SEMAFLEX - Semantic Integration of Flexible Workflow and Document Management

Lisa Grumbach[1], Eric Rietzke[2], Markus Schwinn[2], Ralph Bergmann[1] and Norbert Kuhn[2]

[1] University of Trier, Department of Business Information Systems II, 54286 Trier, Germany
[2] Trier University of Applied Sciences, Location Birkenfeld, Campusallee, 55761 Birkenfeld, Germany

Abstract. Small and medium-sized enterpises need support by process-aware information systems (PAIS) that offer a high degree of flexiblity in workflow execution. Current agile workflow approaches lack acceptance as they introduce a significant overhead for workflow control. Therefore, we propose a new approach for flexible PAIS, based on workflow enactment flexibility by deviation. We enable the potential deviation of the factual workflow from the ideal workflow, while keeping track of the workflow execution. In this paper we describe the proposed SEMAFLEX architecture, which semantically integrates flexible workflow and knowledge-based document management, as well as selected use cases for illustrating the interrelationships among the architectural components.

1 Introduction

Process-Aware Information Systems [1] are essential for efficiency in operational processes for most enterprises in today's business. Large corporations benefit the most, as their business processes and documents are standardized and of large volume. Small and medium-sized enterprises (SME) have different requirements for support concerning their processes. The amount of transactions is much smaller, processes are less standardized [11] and most times weakly structured. The conducted workflows differ significantly due to specific surrounding conditions, peculiar events or coincidences, which implies a need for flexibility. This flexibility might be a competitive advantage concerning large enterprises as ideally a much faster and customized processing of business cases is achieved [7]. Often, current software systems, e.g. Enterprise Resource Planning (ERP) Systems, which are established in large enterprises, cannot satisfy the requirements of SMEs for support of value adding processes, especially due to their lack of flexibility. Another drawback is a significant limitation concerning document management. Business data and documents are managed, but this does not include a semantic analysis and an automatic classification. Thus the content of documents can hardly be used to control the status of processes [2]. In this paper

we present a new approach currently developed within the SEMAFLEX[3] project. Its objective is a more efficient supervision of customized business processes on the basis of a semantic integration of processes and business documents. Dependencies between processes and documents will be identified automatically and used for the automatic identification of the actual workflow. A basic idea of this approach is that the ideal workflow, defined previously, and the actual enacted workflow will be distinguished. Hereby, a new approach for flexible workflow enactment, called Flexibility by Deviation [9] (see also Sect. 2), will be provided, which will allow for deviating from the predefined ideal workflow, but without losing control. Detected deviations from the predefined workflow, will be logged, rated and considered for further control. Deficiency management in construction will serve as application scenario, as we expect a great benefit of the presented approach [6]. In this paper, we present the overall idea as well as a proposal for an architecture for implementing flexibility by deviation based on the semantic integration of document content and workflow execution information. After presenting the basic foundations, the SEMAFLEX concept and architecture are presented. The following description of detailed uses cases will illustrate the components of the architecture as well as their interrelationships.

2 Foundations

Flexible approaches concerning workflow management are discussed since about ten years [10]. Four different concepts are distinguished [9]: "*Flexibility by Design* is the ability to incorporate alternative execution paths within a process model at design time." A major drawback of this approach is that only predictable events might be included in the workflow. The second approach *Flexibility by Change* describes the "ability to modify a process model at runtime". Since every deviation in the workflow requires a manual intervention and remodeling before continuing with the workflow, the acceptance of this approach in practice is rather low. The same applies to *Flexibility by Underspecification*, which enables to postpone the definition of certain unclear parts of the workflow from design time to runtime. The fourth approach, which is implemented in the presented system, is called *Flexibility by Deviation*. It represents the ability to deviate from the prescribed workflow definition at runtime without manually modifying the workflow. Flexibility by deviation has rarely been explored in research so far. Only FLOWer [2] is an implementation of this approach, which is limited to skipping, undoing, and redoing tasks during enactment.

Workflow management is tightly connected with the exchange of documents, which must be organized systematically. Knowledge-based document management allows the semantic analysis and management of documents with the help of various kinds of background knowledge [4]. Semantic technologies enable to regard documents in the entire context of the knowledge base of the enterprise [5]. For example, it is possible to annotate documents semantically and arrange them as semantic net. Previous work in this research area rarely considers the

[3] SEMAFLEX is funded by Stiftung Rheinland-Pfalz für Innovation, grant no. 1158

application context of analysed documents and an explicit context representation is missing. For documents that emerge within a business process controlled by a workflow system, the process context is relevant but also easily available. Business process-oriented knowledge management [3] focusses exactly on the manifold relations between business process and knowledge management, though the latter is regarded predominantly compared to the flexible process enactment of single cases. The virtual office prototype [12] is one of few systems that additionally uses information of process instances for the document analysis.

3 SEMAFLEX Concept and Architecture

The SEMAFLEX concept combines flexible workflow management and knowledge-based document management. As illustrated in Fig. 1 both approaches are semantically integrated on the basis of an ontology, which stores knowledge about documents and workflows. With the help of the document management, incom-

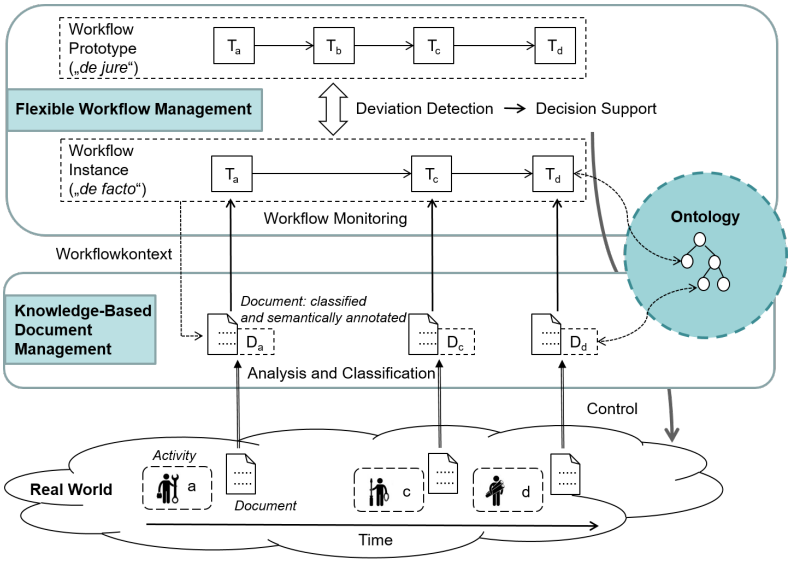

Fig. 1. Concept of SEMAFLEX

ing documents are classified and relevant information is extracted automatically leading to semantic annotations of the documents. This process is based on the conceptual knowledge of the ontology. From the semantic annotation we aim to derive by abductive inference which workflow tasks may have been executed in the real world that caused the observed documents to be present. While specific information in the document might enable to easily determine the workflow instance to which the document belongs to, the semantic description of tasks

will provide knowledge for abductively deriving a hypothesis about actually executed task. Deviations concerning actually enacted tasks (de facto workflow) and defined ideal workflow (de jure workflow) are detected and used for further workflow control. The deviations will be classified w.r.t. their criticality and if necessary, warnings can be issued. In particular, the workflow engine must consider the deviation when determining the further progress of the workflow. For example, it must decide whether a skipped task can be omitted or whether it must be catched up. This requires additional domain specific knowledge about execution constraints on the level of workflow definitions.

To implement this new approach, we propose an architecture on the basis of a semantic integration of knowledge based- document and workflow management. Figure 2 shows the architecture including its single components and their interrelationships. The architecture consists of three main parts. The green

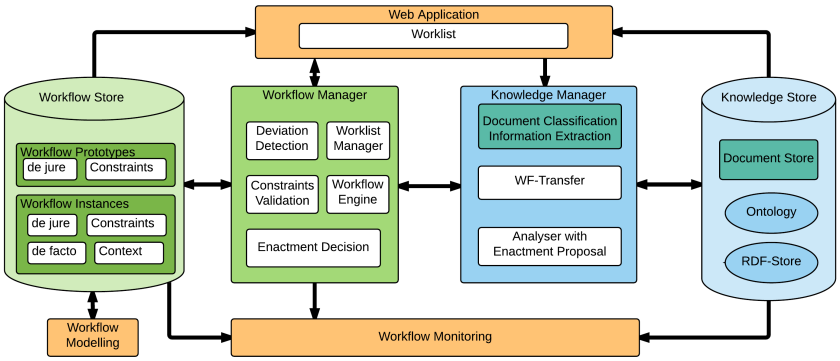

Fig. 2. Architecture of SEMAFLEX

modules are responsible for the general workflow functionality, which basically includes the realization of the approach of Flexibility by Deviation. The blue components realize the semantic integration of documents to represent external process contribution. Every module coloured in orange belongs to a certain user interface. Not only the colours illustrate associated components, but also the layout points out different layers. Whereas the outer parts, i.e. the workflow and the knowledge store, represent the storage layer, the centered components, i.e. the workflow and the knowledge manager, act as application layer. The lower and upper parts incorporate the presentation layer. Smaller modules, depicted in the main modules, take care of specific functionality. The connections between the core components represent the most important kinds of interaction, data flow or activities. In the following every main module and its functionality is described in detail.

Knowledge Manager: The knowledge manager includes all modules that create, extract and process knowledge. With its access to the Knowledge Store

it will expand a simple document store to a knowledge based document management system. The core components combine- extracted data with the predefined conceptual knowledge of the ontology to extract process relevant information within the context of already available knowledge, provided by former documents and by the workflow manager directly. Its main goal is to provide enactment hypotheses to the Workflow manager.

Knowledge Store: The knowledge store contains the stored data like documents, conceptual knowledge, and instantiated knowledge. The document store preserves all documents which were uploaded by the user and offers access to all kind of documents for the purpose of extraction and visualization. The user is granted access to certain documents if they are necessary to enact a task. Another substantial part of the knowledge store is the ontology, which can be divided into two main parts. On the one hand there is knowledge about general concepts, like processes and documents, called upper ontology, which can be applied universally for this workflow-approach. On the other hand a domain ontology is necessary to provide information about domain concepts and specific workflows, with relations between documents and tasks, which is essential for the overall mapping process that incorporates the semantic integration. The instances determined by the extraction module or from notifications of the workflow manager are stored as triples in the RDF store. Beside the metadata and extracted data from documents the RDF Store also reflects the state of all workflow tasks as well as additional information delivered by the workflow manager.

Workflow Manager: The workflow manager covers the application layer concerning workflow functionality and is responsible for the realization of flexibility by deviation. Core components take charge of workflow execution, deviation detection, constraints validation, and managing the task suggestion. The previously mentioned modules are elucidated in the subsequent section with respect to the use cases.

Workflow Store: The workflow store contains all data concerning the workflows, which involves general workflow knowledge, as well as concrete instantiated workflows and their state. Data structures represent the definitions (workflow prototypes), which is the ideal workflow enactment, as well as the executed, traced instances (workflow instances) including the data context. The workflow prototypes comprise the de jure workflow, which is regarded as ideal flow of activities in the context of SEMAFLEX, and constraints which might be constructed additionally. Constraints describe dependencies or requirements between tasks, which should not or must not be violated. As soon as a workflow starts, a new workflow instance will be created on the basis of the corresponding workflow prototype. The current de jure workflow and the constraints of the workflow instance are duplicated and stored in the workflow prototype, as they might be modified over time. The de facto workflow is built step by step as a simple sequence of activities by means of the logged task enactments. The context contains information about the data which is used or generated during the workflow execution. Thus, the user has access to relevant data, while working on tasks.

User Interfaces: There are different types of users, who interact with the system using different interfaces. There will be a graphical user interface for operational users who complete their work following the workflow. Required documents and information, which are necessary to complete certain tasks, are accessible, either loaded from the document or the workflow store. Furthermore, another graphical user interface provides workflow modelling, which should be enabled for users, who have permission to create or adopt workflow prototypes or associated constraints. The third user interface component, which is called Workflow Monitoring, will offer a central overview of all running and terminated workflows. It combines the data from the workflow store and the knowledge store to provide an overview for the real and deviated processes and the impact of and connection to context relevant knowledge.

4 Use Cases

The following use cases derived from the application field of deficiency management in construction [6] will be used to explain the components of the architecture.

The first use case (see Fig. 3, left side) represents a task enactment triggered by the user through the choice of a proposed task like in a standard WFMS. A significant situation in practice might be a deficiency, which is not caused by the enterprise itself, but has to be reported to a subcontractor for remedial actions. The user chooses to send this notification and thus activates the task enactment. Through interaction with the web application, the user selects a task in the worklist which he wants to execute (see 1). The workflow engine is notified about the enactment and updates the de facto workflow, either creating a new instance with the specific task and possibly a corresponding data object, if a new workflow has been started, or appending the new nodes to an existing de facto workflow. This updated information is stored in the workflow store (see 2a). Furthermore the worklist manager updates the suggested tasks in the web application, which might be executed next by the user (see 2b). The selection of these tasks is based upon the data of the de jure workflow and the currently executed task. Besides, the user interfaces are updated concerning the changed workflow status (see 3a). Additionally the knowledge store, specifically the RDF store, is notified about the changing workflow data (see 2c). Another impact of the task enactment, triggered by the workflow engine is the activation of the module deviation detection. This module logs deviations concerning the task enactment with regard to the de jure workflow. Furthermore the deviation detection sets the de jure and the de facto workflow of the workflow instances into relation.

The second use case (cf. right side of Fig. 3) covers the semantic integration of documents and tasks. In deficiency management this might be a received document, which reports the state of the deficiency, with attached picture. Because of extracted information, like customer, contract site, etc. it can be identified as a deficiency acquisition task. As for each deficiency there is one running workflow and as there might be several deficiencies for one contract site, it might be

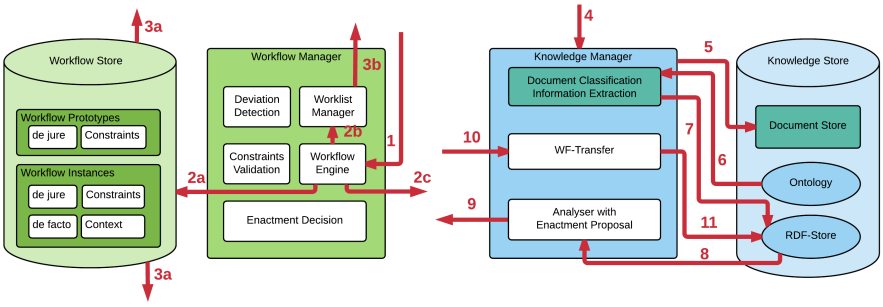

Fig. 3. Processing of Use Case 1 and 2

difficult to determine the corresponding running workflow. Therefore, the user has to manually assign the correct instance. The user transfers a document (see 4), which may be an uploaded pdf-document or a mailed picture, to the web application, which is afterwards processed by the knowledge manager. The first step is to store the document for later accessibility (see 5). Afterwards the module Document Classification and Information Extraction classifies the document based on the definition from the ontology (see 6). The extracted data is stored in the RDF-Store and is linked semantically to available knowledge (see 7). As a new triple is added to the RDF-Store the module Analyser with Enactment Proposal activates an ontology reasoner, which uses modelled inference rules, e.g. property chains of the ontology to check, if this new information can re-enact or start new tasks (see 8). These task candidates are proposed to the workflow manager (see 9). If there are several possible task candidates, the module enactment decision is activated, which involves the user in the mapping process. He might be able to choose the right task instance out of the proposed candidates, as he might be aware of missing information while viewing the corresponding document. An example would be a picture, which represents relevant information. As the extraction module is currently only able to process textual content, visual information is inaccessible and cannot be utilized automatically. If the decision is completed, the chosen task is executed, resulting in an activation of the workflow engine. Constraints are now validated, as the mapping process might have resulted in an undesired state of task enactments. If any constraint is violated, a warning will be send to the user. The warnings are sent to the web application as well as to the workflow monitoring interface. Once an enactment is determined, this new state is send to the knowledge manager (see 10) and subsequently stored in the RDF store (see 11). Such changes will cause another change detection and will start the Analyser again, which results in a loop which ends as soon as no new enactment proposals can be found. The examples reveal that any information might be a contribution to the state of knowledge, whether they come from semantic integrated documents or directly through a user interaction.

5 Conclusion

We presented an architectural concept for a new workflow management approach, which utilizes semantic integration of knowledge-based document and workflow management for the realization of flexibility by deviation. Document classification and information extraction modules are used for the semantic integration, which are ontological-based and use inferencing mechanisms. The workflow component combines imperative and declarative [8] approaches to offer flexibility to the user while preventing him from doing something undesirable. The workflow designer will be able to construct a workflow in an appropriate manner within a range of total flexibility to tight restrictions. Future work will focus on the implementation of the presented approach, followed by an evaluation with business partners, representing the target group of SMEs.

References

1. van der Aalst, W.M.P.: Business process management: A comprehensive survey. ISRN Software Engineering (2013)
2. van der Aalst, W.M.P., Weske, M., Grünbauer, D.: Case handling: a new paradigm for business process support. Data Knowl. Eng. 53(2), 129–162 (2005)
3. Abecker, A., Hinkelmann, K., Maus, H., Müller, H.J.: Geschäftsprozessorientiertes Wissensmanagement. Xpert.press (2002)
4. Bläsius, K., Grawemeyer, B., John, I., Kuhn, N.: Knowledge-based document analysis. In: ICDAR'97, Ulm, Germany, Proceedings. pp. 728–731 (1997)
5. Dengel, A.: Semantische Technologien. Springer Spektrum (2012)
6. Gessinger, S., Bergmann, R.: Potentialanalyse des prozessorientierten Wissensmanagement für die Baubranche. In: LWA 2013. Workshop Proceedings Bamberg, 7.-9. October 2013. pp. 195–202 (2013)
7. Levy, M., Powell, P.: SME flexibility and the role of information systems. Small Business Economics 11(2), 183–196 (1998)
8. Pesic, M., Schonenberg, H., van der Aalst, W.M.P.: Declarative workflow. In: Modern Business Process Automation - YAWL and its Support Environment, pp. 175–201 (2010)
9. Schonenberg, H., Mans, R., Russell, N., Mulyar, N., van der Aalst, W.M.P.: Process flexibility: A survey of contemporary approaches. In: Advances in Enterprise Engineering I, 4th International Workshop CIAO! and 4th International Workshop EOMAS, held at CAiSE 2008, Montpellier, France, June 16-17, 2008. Proceedings. pp. 16–30 (2008)
10. Schonenberg, H., Mans, R., Russell, N., Mulyar, N., van der Aalst, W.M.P.: Towards a taxonomy of process flexibility. In: Proceedings of the Forum at the CAiSE'08 Conference, Montpellier, France, June 18-20, 2008. pp. 81–84 (2008)
11. Supyuenyong, V., Islam, N., Kulkarni, U.R.: Influence of SME characteristics on knowledge management processes: The case study of enterprise resource planning service providers. J. Enterprise Inf. Management 22(1/2), 63–80 (2009)
12. Wenzel, C., Maus, H.: Leveraging corporate context within knowledge-based document analysis and understanding. IJDAR 3(4), 248–260 (2001)