НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Факультет компьютерных наук
Департамент программной инженерии
Курсовая работа
Программа построения регулярного выражения по конечному автомату

Выполнил студент группы БПИ-173
Таранцова Полина Дмитриевна
Научный руководитель:
Доцент департамента программной инженерии , к ф.-м. н.
Каленкова Анна Алексеевна

ОПИСАНИЕ ПРЕДМЕТНОЙ ОБЛАСТИ

Предметная область:

Глубинный анализ процессов (Process mining) **-** область исследований, занимающаяся построением моделей процессов по логам событий.

Задача:

Разработать программу, реализующую построение регулярного выражения по конечному автомату с использованием метода удаления состояний.

ОСНОВНЫЕ ПОНЯТИЯ, ОПРЕДЕЛЕНИЯ, ТЕРМИНЫ

Термин (рус)	Термин (англ)	Определение
Регулярные выражения	Regular expressions	Способ моделирования последовательных процессов, использующий шаблоны выбора, повторения и итерации.
Метод удаления состояний	State removal method	Алгоритм построения регулярного выражения по конечному автомату, предполагающий последовательное удаление состояний автомата и одновременное построение регулярных выражений, соответствующих заменяемым системам переходов.
Конечный автомат	Finite state machine (FSM)	Конечный автомат – это кортеж $FSM = (S, E, T, s_i, S_f)$, где S – конечное множество состояний, E – конечное множество событий, $T \subseteq (S \times E \times S)$ – конечное множество переходов, $s_i \subseteq S$ – начальное состояние, и $S_f \subseteq S$ – множество конечных состояний.
Начальное состояние	Initial state	Стартовое состояние автомата.

ОСНОВНЫЕ ПОНЯТИЯ, ОПРЕДЕЛЕНИЯ, ТЕРМИНЫ

Термин (рус)	Термин (англ)	Определение
Конечные (допускающие) состояния	Accepting states	Состояния, в которых автомат может завершить работу.
Событие (Процесс)	Event	Действие, способное изменить состояние автомата.
Переход	Transition	Событие, связывающее два состояния автомата.
Подпроцесс	Subprocess	Регулярное выражение, которое является частью регулярного выражения, задающего автомат, и может быть заменено одним символом.
Метод построения регулярного выражения из конечного автомата с заменой подпроцессов	Method of constructing regular expression out of an FSM using subprocess changing	Метод построения регулярного выражения из конечного автомата, при котором переходы отождествляются с регулярными выражениями, повторяющиеся регулярные выражения именуются подпроцессами и заменяются ключевыми символами.

КЛЮЧЕВЫЕ СИМВОЛЫ РЕГУЛЯРНЫХ ВЫРАЖЕНИЙ

Обозначение	Название	Определение	Пример
.	Конкатенация	Последовательно извлеченные события.	$a.b$
+	Объединение	Возможность извлечения одного из предложенных событий.	$a+b$
*	Итерация (звезда Клини)	Возможность извлечь событие неограниченное количество раз.	$(a)^*$
()	Скобки	Подпроцесс.	(a)
&	Амперсанд	Переход в единое конечное состояние.	&
ε	Пустой переход	Переход без извлечения события.	ε

ЦЕЛЬ И ЗАДАЧИ РАБОТЫ

Цель работы

Создание программного продукта, реализующего построение регулярного выражения по конечному автомату с использованием метода удаления состояний.

Задачи работы

Создание программного продукта, реализующего возможности

- Построения конечного автомата пользователем, его сохранения и работы с ним;
- Загрузки конечного автомата из внешних источников в формате .tsml и работы с ним;
- Построения регулярного выражения по конечному автомату, сопровождающегося анимацией, демонстрирующей метод удаления состояний;
- Анализа эффективности работы эвристического алгоритма, примененного в данной программе.

АКТУАЛЬНОСТЬ ГЛУБИННОГО АНАЛИЗА ПРОЦЕССОВ

Глубинный анализ процессов (Process mining) позволяет исследовать и усовершенствовать бизнес-процессы в организациях и помогает им эффективнее и экономичнее осуществлять свою деятельность.

Данный программный продукт, в частности, был протестирован на следующих реальных логах событий:

- Финансовой системы банка;
- Исправления ошибок в работе программы;
- Системы бронирования билетов на перелеты

АНАЛИЗ СУЩЕСТВУЮЩИХ РЕШЕНИЙ

Существующие теоретические подходы:

- *Метод удаления состояний*

[Linz, P. An Introduction to Formal Language and Automata/Peter Linz – USA: Jones and Bartlett Publishers, Inc., 2006]

- *Алгоритм выбора удаляемого состояния*

[Delgado, M. Approximation to the smallest regular expression for a given regular language// Delgado, M., Morais, J.// Implementation and Application of Automata/ Domaratzki, M., Okhotin, A., Salomaa, K., Yu, S., eds. - Berlin, Heidelberg: Springer Berlin Heidelberg, 2005 -p. 312–314.]

- *Алгоритм минимизации автомата*

[13. Hopcroft, J. Introduction to Automata Theory, Languages, and Computation / John E. Hopcroft, Jeffrey D. Ullman – 3rd edition – USA: Addison-Wesley, 2001 – 358 p.]

Аналоги программного продукта:

Российских и зарубежных аналогов программного продукта не обнаружено.

ПОДХОД К РЕШЕНИЮ ЗАДАЧИ

Рассмотрим лог событий $E = \{\langle a, b\rangle, \langle a, b, c\rangle, \langle a, d, b\rangle\}$

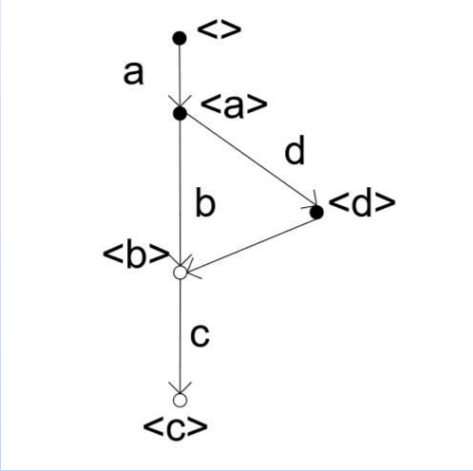


Рис. 1. Конечный автомат, построенный по логу E

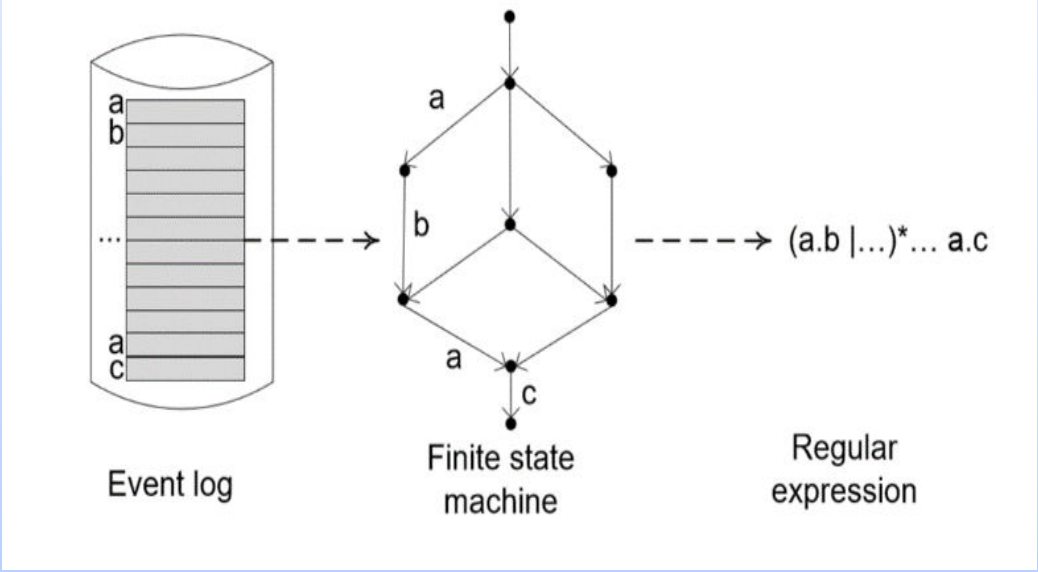


Рис. 2. Разработанный подход.

ПОСЛЕДОВАТЕЛЬНОСТЬ ЗАПУСКАЕМЫХ АЛГОРИТМОВ

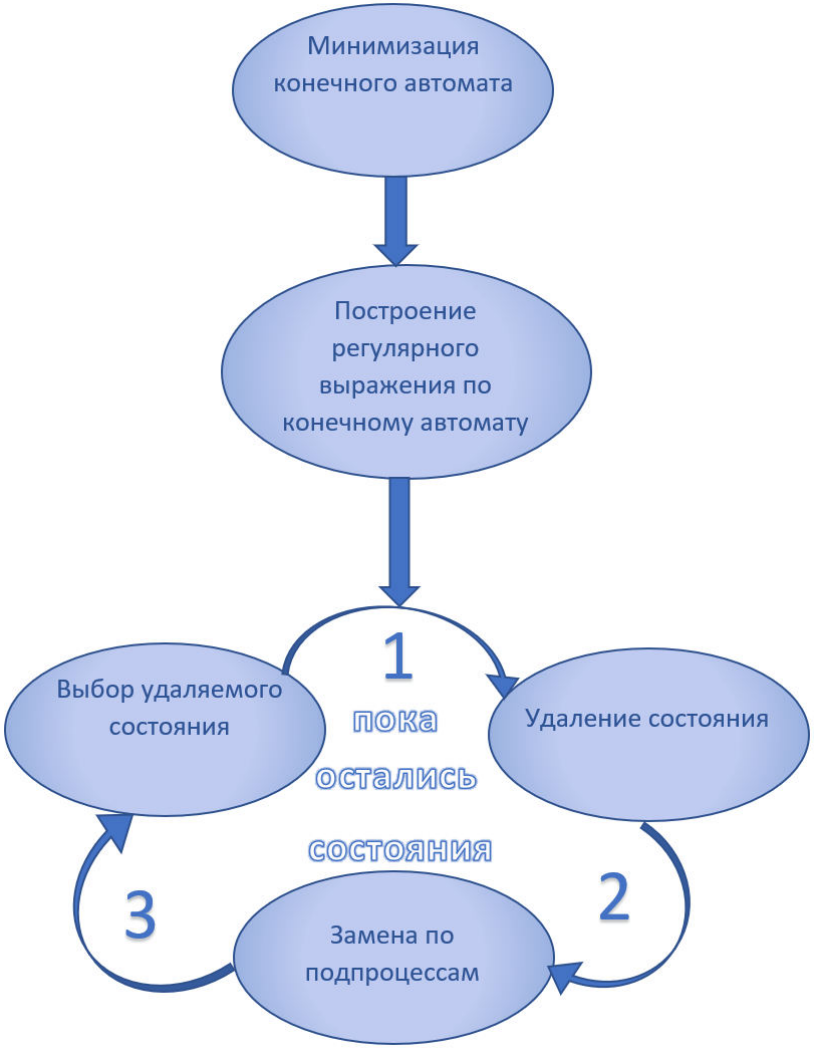


Рис. 3. Последовательность запускаемых алгоритмов.

АЛГОРИТМ УДАЛЕНИЯ СОСТОЯНИЯ

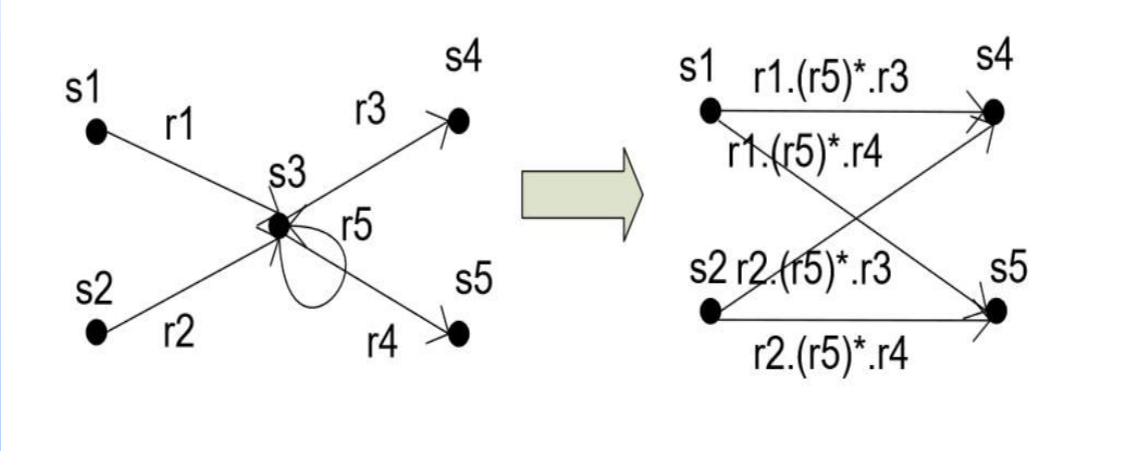


Рис. 4. Технология удаления состояния.

Алгоритм описан в следующем источнике:
[Linz, P. An Introduction to Formal Language and Automata/Peter Linz – USA: Jones and Bartlett Publishers, Inc., 2006]

АЛГОРИТМ ЗАМЕНЫ ПО ПОДПРОЦЕССАМ

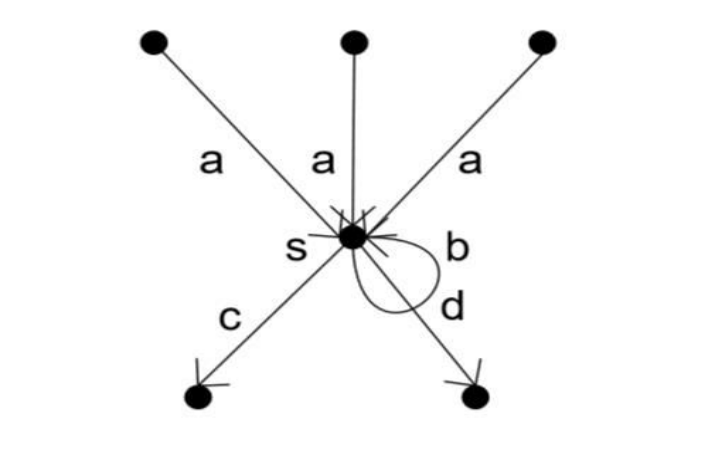


Рис. 5. Часть конечного автомата.

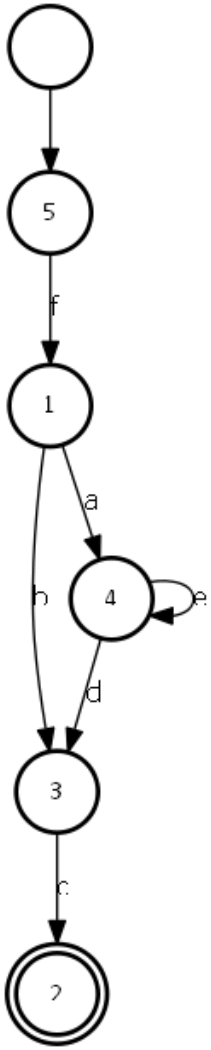
Подпроцессы в данном автомате:

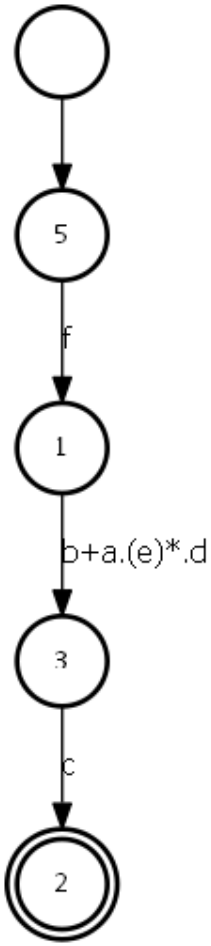
$$a.b^*.d$$
$$a.b^*.c$$
$$a.b^*$$

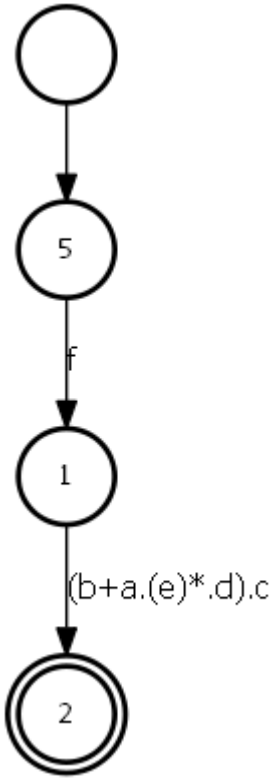
Подпроцессы:
- регулярные выражения, задающие переход;
- отдельные фрагменты этих регулярных выражений;
- префиксы

могут задавать подпроцессы и быть заменены ключевыми символами.

РЕАЛИЗАЦИЯ УДАЛЕНИЯ СОСТОЯНИЙ

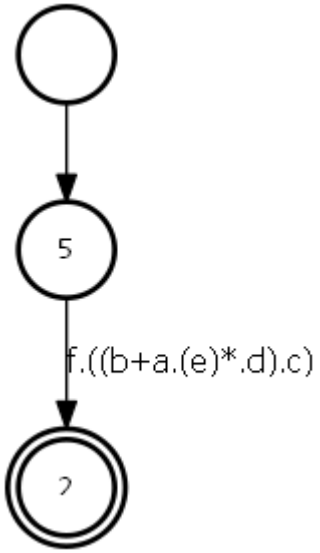






MINED REGULAR EXPRESSION:

f.((b+a.(e)*.d).c)



ТЕХНОЛОГИИ И ИНСТРУМЕНТЫ РЕАЛИЗАЦИИ

Сторонние библиотеки:

- Graphviz.2.38.0.2, 2016

Authors: Arif Bilgin, John Ellson, Emden Gansner, Yifan Hu, Arif Bilgin, Stephen North and others

https://www.nuget.org/packages/Graphviz/2.38.0.2

- QuickGraph.3.6.61119.7, 2011

Author: Jonathan Peli de Halleux

https://www.nuget.org/packages/QuickGraph/

Язык: C# 4.0

Среда программирования: Visual Studio 2017

Технологии:

- .NET Framework 4.0
- ProM
- TSML (XML)











ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

Лог	Количество состояний	Длина регулярного выражения	Длина регулярного выражения после замены подпроцессов	Длина корневого процесса	Средняя длина подпроцессов	Максимальная вложенность
ST	5	48	39	16	2.75	2
FS	7	22	21	14	2.33	1
FS	9	66	54	32	2.2	2
FS	12	397	175	99	3.8	3
FS	13	747	207	97	3.93	5
FS	19	2480	490	214	6.27	6
BS	5	19	16	11	2.5	2
BS	7	569	106	16	5.625	4
BS	9	11000	337	114	5.87	7
BS	11	41953	643	135	9.24	8

СТАТЬЯ ПО РЕЗУЛЬТАТАМ РАБОТЫ

Статья:

Tarantsova P., Kalenkova A. A. Constructing Regular Expressions from Real-life Event Logs, in: Analysis of Images, Social Networks and Texts. 7th International Conference, AIST 2018

Пути дальнейшей работы

1. *Участие в научной конференции AIST - 2018*
2. *Внедрение программного продукта: возможность использования на занятиях по дисциплине «Автоматы и модели процессов» факультета компьютерных наук НИУ ВШЭ*
3. *Разработка алгоритма синтеза алгебраических моделей процессов*

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Hopcroft, J. Introduction to Automata Theory, Languages, and Computation / John E. Hopcroft, Jeffrey D. Ullman – 3rd edition – USA: Addison-Wesley, 1979 – 358 p.
2. Neumann, C. Converting Deterministic Finite Automata to Regular Expressions [Электронный ресурс] – Режим доступа: http://liacs.leidenuniv.nl/~bonsanguemm/FI2/DFA_to_RE.pdf, свободный. (дата обращения: 24.02.18)
3. van der Aalst, W.M.P. Process Mining: Data Science in Action / Wil van der Aalst- 2nd edition – Heidelberg, Germany: Springer, 2016.
4. Leemans, S. Discovering Block-Structured Process Models from Event Logs - A Constructive Approach/ Leemans, S., Fahland, D., van der Aalst, W// Application and Theory of Petri Nets and Concurrency/ Colom, J., Desel, J., eds.// 34th International Conference, PETRI NETS 2013 - Milan, Italy// Volume 7927 of Lecture Notes in Computer Science - Berlin, Heidelberg, Germany: Springer, 2013 – p. 311–329.
5. Leemans, S. Discovering Block-Structured Process Models from Incomplete Event Logs/ Leemans, S., Fahland, D., van der Aalst, W// Application and Theory of Petri Nets and Concurrency/ Ciardo, G., Kindler, E., eds.// 35th International Conference, PETRI NETS 2014 - Tunis, Tunisia// Volume 8489 of Lecture Notes in Computer Science. Springer International Publishing – Cham, Germany: 2014- p. 91–110.
6. van der Aalst, W.M.P. Process mining: a two-step approach to balance between un- derfitting and overfitting/ van der Aalst, W.M.P., Rubin, V., Verbeek, H.M.W., van Dongen, B.F., Kindler, E., Günther, C.W.// Software & Systems Modeling 9(1) – 2008 –
7. p. 87.
8. Kleene, S.C. Representation of events in nerve nets and finite automata/ Kleene, S.C// Automata Studies/ Shannon, C., McCarthy, J., eds - Princeton, NJ, USA: Princeton University Press, 1956 – p. 3–4.1
9. Brzozowski, J.A. Derivatives of regular expressions/ Brzozowski, J.A.// J. ACM 11(4) - October 1964 – p. 481–494.
10. Linz, P. An Introduction to Formal Language and Automata/Peter Linz – USA: Jones and Bartlett Publishers, Inc., 2006.
11. Delgado, M. Approximation to the smallest regular expression for a given regular language// Delgado, M., Morais, J.// Implementation and Application of Automata/ Domaratzki, M., Okhotin, A., Salomaa, K., Yu, S., eds. - Berlin, Heidelberg: Springer Berlin Heidelberg, 2005 -p. 312–314.
12. van Dongen, B.F. The prom framework: A new era in process mining tool support/ van Dongen, B.F., de Medeiros, A.K.A., Verbeek, H.M.W., Weijters, A.J.M.M., van der Aalst, W.M.P.// Applications and Theory of Petri Nets 2005/ Ciardo, G., Darondeau, P., eds. - Berlin, Heidelberg: Springer Berlin Heidelberg, 2005 – p. 444–454.

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Спасибо за внимание!

Таранцова Полина Дмитриевна,
pdtarantsova@edu.hse.ru

Москва - 2018