business analytics. Within the field of knowledge management, he particularly focuses on the application of text mining techniques for organizational learning purposes.

www.shortbio.org/matthies.benjamin@fh-swf.de