

УДК 007.5; 004.85

Е.О. Богатов

*Харьковский национальный университет радиоэлектроники, Харьков*

## МЕТОД ФОРМИРОВАНИЯ ОГРАНИЧЕНИЙ В ЗАДАЧЕ СТРУКТУРИЗАЦИИ ЖУРНАЛОВ РЕГИСТРАЦИИ СОБЫТИЙ

*В статье предложен метод формирования ограничений при подготовке данных для process mining. Предложенный метод направлен на построение такой совокупности ограничений для исходных данных из журналов регистрации событий, которые позволили бы сформировать модели подпроцессов, связанных с отдельными элементами организационной структуры и обработкой заданных объектов.*

**Ключевые слова:** бизнес-процесс, модель журнала регистрации событий, отношения порядка.

### Введение

**Актуальность.** Существующие на сегодняшний день методы интеллектуального анализа бизнес-процессов (БП), направлены на построение моделей таких процессов на основе анализа их журналов регистрации событий (ЖРС), которые в свою очередь генерируются соответствующей информационной системой. Данные методы позволяют построить «черновую» модель бизнес-процессов (или уточнить существующую), выявить потенциал для оптимизации и «узкие места» бизнес-процессов. Результаты process mining дают возможность организовать эффективный контроль выполнения бизнес-процессов и, при необходимости, их реинжиниринг.

К структуре журналов регистрации событий выдвигается ряд требований, которые зачастую выполняются в автоматическом либо автоматизированном режимах процессно-ориентированными информационными системами. Однако, существует значительное количество организаций, в которых бизнес-процессы слабо структурированы либо не определены. При этом обычно используются корпоративные информационные системы, ориентированные на функциональное управление, что вызывает трудности в описании и построении моделей существующих процессов. Текущие действия в данных информационных системах протоколируются с метками времени вне зависимости от того, к какому процессу и объекту принадлежит фиксируемое действие. Таким образом, информация, которая содержится в регистрационных файлах корпоративных систем, является достаточной для построения процессной модели, отражающей реально выполняющиеся бизнес-процессы. Однако указанная в данных журналах информация не упорядочена в соответствии с существующими требованиями в process mining, что значительно затрудняет анализ рассматриваемого БП, а также снижает адекватность и репрезентативные свойства полученной модели.

В более ранних работах [6] были представлены модели структурированного и неструктурированно-

го журналов регистрации событий, которые являются основой для уточнения метода предварительно обработки журналов регистрации событий [4]. Однако данный метод подразумевает указание на одном из этапов зависимостей между элементами бизнес-процессов. В роли таких зависимостей могут выступать ограничения, накладываемые на последовательность исполнения/реализации рассматриваемого бизнес-процесса.

Вышеизложенное определяет актуальность разработки метода формирования ограничений для исходного набора событий при решении задач process mining.

**Анализ научных публикаций.** В области constraint programming рассматриваются вопросы решения задач с помощью универсального решателя при условии формирования корректного набора ограничений [1 – 3]. Было предложено решение ряда задач:

- проверка допустимости сформированного набора задач, задающих ограничения;
- поиск допустимого набора задач, задающих ограничения, с максимальными и минимальными значениями переменных;
- оптимизация сформированного набора задач и т.д.

Однако вопросы формирования ограничений при построении моделей бизнес-процессов методами process mining детально не исследованы.

Таким образом, очевидна необходимость разработки метода формирования набора ограничений на исполнение БП, что позволит автоматизировать процесс формирования набора ограничений.

**Постановка задачи.** Задача заключается в разработке метода формирования набора ограничений, который позволит описать разные, на качественном уровне, модели анализируемого бизнес-процесса, представленные следующими проекциями: организационной, функциональной и временной. К исходным данным задачи следует отнести журнал регистрации событий, который содержит в общем случае неупорядоченное описание событий, происходящих на объекте управления с указанием меток времени.

Необходимо представить в формализованном виде ограничения, накладываемые структурными элементами бизнес-процесса на его исполнение, разработать метод формирования набора ограничений, который позволит выявить зависимости между каждым из структурных элементов бизнес-процесса.

### Метод формирования набора ограничений

Вопросы формирования ограничений и их дальнейшей обработки рассматриваются в области программирования задач (constrain programming). Согласно данной парадигме, под ограничением следует понимать задание описательной характеристики, области допустимых значений, в которых может состоять рассматриваемый объект. Более широко решаемая в рамках парадигмы задача представлена в классическом описании [3]:

– общее представление задачи формирования ограничения имеет следующий вид  $P = \langle X, D, C \rangle$ , где  $X$  –  $n$ -размерное множество переменных  $X = \langle x_1, x_2, \dots, x_n \rangle$ ,  $D$  –  $n$ -размерное множество доменов  $D = \langle d_1, d_2, \dots, d_n \rangle$ ,  $C$  – это  $t$ -размерное множество задач  $C = \langle C_1, C_2, \dots, C_t \rangle$ ;

– каждый домен может принимать одно из множества значений  $X$ , т.е.  $x_i \in D_i$ ;

– задачей  $C_j$  описывается следующим множеством  $\langle R_{S_j}, S_j \rangle$ , где  $R_{S_j}$  – множество связей переменных на пространстве  $S_j$ , т.е. подмножество  $D' \subset D$ , по сути, под задачей подразумевается описание допустимого набора переменных на пространстве размерностью в  $n$  доменов.

Парадигма программирования задач подразумевает решение следующих задач:

– поиск минимум одного допустимого размещения переменных в каждом из доменов, которое удовлетворяет всем поставленным задачам, если таковое существует;

– поиск всех допустимых наборов значений удовлетворяющих всем поставленным задачам;

– поиск допустимых наборов с максимальными либо минимальными значениями переменных в каждом из доменов, удовлетворяющих всем задачам.

Рассмотрим более подробно решение задачи поиска размещения, удовлетворяющего поставленным задачам, парадигмы программирования задач [1, 2]: пусть имеется конечный алфавит переменных, который определяется как множество переменных  $X = \langle x_1, x_2, \dots, x_n \rangle$ ; пусть имеется множество объектов  $Obj = \langle obj_1, obj_2, \dots, obj_n \rangle$  и каждый объект может принимать одно из множества значений  $X$ , определяемых как  $x_i \in obj_j$ .

Множество всех возможных состояний, которые могут принимать объекты, представляется

через  $Obj \rightarrow X$ , обозначим как  $x^{Obj}$ . Для формирования ограничений на группы объектов выделяют понятие домена/группы объектов, которое определим как  $Obj' = (Obj'_1, Obj'_2, \dots, Obj'_n)$ ,  $Obj' \subset Obj$ , таким образом, значения, принимаемые в данном подмножестве, определяются как  $x(Obj')$ . Данное обозначение  $x(Obj')$  следует понимать как функцию формата, являющуюся сужением функции  $\bar{x} : Obj \rightarrow X$ , на подмножество  $T' \subset T$ . Функция, описывающая допустимость выбранного набора ограничений, определяется как  $G : x^{Obj} \rightarrow \{0, 1\}$ . Которая для каждой функции определяет её допустимость  $G(D_i \rightarrow x) = 1$ , или недопустимость.

Пусть задано условие  $g(Obj')^{(x(Obj'))}$ , которое является ограничением на существование объекта  $Obj'$  с заданной переменной  $x$ . Таким образом, задача проверки допустимости размещения описывается следующим выражением:

$$G(\bar{x}) = \bigvee_{T' \in T} g(Obj')^{(x(Obj'))} = 1. \quad (1)$$

Классическое представление парадигмы поиска задач имеет много схожего с задачей формирования ограничений на исполнение бизнес-процесса. В бизнес-процессе, как и в парадигме программирования задач можно выделить множество переменных, в роли которых выступает множество значений, принимаемых анализируемым структурным элементом. К множеству доменов следует отнести то же множество элементов анализируемого структурного элемента бизнес-процесса. Таким образом, количество переменных структурного элемента БП, который анализируется, определяет количество элементов в множестве доменов, что позволяет поставить в соответствие количества переменных к количеству доменов. В качестве сформулированных задач выступают отношения порядка, которые определяют  $R_{S_j}$  как связи между переменными в определенных доменах, а  $S_j = 2$ , т.к. осуществляется описание отношений между двумя рядом стоящими элементами через отношения порядка (рис. 1).

Вышеизложенное свидетельствует о возможности проверки набора сформированных отношений порядка на соответствие анализируемому БП и позволяет сформировать набор ограничений.

$$X = \langle A, B, C, D, E \rangle \quad D = \langle D_1, D_2, D_3, D_4, D_5 \rangle$$

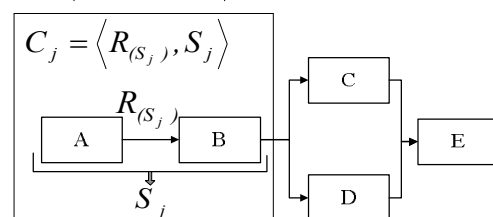


Рис. 1. Описание модели БП в соответствии

с парадигмой решения задач

БП имеет следующие уровни представления: уровень бизнес-правил; уровень процесса; уровень объектов. Задача формирования набора ограничений имеет непосредственную связь с уровнем бизнес-правил, т.к. на данном уровне описывается порядок наступления структурных элементов бизнес-процесса.

Также следует отметить, что бизнес-правила являются основным и наиболее важным механизмом управления БП. В настоящее время реализация бизнес-процессов (БП) основывается преимущественно на использовании описательных или нормативных моделей таких процессов. Однако, помимо описания собственно последовательности работ бизнес-процесса, при функционировании любого предприятия используются также набор политик, выраженных в виде множества бизнес-правил. Последние фактически определяют алгоритмы распределения ресурсов процесса и отражают некоторые характерные особенности и свойства текущей деятельности соответствующего предприятия. Иными словами, правила определяют возможные действия в различных ситуациях, возникающих при выполнении каждого из бизнес-процессов предприятия.

Указанные политики и правила являются формой выражения знаний о БП и в значительной степени формируются на основании опроса исполнителей процесса. В то же время, противоречивость знаний о БП, представленная его исполнителями, требует решения проблемы оценки несоответствий бизнес-правил между собой, а также оценки несоответствий указанных правил и спецификации процесса.

Такого рода проверку можно осуществить за счет применения техник программирования парадигмы решения задач к ЖРС. Формирование набора ограничений на исполнение бизнес-процесса осуществляется на основе анализа журнала регистрации событий.

Журналы регистрации событий неструктурированного бизнес-процесса включает следующие обязательные элементы [4]:

- временной параметр;
- событие, отражающее действие процесса;
- объект, с которым выполняется данное действие процесса;
- исполнитель действия.

Событие процесса  $e_n$  характеризуется следующим набором параметров:

$$e_n = (t_n^*, t_n^{**}, obj_s, isp_p), \quad (2)$$

где  $t_n^*, t_n^{**}$  – моменты времени начала и завершения события  $e_n$ ;  $obj_s$  – объект, с которым работает процедура процесса, отраженная в ЖРС в виде события  $e_n$ ;  $isp_p$  – исполнитель процедуры процесса, отраженной в ЖРС в виде события  $e_n$ .

При построении модели бизнес-процесса ключевым параметром является временной, который

непосредственно определяет порядок наступления событий в рамках рассматриваемой ситуации. В то время, как порядок наступления других элементов процесса представлен неявно, хотя и оказывает непосредственное влияние на временной порядок. Таким образом, журнальные файлы процессно-ориентированных систем можно описать в следующих проекциях: временной, объектной и структурной.

Порядок выполнения процедур бизнес-процесса задается отношением порядка, которое, по сути, является бизнес-правилом на исполнение БП относительно времени  $\prod_t (e_{t_k}, e_{t_l})$  для каждой пары соседних элементов  $(e_{t_k}, e_{t_l})$ :

$$c_i = \{E_j\}, E_j \subseteq E, (\forall e_{t_k}, e_{t_l} \in E_j) \exists \prod_t (e_{t_k}, e_{t_l}), \quad (3)$$

где  $E$  – множество событий в ЖРС;  $e_{t_k}, e_{t_l}$  – события, которые произошли (начали выполняться) в моменты времени  $t_k$  и  $t_l$ .

Данное отношение порядка определяет, что между событиями  $e_{t_k}, e_{t_l}$  на шкале времени в ЖРС не существует промежуточных событий:

$$\begin{aligned} & (\forall e_{t_k}, e_{t_l} \in E_j) (\neg \exists e_{t_m} \in E_j) \\ & ((t_k < t_l) \wedge (t_l > t_m) \vee (t_m > t_k)) \Rightarrow \\ & \prod_t (e_{t_k}, e_{t_l}). \end{aligned} \quad (4)$$

Отношение порядка по времени накладывает ограничения на последовательность обработки объектов и событий БП. Так, упорядоченность по объектам, а затем по времени позволяет сформировать ситуации (и выделить подмножества  $E_j$ ), связанные с последовательными цепочками действий по обработке отдельных объектов без учета организационной структуры предприятия

Аналогичным образом определяется формализованное представление отношений порядка для других проекций.

Отметим, что отношение порядка также существуют между исполнителями действий БП. Данное отношение порядка позволяет отразить иерархическую организационную структуру предприятия  $Os$ , а также задать отношение «начальник-подчиненный» для каждой пары исполнителей  $(isp_p, isp_r)$ . Следовательно, для событий  $e_n$  можно задать отношение порядка на основе организационной структуры предприятия  $\prod_{Os} (isp_p, isp_r)$ :

$$\begin{aligned} & (\forall e_{isp_p}, e_{isp_r} \in E_j) (\neg \exists e_{isp_k} \in E_j) \\ & ((isp_p < isp_r) \wedge (isp_r > isp_k) \wedge (isp_k > isp_p)) \Rightarrow \\ & \prod_{Os} (e_{isp_p}, e_{isp_r}). \end{aligned} \quad (5)$$

Упорядочивание на основе  $\prod_{Os} (isp_p, isp_r)$ , а за

тем по времени позволяет осуществить привязку полученных в результате интеллектуального анализа процессов моделей БП к существующей организационной структуре. При этом появляется возможность выстроить иерархию подмножеств  $E_j$  на основе семантики «начальник - подчиненный». Более того, набор подмножеств  $E_j$ , соответствующий начальнику, будет отражать процессы, которые выполняются в подчиненных ему подразделениях.

Формирование пар, описывающих отношения порядка исполнителей БП, позволят представить уровень организационной структуры предприятия. Также следует отметить, что упорядочивание БП по исполнителям приводит к изменению структуры ЖРС и как следствие влияет на выходную модель БП. Формируемая в результате упорядочивания по исполнителям модель БП ориентирована на отображение последовательности участия каждого из существующих исполнителей в различных этапах БП

Анализ существующих бизнес-процессов, а также журналов регистрации событий показывает, что объекты, с которыми оперирует процесс, семантически связаны между собой, и потому представляется возможным выделение отношения порядка между объектами БП. Отношение порядка на основе обрабатываемых объектов  $\prod_{Obj} (e_{obj_q}, e_{obj_d})$  имеет вид:

$$\begin{aligned} & (\forall e_{obj_q}, e_{obj_d} \in E_j) (\neg \exists e_{obj_y} \in E_j) \\ & ((obj_q < obj_d) \wedge (obj_d > obj_y) \wedge (obj_y > obj_q)) \quad (6) \\ & \Rightarrow \prod_{Obj} (e_{obj_q}, e_{obj_d}). \end{aligned}$$

На основе высказанного можно выделить следующие стратегии анализа ЖРС, каждая из которых соответствует структурному элементу БП и проекции представления БП. Выбор конкретной стратегии зависит от поставленных перед аналитиком задач:

– при формировании модели отображающей последовательность наступающих в БП событий, необходимо в первую очередь выполнить формирование набора ограничений на исполнение событий БП в форме отношений порядка по времени  $\prod_t (e_{t_k}, e_{t_l})$ ;

– при формировании модели отображающей иерархическую подчиненность исполнителей, либо очередность участия каждого из исполнителей процесса, следует выполнить формирование набора ограничений на последовательность выполнения исполнителями работ в форме отношений порядка по исполнителям  $\prod_{Os} (e_{isp_p}, e_{isp_r})$ ;

– при формировании модели, отображающей структуру объектов, очередность их формирования и зависимость одного от другого, следует выполнить формирование набора ограничений на последовательность участия объектов в рассматриваемом БП в форме отношений порядка по объектам

$$\prod_{Obj} (e_{obj_q}, e_{obj_d}).$$

Предлагаемый метод формирования ограниченной задаче структуризации журналов регистрации событий включает следующие этапы:

- 1) выбора информации из ЖРС за заданный временной интервал;
- 2) выбор стратегии анализа ЖРС в соответствии с поставленными задачами;
- 3) анализ структурных элементов ЖРС на наличие ограничений;
- 4) отсев единичных закономерностей.

Исходными данными метода являются ЖРС. Первый этап данного метода подразумевает выбор структурных элементов ЖРС за заданный временной отрезок. Также следует отметить, что важным условием для перехода на следующий этап данного метода - наличие событий ЖРС находящихся в заданном временном промежутке. В противном случае, если результат выборки записей из ЖРС за заданный временной промежуток равен пустому множеству, переход на следующий этап метода невозможен

$$L = \forall (e_i, t_i, u_i, obj_j) (t_s < t_i < t_e). \quad (7)$$

На втором этапе выполняется выбор структурного элемента ЖРС, закономерности исполнения которого предполагается отыскать. В формализованном виде данный этап имеет следующий вид:

$$Isp = isp_1, isp_2, \dots, isp_n, Isp \subseteq Isp_L, \quad (8)$$

где  $Isp$  – отобранное множество исполнителей;  $Isp_L$  – множество исполнителей всего БП.

$$Obj = obj_1, obj_2, \dots, obj_k, Obj \subseteq Obj_L, \quad (9)$$

где  $Obj$  – отобранное множество объектов;  $Obj_L$  – множество объектов всего БП.

$$E = e_{t_1}, e_{t_2}, \dots, e_{t_n}, E \subseteq E_L, \quad (10)$$

где  $Obj$  – отобранное множество событий;  $Obj_L$  – множество событий всего БП.

Третий этап подразумевает поиск одного из типов отношений порядка

$$\prod_{Os} (e_{isp_p}, e_{isp_r}), \prod_{Obj} (e_{obj_q}, e_{obj_d}) \text{ или } \prod_t (e_{t_k}, e_{t_l}).$$

На основе выявленных зависимостей формируются отношения порядка для анализируемого структурного элемента БП. В формализованном виде данный этап имеет вид:

$$\forall (e_{isp_1}, e_{isp_2}, \dots, e_{isp_p}) (\exists \prod_{Os} (e_{isp_{p-1}}, e_{isp_p})), \quad (11)$$

$$\forall (e_{obj_1}, e_{obj_2}, \dots, e_{obj_q}) (\exists \prod_{Obj} (e_{obj_{q-1}}, e_{obj_q})), \quad (12)$$

$$\forall (e_{t_1}, e_{t_2}, \dots, e_{t_k}) (\exists \prod_T (e_{t_{k-1}}, e_{t_k})). \quad (13)$$

Также следует отметить, что выявленные зависимости дополняются известными на момент проведения анализа бизнес-правилами о БП:

$$Vp = \forall \prod_{Os} (e_{isp_p}, e_{isp_r}) \wedge \forall Vp (e_{isp_p}, e_{isp_r}), \quad (14)$$

$$Vp = \bigvee_{Obj} (e_{obj_q}, e_{obj_d}) \wedge \bigvee Vp(e_{obj_q}, e_{obj_d}), \quad (15)$$

$$Vp = \bigvee_t (e_{t_k}, e_{t_l}) \wedge \bigvee Vp((e_{t_k}, e_{t_l})). \quad (16)$$

На четвертом этапе выполняется проверка допустимости существования сформированного набора ограничений для анализируемого ЖРС. В силу того, что на предыдущем этапе осуществляются объединение выявленных отношений порядка и ранее известных бизнес-правил анализируемого ЖРС, имеется вероятность, что дополненные знания могут выходить за временные рамки вошедших в выборку записей ЖРС.

На данном этапе целесообразно применить ранее упомянутую парадигму программирования задач, для проверки применимости сформированного набора правил на полученном на первом этапе наборе записей ЖРС.

Отбор событий  $E_j$  по критерию пересечения связанных с ними множеств объектов и исполнителей создает условия для решения задачи выделения бизнес-процессов по границам подразделений. Решение этой задачи позволяет на уровне подразделений устранить противоречие между «горизонтальной», «сквозной» структурой бизнес-процессов и вертикальной иерархией подразделений. Очевидно, что отобранные таким образом события необходимо упорядочить одним из рассмотренных способов. Предложенный метод формирования ограничений применим при решении задачи формирования структурированного ЖРС по средствам метода предварительной обработки журналов регистрации событий, который подразумевает выполнение следующих этапов:

- в заданном интервале времени, и связанных с этими событиями объектов;
- отбор подмножеств событий для заданного набора объектов, с которыми оперирует процесс;
- сортировка событий по времени, а затем на основе причинно следственных связей между объектами;
- выявление начальных и конечных событий будущих последовательностей для каждого отобранного подмножества;
- формирование процессно-ориентированного журнала на основе частично упорядоченных последовательностей событий.

#### МЕТОД ФОРМУВАННЯ ОБМЕЖЕНЬ В ЗАДАЧІ СТРУКТУРИЗАЦІЇ ЖУРНАЛІВ РЕЄСТРАЦІЇ ПОДІЙ

Є.О. Богатов

*У статті запропоновано метод формування обмежень при підготовці даних для process mining. Запропонований метод спрямований на побудову такої сукупності обмежень для вихідних даних з журналів реєстрації подій, котрі дозволили б сформувати моделі підпроцесів, пов'язаних з окремими елементами організаційної структури та обробкою заданих об'єктів.*

**Ключові слова:** *бізнес-процес, модель журналу реєстрації подій, відношення порядку, обмеження.*

#### METHOD OF FORMING CONSTRAINTS FOR STRUCTURING EVENT LOGS

E.O. Bogatov

*Method of forming constrains of data preprocessing for process mining was proposed in article. The method that have been proposed was oriented on building set of constrains for output data of logs, that have allowed to build of sub-process models, that were relate with distinct elements of organizational structure and processing of specified objects.*

**Keywords:** *business process, log file models, order relation, consrains.*

Таким образом, применение предложенного в статье метода позволит выделить причинно-следственные связи между структурными элементами БП, необходимые на третьем этапе метода предварительного анализа БП.

#### Выводы

Впервые предложен метод формирования ограничений в задаче структуризации журналов регистрации событий, который задает ограничения на множество событий, отражающих выполнение неструктурированных бизнес-процессов, на основе отношений порядка по времени, по иерархической организационной структуре и по обрабатываемым процессом объектам. Метод обеспечивает возможность для формирования структурированных подмножеств исходных данных, позволяющих средствами process mining построить модели подпроцессов, связанных с отдельными элементами организационной структуры и обработкой заданных объектов.

#### Список литературы

1. Schlesinger M.I. *Solution to Structural Recognition (MAX,+)-problems by their Equivalent Transformations. Part 1. In Russian: Решение (MAX,+)-задач структурного распознавания с помощью их эквивалентных преобразований. Часть 1 / M.I. Schlesinger // Control Systems and Computers. – Kiev. – 2007(1). – P. 3-15.*
2. Schlesinger M.I. *Solution to Structural Recognition (MAX,+)-problems by their Equivalent Transformations. Part 2. In Russian: Решение (MAX,+)-задач структурного распознавания с помощью их эквивалентных преобразований. Часть 2 / M.I. Schlesinger // Control Systems and Computers. – Kiev. – 2007(1). – P. 3-15.*
3. Rossi F. *Handbook of Constraint Programming / F. Rossi, P. van Beek, T. Walsh. – Pisa: Elsevier Science, 2006. – 969 p.*
4. Левыкин В. М. *Метод предварительной обработки журналов регистрации событий в задачах интеллектуального анализа процессов / В.М. Левыкин, С.Ф. Чалый, Е.О. Богатов, М.И. Дикусар // Системи управління, навігації та зв'язку. – К.: ДП «ЦНДІ НГУ», 2011. – Вип 4 (20). – С. 203-207.*

Поступила в редколлегию 1.02.2013

**Рецензент:** д-р техн. наук, проф. С.Ю. Шабанов-Кушнаренко, Харьковский национальный университет радиоэлектроники, Харьков.