

Скорость падающей и преломленной волн можно записать так:

$$\frac{\sin \theta_{iS}}{U_{LA}} = \frac{\sin \theta'_{iS}}{U_{TA}} = \frac{\sin \theta_{rS}}{U_{LB}} = \frac{\sin \theta'_{rS}}{U_{TB}}, \quad (11)$$

где θ_{iS} и θ'_{iS} – соответственно углы отраженной продольной и падающей/отраженной поперечной волн; θ_{rS} и θ'_{rS} – углы проходящих продольной и поперечной волн.

Выводы. Таким образом, исследование воздействия ультразвуковых волн на элементы коры и оценка степени поглощения волн позволят определить наиболее рациональные параметры генератора и ультразвуковых инструментов для большей степени очистки дерева от коры.

Литература

1. *Гаспарян Г.Д.* Разработка и обоснование параметров установки для окорки лесоматериалов ультразвуком: дис. ... канд. техн. наук. – Братск, 2005. – 160 с.
2. *Гаспарян Г.Д.* Теоретические и экспериментальные исследования воздействия ультразвуковых волн на кору лесоматериалов с целью его окорки. – Деп. в ВИНТИ №1399-В2006, 2006.
3. *Балдев Р., Раджендран В., Паланичами П.* Применение ультразвука. – М.: Техносфера, 2006. – 576 с.
4. *Bergmann L.* Ultrasonics, Wiley. – New York, 1988.
5. *Blitz J.* Fundamentals of Ultrasonics, 2nd Edn. Butterworth London. – 1967.



УДК 004.91

А.А. Яровая, А.С. Марченко

ПРИМЕНЕНИЕ PROCESS MINING К АНАЛИЗУ ДАННЫХ ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА

В работе предложен вариант использования Process mining для реализации аналитической обработки потоков данных электронного документооборота. Представлены алгоритмы структурной обработки данных и структуризации потоков данных архива электронного документооборота. Приведен пример оценки функционального взаимодействия подразделений внутри организации.

Ключевые слова: моделирование бизнес-процессов, идентификация процессов, аналитическая обработка процессов, журнал событий.

А.А. Yarovaya, A.S. Marchenko

PROCESS MINING APPLICATION TO THE ANALYSIS OF ELECTRONIC DOCUMENT FLOW DATA

The variant of the Process mining use to implement the analytical processing of the electronic document data streams is suggested in the work. Algorithms for data structural processing and data stream structuring of archive electronic documents are presented. The assessment example of the division functional interaction within the organization is given.

Key words: business process modeling, process identification, process analytical processing, event log.

Моделирование бизнес-процессов – одно из наиболее динамично развивающихся направлений системного анализа. С точки зрения улучшения управляемости организаций и соответствия сертификатам стандарта серии ISO 9000 разработка модели бизнес-процессов организации является одним из приоритетных вопросов деятельности организаций. Процесс построения формальной модели бизнес-процессов, состоящей из взаимосвязанных операций и потоков работ, является одним из основных методов моделирования бизнес-процессов.

Основной целью моделирования является систематизация знаний об организации и ее бизнес-процессах в графической форме, аналитическая обработка и анализ полученной информации [1–4].

Идея анализа процессов заключается в том, чтобы выявить, отследить и произвести улучшение реальных (а не предполагаемых) процессов путем извлечения знаний из журналов событий, доступных в современных информационных системах. Для анализа процессов необходимо осуществить следующую последовательность мероприятий [3, 4]:

- идентификация процессов (извлечение моделей процесса из журналов событий);
- проверка соответствия (мониторинг отклонений путем сравнения моделей и журналов событий);
- интеллектуальный анализ социальных сетей и организаций;
- автоматизированное построение имитационных моделей;
- расширение и пересмотр моделей;
- ситуационный прогноз;
- выработка рекомендаций на основе предшествующего опыта организации.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Структурирование и обработка данных электронного документооборота.
2. Составление журнала событий.
3. Аналитическая обработка процессов с помощью программного обеспечения ProM 6 (Process mining) [7, 8].

В качестве объекта исследований принята организация с вертикальной иерархической организационной структурой и слабоформализованными связями на горизонтальном уровне. В качестве исходных данных для построения исследования используются архивные данные электронного документооборота за период 2000–2008 годов. Отправной точкой для анализа процессов является журнал событий, сформированный из вышеуказанных данных системы электронного документооборота. В целях структурирования данных, структурной идентификации функциональной модели организации будут использоваться методы анализа Process mining на основе объективных данных потоков электронных документов.

Все методы анализа процессов предполагают, что возможно произвести последовательную запись событий таким образом, чтобы каждое событие было сопоставлено с операцией (четко определенным шагом в процессе) и относилось к конкретной ситуации или экземпляру процесса. В связи с вышеизложенным первостепенной задачей является автоматизированное извлечение данных из системы документооборота для идентификации структуры и построения журнала событий и дальнейшего анализа. Для решения задачи извлечения статистических данных из баз электронного документооборота используется комплекс алгоритмов, разработанный и реализованный в виде программы для ЭВМ (свидетельство РосПатента о регистрации № 2007 61 1443 от 05.04.2007 г.).

Основная структура данных, извлекаемых из системы электронного документооборота с разбивкой по основным типам документов, представлена в таблицах 1–3.

Таблица 1

Исходный вариант структуры заголовков журналов для входящих документов

Вид документа	Вид доставки	Рег. индекс	Дата	В ответ на (номер)	В ответ на (дата)	Рег. №-префикс	Рег. № ответа	Дата ответа	Куда направлено	Получатель	Заголовок	Корреспондент	Информация
---------------	--------------	-------------	------	--------------------	-------------------	----------------	---------------	-------------	-----------------	------------	-----------	---------------	------------

Исходный вариант структуры заголовков журналов для внутренних документов

Вид документа
Гриф доступа
Рег. индекс подразделения
Рег. индекс документа
Дата регистрации
Листов в экз.
Приложения
Документ подписан
Составитель документа
Куда направлено
Получатель
Примечания
Заголовок
Статус
Дата постановки на контроль
Дата снятия с контроля
Срок исполнения
Причина снятия с контроля
Исполнитель
Подразделение

Таблица 3

Исходный вариант структуры заголовков журналов для исходящих документов

Вид документа
Гриф доступа
Рег. индекс
Дата
В ответ на (номер)
В ответ на (дата)
Экземпляры
Листы в экз.
Приложения
Куда направлено
Получатель
Заголовок
Кем подписано
Исполнитель
Рег. индекс1
Рег. индекс2

Сформированные из системы электронного документооборота данные нуждаются в структурировании и дальнейшем преобразовании, поскольку полученная информация представлена в виде, не допускающем дальнейшей однозначной идентификации и обработки. А именно, кроме наличия большого количества второстепенной информации, проблемным является наличие информации о нескольких получателях в одном текстовом поле в виде списка всех получателей через разделительные символы, существует разделение частей идентификатора и закрепление его за несколькими адресатами. В качестве первичного идентификатора выбраны номер и дата документа. В структуру номера при этом заложена информация о номенклатурной принадлежности документа (тип документа, структурная принадлежность инициатора документа и его порядковый номер). Номер при выгрузке из архивных данных разделен по составляющим и может быть привязан к нескольким адресатам. Первостепенной является задача приведения выгруженных данных к структурированному виду с выделением однозначно определенного идентификатора (номера и даты документа) и ключевых полей (корреспондент и адресат документа). Последние также должны быть извлечены из текстовых информационных полей карточек электронных документов.

Для решения задачи структурирования данных, преобразования их в журнал событий и однозначной идентификации ключевых ресурсов процессов разработан алгоритм приведения исходных данных в соответствие и проведена обработка данных с помощью программных средств. Для этой цели используются скрипты, написанные на языке VBA (Virtual Basic for Application) и реализующие следующие задачи:

- преобразование ключевых идентификаторов документов;
- обработка информации об инициаторах, ресурсах процессов, получателях документов в соответствии со структурой журнала событий;
- очистка исходных данных и сформированного журнала событий от второстепенной информации (сопровождающих текстовых заметок, созданных по ходу жизненного цикла документа и содержащих информацию об особенностях прохождения этого цикла для конкретного документа).

Результат структурирования и идентификации данных для журнала событий по основным типам документов (для входящих, внутренних и исходящих документов) представлен в таблицах 4–6.

Таблица 4

Структурированный журнал событий для входящих документов

Вид документа	Вид доставки	Идентификатор	Дата	В ответ на (номер)	В ответ на (дата)	Рег.№ ответа	Дата ответа	Куда направлено	Заголовок	Корреспондент
---------------	--------------	---------------	------	--------------------	-------------------	--------------	-------------	-----------------	-----------	---------------

Таблица 5

Структурированный журнал событий для внутренних документов

Вид документа	Гриф доступа	Рег. индекс подразделения	Рег. индекс документа	Дата регистрации	Листы в экз.	Приложения	Документ подписан	Составитель документа	Куда направлено	Примечания	Заголовок	Дата постановки на контроль	Дата снятия с контроля	Причина снятия с контроля
---------------	--------------	---------------------------	-----------------------	------------------	--------------	------------	-------------------	-----------------------	-----------------	------------	-----------	-----------------------------	------------------------	---------------------------

Таблица 6

Структурированный журнал событий для исходящих документов

Вид документа	Гриф доступа	Идентификатор	Дата	В ответ на (номер)	В ответ на (дата)	Экземпляры	Листы в экз.	Приложения	Куда направлено	Заголовок	Кем подписано	Исполнитель
---------------	--------------	---------------	------	--------------------	-------------------	------------	--------------	------------	-----------------	-----------	---------------	-------------

Журналы событий могут быть использованы для проведения трех типов мероприятий по анализу процессов. Первый тип – это структурная идентификация. Метод структурной идентификации создает модели на основе журнала событий, без использования какой-либо априорной информации о процессе. Структурная идентификация процесса – самый известный метод анализа процессов. Второй тип анализа процессов – оценка соответствия процесса. Метод заключается в сравнении существующей модели процесса с журналом событий этого же самого процесса. Оценка соответствия позволяет оценить, насколько соответствуют друг другу реальное функционирование процесса, описанное журналом событий, и формальная модель этого процесса. К третьему типу мероприятий по анализу процессов относятся действия по повышению эффективности. Идея действий этого типа в том, чтобы расширить или улучшить существующую модель процесса на основании использования информации о процессе, содержащейся в журналах событий. Учитывая то, что оценка соответствия процесса направлена на сглаживание противоречий между моделью процесса и его реальным состоянием, этот третий тип анализа процессов направлен на изменение или расширение априорно описанной формальной модели [5, 7].

Для решения задачи структурной идентификации обработанные журналы, полученные из архивных данных системы электронного документооборота, необходимо преобразовать к одному из типов EventLog, используемых для автоматизированной обработки Process Mining. Тип EventLog имеет следующую структуру [6]:

Таблица 7

Структура используемого типа EventLog*

Case	Activity	Timestamp	Resource
------	----------	-----------	----------

*Case – экземпляр процесса; Activity – событие в пределах экземпляра; Timestamp – время начала или завершения; Resource – кто участвует в событии (Activity).

Для анализа процессов событий предполагается использовать программное обеспечение ProM6, зарекомендовавшее себя в качестве отличного средства для аналитической обработки процессов. ProM6 является общей системой с открытым исходным кодом для реализации анализа процессов в стандартной форме посредством всевозможных плагинов. В настоящее время эта структура имеет более 400 плагинов для процессов добычи, анализа, мониторинга и преобразования данных. Система ProM 6 работает с журналами событий в XES или MXML формате.

Для аналитической обработки процессов (Process mining) при помощи программного обеспечения ProM сформированный журнал событий необходимо преобразовать в соответствующий формат (XES) [8]. Такое преобразование форматов журнала событий производится с помощью сторонних программных средств, таких, например, как Nitro.

После преобразования очищенного и структурированного журнала с помощью программного обеспечения Nitro получен преобразованный журнал EventLog указанного типа, пригодный для программной обработки Process mining следующего вида:

Таблица 8

Структура преобразованного EventLog

Case (идентификатор)	Activity (заголовок)	Timestamp (дата)	Resource (куда направлено, подразделение)	Other (остальные поля журнала)
-------------------------	-------------------------	---------------------	---	--------------------------------------

После аналитической обработки полученного журнала событий на пробной части данных была реализована задача идентификации структурного взаимодействия программным обеспечением ProM с помощью плагина "Mine for a Handover-of-Work Social Network" [8].

Полученная схема взаимодействия, называемая "Social Network", в которой круги являются resource, представлена на рисунке.

Размеры кругов зависят от весового коэффициента во взаимодействии подразделений: чем больше круг, тем больше процессов было инициировано и адресовано подразделению. Стрелки направлены от инициатора взаимодействия к адресату, двунаправленные стрелки между подразделениями говорят о том, что оба они являются и корреспондентами, и адресатами по отношению друг к другу.

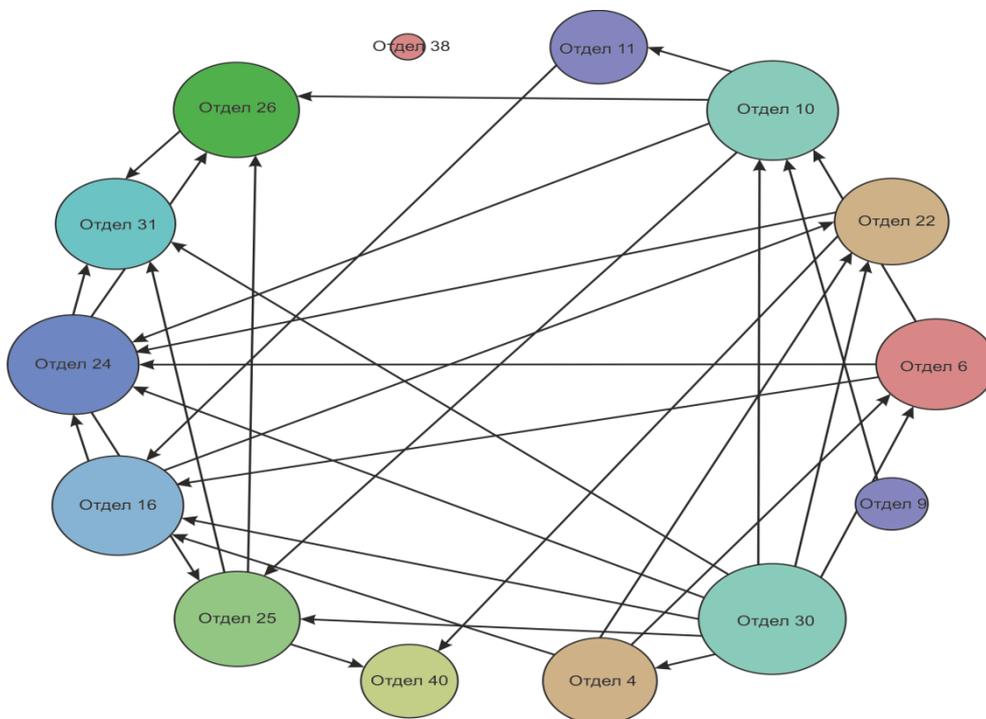


Схема взаимодействия

Таким образом, на части данных получена картина реального взаимодействия между подразделениями, обусловленная имеющимися потоками электронных данных и анализом содержимого этих потоков.

В дальнейшем предполагается решать задачу структурной идентификации данных на полном массиве архива данных электронного документооборота, задачу исследования функционального взаимодействия внутри организации – на основе полных данных архива. И кроме того, при наличии описанных реальных процессов предполагается рассмотреть задачу сопоставления заданий и данных архивных данных электронного документооборота с операциями и выявления процессов. В качестве методов анализа будут использоваться Process mining, программное обеспечение ProM применительно к объективным данным потоков электронных документов.

Литература

1. *Калянов Г.Н.* Моделирование, анализ, реорганизация и автоматизация бизнес-процессов: учеб. пособие. – М.: Финансы и статистика, 2006. – 240 с.
2. *Замятина О.М.* Моделирование и комплексный анализ бизнес-процессов предприятия. – Томск, 2004. – 174 с.
3. *Репин В.В., Елиферов В.Г.* Процессный подход к управлению. Моделирование бизнес-процессов. – М.: Стандарты и Качество, 2004. – 408 с.
4. *Мышкис А.* Элементы теории математических моделей. – Изд. 3-е, испр. – М.: КомКнига, 2007. – 192 с.
5. *Claes, Jan, Geert Poels.* Process Mining and the ProM Framework: An Exploratory Survey // In Lecture Notes in Business Information Processing. Berlin, Germany: Springer. – 2012.
6. *Aalst W.M.P. van der, Hofstede A.H.M. ter & Weske M.* Business Process Management: A Survey. International Conference on Business Process Management (BPM 2003), Springer-Verlag, Berlin. – 2003. – LCNS Vol. 2678. 1–12.
7. *Aalst W.M.P. van der.* Process mining: Discovery, Conformance and Enhancement of Business Processes. Springer-Verlag, Berlin, 2011.
8. URL: <http://www.processmining.org>.

