Human-Centered Process Engineering Based on Content Analysis and Process View Aggregation

Sonja Kabicher and Stefanie Rinderle-Ma

University of Vienna, Faculty of Computer Science
Vienna, Austria
{sonja.kabicher,stefanie.rinderle-ma}@univie.ac.at

Abstract. In the context of business process modeling, the transformation of process information elicited from process participants into process models often remains a black box. This paper presents a method that supports the designer to extract a formal business process model from natural language captured in written form in a transparent and traceable way. To-do's of process participants are examined by means of the qualitative content analysis in order to identify essential process elements and to handle different levels of abstraction and labels used for describing operations. Results of the analysis serve as the basis for shaping process view logs which can be aggregated and merged into entire process models. The conduction of the method is described in a case study.

Keywords: Process design, Process views, Qualitative Content Analysis.

1 Introduction

Business process modeling is a widely used and accepted technique for capturing, understanding, and analyzing business processes in organizations. One main goal of business process modeling is to visualize processes (1) how they are actually performed (lived process), and (2) how they should be performed (intended process model). To strengthen organizations' effectiveness (doing the right things) and efficiency (doing things right), internal transparency and awareness of the organizations' business processes is needed. From the workflow perspective, business process modeling plays a key role in the phases design and diagnosis.

In the design phase, designers usually elicit information about the business process e.g., during workshops and interviews with selected employees (e.g. employees assigned to particular roles, middle and top management) [1]. At this stage, the processes are known and lived in the organization and often exist, at least process fragments if at all, simply as mental models in the heads of the member staff. The process designers face the challenge to transform implicit processes into explicit process models. In the diagnosis phase, which takes place some time after the enactment of the (new or improved) workflow system, the task of the designers is to find out how performers of the process work with the new system, as well as where and why they diverge from the intended process [2]. Common practices are, e.g., interviews, observations and process mining of event logs. Results are used to understand

H. Mouratidis and C. Rolland (Eds.): CAiSE 2011, LNCS 6741, pp. 467–481, 2011.

work practice and to refine the process definition, or in other words, to improve the workflow system's support of employees in performing their business.

In this work, we present a technique of identifying lived business processes by means of To-do's of the performers in a process. Coming from a human-centric perspective of business processes and workflows, we argue that a bottom-up approach of capturing and documenting actually performed business processes leads to a more faithful process model to reality than a top-down approach. In a bottom-up approach, performed activities and tasks are collected from the performers of the process and are mapped to the process model. Thus, the process model is implicitly modeled by all process participants, whereas in a top-down approach the process model is mainly defined considering the knowledge about the company's processes available at various management levels, e.g. top-level, middle-, and lower management (depending on the company's organizational structure) and the collected data is optionally complemented with process information elicited from samples of further process participants. The top-down approach seems to be often used in practice when business processes need to be, e.g. defined and redesigned, as it is a question of time and costs to which extend all participants of the process are actively involved in the process elicitation phase.

We assume that personal To-do's implicate information about process fragments that are very close to reality and which include information about activities, tasks, roles, agents, decisions, delegations, time, data, and tools. By *personal To-do's* we understand activities and tasks listed by an individual organizational member in order to organize his or her work. These lists include activities and tasks (complemented with additional data if available) sorted in treated chronological order. We consider To-do lists as a resource of the individual used to perform a particular work task. We use the terms *activity* and *task* according to the definition of BPMN [3].

We use the *qualitative content analysis* as a technique to capture data of the To-do lists in a transparent, structured, uniform and comprehensive way. The content analysis supports the process designer to deal with (1) systematic identification of process elements, (2) homogenization of labels, and (3) granularity of activities. We show how the output of the analysis can be used to model or mine views of individual process participants and to aggregate and merge process fragments resulting from To-do's into an entire business process. By the term "process view" we understand a part of a process (process fragment) that is performed by a single organizational member. The method presented in the paper is evaluated in a case study.

The work is structured as described in the following. The next section presents the Business Process Model Extraction (BPME) Method that allows a transparent and traceable transformation of process information communicated in natural language to a formal process model. In Section 3 the qualitative content analysis and its potential to deal with challenges particularly arising in the context of business process modeling (level of abstraction, labeling and identification of essential process elements) are discussed. Section 4 focuses on the preparation of process view logs based on the output of the qualitative content analysis and the mining of the logs into one entire process model. The case study which was conducted for the teaching process in a real-world setting is presented in section 5. In Section 6 related work is reflected and Section 7 concludes our work.

2 Challenges and Overall Methodology

The elaboration of business process models is usually performed according to a common practice. In general, interviews and workshops are conducted in order to ascertain process information which is than interpreted by the designer by means of, e.g. visual and textual illustrations of the process [4]. So far, the step between information gathering and the presentation of the process models is handled like a black box. In this step, the designer's task is to transform mainly textual descriptions into natural language of process participants (humans) into a formal process model.

Our work concentrated on the analytical and designing step between data collection and the visualization of the process model. Our main research question was how to translate various process elements captured in To-do lists of individuals into a process model in a structured and traceable way. The following challenges are:

- How to extract process elements, like activities, tasks, roles, agents, time, data, tools, and decisions, from To-do lists of single organizational members in a structured and traceable way?
- How to handle different labeling of To-do's?
- How to deal with different granularity of To-do's?
- How to find reference points between the process views (connection, participation and delegation)?
- How to deal with special cases identified in To-do's?

Our approach included the following research instruments and techniques:

- Personal face-to-face interviews to elicit To-do lists from process participants.
- Qualitative content analysis to examine gathered To-do lists in order to identify key process elements and terms, and to transform the To-do's into process view logs.
- Mining of process view logs and the process model.

We propose a reusable semi-structured method that supports the designer to extract process models from text material (e.g. interviews, workshop protocols) in a traceable way. Traceability is supported by the use of the qualitative content analysis to code and categorize the gathered textual information and thus to identify and tag essential process elements. Results of the qualitative content analysis serve as the basis for the development of the process model which can be created by modeling or by transforming the data into process view logs.

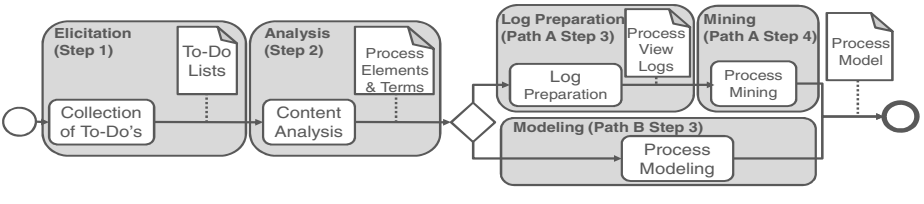

Fig. 1. Business Process Model Extraction (BPME) Method

Fig. 1 illustrates the steps of our method. In a first step, lived processes are elicit from process participants. Second, collected information is analyzed and prepared for further processing. There are two alternative ways of continuing: path A includes the capturing of collected process data in fictive logs (Step 3) which were then mined (Step 4). Choosing path B, the designer continues with manually designing process models. In this work we focus on the analysis (Step 2) and the steps of path A (log preparation and mining). In the following we describe the different steps at an abstraction level. In the Sections 3 and 4 the steps will be discussed in more detail.

Elicitation. The elicitation of process information is an important step as it determines the quality of the gathered material. In our view, an organizational member knows best what activities he or she actually performs in his or her job and how and when this work is done. To-do lists can be used to capture all the things that need to get done in a logical and also chronological order. To-do lists might also represent practical processes of work of single persons. For example, particularly new employees may notice their To-do's for particular scenarios (e.g. "What is usually to do?", and "What is to do in a special case?") during job instruction. Experienced employees often have their To-do lists in a particular context of work in their heads. There are several task management tools that support the organization of To-do's. If used by employees, such tools can be considered as a source of information.

Analysis. The To-do lists are examined with the qualitative content analysis. The method supports the designers to identify key process elements (e.g. activities, tasks, roles, agents, time, data, and tools) and terms (e.g. pre- and post conditions).

Log preparation. The process views are then transformed into process view logs. The logs illustrate personalized views of the process. Different views on the process can be generated by subsuming To-do's of process members with particular characteristics e.g. acting in the same roles, but also members of the whole process.

Mining. In contributions related to process mining, processes are commonly considered to be already performed. The resulting process model is automatically derived from generated logs. These logs include the IDs of each process instance and each event within the process instances. In a human-centric approach of identifying the business process model, as described in this work, processes must not be performed at the time of process mining. The process model is constructed on the basis of To-do lists of roles (humans) that are involved in the process.

3 Qualitative Content Analysis

In this section the qualitative content analysis and its potential in the context of business processes are described. Challenges of analyzing To-do's of process members are discussed, like unique labeling and the level of abstraction of To-do's.

3.1 Qualitative Content Analysis in General

The qualitative content analysis is a rule-guided technique to systematically analyze text [5]. Typically, the content analysis includes the steps: unitization, categorization, and coding [6]. Unitization means to divide the textual material into units of analysis (e.g. sections, thematic units, or syntactical units) [7]. In the categorization step a category scheme is created. The coding step includes the assignment of each unit of

analysis to a category. Ideally, more than one coder performs the content analysis in order to assure inter-coder reliability of findings. There are tools that support the qualitative analysis of text material, e.g. the QDA Miner [8]. Such tools offer functionalities like coding retrieval, coding frequency, and coding co-occurrences. Existing experiments support automation of the analysis in order to reduce the high manual coding effort of large text material [9].

3.2 Qualitative Content Analysis in the Context of Business Processes

In the context of business process modeling, the qualitative context analysis offers a possibility to analyze textual material, e.g. interviews and workshop protocols collected during the phase of eliciting process information, in a structural and traceable way. Traceability is supported by coding text material, for example process views of individuals, or "process stories", and thus to identify, tag, and cluster essential process elements. Some questions that can be answered with the qualitative content analysis particularly in the context of business processes are:

- What are the tasks and activities of a particular process view?
- What tasks and activities can be subsumed according to their meaning?
- What persons and roles are involved in particular tasks and activities?
- What tools are used to perform particular tasks and activities?
- What kind of data is transferred, where, why and how?
- Are there decisions in the process view?
- Are there activities performed by several persons and roles?
- Are there activities in the process view that are delegated?

Results of the qualitative content analysis can be interpreted in several ways, but the final output of our method is always a business process model. The results may be manually modeled (resulting in process models, or process scenarios [1]), or further translated into logs. The latter alternative is described in more detail in Section 4.

3.3 Challenges of Qualitatively Analyzing To-Do's

In this section we address challenges the designers face when they examine To-do's of individual process members that belong to a particular process under investigation. We propose the elicitation of process information by means of To-do lists of individual process members. Several preconditions need to be considered when a qualitative content analysis of To-do's of individual process members is conducted.

Precondition 1: Clearly define the process under investigation.
Precondition 2: Identify at least one process member.
Precondition 3: Collect information: Ask process participants in individual exchange to list their To-do's in the process of investigation. Ask for the usual case and for special cases.
Precondition 4: Capture information. If necessary transform information into text.

The procedure of examining To-do's of individual process members based on the content analysis is described in the following. We discuss the steps specifically important for extracting business process models in more detail in Sections 3.4-3.6.

Step 1: Consider each To-do list of a process member as a case. For analysis use the entire text material.
Step 2: If the roles of the process members are known, group the To-do lists of the same roles and analyze the cases considering their group affiliation. Cases of the same group are analyzed in sequence. Considering the roles during the analysis supports a faster identification of possibly identical To-do's and similar labeling. The designer develops a particular understanding of the context of work. If the roles are not known, the designer may identify other hints towards a particular ordering in analyzing the cases, e.g. data transfer or delegation to other persons mentioned in the To-do lists which allows a sequential analysis of the process flow views.
Step 3: Define the unit of analysis. The unit of analysis may be, e.g. a set of semantically related To-do's, one To-do (thematic unit) or a sentence. To face the challenge of a uniform level of granularity of the To-do's, compare with Section 3.5.
Step 4: Paraphrase the units of analysis. Reduce text to the key essence of the unit of analysis. In this phase, the identification of process elements takes place (compare with section 3.6).
Step 5: Label or code To-do's. There are mainly two alternatives to perform labeling (compare with section 3.4): inductive or deductive labeling.
Step 6: Evaluate results. For example, list all units of analysis with the same code, analyze the codes' frequency, the number of cases in which the codes were found, etc.

3.4 Homogeneity of Activity Labels

The labeling of activities as recorded by process members might vary. Synonyms and homonyms need to be handled in order to design an understandable process model [10]. The content analysis supports two ways of determining unique labels [5]: the inductive and the deductive elaboration of labels.

Inductive elaboration of labels. Performing the inductive elaboration of labels means to derive labels from the text material. The labels are elaborated in a process of generalization. This procedure is used to attain a close reflection of the language and terms used in the material. Labels should reflect the content of the text material as good as possible. The inductive elaboration of labels can be complemented by the frequency measure of terms used in the material [11]. A visual representation of the terms' frequency offer word clouds. When the frequency measure is used to label operations misinterpretation may arise due to technical terms used by persons working in different roles and unequally long text segments. Therefore, designers always need to be aware of the content of the text material.

Deductive determination of labels .The deductive determination of labels implies that the labels are defined a priori. This means that the labels are specified before the text material is analyzed. In this case, terms are considered that are, e.g., commonly used in practice, in a particular sector or in a company. Formulations actually used by process members are not explicitly considered in the phase of labeling process steps.

3.5 Dealing with Different Granularity

The descriptions of the To-do's listed by the process members might vary in their granularity. *Summarization* is a technique that supports the structured and meaningful aggregation of activities according to their semantic relation [5].

The goal of the summarization is to reduce material towards a particular level of abstraction and thus supports the designer to find related tasks and activities in the text material. The technique includes a stepwise generalization of the text material. In a first step, the text segments are paraphrased which includes the omission of decorating text passages, the translation into a consistent level of language and the

transformation of the text segments into a short form. The second step includes the generalization of the paraphrases towards a specified level of abstraction (aggregation of tasks). A second generalization can be done in order to subsume related activities to process fragments (aggregation of activities).

3.6 Identification of Process Specifications

The identification of process elements is appropriate in the paraphrasing step. In this step the unit of analysis is reduced to its key essence. The key essence of a To-do is the performed task or activity, and is captured in the code by a meaningful noun and a verb [12]. The labels serve as labels of the tasks and activities in the process model. Reference points important for process view aggregation are listed in the following. We suggest adding reference points as variables to the cases. Variables specify the properties associated with the case.

- Hints to data transfer (data_input and data_output) or events (e.g. eMail, Mail, PhoneCall) between two or more process participants (connection). *E.g. To-do of the AdministrativeCooperator: "Mail lecturing contract to lecturer". To-do of a Lecturer: "Receipt of the lecturing contract in my post office box".*
- Hints to group tasks and activities (participation). *E.g. To-do of the Lecturers and the TeachingCoordinators: "Participation in coordination meeting".*
- Hints to decisions. *E.g. To-do of a Lecturer: "Either I contact the Administrative-Cooperator if I want to book the lecture hall X or I contact the Secretary if I want to book the lab."*
- Hints to delegation. It can be assumed that the To-do's are performed by the process member who listed the To-do's. Otherwise a hint to another person is given. *E.g. To-do of a ModuleCoordinator: "A lecturer of our lecturer team books a lecture hall for all of us".*
- Hints to tools used. It might be the case that process participants rather mention the particular tool name (e.g. Fronter) than the general term (e.g. learning platform). *E.g. To-do of a Lecturer: "Then I enter the grades of the students into iswi."*
- Hints to reoccurring activities. *E.g. To-do of a Lecturer: "Reoccurring task: conduction of the units".*
- Time notes. We assume that To-do list already reflect a sequential order of the To-do's. The first To-do mentioned in the list is the first To-do that is performed by the person in the process under investigation. Explicitly mentioned time notes may be vague but support the designer to sequentially order tasks and activities when subsuming all tasks and activities of the various process members. *E.g. To-do of a Lecturer: "At the day of the unit I print the attendance list.", or "Before the semester starts I plan the course".*
- Name of the agent. The name of the process member is captured as well.

Furthermore, process members may mention comments that refer to activities performed by other process members or to the general process. *E.g. Mentioned by a Lecturer: "The AdministrativeCooperator books the lecture hall in the system".* Such information might be important for the designer when he or she composes the

logs. We suggest to label such statements as *comments* to recall the data if necessary. In order to be able to transform results of the qualitative content analysis into process view logs, post conditions of the analysis are listed in the following.

Post condition 1: *Each To-do of each person is considered as a case. A case is the unit of analysis.*
Post condition 2: *Each case is labeled. The labels reflect the key essence of the To-do. The labels illustrate the labels of tasks and activities in the process model.*
Post condition 3: *A precise coding guideline is elaborated that support comprehensibility and traceability.*
Post condition 4: *Each case is complemented by variables. Variables necessary for process modeling or log preparation are: Agent (name of the process member), Agent Position (role of the process member), Tool (tool name), DataInput (data name), DataOutput (data name), OrganizationOfWork (if delegation name of process member, else (direct work) no entry), Connection (person name or role label), Participation (person name or role label), Decision (String), Recurrence (Boolean), Event (event type), TimeNote (String). Further variables might be, e.g. Granularity (GranularityLevel), FirstGeneralizationLabel (String), SecondGeneralizationLabel (String).*
Post condition 5: *The beginning and the end of the entire process are identified (e.g. from time notes or from the semantic context).*

According to the BPME method, the designer can choose two alternative paths after analyzing the data. The process view models and the entire process model can be modeled either (1) manually, (2) using a tool that supports scenario-based process modeling like GRETA [1], or (3) the designer transforms the captured process view information into process view logs and entire process logs. In the following we focus on the latter. We argue that the extraction of the business process model out of process view logs offers a structured and traceable way of illustrating the identified process model. We present guidelines for preparing logs based on results of the qualitative content analysis of process member To-do's in the next section.

4 Preparation of Logs and Process Mining

We want to know how the processes of the individual process members as well as the whole process look like. In order to capture all possible variants of performing process paths, process view logs for each process participant are prepared for mining. The preparation of the logs is based on the results of the qualitative content analysis of To-do lists of process members. We suggest the designer to choose the level of abstraction that is included in the original To-do lists (results of the paraphrasing in the qualitative content analysis), as the To-do's reflect the "mental process steps" of each process participant. For each To-do list a process view log is elaborated. One output of the log preparation step are logs that reflect To-do's of a process member in a particular thematic process (e.g. sale process of a particular organization). The process view logs reflect process fragment scenarios. The most challenging task for the designer is now to transform reference points (e.g. operations that are connected to a message event between process participants (send and receive), operations that are performed by several process participants at the same time, like meetings, and time notes) into concrete (artificial) start- and end date and time stamps. The procedure of considering reference points in the process view logs is described in the following.

Step 1: Highlight reference points that are bounded to time. Such reference points are the variables Connection, Participation, DataInput, DataOutput, Event, Recurrence, Tool and TimeNote. Connection and Participation: It may be the case that there are links identified to persons or roles whose To-do lists were not gathered in the elicitation phase (compare with Fig.2). This is an indication for the designer that not all process participants were considered in the ascertainment of the process.

DataInput, DataOutput and Events: A question that need to be answered by the designer is if events, such as "send document x to person B" and "receive document x from person A" are considered in the process models as operations. The overall question to be answered in the first step is for what purpose is the process model designed. If it is designed to illustrate the actual process as perceived by the process members, than events as illustrated in the example may be integrated as operations (particularly if the event of person B "receive document x from person A" triggers the process of person B). If the process model is designed as a basis for the implementation of a workflow system, then such events may not be considered in the process model, as in a workflow systems all data material is available for process participants.

Step 2: Aggregation of tasks. If the level of abstraction considered by the designer is identical to the abstraction level of the original To-do's, then there might be detailed operations across process views (e.g."Writing on PhD thesis", "Exchange with mentor", "Enrolling in creative writing workshop") less detailed operations probably implicitly subsuming the detailed operations (e.g. "Elaboration of PhD thesis"). In such a case the designer has to decide if, e.g. such creative tasks should all be considered in the process model. An option for the designer is use by default the higher level task and to additionally consider these detailed operations that were (n/2+1)-times mentioned across process views (n=number of process participants).

Step 3: Set order of execution. Transfer explicit time notes mentioned in the To-do list and transform them into concrete date and time. Estimate the duration of the operation. If there is no explicit time note mentioned for a particular To-do, then first check time notes of reference points in To-do lists of other process members. Choose these time notes as support to set a concrete date and time and consider at the same time already set time specifications of other To-do's in the list. The sequential ordering of the To-do's in the list should be reflected in the time stamps. If necessary, the designer may also consider Comments for specifying time and date of operations, or directly contact the respective process member.

Connection and participation: Adjust time stamps according to connection and participation hints. A connection considers work- or data transfer to another process member. Participation considers operations that are collectively performed (e.g. meetings).

Recurrence of tasks. If there are hints that a tasks is executed several times (loop) then the designer has to decide if the loop is considered in the log as loop or if each execution of the operation is captured as a particular step in the process model.

Step 4: Check the agent of the task. If there are indications that an operation is delegated to another process participant, than the operations are still considered as operations of the delegator. In the field "Agent" the name of the delegatee is entered. The tasks of the delegatee remain as tasks of the process member in his or her process view log.

Step 5: Set number of process instances (cases). The number of process instances in a process view is determined by the number of the decision paths. If there are > 1 decisions, the product of the decision paths determines the maximum number of process instances. So, if there is decision n including m paths and decision d including e paths considered in one process view, than the designer needs to build a maximum of m x e process instances.

Step 6: Explicate conditions and rules. Particular decision paths must not or cannot be considered in the same process instance. Therefore, rules for joining decision paths (if decision ≥ 1) need to be defined. The designer keeps overview when the operations are organized into operations that are always executed and operations that are only executed if a particular condition occurs. Thus, the designer obtains "operation blocks" that can be merged into one process instance by considering merging rules and condition.

Step 7: Assemble logs. Illustrate common and alternative process paths as process view instances. Consider thereby the entire process view (from start to the end) as one process instance. If there are decisions included, then transform the common and alternative process paths into different instances.

The next task is to prepare the log instances of the entire process including all process views. The output is a process model of the entire process that captures possible variations of process paths by considering the allowed decision combinations. For preparing such logs, the following preconditions need to be considered.

Precondition 1: The process that is analyzed is clearly defined and has a common thread (e.g. in a sales process, the common thread is the order on which the process participants work on throughout the whole process).

Precondition 2: The process views are already transformed into process view logs. The process view logs reflect common and alternative process paths.

The steps necessary to merge process view logs in such way that they reflect the whole process with its alternative paths are presented in the following. In addition to the mentioned steps below, exchange with the process participants during workshops might support the analyst to get an immediate "big picture" of the whole process and to reduce errors and misinterpretations in the whole process model.

Step 1: Determine the number of process instances (cases). The process instances illustrate possible entire process path. Elaborate per process instance which decision path of what process member is considered. Consider rules and conditions when assembling the process view instances.

Step 2: Bring the process fragment scenarios together and sort the operations according to the time stamps.

In this work our intension is not to present a new mining algorithm for process view mining. Our goal is rather to discuss the purpose and benefits of mining prepared logs. Mining of the process view logs and the entire process supports the designer in constructing and visualizing process models from data analyzed and prepared in a structured and traceable way. The mining of each single log of the individual process

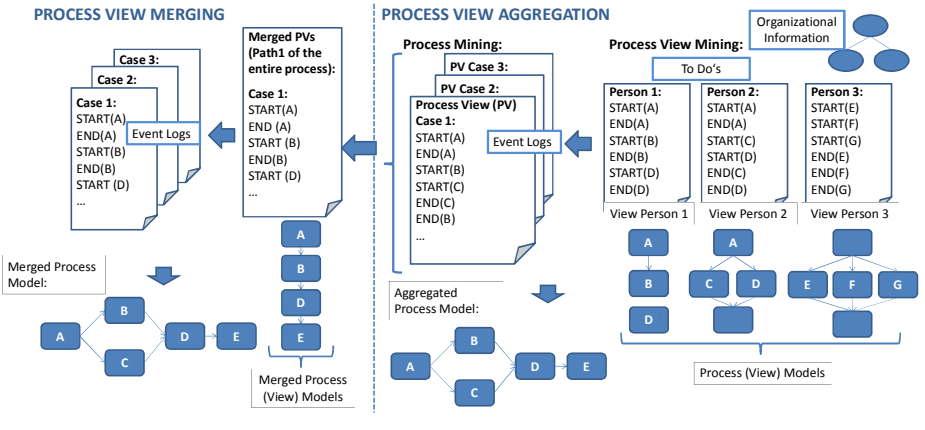

Fig. 2. Process view aggregation vs. merging

participants offers for each process participant a personalized process view. The designer obtains manageable process views. The process views can be aggregated or merged into one process model. Aggregation of process views means to consider each process view as an instance of the entire process in the logs, whereas merging of process views means to consider each process view as a part of an instance of the entire process in the logs (compare with Fig. 2).

5 Case Study

In this section we present a case study in which we applied the BPME method to extract the process model of teaching from To-do's of selected process participants at the Faculty of Computer Science, University of Vienna. For our investigation we considered both the administrative and creative process of teaching. A bottom-up approach was pursued to find out what activities the participants of the process perform to enable and support teaching at the faculty.

5.1 Collection of To-Do's in the Teaching Process

In face-to-face interviews selected persons working in key positions of the teaching process were asked to list their activities in the process under investigation similar to a scheduled To-do list. The interviews were conducted at the Faculty of Computer Science at the University of Vienna in October 2010. 12 persons participated in the survey. The selected persons were directly asked to participate in the study. The participation in the survey was voluntary. A question of the semi-structured interview was to specify the own role in the process of teaching. Some interviewees acted in several roles, e.g. the role of the module coordinator and the lecturer. The interviewer asked for the To-do's of each role successively and recorded the To-do list per role in separate protocols. Altogether, the activities appropriate to the following roles in the teaching process were collected (the cases illustrate persons): administrative cooperator (case 1), director of study program (case 2), lecturers (cases 3, 4, 5, 6, 7, 8, and 16), lecturer team coordinator (case 9), module coordinators (cases 10, 11), secretary (case 12), teaching coordinators (cases 13, 14), and technical staff (case 15).

5.2 Content Analysis of To-Do's, Log Preparation, and Mining

The protocols of the interviews were examined with the content analysis. Most of the activities were described in full sentences rather than by means of cues that are more common in personal To-do lists [13, p. 736]. The text material comprised 16 protocols and a total of 7,814 words. A case comprised between 65 and 1077 words. We conducted the content analysis with the demo version of the QDA Miner [8]. Analyzing functionalities, like coding frequency allow recognizing reference points among process participants. Fig. 3 illustrates an extract of the coding frequency analysis. "Inquire date and lecture hall", "Book date and lecture hall", and "Check scheduling conflict" are aggregations of more detailed tasks mentioned as To-do's. These aggregated tasks were subsumed to "reservation of lecture hall". "Inquire date and lecture hall" includes two reference points (connections) among process participants.

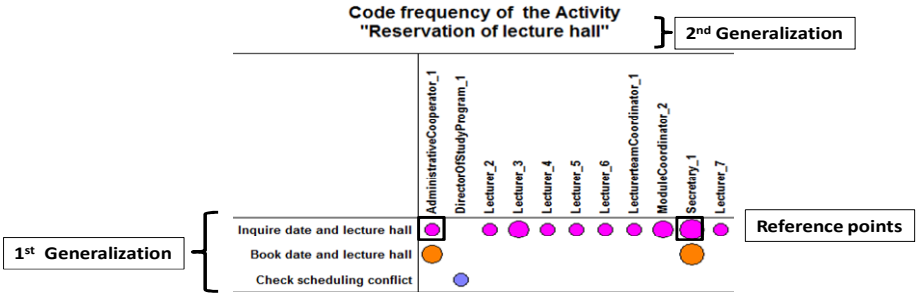

Fig. 3. Reference points among operations

Whereas the lecturers, module coordinators and lecturer team coordinators request for lecture halls, the administrative cooperator and the secretary receive the requests and book the halls.

Afterwards, each single To-do of the 399 To-do's was coded at the lowest possible level of abstraction already included in the To-do by paraphrasing. 195 labels reflecting the key essence of the To-do's were inductively derived. 77 hints to connections among process participants, 39 hints to operations with participation of several process participants (e.g. meetings), 49 hints to decisions (each operation of the alternative path was marked as a decision hint), 40 hints to events (e.g. eMail, phone calls, or messages), 12 hints to delegation, and 43 time notes could be identified. 8 tools could be extracted out of the To-do's as well as 30 different data material (e.g. "teaching scheduling sheet", "numbers of exams of the last semester", documented course schedule") could be identified.

In a next step we elaborated the process view logs per process participant (per role). The main challenges were (a) to define the artificial time stamps and (b) the inclusion of all possible combinations of process paths as these issues were the critical points for the quality of the mining result. Based on the number and conditions of the decisions we elaborated possible process paths per role and captured each possible path as a case (process instance) of each process view. The alpha mining algorithm was used for the transformation of the individual process view logs into process view models. We elaborated process logs in which each case illustrates an entire path of the entire process, and considered the course (with unique course ID) as the "common thread" throughout the process. To organize all the possible paths of the process per case the selected paths of the process views were documented. Log segments were compound to a possible way of the process. The heuristics mining algorithm was used to analyze the entire process logs. Fig. 4 illustrates a comparison of the result of applying the heuristic mining algorithm in ProM [18] of the collected, analyzed, and merged To-do's to a process model (left and middle part) and a manually modeled process segment in which BPMN notation was used (right part). The BPMN model segment was designed by the process-aware designer before the To-do's of the process participants were analyzed and mined. Both segment models gather the same activities (prepare, conduct, and post process unit).

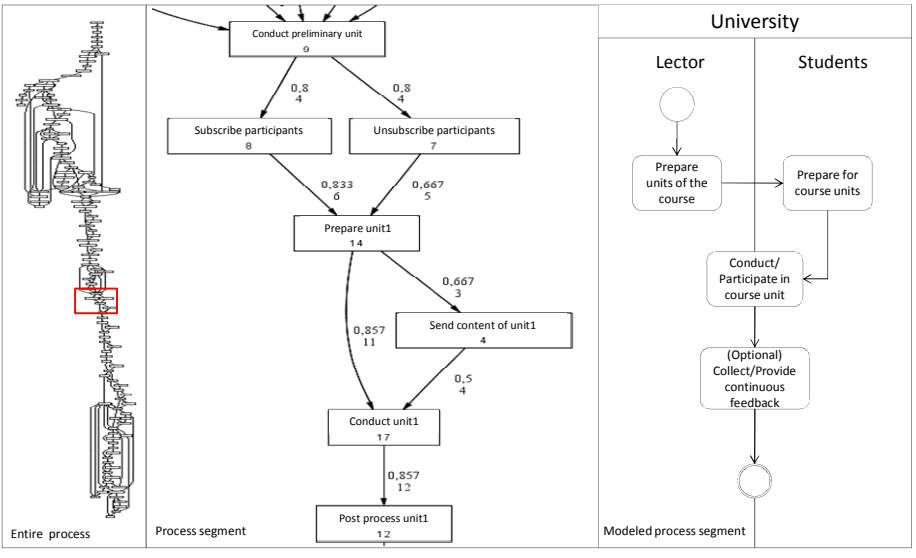

Fig. 4. Comparison of the merged process logs analyzed with the heuristics miner and a manually designed BPMN process model

Whereas the BPMN model provides more details about the participating roles due to its notation, the process segment of the mined To-do's offers information about how often the operations were considered in the logs (and thus in the To-do's of the process participants) and provides a more precise insight into the actual operations (e.g. several process participants explicitly mentioned the conduction of the preliminary unit, operations that were performed to fix the participants of the course, and operations among lecturers before a course unit was conducted).

6 Related Work

Process design is one of the main phases of the business process life cycle. There are basically two different approaches of process design that can be distinguished. First, business process (re-)engineering refers to the systematic process exploration based on methods such as interviews and the subsequent modeling of as-is processes, followed by an optimization of these processes (to-be processes). Much work has been spent on questions such as process modeling notations [17]. Only few approaches tackle process redesign [4]. Although there are known common practices to elicit process models, like questionnaires and workshops, no systematic approach for process exploration and design based on activities and tasks representing practical processes of work of single persons sorted in treated chronological order (To-Do's) has been presented yet. We utilize existing process mining algorithms [18] for our work not only to mine instances of the whole process but also operations that are performed by a single agent of the process (process view logs).

Commonly, the challenge of finding an appropriate level of abstraction is faced with elimination and aggregation [14] [15]. A technique considering the semantic relation between activities is presented in [16]. The meronymy-based aggregation uses business process domain ontologies for abstraction. To enable aggregation, process model activities are matched to an ontology. The technique offer transparency and traceability in determining a particular abstraction level. However, preconditions of this aggregation are a predefined ontology and the use of domain specific labels.

[1] proposes a scenario-based modeling paradigm which considers a process model as a set of scenarios. While [1] considers all process views and divide them into process fragments, we consider each process view of each process participant.

7 Conclusion

As so far the procedure of process models extraction from process information elicited from process participants was handled like a black box, our main research interest of this work was to find out how to translate various process elements captured in To-do lists of individuals into a process model in a structured and traceable way. We proposed the Business Process Model Extraction (BPME) method that supports a transparent way of transforming process information in natural language into formal process models. We discussed the potential of the qualitative content analysis in dealing with differing levels of abstraction, heterogeneous labeling and the identification of essential process elements. We presented guidelines for using the qualitative content analysis as a designer's tool to extract process information from To-do lists of process participants in a structured, documented and traceable way. In order to support transparency and traceability up to the final process model, we presented guidelines to transform results of the content analysis into view process logs and discussed two opportunities that lead to an entire process model by mining: the aggregation and the merging of process views. In the case study we showed that the BPME method could be successfully performed. The main challenges faced were the determination of artificial time stamps in the process view logs that were crucial for process mining and the elaboration of the possible process paths based on the decision paths in each process view. The use of the BPME method offered the following outputs: (a) a documented and traceable procedure of transforming To-do lists into the process model, (b) illustrative personalized process view models, and (c) an illustrative process model reflecting the To-do's of the process participants. In future work we will extend our work on process view mining. Furthermore, we will offer more details about the challenges of the elicitation step of the BPME method, and the validation of the resulting models that should support our approach.

References

1. Fahland, D., Weidlich, M.: Scenario-based process modeling with Greta. In: La Rosa, M. (ed.) Proceedings of BPM Demonstration Track. CEUR-WS.org, Hoboken, vol. 615 (2010)
2. Hammori, M., Herbst, J., Kleiner, N.: Interactive workflow mining - requirements, concepts and implementation. Data & Knowledge Engineering 56, 41–63 (2006)

3. OMG: Business Process Model and Notation (BPMN) Version 2.0 (2010)
4. Reijers, H.A.: Process Design and Redesign. In: Dumas, M., van der Aalst, W., ter Hofstede, A.H.M. (eds.) Process Aware Information Systems: Bridging People and Software Through Process Technology. Wiley-Interscience, Hoboken (2005)
5. Mayring, P.: Qualitative Inhaltsanalyse. Beltz, Weinheim (2003)
6. Srnka, K., Koeszegi, S.: From Words to Numbers: How to Transform Qualitative Data into Meaningful Quantitative Results. Schmalenbach Business Review 59, 29–57 (2007)
7. Strijbos, J.-W., Martens, R.L., Prins, F.J., Jochems, W.M.G.: Content analysis: What are they talking about? Computers & Education 46(1), 29–48 (2006)
8. Provalis Research, http://www.provalisresearch.com/
9. Nastase, V., Koeszegi, S., Szpakowicz, S.: Content Analysis Through the Machine Learning Mill. Group Decision and Negotiation 16, 335–346 (2007)
10. Rinderle-Ma, S., Reichert, M., Jurisch, M.: Equivalence of Web Services in Process-Aware Services. In: IEEE 7th International Conference on Web Services, pp. 501–508. IEEE, Los Angeles (2009)
11. Derntl, M., Neumann, S., Griffiths, D., Oberhuemer, P.: ISURE - Report on usage and recommendations of specification for instructional model, ICOPER Deliverable D3.2 (2010)
12. Mendling, J., Recker, J., Reijers, H.A.: On the Usage of Labels and Icons in Business Process Modeling. International Journal of Information System Modeling and Design 1, 40–58 (2009)
13. Bellotti, V., Dalal, B., Good, N., Flynn, P., Bobrow, D.G., Ducheneaut, N.: What a To-Do: Studies of Task Management Towards the Design of a Personal Task List Manager. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 735–742. ACM, New York (2004)
14. Polyvyanyy, A., Smirnov, S., Weske, M.: The Triconnected Abstraction of Process Models. In: Proceedings of the 7th International Conference on Business Process Management, pp. 229–244. Springer, Ulm (2009)
15. Bobrik, R., Reichert, M., Bauer, T.: View-based process visualization. In: Proceedings of the 5th International Conference on Business Process Management, pp. 88–95. Springer, Brisbane (2007)
16. Smirnov, S., Dijkman, R., Mendling, J., Weske, M.: Meronymy-Based Aggregation of Activities in Business Process Models. In: Parsons, J., Saeki, M., Shoval, P., Woo, C., Wand, Y. (eds.) ER 2010. LNCS, vol. 6412, pp. 1–14. Springer, Heidelberg (2010)
17. van der Aalst, W., ter Hofstede, A.H.M., Kiepuszewski, B., Barros, A.P.: Workflow patterns. Distributed and Parallel Databases 14, 5–51 (2003)
18. van der Aalst, W., Reijers, H.A., Weijters, A., Dongen, B.F., Alves de Medeiros, A.K., Song, M., Verbeek, H.M.W.: Business process mining: An industrial application. Information Systems 32, 713–732 (2007)