GME: A Contemporary Approach Workflow Process Improvement of Software by Uncovering hidden Transactions of a healthcare legacy application

Shashank Sharma and Dr. Sumit Srivastava
MU Jaipur, shashanksharmaaa@gmail.com, sumit.srivastava@jaipur.manipal.edu

Abstract - **In organization numbers are increasing day by day with a drastic pace which prefers the extraction of the workflow of processes to interpret the operational processes. For a viably and sorted out approach to drive the development in the realm of digitization is utilized by the approach of work process extraction. The work process extraction/mining are otherwise called process mining. The goal of Workflow mining is to get the extraction of data of an association's method of business by changing over the logs of occasion information recorded in association's frameworks. This impact to the enhance conformation of processes to organization regulation where workflow mining approach for analysis is actualized. Work process Mining strategies are absolutely rely upon the nearness of framework occasion log information. We accept to involve setting various endeavors on building our strategies or frameworks to record the greater part of the old information. The urge to comprehend and expand their procedures of businesses entails the process exploration practices. This paper displays a philosophy how programming occasion log information is analyzed to grasp and advance the product work process by utilizing arrangement which best in class utilized as a part of the product code clone streamlining for the human services area application.**

***Keywords*: Workflow net, Genetic Mining for Events(GME), Genetic Algorithm(GA), Process Mining, Process workflow, Petri net, IoE (Internet of Events), Clone Optimization, workflow, petrinet, BPM(Business Process Model), MXML(Mining XML)**

I. INTRODUCTION

Workflow mining a.k.a. Process Mining is a considerably new and emerging area of academic research within data analytics. The key objective here is to deploy workflow-related data in directive to obtain pertinent info and knowledge by employing data analytic algorithms and determining a workflow model. This segment discusses the idea of how an log from events is the foundation of exploration along with other main building block of process mining. Work process mining articles to channel the hole between enormous information investigation and conventional business work process/process administration. This field can essentially be ordered into (1) Workflow revelation, (2) conformance checking and (3) improvement [1]. This permits the extraction of bits of knowledge about the by and large and inward conduct contained in any given procedure. Work process disclosure procedures underscore on utilizing the occasion information in order to decide work process models. Conformance checking strategies emphasize on supporting the occasion information on a work process model to confirm how well the model fits the information and the other way around [3]. In spite of the fact that enlargement methods utilize occasion information and work process models to repair or increase the work process show. Hence, workflow mining provides the conduit the gap between data mining and machine learning practices and the business process management discipline.

A. The Event Log as the Linchpin of reasoning

Data is a crucial building block in various discovery domains. Process Mining uses data that is accounted by event logs. A sample log is shown in table 1. An event log in terms of this situation can be defined as the process of recording of an action instance on the system. Action or activity instances are units of work that are registered by the system when work is piloted in the situation of a assured process. Statuses of activity or action are specified to a set of languages that are fixed to the workflow modeling hypothesis. Various workflow modeling hypothesis have unique implementation standards that are governed by their state evolution diagrams. A simple state evolution diagram has been explained in Figure 1. More elaborate diagrams have been described in process modeling literature. Consider a case that where the life phase action of Business process object notation is characterize by the Object Management Group (OMG) [7]. Adding to this, different process displaying ideal models, for example, YAWL (Yet Another Workflow Language), case taking care of [9] and explanatory methodologies (EM – BrA2CE [4], Declare [8]) which proposes unequivocal however undifferentiated from base outlines for their comparing delineation semantics.

TABLE 1: EXAMPLE OF RUN OF THE MILL EVENT LOG

Log ID	Event Trem	Event class	Discoverer	Time stamp	Extra Data
011	Make order form	Start	Employee A1	19-10-1955 15:15:02	….
012	Make order form	Complete	Employee A2	19-10-1955 05:14:01	….

It can get fascinating to experience different establishment semantics and state progress outlines, for work process mining and work process investigation; in any case, the real information is what is vital. Process Data is generally pooled from various vaults that are gotten from CRM, ERP, WFM and other assorted data frameworks. This outcome in trouble while recognizing hypothetical establishment of an occasion as far as state progress outlines and

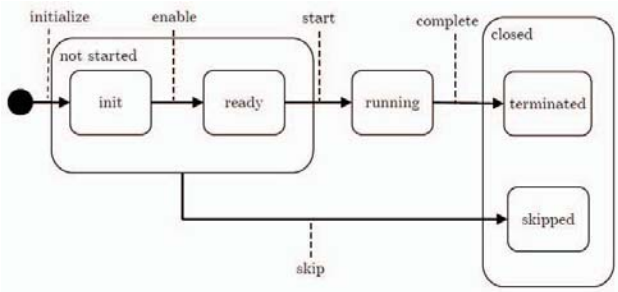


Fig. 1: DEPICTION OF STATE TRANSFORMATION OF A PROCESS

genuine information found practically speaking. In different divisions, for example, CRM(customer relationship administration), item improvement, monetary administrations and so forth business work process depend intensely on heritage data frameworks or low work process arranged data frameworks. Further, the enrolled business process information is very much characterized with the end goal that exclusive a specific kind of state progress (e.g. Achievement of a movement event) can be mined by means of. real information pieces. Work process Mining can be viewed as most productive in measured conditions where business data frameworks have more extensive alternatives of conduct, here; the change of open information into an occasion log is regularly a critical undertaking.

1.2 How Petri net is different from Workflow net?

Petri net basically deals with mathematical modelling where as work flow net deals with XML modelling. Petrinet basically deals with low level work flow where as work flow net deals with high level work flow. Petrni net basically used to represent theoretical concepts of computer science whereas work flow net basically used for business process/ management modelling.

Filtration & extraction of event logs through business information systems is usually drifted out by text-based data scripts. In actuality, process data is more often distributed over various data sources and it is painstaking to define the precise scope of process that is studied. Further, a comprehensive ETL-phase is required before a concrete analysis is initiated. Adding to which, the data must be in an event-log storage format.

One of the initial approaches to storing log based on event performed/ executed is in the MXML format. From the date of its inception i.e. 2003, this fact based standard is used i.e. MXML format because it is highly integrated with the ProM-framework which is a framework used for academic purpose for workflow Process improvement. It was not as of not long ago that IEEE team was favored rather than MXML for process/ workflow Mining. Then after the new standard has come into existence which enhances the efficiency & make this process easy & effective which is known as XES(eXtensible Event Stream) [6]. Logs behaves, as a foundation of the workflow mining, it is significant to express the variable necessities to which an occasion log must approve. The taking after three suppositions are mandatory and basic:

- Activity instance must be well defined for the workflow instance of an event which specified by a unique activity name
- Unique process instances or ID which is used by cases must be referred by an events
- Ordered Timestamp must be recorded by the events

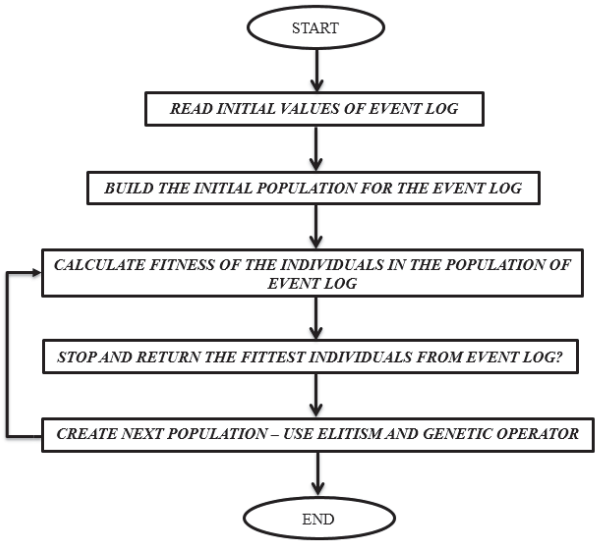
II. LITERATURE REVIEW

As Agrawal et al. [12] and Pinter et al. [15] are the founders who originally able to extract the chances of getting to explore the split/join connections on the extracted models. The main idea behind to focus on model to extract work flow pattern[14] and each point in this system has an OR-split/join semantics. Without a doubt, each planned roundabout segment in the model has a boolean limit that evaluates to certified or false after an errand is executed. The appraisal of the boolean conditions sets what number of branches are authorized after an endeavor is executed. Cook et al. [13] have the primary approach that does not concentrate on a whole mined model. Methodology looking for the most progressive approach for the workflow model. In actuality, all over they do mine a whole procedure appear, however that is not their essential point. The builds up that can't be mined by all frameworks are circles, sans non choice, vague assignments moreover, duplicate endeavors. Grecco et al. [16] not extract any of the loops through his research. The validation behind states that their approach is not able to extract meaningful understanding from the event logs after applying proposed algorithm. They tried to visualize the differentiating behaviour of the mined models

through various fundamental algorithm and tried to analyse the different behaviour. Cycles or loops made this approach ineffective & efforts goes in the vain. Some different procedures can't mine self-assertive loops in light of the fact that their model documentation (or portrayal) does not bolster this sort of circle. The key inspiration driving why most methodologies can't mine non-neighborhood sans non-choice is that the larger part of their mining computations rely upon neighborhood information in the logs. The systems that don't mine nearby without non decision can't do as such on the grounds that their portrayal does not bolster such a develop. Generally the system depends on a piece organized documentation, as Herbst et al. also, Schimm. Skip errands are not mined because of portrayal restrictions too. Split/join imperceptible undertakings are not mined by numerous methods, with the exception of for Schimm and Herbst et al.. As a matter of fact, we likewise don't focus at finding such sort of assignments. Nevertheless, it is routinely the case that it is possible to fabricate an exhibit with no split/join imperceptible errands that impart a comparable lead to the approach with false or wrong assignments for the split/join constructs. Copy undertakings are definitely not mined in light of the fact that numerous methods expect that the mapping between the undertakings what's more, their marks is injective. By the day's end, the names are exceptional per task. The principle systems that mine duplicate endeavors are Cook et al. [13] for progressive structures just, and Herbst et al. [17] for both back to back and parallel methodology. We don't consider Schimm [18] to mine strategy models with duplicate errands since his approach expect that the revelation of the duplicate endeavors is done in a pre-getting ready stride. This progression identies all the copies and ensures that they have one of a kind identiers when the occasion log is given as contribution to the mining calculation. As a matter of fact, every one of the procedures that we survey here would handle copy undertakings if this same pre-preparing step would be done before the sign in given as contribution to them.

III. PROPOSED ALGORITHM WITH PROCEDURE AND A SAMPLE CASE STUDY

The proposed process efficient GME (Genetic Mining for Events) is based on GA and process mining of event logs. In this, function of fitness work is registered by the nature of a person. The nature of an individual is fundamentally set by its replaying of the log follow. The entire flow of GME is shown in Flow chart 1. The steps of the proposed algorithm are written below:



FLOWCHART 1: PROPOSED ALGORITHM FOR GME

In order to apply an innate/hereditary count, we need to address people. Each individual thinks about to a possible methodology model and its depiction should be anything other than hard to manage. Our underlying thought was to speak to forms straightforwardly by Petri nets. Lamentably, Petri nets end up being a less helpful way to speak to forms in this unique circumstance. The principal reason is that in Petri nets there are places whose nearness can't be gotten from the log, i.e., events simply imply the dynamic parts of the net (changes). Because of this it ends up being harder to make a hidden masses, describe inherited directors (half breed and change), and depict mixes of AND/OR-parts/joins. Note that given a log, it is definitely not hard to discover the activities and along these lines, the progressions that exist in the Petri net. Be that as it may, implementing certain routings by just associating changes through spots is intricate (if certainly feasible). In this manner, we will utilize an alternate interior portrayal.

Table 2 demonstrates the inside portrayal of a specific individual which utilizes by mentioned hereditary mining approach. This purported causal network characterizes the relations of causal type between the exercises and if there should arise an occurrence of various info or yield exercises, the rationale is delineated. Consider for instance the line beginning with A. This column demonstrates that there is a not a causal connection amongst A and A (take note of the initial 0 in the column), in any case, there is a causal association continues among A and B (observe the underlying 1 in this segment). The following two sections in the segment show that there are also causal relations amongst A and C and A and D. The last component in the line demonstrates the directing rationale, i.e., B ∨ C ∨D demonstrates that A is trailed by B, C, or D. The section named "Yield" demonstrates the rationale relating an action to causally following exercises. The primary line underneath

"Info" demonstrates the rationale relating a movement to causally going before exercises. Note that the data province of A is substantial, i.e., no data required. Activity G has E ∧ F as data condition, i.e., both E and F need to complete a particular true objective to enable G. Development H has B ∨ C ∨ G as data condition, i.e., B, C, or G needs to complete the process of remembering the true objective to enable H.

TABLE 2: CAUSAL MATRIX MADE AND UTILIZED FOR PORTRAYAL

	INPUT								
	true	*A*	*A*	*A*	*D*	*D*	$E \wedge F$	$B \vee C \vee G$	
→	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*OUTPUT*
A	0	1	1	1	0	0	0	0	$B \vee C \vee D$
B	0	0	0	0	0	0	0	1	*H*
C	0	0	0	0	0	0	0	1	*H*
D	0	0	0	0	1	1	0	0	$E \wedge F$
E	0	0	0	0	0	0	1	0	*G*
F	0	0	0	0	0	0	1	0	*G*
G	0	0	0	0	0	0	0	1	*H*
H	0	0	0	0	0	0	0	0	*true*

Subsequently after windup with introduction & groundwork we come back to the objective of this paper: hereditary process mining. With a specific end goal to apply a hereditary calculation we have to speak to people. Every individual compares to a conceivable procedure model and its portrayal ought to be anything but difficult to deal with. Our underlying thought was to speak to forms straightforwardly by Petri nets. Tragically, Petri nets end up being a less helpful way to speak to forms in this specific circumstance. So to overcome from this problem we use causal matrix[11,12].

TABLE 3: CONCISE & CONCRETE ENCODING FORMAT OF AN INDIVIDUAL IN TABLE 2

ACTIVITY	*INPUT*	*OUTPUT*
A	{}	{{B,C,D}}
B	{{A}}	{{H}}
C	{{A}}	{{H}}
D	{{A}}	{{E},{F}}
E	{{D}}	{{G}}
F	{{D}}	{{G}}
G	{{E},{F}}	{{H}}
H	{{B,C,G}}	{}

A matrix of causality is a row, X with elements named as (M, N, O, P) where
- M comprises of activities with type of finite sets,
- $N \subseteq M \times M$ is the relation of causality,
- $O \in A \rightarrow P(P(M))$ is the function of input condition type,3
- $O \in A \rightarrow P(P(M))$ is the function of output condition type, such that
- $N = \{(m_1, m_2) \in M \times M \mid m_1 \in O(m_2)\}$,4
- $N = \{(m_1, m_2) \in M \times M \mid m_2 \in P(m_1)\}$,
- $\forall_{m \in M} \forall_{QQs^R \in O(m)} Q \cap Q^L = \emptyset \Rightarrow Q = Q^L$,
- $\forall_{m \in M} \forall_{Q,Q^R \in P(m)} Q \cap Q^L = \emptyset \Rightarrow Q = Q^L$,
- $N \cup \{(m_o, m_i) \in M \times M \mid m_o {}^{N}\bullet = \emptyset \wedge {}^{N}\bullet m_i = \emptyset\}$ is a connected graph of strong type.

A. Binding of Petrinet and Matrix of Causality

Binding of a petri net through the framework of causality network included in light of the fact that it requires to "explore locations" and along with this need to explore more on lattice that supposed to be more detailed than its previous counter parts [11, 12].

Let X = (M, N, O, P) be a matrix of causality representation type. The binding of Petri net is a row $\Pi^{S}_{X \rightarrow TS}$ (X) = (T, U, V), where

- $T = \{w, y\} \cup \{w_{U,Q} \mid U \in M \wedge q \in O(U)\} \cup \{P_{u,q} \mid U \in M \wedge Q \in P(U)\}$,
- $U = M \cup \{Z_{U1, U2} \mid (U_1, U_2) \in N\}$,
- $V = \{(w, u) \mid u \in M \wedge {}^{N}\bullet u = \emptyset\} \cup \{(u, y) \mid U \in M \wedge U {}^{N}\bullet = \emptyset\} \cup \{(w_{u,q}, U) \mid U \in M \wedge q \in W(U)\} \cup \{(U, P_{u,w}) \mid U \in M \wedge w \in P(u)\} \cup \{(P_{u1,q}, Z_{u1,u2}) \mid (u1, u2) \in N \wedge q \in P(U_1) \wedge U_2 \in Q\} \cup \{(Z_{u1,u2}, W_{u2,q}) \mid (u_1, u_2) \in N \wedge q \in W(U_2) \wedge U_1 \in Q\}$.

In this first we take the event log of hospital case of healthcare information system. In this particular event log we have 42 events or events flow. Our aim is to classify the events on the basis of type of events. We start with event log then we select MXML legacy classifier & process discovery algorithm for the xtraction of petri net. Along with this we apply some user specified constraints to get expected result.

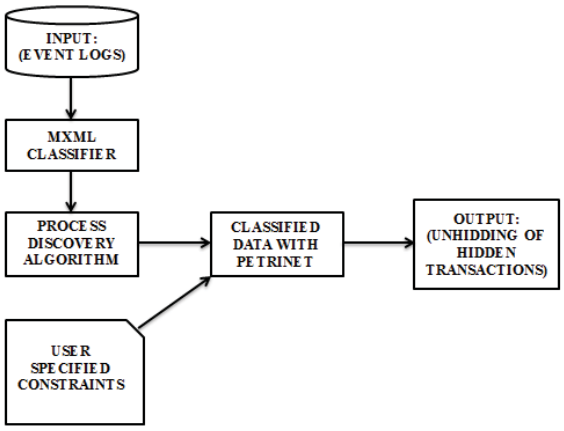


Fig. 2: PROCEDURE FOR GME

B. Comparison Analysis

For verifying & validating the effectiveness of GME, we used following standard algorithm that are used for workflow management as well as analyse & compare it with proposed GME. In figure number 6 you find black boxes (dark or complete black) which referred to hidden transactions, which shows that other or previous algorithms are unable to extract or locate hidden transactions although algorithms are able to classify the event log.

TABLE 4
COMPARISON OF RESULT FOR VARIOUS ALGORITHMS

Name of Algorithm	Input Format	Output Format	Intermediate Output	Able to Unhide hidden transactions
Alpha	MXML Log File	Petrinet	No	No
Alpha++	MXML Log File	Petrinet	No	No
Tsinghua-Alpha	MXML Log File	Petrinet	No	No
GME	MXML Log File	Petrinet	Heuristic Net	Yes

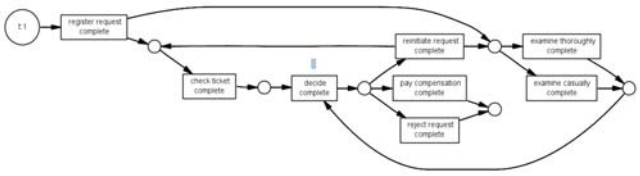


Fig. 3: ALPHA ALGORITHM

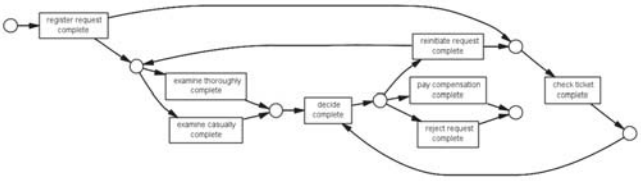


Fig. 4: ALPHA++ ALGORITHM

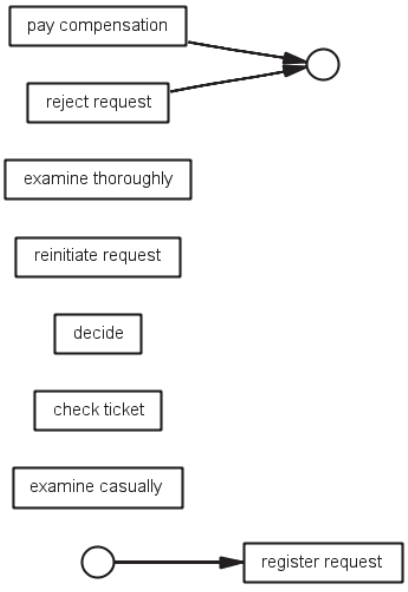


Fig. 5: TSINGHUA-ALPHA ALGORITHM

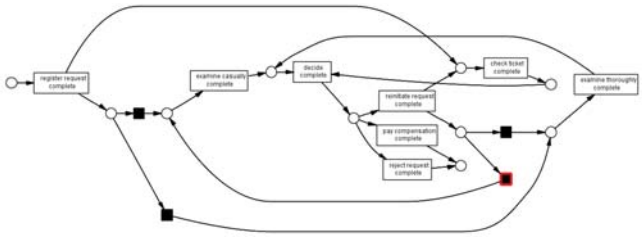


Fig. 6: OUTPUT PETRI-NET FROM PROPOSED GME ALGORITHM

IV. RESULT & OUTCOMES

In this, we received classified result for that particular hospital case. We got 8 classified event classes. For this classification, we used 2 criteria i.e. on the basis of event functionality & other one on the basis of event type. Along with this, complete classification is further sub-divided into 3 parts i.e. start event (count is 1), originators (count is 6) & end event (count is 2). After interpreting the petri net, we find that which work flow process model we need to work upon to improve the process of information system. Please find below the outcomes.

TABLE 4: CLASSIFIED EVENT LOG DATA

Model element	Event type	Occurrences (absolute)	Occurrences (relative)
check ticket	complete	9	21.429%
decide	complete	9	21.429%
register request	complete	6	14.286%
examine casually	complete	6	14.286%
reinitiate request	complete	3	7.143%
examine thoroughly	complete	3	7.143%
pay compensation	complete	3	7.143%
reject request	complete	3	7.143%

V. CONCLUSION & FUTURE WORK

"Workflow scientist" desires to possess particular/ exact to initiate innovation in a progressively digitalized ecosphere. In this paper, we just only conceptualized our idea with a small case study. This enables us to investigate the operational procedure work process of healthcare information systems that frameworks under genuine situations, and utilize extraction systems for procedures to get particular and perceived programming change models. This paper oriented on associate the organized mining approach for the event classes from event data for the software process improvement by using the Petri nets flow model approach. In the future, we aim at conducting additional experiments using different variety of event log data sets. A reasonable succeeding phase is to progress with tool support for domain-based information management systems.

REFERENCES

[1] W.M.P. van der Aalst. Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer-Verlag, Berlin, 2011
[2] W.M.P. van der Aalst, A.J.M.M. Weijters, and L. Maruster. Workflow Mining: Discovering Process Models from Event Logs. IEEE Transactions on Knowledge and Data Engineering, 16(9):1128-1142, 2004

[3] Adriansyah, A., van Dongen, B.F., van der Aalst, W.M.P.: Towards robust conformance checking. In: Business Process Management Workshops, Lecture Notes in Business Information Processing, vol. 66, pp. 122-133. Springer Berlin Heidelberg 2011
[4] S. Goedertier. Declarative Techniques for Modeling and Mining Business Processes. Phd thesis, Katholieke Universiteit Leuven, Faculty of Business and Economics, Leuven, September 2008
[5] W.M.P. van der Aalst. Data Scientist: The Engineer of the Future. In K. Mertins, F. Benaben, R. Poler, and J. Bourrieres, editors, Proceedings of the I-ESA Conference, vol. 7 of Enterprise Interoperability, pages 13-28. Springer-Verlag, Berlin, 2014
[6] C. W. Günther. XES Standard Definition. www.xes-standard.org
[7] Object Management Group (OMG). Business Process Modeling Notation (BPMN) – Spectification. OMG Document – formal/2011-01-03, January 2011
[8] M. Pesic and W. M. P. van der Aalst. A Declarative Approach for Flexible Business Processes Management. In J. Eder and S. Dustdar, editors, Business Process Management Workshops, 4103 of Lecture Notes in Computer Science, pages 169–180. Springer, 2006
[9] W. M. P. van der Aalst, M. Weske, and D. Grünbauer. Case handling: A new paradigm for business process support. Data Knowl. Eng., 53(2):129–162, 20052010
[10] T. Calders, C. Guenther, M. Pechenizkiy, and A. Rozinat. Using Minimum Description Length for Process Mining. In ACM Symposium on Applied Computing (SAC 2009), pages 1451{1455. ACM Press, 2009
[11] R. Agrawal, D. Gunopulos, and F. Leymann. Mining Process Models from Workflow Logs. In Sixth International Conference on Extending Database Technology, pages 469–483, 1998.
[12] J. Dehnert and W.M.P. van der Aalst. Bridging the Gap Between Business Models and Workflow Specifications. International Journal of Cooperative Information Systems, 13(3):289–332, 2004.
[13] J.E. Cook, Z. Du, C. Liu, and A.L.Wolf. Discovering Models of Behavior for Concurrent Workflows. Computers in Industry, 53(3):297{319, 2004.
[14] IBM. IBM MQSeries Workow - Getting Started With Buildtime. IBM Deutschland Entwicklung GmbH, Boeblingen, Germany, 1999.
[15] S.S. Pinter and M. Golani. DiscoveringWorkflow Models from Activities Lifespans. Computers in Industry, 53(3):283{296, 2004.
[16] G. Greco, A. Guzzo, L. Pontieri, and D. Sacca. Mining Expressive Process Models by
Clustering Workflow Traces. In H. Dai, R. Srikant, and BIBLIOGRAPHY 365 C. Zhang, edi
tors, PAKDD, volume 3056 of Lecture Notes in Computer Science, pages 52{62. Springer,
2004.
[17] J. Herbst and D. Karagiannis. Workow Mining with InWoLvE. Computers in Industry, 53(3):245{264, 2004.
[18] G. Schimm. Mining Exact Models of Concurrent Workflows. Computers in Industry, 53(3):265{281, 2004.
[19] J. Dehnert and W.M.P. van der Aalst. Bridging the Gap Between Business Models and Workflow Specifications. International Journal of Cooperative Information Systems, 13(3):289–332, 2004.