



International Journal of Engineering Research and Science & Technology

ISSN : 2319-5991
Vol. 2, No. 1
February 2013



www.ijerst.com

Email: editorijerst@gmail.com or editor@ijerst.com

Research Paper

INTERACTIVE TEMPORAL MINING OF WORKFLOW LOGS

Yasmin Shaikh^{1*} and Sanjay Tanwani^{2*}

*Corresponding Author: **Yasmin Shaikh**, ✉ yshaikh.scs@dauniv.ac.in

Data mining also known as knowledge discovery is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cut costs, or both. The data mining techniques used to analyze the business processes are called Process Mining techniques. The basic idea is to extract knowledge from event logs recorded by an information system. Workflow is the record of who does what and when. Process Workflow Management Systems aims at improving the efficiency within the business by automating many of its processes. Workflow mining is the process of analyzing the workflow event logs to produce useful knowledge about a process. The workflow mining techniques extract the frequent workflow patterns. The timestamps contained in process logs are used to define the sequential ordering of the performed tasks. These models reveal sequential patterns of tasks but do not reveal the temporal distance between the tasks. To overcome this limitation, time is explicitly included in extracted patterns i.e. temporal mining is used. In this work, a novel technique is developed for the analysis of process workflow logs using the Temporal Mining and Information Control Net (ICN) workflow modeling language. Such a technique is useful in analyzing process logs in all the process perspectives including the temporal component revealing the temporal distance between the events. Using the result of the analysis, the management can streamline the responsibilities of each employee and clarify the roles of every employee and machine within the process. Also, the company can easily identify where improvements can be made to increase efficiency and to improve the quality of the product or service. By doing this, the decision makers would be able to eliminate problems such as “bottlenecks” and “redundancies” within the system.

Keywords: Workflow Mining, temporal mining, Information Control Net modeling language

INTRODUCTION

Data mining is extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) information or patterns from data in large

databases. Data mining is the core of knowledge Discovery in Databases process. Workflow is the record of who does what and when. Workflow is a term used to describe the tasks, procedural

¹ International Institute of Professional Studies, Devi Ahilya Vishwavidyalaya, Indore.

² School of Computer Science & IT, Devi Ahilya Vishwavidyalaya, Indore.

steps, organizations or people involved, required input and output information, and tools needed for each step in a business process.

Workflow management systems assist in execution, monitoring and management of a process. It is a system of overseeing the process of passing information, documents, and tasks from one employee or machine within a business to another. It also helps businesses find ways to improve their production or service process.

Workflow Mining

Workflow mining is defined as the rediscovery of an explicit workflow model given a workflow event log. In (Michele Belingerio *et al.*, 2009), the rediscovered workflow model can be then compared to the original control flow model to find the differences between how work is perceived to be done and how work is actually being done. This comparison is called Delta analysis. This comparison is helpful in performance analysis; finding out whether the workflow model is better than original model.

The workflow analysis is the tool that allows the organization to streamline its processes and realize the associated savings. For example, an insurance company could use a workflow automation application to ensure that a claim was handled consistently from initial call to final settlement. The workflow application would ensure that each person handling the claim used the correct online form and successfully completed their step before allowing the process to proceed to the next person and procedural step.

Temporal Workflow Mining

Temporal Mining is the discovery of interesting patterns from the data with some temporal component in it. Workflow mining aims at finding

the control flow, i.e., sequence of tasks being performed. The temporal information in the log is used only to order the tasks. But the temporal information associated with logs in the form of timestamps can help in distinguishing among different temporal behaviors.

Process Perspectives

Process mining research is concerned with the extraction of knowledge about a (business) process from its process execution logs. Process mining strives to gain insight into various perspectives, such as *functional*, *behavioral*, *informational*, *organizational* and *operational*.

The functional perspective of workflow defines what has to be done; the task. The behavioral perspective of workflow defines when tasks are done, i.e., control flow of a process. The informational perspective of workflow defines which data is processed by a process, i.e., data flow of a process. The organizational perspective of workflow defines who performs the task.

The operational perspective of workflow defines how a task is done, what are the resources required to complete a task.

This paper presents a novel technique for analysis of workflow data using TAG and ICN workflow modeling language which can be used for identifying critical points within the processes needing improvement for efficiency and quality. In Section 2, we present a literature review. In Section 3, scope of work is defined. In Section 4, research methodology is discussed which illustrates the use of ICN modeling language with temporal mining of workflow logs. In Section 4, a case study is presented based on the proposed technique. In Section 6, results are discussed. Finally in Section 7 and 8 we present our conclusions and area of further work.

LITERATURE REVIEW

In Giannotti *et al.* (2006) propose notion of temporally annotated sequential patterns that are an extension of the sequence mining paradigm, where each transition in a sequential pattern is annotated with a typical transition time derived from the source data. Frequent Sequential Pattern (FSP) mining deals with the extraction of frequent sequences of events from datasets of transactions; those, in turn, are time-stamped sequences of events (or sets of events) observed in some business contexts: customer transactions, patient medical observations, web sessions, trajectories of objects moving among locations. Time in FSP is typically used as a user-specified constraint for the purpose of either preprocessing the input data into ordered sequences of (sets of) events or as a pruning mechanism to shrink the pattern search space and make computation more efficient. In either case, time is forgotten in the output of FSP. Sequential pattern mining is extended in order to derive relevant/useful temporal information from input data. A form of sequential patterns is introduced that is annotated with temporal information representing typical transition times between the events in a frequent sequence. Such a pattern is called Temporally-Annotated Sequence (TAS) in short.

TAS is based on the combination of two concepts: frequency and temporal similarity. For instance, we use a TAS $a \rightarrow b$ to denote the fact that there is a sufficient number of sequences in the source dataset that contains $a \rightarrow b$ as a subsequence and that are temporally similar to $a \rightarrow b$, i.e., the time for getting from a to b is close to t . Therefore, a TAS $a \rightarrow b$ indicates that t is a typical time for getting from a to b .

In Giannotti *et al.* (2006) a new algorithm for mining frequent TAS, that is efficient and correct and complete w.r.t. the formal definition of TAS is proposed. A new way for concisely representing sets of frequent TAS's, making them readable for the user is also proposed in the paper.

In Aubrey (2006), Comprehensive Workflow Mining is proposed to broaden the scope of workflow mining. The paper introduces a concept of Information Control Net (ICN) workflow modeling language. Also, mapping of the mathematical model to graphical model is presented in the paper. The proposed model includes all the five process perspectives namely: *functional*, *behavioral*, *informational*, *organizational* and *operational*.

SCOPE OF THE WORK

The proposed work covers the workflow mining. It is intended to derive temporal patterns from workflow, which may have relevance in several applications such as:

- i) In web log analysis, different categories of users (experienced vs. novice, interested vs. uninterested, robots vs. humans) might react in similar ways to some pages, i.e., they follow similar sequences of web access—but with different reaction times.
- ii) In medicine, the relationship in time between the onset of patients' symptoms, drug consumption, and response to treatments.
- iii) In transportation, the time of traveling between cities using different routes may help in optimizing the transportation time.

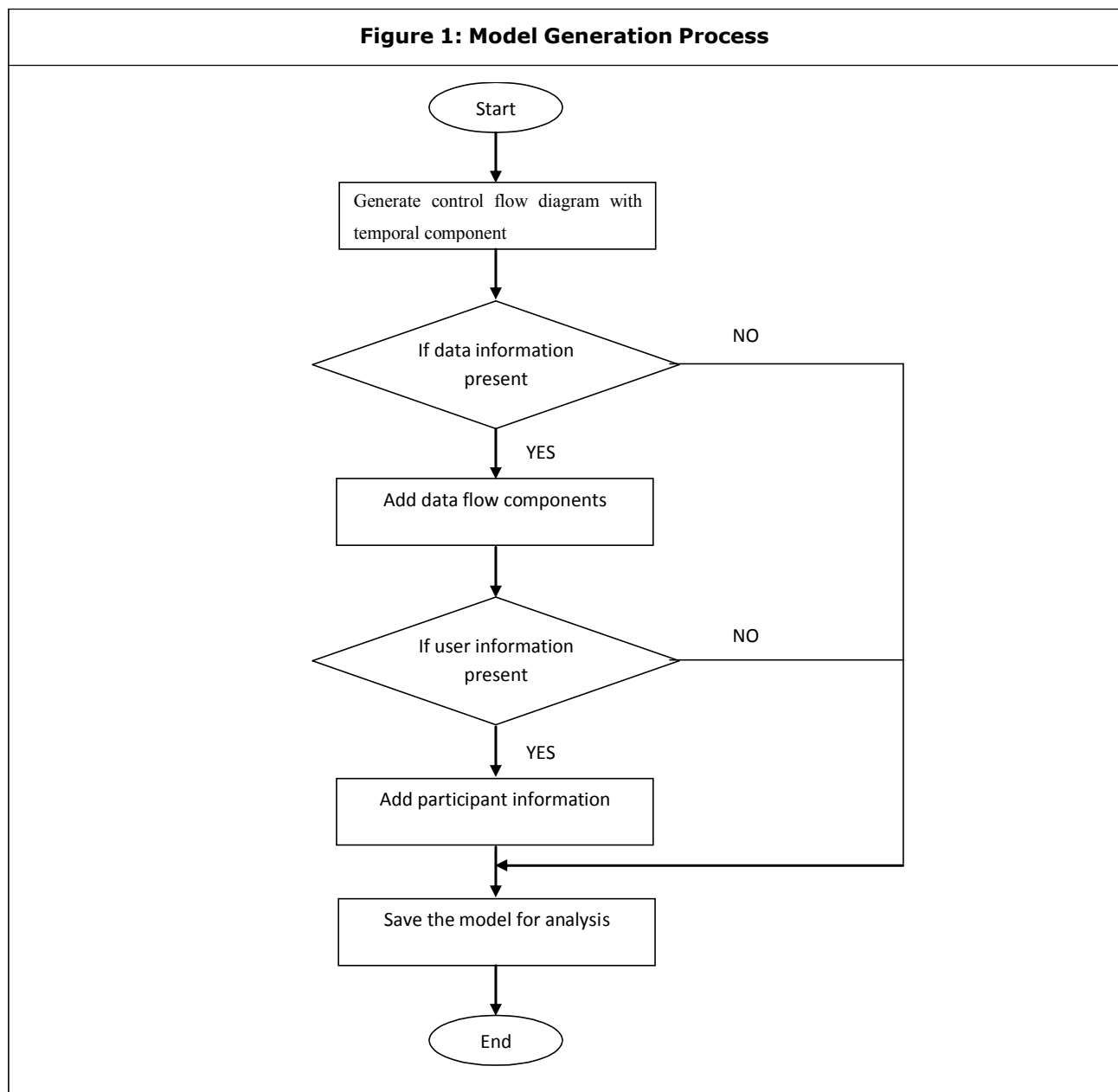
Thus the proposed work is useful in finding interesting patterns in processes where time is crucial issue.

RESEARCH METHODOLOGY

A generic process mining approach incorporating TAS based mining and comprehensive workflow mining will be used as depicted in Figure 1. First of all control flow diagram with temporal component is generated. The diagram exhibits sequence of events and the temporal distance between the events. If source and destination of data are

included in the log then the data stores' information is also included. Also, if the performer of the task is mentioned in the log, the participants are also depicted in the diagram.

The patterns are extracted including all the five perspectives of process and the time can be explicitly included in the extracted knowledge, thus temporal information becomes the most



important outcome of the analysis process. It could help in anomaly detection and highlighting situations in which some entities are faster (or slower) than others in performing certain tasks or situations in which resources take too much time to perform certain tasks. This makes it possible to distinguish between apparently identical process executions that are performed with different transition times between consecutive tasks.

The workflow analysis based on proposed technique illustrates how we should be doing things and shows the changes needed to maximize resources (both personnel and material).

Behavioral Perspective with Temporal Component

In Aubrey *et al.* (2006), Information Control Net (ICN) workflow modeling language, the control flow model is used to specify Behavioral perspective of workflow. An ICN control flow model with temporal component can be defined as a pair (T, d) where:

- T is a finite set of tasks.
- C_t is the time component that represents the temporal distance between the completions of sequential tasks.
- $\delta: \delta_i \cup \delta_o$ where,
 - $\delta_o: T, C_t \rightarrow \mathcal{P}(T)$, C_t is a multi-valued mapping of a task to a set of sets of successor tasks with the temporal distance between each pair of tasks.
 - $\delta_i: T, C_t \rightarrow \mathcal{P}(T)$, C_t is a multi-valued mapping of a task to a set of sets of predecessor tasks with the temporal distance between each pair of tasks.

Sequential Pattern Using Reverse Reengineering

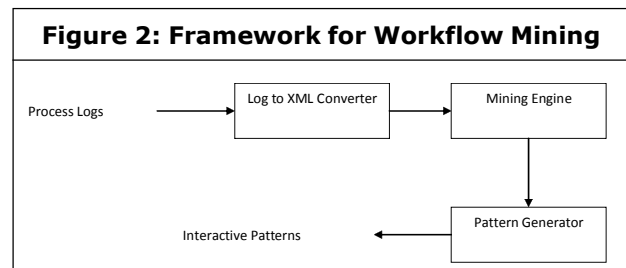
Construct a graphical representation of the log using reverse reengineering (Michele *et al.*, 2009). It results in a sequential pattern with temporal annotations.

Interactive and Iterative Interpretation

Allow the user to interact with the frequent sequential patterns for analyzing them. The user interaction procedure allow the user to extract a subset of the possible transformations, select particular branch of graph to explore, backtrack to previous node of the graph, view possible alternatives and select the best alternative in order to optimize the process execution.

Framework for Mining Process Logs

A framework is developed for mining workflow logs as depicted in Figure 2. The workflow log is converted into XML format (Rubin *et al.*, 2007). The XML file is given as input to mining engine. The mining engine uses an algorithm prefix span (Pie *et al.*, 2001), to generate the frequent sequences. Using frequent sequences, frequent patterns are generated using ICN modeling language. For enabling user interaction with the generated pattern, the patterns are made interactive using the concept of Interactive Poset Graph Navigation (Michele *et al.*, 2009). To apply the same framework in other, only the pattern generation method need to be modified.



CASE STUDY

The example event log has five fields: Case Id; the process instance, activity; the event, event type, participant, and time stamp. The control-flow and organizational perspective can be mined using this log. The process is taken from a travel agency, where a driver is fined an amount because of breaking rules/discipline. The process requires a fine to be entered in the system. After fine is entered into the system, the bill is sent to the driver. If the driver does not pay the bill within one

month, a reminder is sent to the driver. When the bill is paid, the case is archived.

Results of Case Study

- From the control-flow and organizational graph, it can be concluded that:
- “Anne” files fine and transfers work to both “Mary” and “John”.
- “Mary” never has to send a reminder more than once, while “John” does not seem to perform good.
- Also, “Mary” performs a task faster than “John”.

Table 1: The Event Log for Case Study

Case Id	Activity	Event Type	Participant	Time Stamp
1	File Fine	START	Anne	20-07-2004 14:00:00
2	File Fine	START	Anne	20-07-2004 15:00:00
1	Send Bill	SEND	Mary	20-07-2004 15:05:00
2	Send Bill	SEND	Mary	20-07-2004 15:07:00
3	File Fine	START	Anne	21-07-2004 10:00:00
3	Send Bill	SEND	John	21-07-2004 14:00:00
4	File Fine	START	Anne	22-07-2004 11:00:00
4	Send Bill	SEND	John	22-07-2004 11:10:00
1	Process Payment	PROCESS	Mary	24-07-2004 15:05:00
1	Close Case	COMPLETE	Mary	24-07-2004 15:06:00
2	Send Reminder	REMIND	Mary	20-08-2004 10:00:00
3	Send Reminder	REMIND	John	21-08-2004 10:00:00
2	Process Payment	PROCESS	Mary	22-08-2004 09:05:00
2	Close Case	COMPLETE	Mary	22-08-2004 09:06:00
4	Send Reminder	REMIND	John	22-08-2004 15:10:00
3	Send Reminder	REMIND	John	21-09-2004 10:00:00
4	Send Reminder	REMIND	John	22-09-2004 17:10:00
4	Process Payment	PROCESS	John	29-09-2004 14:01:00
4	Close Case	COMPLETE	John	29-09-2004 17:30:00
3	Send Reminder	REMIND	John	21-10-2004 10:00:00
3	Process Payment	PROCESS	John	25-10-2004 14:00:00
3	Close Case	COMPLETE	John	25-10-2004 14:01:00

Conclusions on the Basis of Result

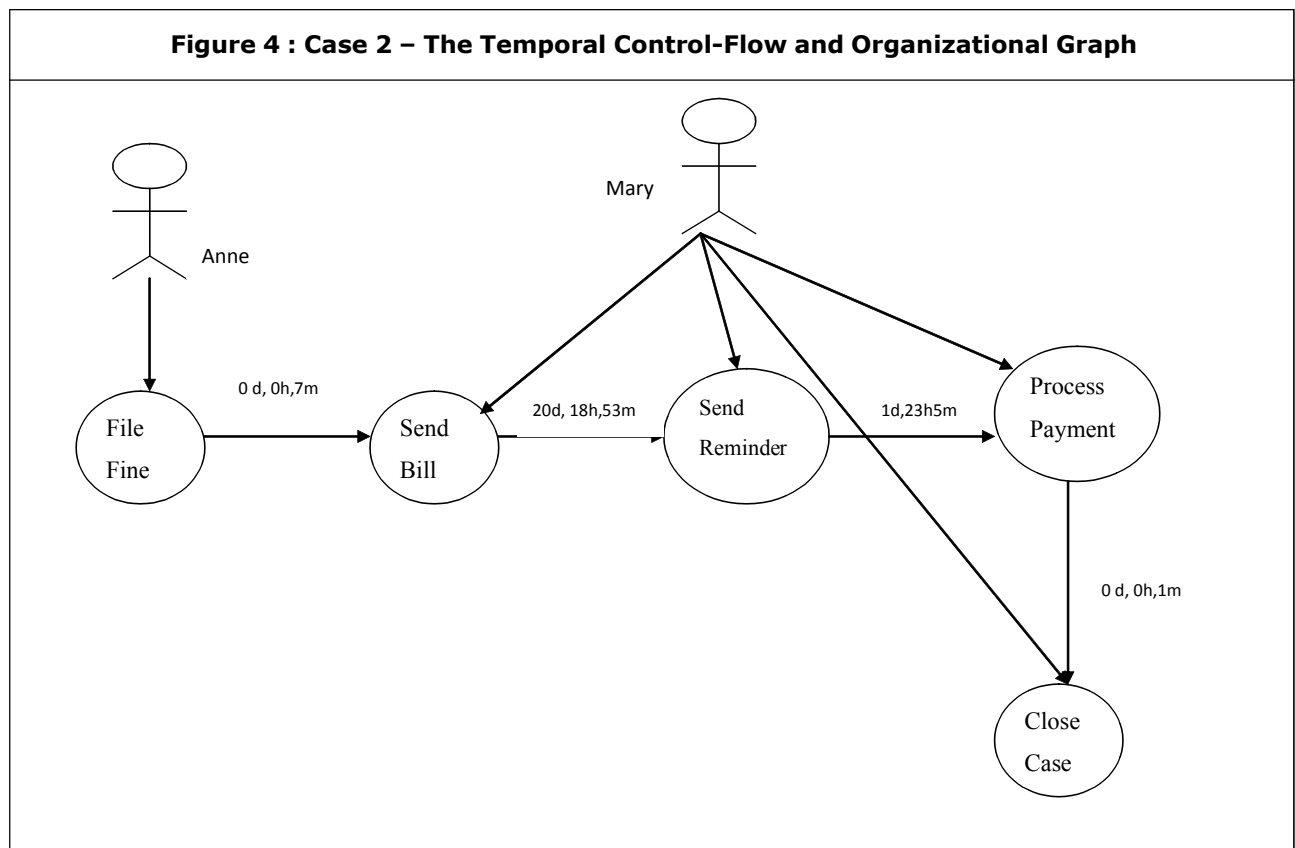
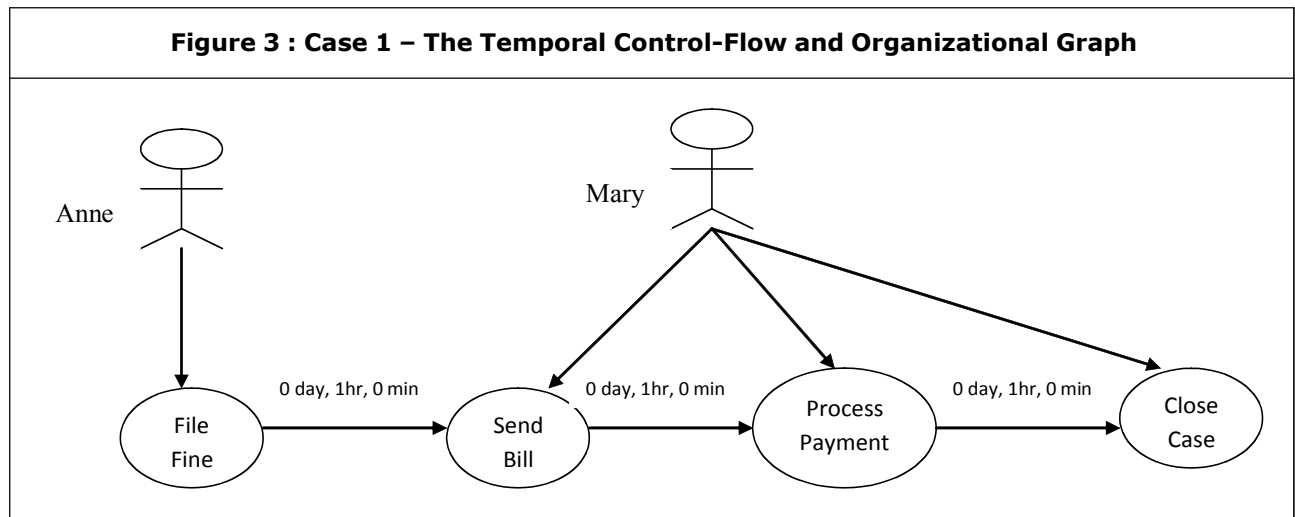
- The manager could talk to “Mary” and check if has another approach to send “reminders” that “John” could benefit from.

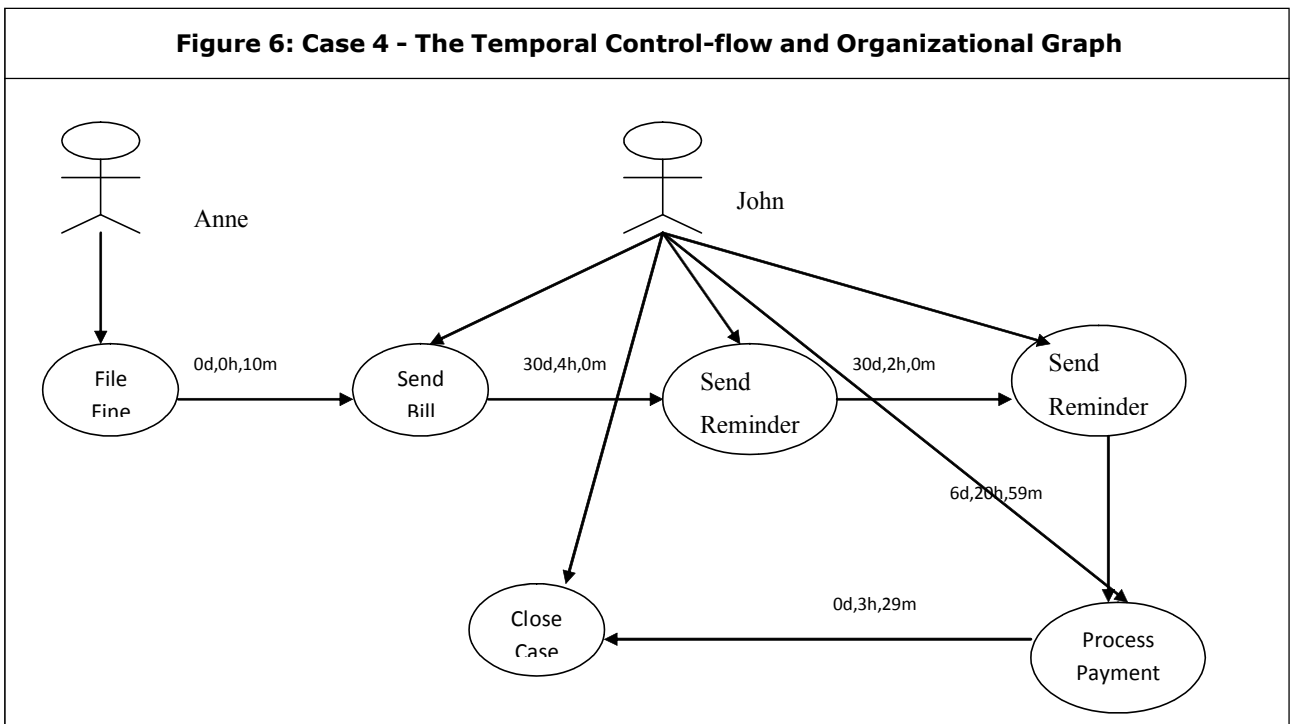
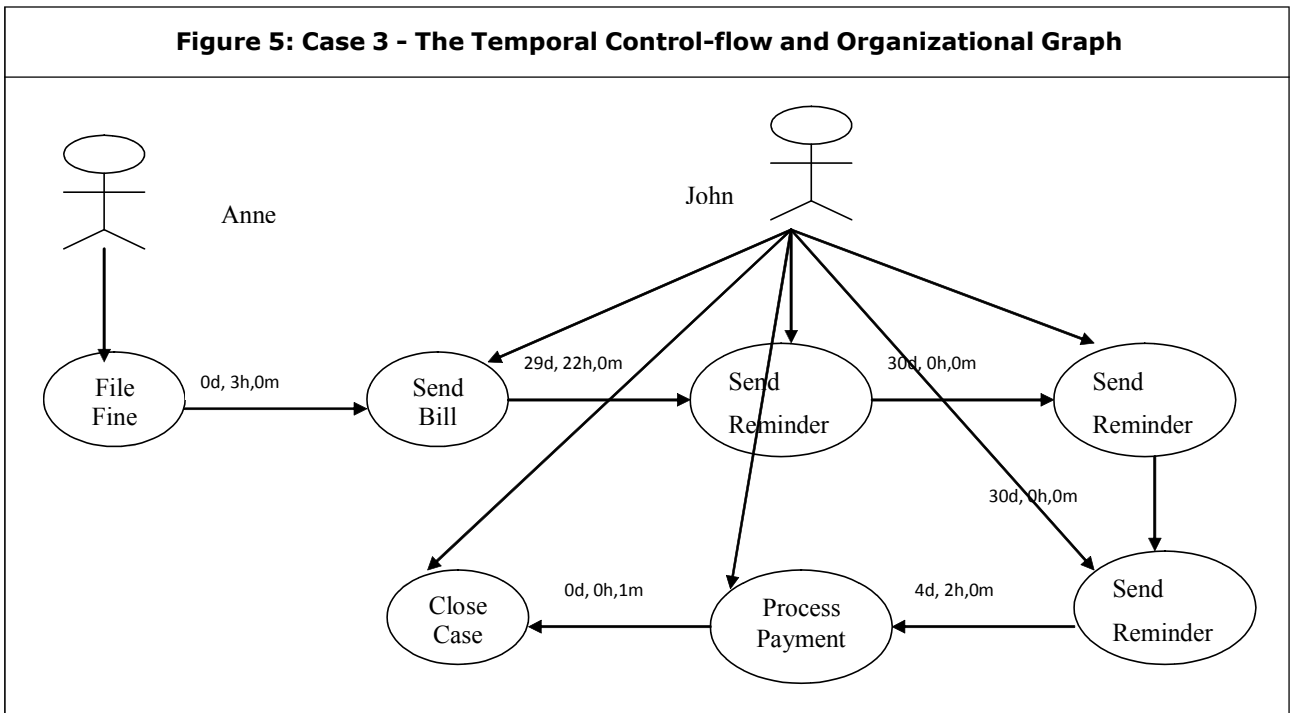
This can help in making good practices a common knowledge in the organization.

RESULTS AND DISCUSSION

The outcome of proposed work can be summarized in following points:

- A novel unique mining method that extensively takes into consideration the time information





- as well as organizational, informational and operational perspective of a process.
- Development of framework to implement methodology of process mining discussed above.
- Using the patterns, mutually independent activities for parallel execution can be identified.
- Interactive framework facilitating the user to extract patterns and explore suggestive

changes in workflow leading to positive improvement in process.

FUTURE ENHANCEMENT

The proposed technique can be applied on data collected from various organizations. More realistic and novel patterns can be generated which may then be helpful in improving the business processes.

CONCLUSION

The proposed work provides an interactive framework for mining workflow graphs from process logs. Using this framework the user can perform a temporal analysis by means of a TAS-based mining paradigm. The interactive approach allows useful temporal patterns to be identified from process workflow logs in an efficient manner. Applying this framework to the different domains such as web logs, logistics data, medical data, a wide variety of associations and patterns may be derived.

ACKNOWLEDGMENT

The researchers wish to express their deepest gratitude and warmest appreciation to Dr. A.K. Ramani, Head, School of Computer Science & IT, Devi Ahilya Vishwavidyalaya, Indore, for giving us an opportunity to carry out this research in the department and for helping us very willingly as often as we wanted. Without his cooperation and motivation this research would not have been possible.

REFERENCES

1. Aubrey J Rembert (2006), *Comprehensive Workflow Mining*, ACM-SE 44: Proceedings of the 44th Annual Southeast Regional Conference, March.
2. Ekkart Kindler, Vladimir Rubin and Wilhelm Schäfer (2006), *Incremental Workflow Mining for Process Flexibility*, http://lamswww.epfl.ch/conference/bpmds06/program/Kindler_11.pdf
3. Fosca Giannotti, Mirco Nanni and Dino Pedreschi (2006), *Efficient Mining of Temporally Annotated Sequences*, In Proceedings of the 6th SIAM International Conference on Data Mining.
4. Fosca Giannotti, Mirco Nanni, Dino Pedreschi and Fabio Pinelli (2006), *Mining Sequences with Temporal Annotations*, In Proceedings of the 2006 ACM Symposium on Applied Computer Science (SAC), pp. 593-597.
5. Michele Belingerio, Fabio Pinelli, Micro Nanni and Fosca Giannotti (2009), *Temporal Mining for Interactive Workflow Data Analysis*, Proceedings of the 15th ACM SIGKDD international conference on Knowledge Discovery and Data Mining, KDD'09, June 28 – July 1, 2009, Paris, France.
6. Pei J *et al.* (2001), *Prefixspan: Mining Sequential patterns by Prefix Projected Growth*, In ICDE, pp. 215-225.
7. Rubin V, Gunter C W, Wil M P van der Aalst, E Kindler, van Dongen B F and Schafer W (2007), *Process Mining Framework for Software Processes*, Proceedings of the 2007 International Conference on Software Process, May.



International Journal of Engineering Research and Science & Technology

Hyderabad, INDIA. Ph: +91-09441351700, 09059645577

E-mail: editorijerst@gmail.com or editor@ijerst.com

Website: www.ijerst.com

