



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

# Process mining — Методы синтеза и анализа моделей процессов

**Алексей Мицюк**

м.н.с. лаборатории ПОИС ФКН

Семинар САЕ «Математика, компьютерные науки и информационные технологии»  
МИЭМ им. А.Н.Тихонова НИУ ВШЭ  
Москва, 13 февраля 2017

# Данные — это нефть информационного общества

Select Administrator: Windows PowerShell

```
PS C:\Windows\system32> Get-EventLog -LogName application -Source MSSQLSERVER
```

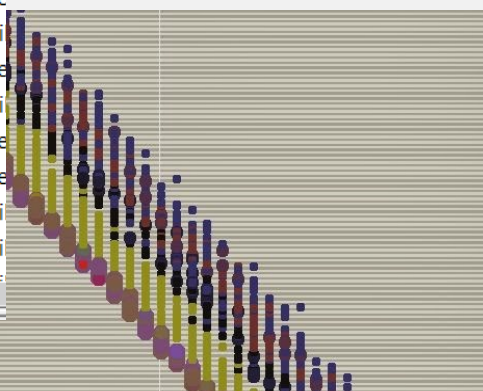
| Index | Time         | EntryType   | Source      | InstanceID | Message                                            |
|-------|--------------|-------------|-------------|------------|----------------------------------------------------|
| 11282 | Sep 09 19:59 | Information | MSSQLSERVER | 1073760856 | SQL Trace was stopped due to server shutdown. I... |
| 11281 | Sep 09 19:59 | Information | MSSQLSERVER | 1073758972 | SQL S...                                           |
| 11280 | Sep 09 19:59 | Information | MSSQLSERVER | 1073748351 | .NET                                               |
| 11279 | Sep 09 19:59 | Information | MSSQLSERVER | 1073751513 | Serv...                                            |
| 11278 | Sep 09 16:14 | Information | MSSQLSERVER | 1073759714 | A sig...                                           |
| 11274 | Sep 09 15:13 | Information | MSSQLSERVER | 1073759714 | A sig...                                           |
| 11273 | Sep 09 14:44 | Information | MSSQLSERVER | 1073759714 | A sig...                                           |
| 11272 | Sep 09 14:38 | Information | MSSQLSERVER | 1073759714 | A sig...                                           |
| 11271 | Sep 09 14:34 | Information | MSSQLSERVER | 1073759714 | A sig...                                           |
| 11268 | Sep 09 14:28 | Information | MSSQLSERVER | 1073759714 | A sig...                                           |
| 11267 | Sep 09 14:24 | Information | MSSQLSERVER | 1073759714 | A sig...                                           |
| 11266 | Sep 09 14:19 | Information | MSSQLSERVER | 1073759714 | A sig...                                           |
| 11261 | Sep 09 14:13 | Information | MSSQLSERVER | 1073759714 | A sig...                                           |
| 11260 | Sep 09 14:06 | Information | MSSQLSERVER | 1073757281 | Conf...                                            |
| 11259 | Sep 09 14:03 | Information | MSSQLSERVER | 1073749952 | Using...                                           |
| 11258 | Sep 09 14:03 | Information | MSSQLSERVER | 1073749914 | Atten...                                           |
| 11257 | Sep 09 14:03 | Information | MSSQLSERVER | 1073749952 | Using...                                           |
| 11256 | Sep 09 14:03 | Information | MSSQLSERVER | 1073749914 | Atten...                                           |
| 11255 | Sep 09 14:03 | Information | MSSQLSERVER | 1073749952 | Using...                                           |
| 11254 | Sep 09 14:03 | Information | MSSQLSERVER | 1073749914 | Atten...                                           |
| 11251 | Sep 09 14:02 | Information | MSSQLSERVER | 1073745232 | Recou...                                           |
| 11250 | Sep 09 14:02 | Information | MSSQLSERVER | 1073751512 | Serv...                                            |
| 11249 | Sep 09 14:02 | Information | MSSQLSERVER | 1073748077 | Connec...                                          |
| 11248 | Sep 09 14:02 | Information | MSSQLSERVER | 1073751490 | The I...                                           |
| 11247 | Sep 09 14:02 | Information | MSSQLSERVER | 1073751490 | The S...                                           |
| 11246 | Sep 09 14:02 | Information | MSSQLSERVER | 1073758261 | Start...                                           |
| 11245 | Sep 09 14:02 | Information | MSSQLSERVER |            |                                                    |
| 11244 | Sep 09 14:02 | Information | MSSQLSERVER |            |                                                    |
| 11243 | Sep 09 14:02 | Information | MSSQLSERVER |            |                                                    |
| 11242 | Sep 09 14:02 | Information | MSSQLSERVER |            |                                                    |
| 11241 | Sep 09 14:02 | Information | MSSQLSERVER |            |                                                    |
| 11240 | Sep 09 14:02 | Information | MSSQLSERVER |            |                                                    |

| No. | Source | Event              | Severity | Script | Alarm          | Situation | Action | Options |                    |
|-----|--------|--------------------|----------|--------|----------------|-----------|--------|---------|--------------------|
| 1   | * Any  | * SYS_NODE_DOWN    | * Any    |        | ? %m           | - None    | - None | - None  | Show               |
|     | * Any  | * SYS_NODE_UP      | * Any    |        | ✓ NODE_DOWN_%i | - None    | - None | - None  | Ackno when         |
|     | * Any  | * SYS_SERVICE_DOW  | * Any    |        | ? %m           | - None    | - None | - None  | Show service state |
|     |        | * SYS_SERVICE_UNKI |          |        |                |           |        |         |                    |

| Дата и время       | Источник              | Код события | Категория задачи     |
|--------------------|-----------------------|-------------|----------------------|
| 15.02.2015 1:51:56 | Microsoft Windows ... | 4624        | Вход в систему       |
| 15.02.2015 1:41:55 | Microsoft Windows ... | 4648        | Вход в систему       |
| 15.02.2015 1:32:38 | Microsoft Windows ... | 4648        | Вход в систему       |
| 15.02.2015 1:31:34 | Microsoft Windows ... | 4648        | Вход в систему       |
| 15.02.2015 1:30:58 | Microsoft Windows ... | 4648        | Вход в систему       |
| 15.02.2015 1:30:48 | Microsoft Windows ... | 4648        | Вход в систему       |
| 15.02.2015 1:15:58 | Microsoft Windows ... | 4648        | Вход в систему       |
| 15.02.2015 1:15:39 | Microsoft Windows ... | 4648        | Вход в систему       |
| 2015 1:07:12       | Microsoft Windows ... | 6281        | Целостность систе... |
| 2015 0:49:24       | Microsoft Windows ... | 4648        | Вход в систему       |
| 2015 0:37:45       | Microsoft Windows ... | 6281        | Целостность систе... |
| 2015 0:31:36       | Microsoft Windows ... | 4648        | Вход в систему       |
| 2015 0:26:10       | Microsoft Windows ... | 6281        | Целостность систе... |
| 2015 0:23:27       | Microsoft Windows ... | 4648        | Вход в систему       |
| 2015 0:23:08       | Microsoft Windows ... | 4648        | Вход в систему       |

Система: Событий: 41 027

| Уровень  | Дата и время        | Источ...  |
|----------|---------------------|-----------|
| Сведения | 24.03.2015 16:53:06 | Servic... |
| Ошибка   | 24.03.2015 16:41:24 | Distri... |
| Сведения | 24.03.2015 16:40:59 | Kerne...  |
| Сведения | 24.03.2015 16:40:59 | Kerne...  |
| Ошибка   | 24.03.2015 16:40:57 | Kerne...  |
| Сведения | 24.03.2015 16:40:57 | Kerne...  |
| Сведения | 24.03.2015 16:40:55 | Kerne...  |
| Ошибка   | 24.03.2015 16:40:53 | Distri... |
| Сведения | 24.03.2015 16:40:30 | Kerne...  |
| Ошибка   | 24.03.2015 16:35:55 | Distri... |
| Сведения | 24.03.2015 16:35:48 | Kerne...  |
| Сведения | 24.03.2015 16:35:47 | Kerne...  |
| Ошибка   | 24.03.2015 16:35:25 | Distri... |
| Ошибка   | 24.03.2015 16:12:45 | Distri... |



|                        |             |                                   |
|------------------------|-------------|-----------------------------------|
| Time sort (metrics):   | None        | The Application Experience ser... |
| hours                  | None        | The Multimedia Class Schedul...   |
| Component Type: Instar | None        | The Multimedia Class Schedul...   |
| # of components: 1000  | None        | The Application Experience ser... |
| items                  | None        | The Application Experience ser... |
| time(first)            | 1/1/07 1:00 | The Application Experience ser... |
| time(end)              | 2/20/07 2:3 | The Application Experience ser... |
| avg spread             | 172.57737   | The Multimedia Class Schedul...   |
| min spread             | 86.38333    | The Multimedia Class Schedul...   |
| max spread             | 250.9       | The Multimedia Class Schedul...   |
| Component ID: 1        | None        | The Multimedia Class Schedul...   |
| # of dots: 14          | None        | The Multimedia Class Schedul...   |
| items                  | None        | The Multimedia Class Schedul...   |
| time(first)            | 1/1/07 1:00 |                                   |
| time(end)              | 1/9/07 5:14 |                                   |
| avg. interval          | 15.09487    |                                   |
| min interval           | 0.0         |                                   |
| max interval           | 122.11667   |                                   |
| Component ID: 10       |             |                                   |
| # of dots: 14          |             |                                   |



# Журналы событий — специальный вид данных

Select Administrator: Windows PowerShell

```
PS C:\Windows\system32> Get-EventLog -LogName application -Source MSSQLSERVER
```

| Index | Time         | EntryType   | Source      | InstanceID |
|-------|--------------|-------------|-------------|------------|
| 11282 | Sep 09 19:59 | Information | MSSQLSERVER | 1073760856 |
| 11281 | Sep 09 19:59 | Information | MSSQLSERVER | 1073758972 |
| 11280 | Sep 09 19:59 | Information | MSSQLSERVER | 1073748351 |
| 11279 | Sep 09 19:59 | Information | MSSQLSERVER | 1073751513 |
| 11275 | Sep 09 16:14 | Information | MSSQLSERVER | 1073759714 |
| 11274 | Sep 09 15:13 | Information | MSSQLSERVER | 1073759714 |
| 11273 | Sep 09 14:44 | Information | MSSQLSERVER | 1073759714 |
| 11272 | Sep 09 14:38 | Information | MSSQLSERVER | 1073759714 |
| 11271 | Sep 09 14:34 | Information | MSSQLSERVER | 1073759714 |
| 11268 | Sep 09 14:28 | Information | MSSQLSERVER | 1073759714 |
| 11267 | Sep 09 14:24 | Information | MSSQLSERVER | 1073759714 |
| 11266 | Sep 09 14:19 | Information | MSSQLSERVER | 1073759714 |
| 11261 | Sep 09 14:13 | Information | MSSQLSERVER | 1073759714 |
| 11260 | Sep 09 14:06 | Information | MSSQLSERVER | 1073757281 |
| 11259 | Sep 09 14:03 | Information | MSSQLSERVER | 1073749952 |
| 11258 | Sep 09 14:03 | Information | MSSQLSERVER | 107374914  |
| 11257 | Sep 09 14:03 | Information | MSSQLSERVER | 1073749952 |
| 11256 | Sep 09 14:03 | Information | MSSQLSERVER | 107374914  |
| 11255 | Sep 09 14:03 | Information | MSSQLSERVER | 1073749952 |
| 11254 | Sep 09 14:03 | Information | MSSQLSERVER | 107374914  |
| 11251 | Sep 09 14:02 | Information | MSSQLSERVER | 1073745232 |
| 11250 | Sep 09 14:02 | Information | MSSQLSERVER | 1073751512 |
| 11249 | Sep 09 14:02 | Information | MSSQLSERVER | 1073748077 |
| 11248 | Sep 09 14:02 | Information | MSSQLSERVER | 1073751490 |
| 11247 | Sep 09 14:02 | Information | MSSQLSERVER | 1073751490 |
| 11246 | Sep 09 14:02 | Information | MSSQLSERVER | 1073758261 |
| 11245 | Sep 09 14:02 | Information | MSSQLSERVER |            |
| 11244 | Sep 09 14:02 | Information | MSSQLSERVER |            |
| 11243 | Sep 09 14:02 | Information | MSSQLSERVER |            |
| 11242 | Sep 09 14:02 | Information | MSSQLSERVER |            |
| 11241 | Sep 09 14:02 | Information | MSSQLSERVER |            |
| 11240 | Sep 09 14:02 | Information | MSSQLSERVER |            |

## Источники:

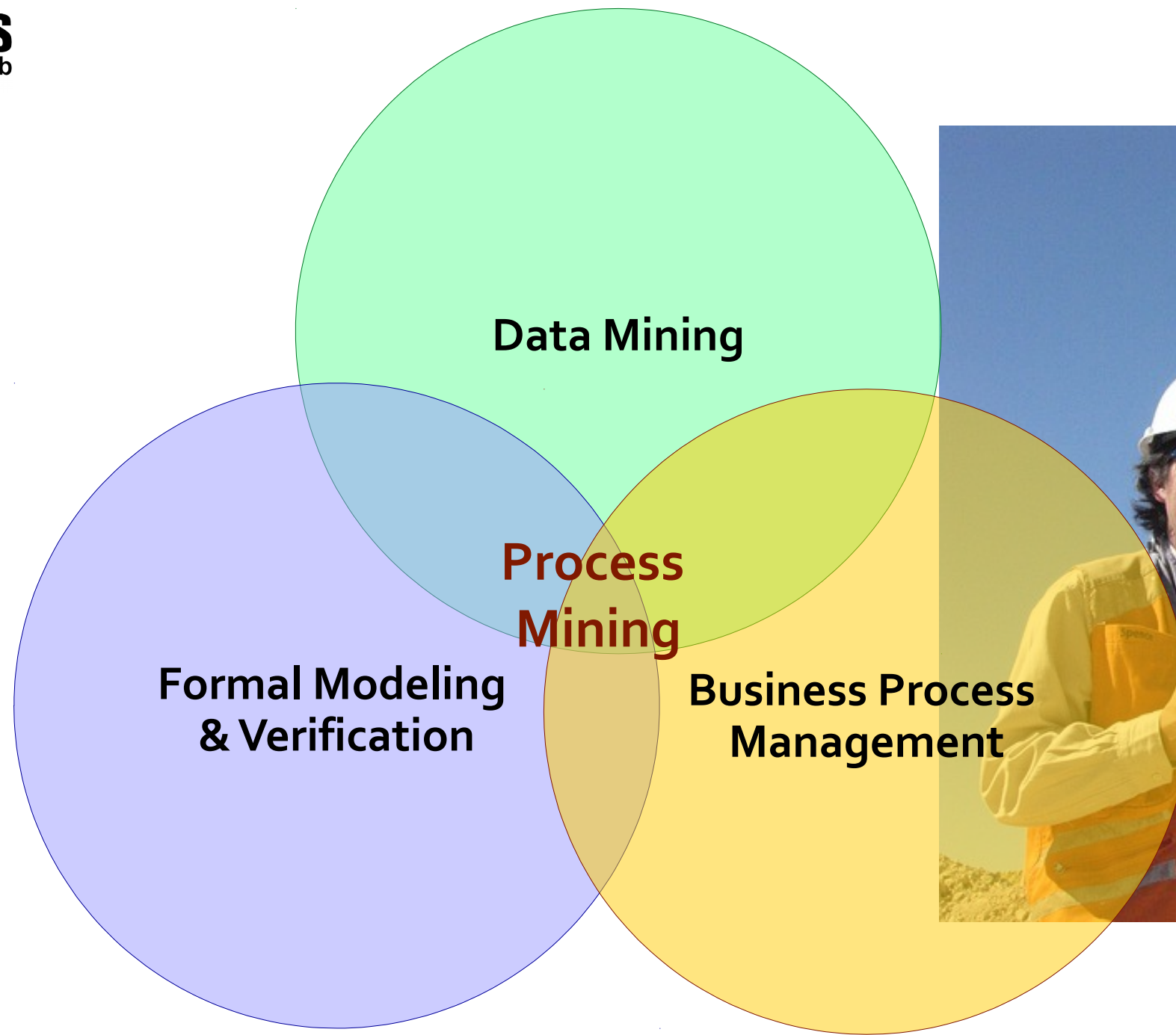
- Информационные системы
- Сетевое оборудование
- Бизнес-процессы
- Банковские транзакции
- Промышленные системы
- Образование и медицина
- Транспорт
- Устройства интернета вещей

## Как анализировать такие данные?

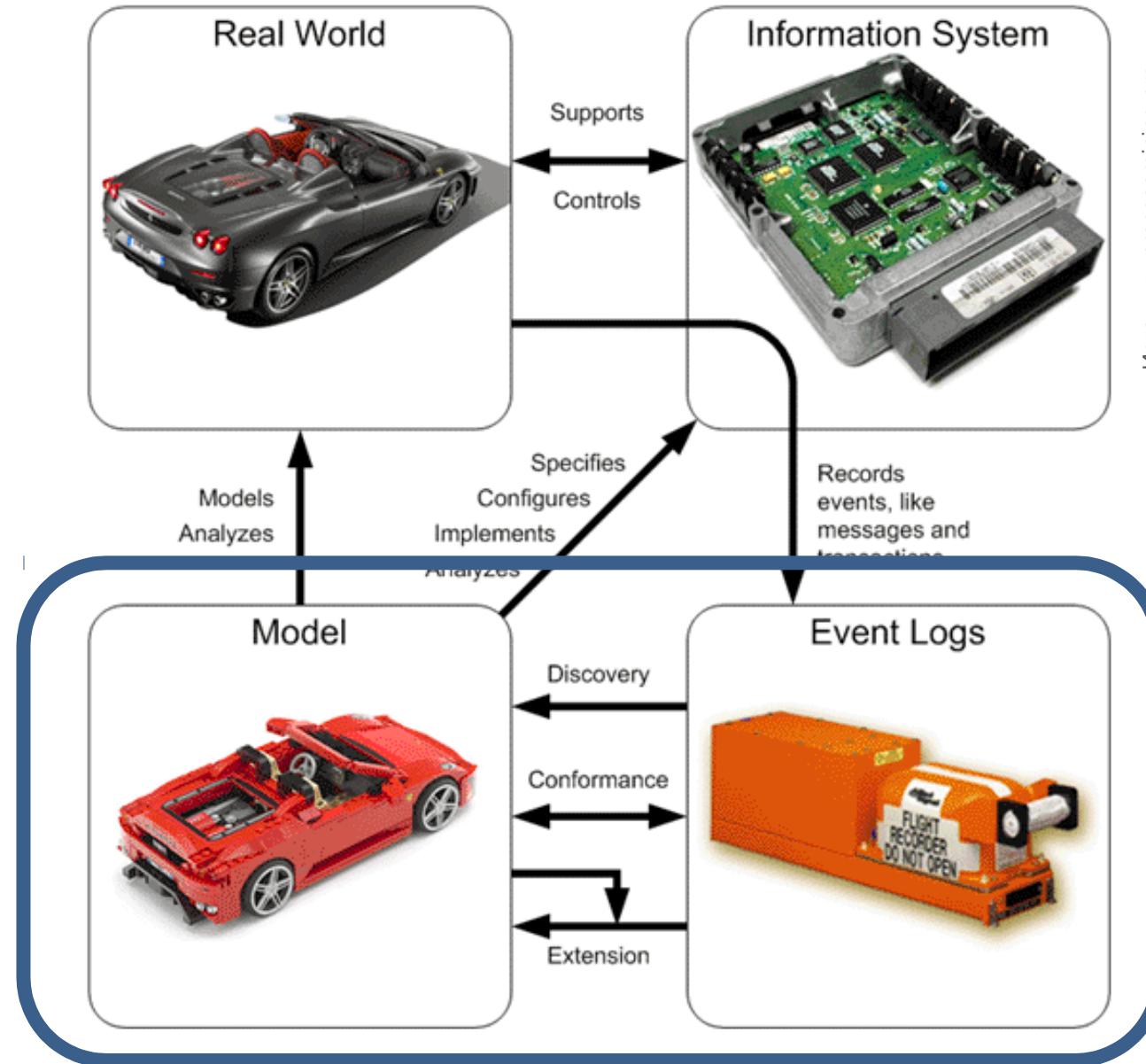
Источник: wikipedia.org

SCHOOL OF ECONOMICS  
HIGHER EDUCATION AND RESEARCH CENTER  
NATIONAL RESEARCH UNIVERSITY

(с) Алексей Мицюк, 2017



# Process mining



Источник: processmining.org



# Три основные класса задач process mining

Извлечение моделей (Process Discovery)

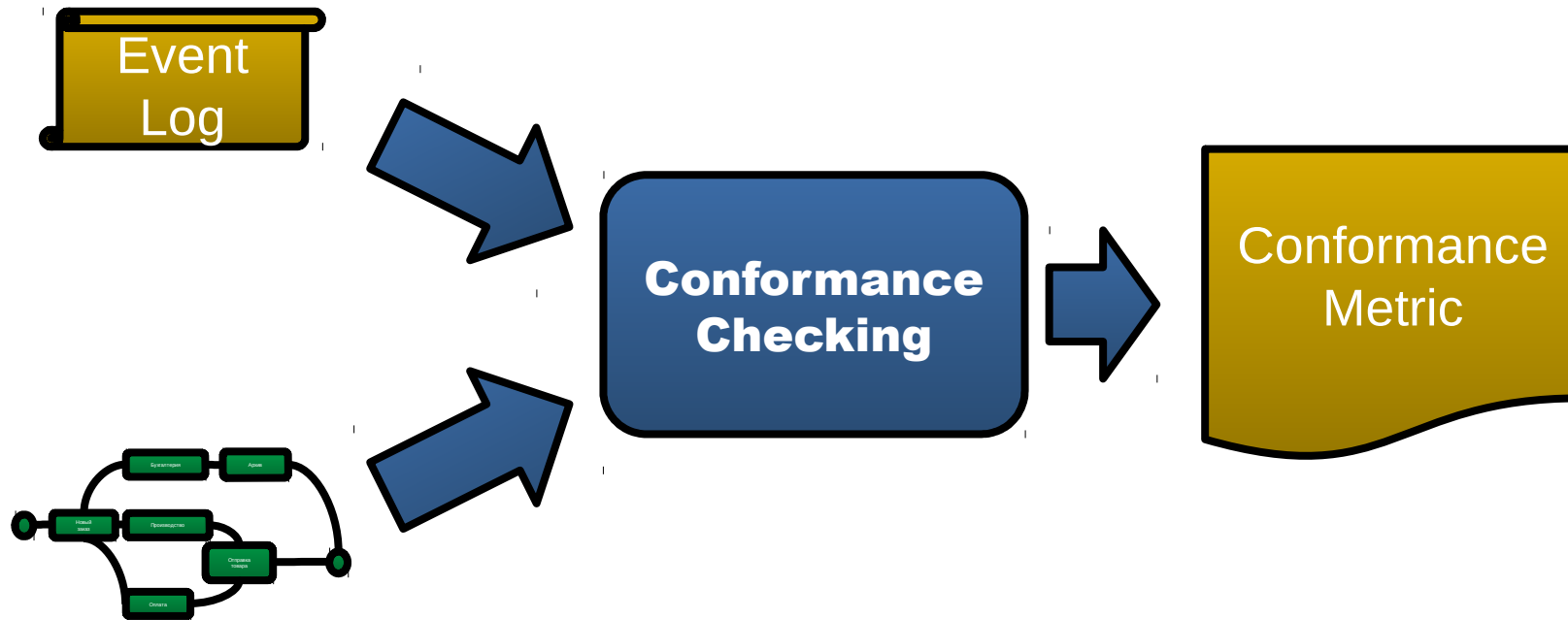
Проверка соответствия (Conformance Checking)

Усовершенствование (Process Enhancement)

# Извлечения модели процесса

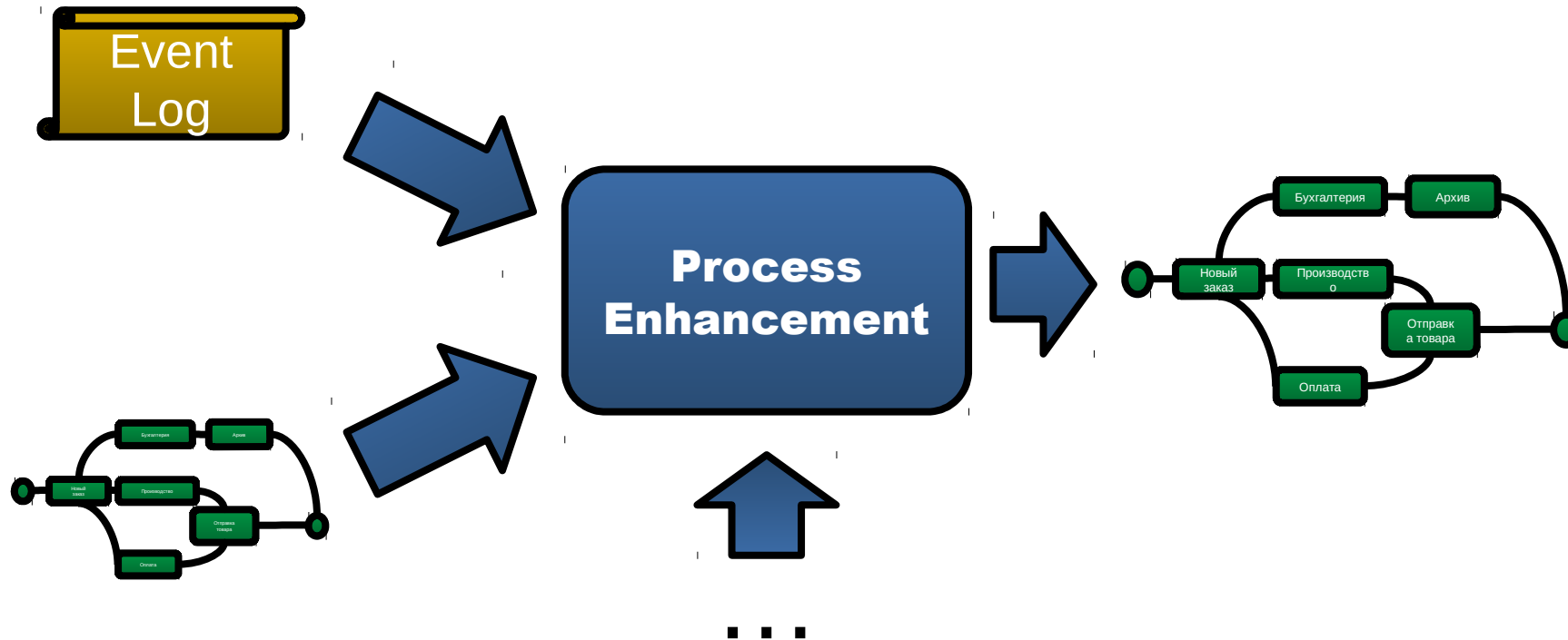


# Проверка соответствия





# Усовершенствование



# Используемые нотации

- Конечные автоматы (FA)
- Системы переходов (Transition Systems)
- Сети Петри (Petri Nets)
- Сети потокой работ (Workflow nets)
- BPMN (Business Process Model and Notation)
- BPEL (Business Process Execution Language)
- YAWL (Yet Another Workflow Language)
- EPC (Event-driven Process Chains)
- Деревья процессов (Process Trees)
- Fuzzy and Heuristic Models
- Социальные графы (Social Graphs)

INSURANCE\_PROG  
complete  
0,002

INSURED\_PERSON-CH  
complete  
0,106

FARE\_DETAIL-CLICK  
complete  
0,050

BIRTHDAY-FILL  
complete  
0,203

INSURANCE\_AGREEM  
complete  
0,004

WINDOW-UNLOAD  
complete

DOCCOUNTRY-SELEC  
complete

MIDDLE-FILL  
complete  
0,005

INSURANCE\_AGREEM  
complete  
0,035

DOCNUMBER-CLEAR  
complete  
0,003

DOCEXPURE-CLEAR  
complete  
0,005

DOCEXPURE-FILL  
complete

C\_PHONE\_NUMBER-C  
complete  
0,000

# Альфа-алгоритм: исходные данные

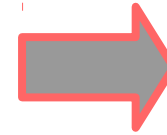
Автор алгоритма: prof. Wil van der Aalst

|       |   |   |            |       |              |
|-------|---|---|------------|-------|--------------|
| 89723 | 1 | A | 2017-01-03 | 23:01 | -----//----- |
| 89733 | 2 | A | 2017-01-03 | 23:01 | -----//----- |
| 89567 | 3 | A | 2017-01-03 | 23:02 | -----//----- |
| 89587 | 3 | B | 2017-01-03 | 23:04 | -----//----- |
| 89592 | 1 | B | 2017-01-03 | 23:05 | -----//----- |
| 89392 | 1 | C | 2017-01-03 | 23:06 | -----//----- |
| 89692 | 2 | C | 2017-01-03 | 23:06 | -----//----- |
| 89791 | 4 | A | 2017-01-03 | 23:07 | -----//----- |
| 89543 | 2 | B | 2017-01-03 | 23:09 | -----//----- |
| 89591 | 2 | D | 2017-01-03 | 23:10 | -----//----- |
| 89299 | 5 | E | 2017-01-03 | 23:10 | -----//----- |
| 89395 | 4 | C | 2017-01-03 | 23:11 | -----//----- |
| 89322 | 1 | D | 2017-01-03 | 23:12 | -----//----- |
| 89651 | 3 | C | 2017-01-03 | 23:14 | -----//----- |
| 89992 | 3 | D | 2017-01-03 | 23:15 | -----//----- |
| 89712 | 4 | B | 2017-01-03 | 23:18 | -----//----- |
| 89898 | 5 | F | 2017-01-03 | 23:19 | -----//----- |
| 89694 | 4 | D | 2017-01-03 | 23:22 | -----//----- |

Описание алгоритма: Wil M.P. van der Aalst, and Boudewijn F. van Dongen. "Discovering Petri nets from Event logs." Transactions on Petri Nets and Other Models of Concurrency VII. Springer Berlin Heidelberg, 2013. 372-422.

# Альфа-алгоритм: исходные данные

| Event id |   |   | Timestamp        |              | Additional attributes |
|----------|---|---|------------------|--------------|-----------------------|
| 89723    | 1 | A | 2017-01-03 23:01 | -----//----- |                       |
| 89733    | 2 | A | 2017-01-03 23:01 | -----//----- |                       |
| 89567    | 3 | A | 2017-01-03 23:02 | -----//----- |                       |
| 89587    | 3 | B | 2017-01-03 23:04 | -----//----- |                       |
| 89592    | 1 | B | 2017-01-03 23:05 | -----//----- |                       |
| 89392    | 1 | C | 2017-01-03 23:06 | -----//----- |                       |
| 89692    | 2 | C | 2017-01-03 23:06 | -----//----- |                       |
| 89791    | 4 | A | 2017-01-03 23:07 | -----//----- |                       |
| 89543    | 2 | B | 2017-01-03 23:09 | -----//----- |                       |
| 89591    | 2 | D | 2017-01-03 23:10 | -----//----- |                       |
| 89299    | 5 | E | 2017-01-03 23:10 | -----//----- |                       |
| 89395    | 4 | C | 2017-01-03 23:11 | -----//----- |                       |
| 89322    | 1 | D | 2017-01-03 23:12 | -----//----- |                       |
| 89651    | 3 | C | 2017-01-03 23:14 | -----//----- |                       |
| 89992    | 3 | D | 2017-01-03 23:15 | -----//----- |                       |
| 89712    | 4 | B | 2017-01-03 23:18 | -----//----- |                       |
| 89898    | 5 | F | 2017-01-03 23:19 | -----//----- |                       |
| 89694    | 4 | D | 2017-01-03 23:22 | -----//----- |                       |



Case 1: <A, B, C, D>  
Case 2: <A, C, B, D>  
Case 3: <A, B, C, D>  
Case 4: <A, C, B, D>  
Case 5: <E, F>



Variant 1: <A, B, C, D>  
Variant 2: <A, C, B, D>  
Variant 3: <E, F>

Case id

Event class



# Альфа-алгоритм: четыре отношения

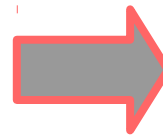
$x > y$  – (последователь) есть случай,  
когда  $y$  следует непосредственно за  $x$

$x \rightarrow y$  – (причина-следствие)  $x > y$  И НЕ  $y > x$

$x || y$  – (параллелизм)  $x > y$  И  $y > x$

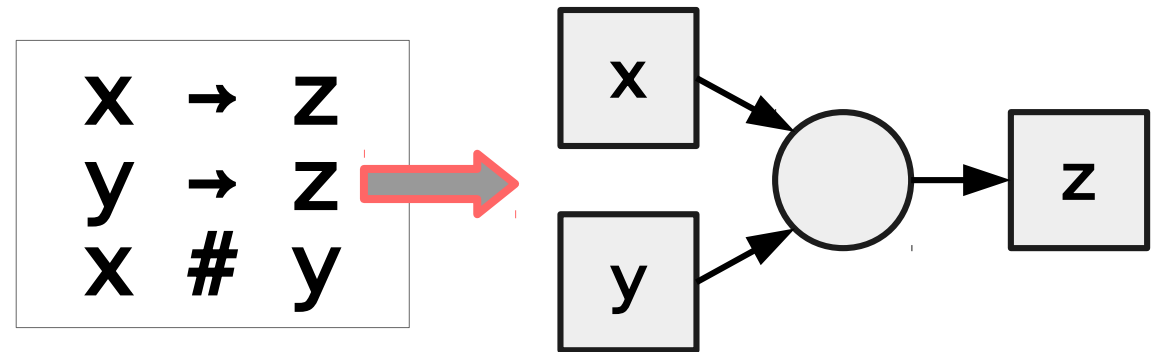
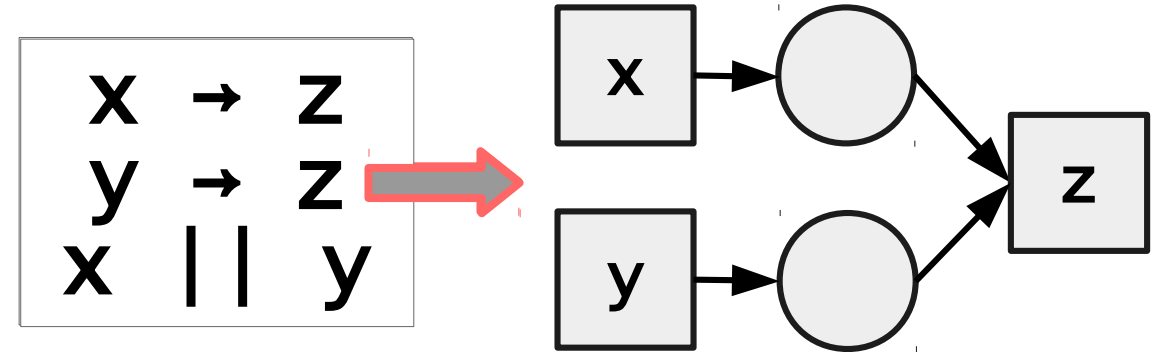
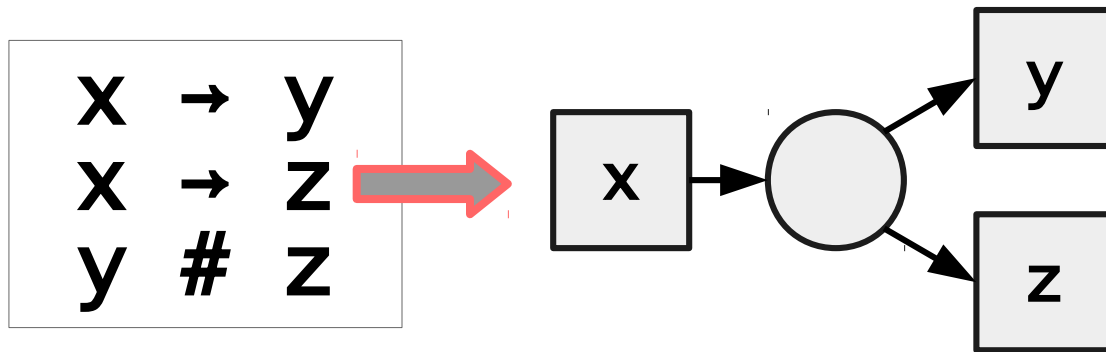
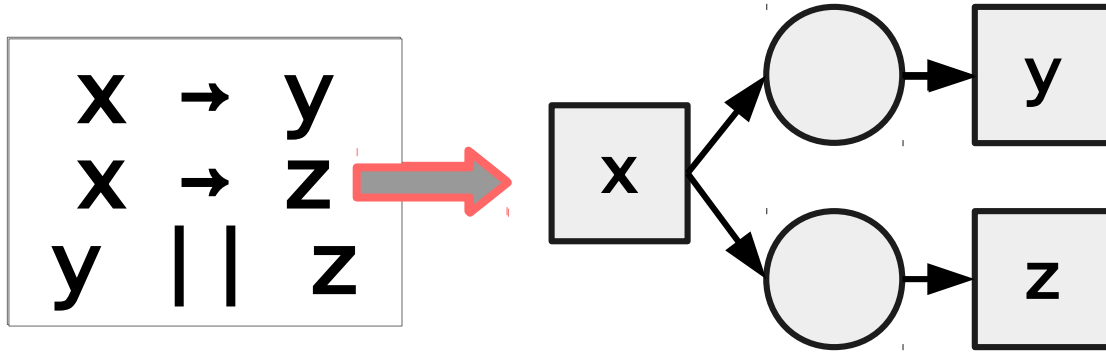
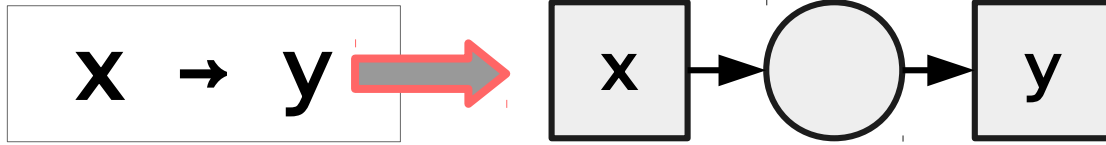
$x \# y$  – (выбор) НЕ  $x > y$  И НЕ  $y > x$

Variant 1:  $\langle A, B, C, D \rangle$   
Variant 2:  $\langle A, C, B, D \rangle$   
Variant 3:  $\langle E, F \rangle$



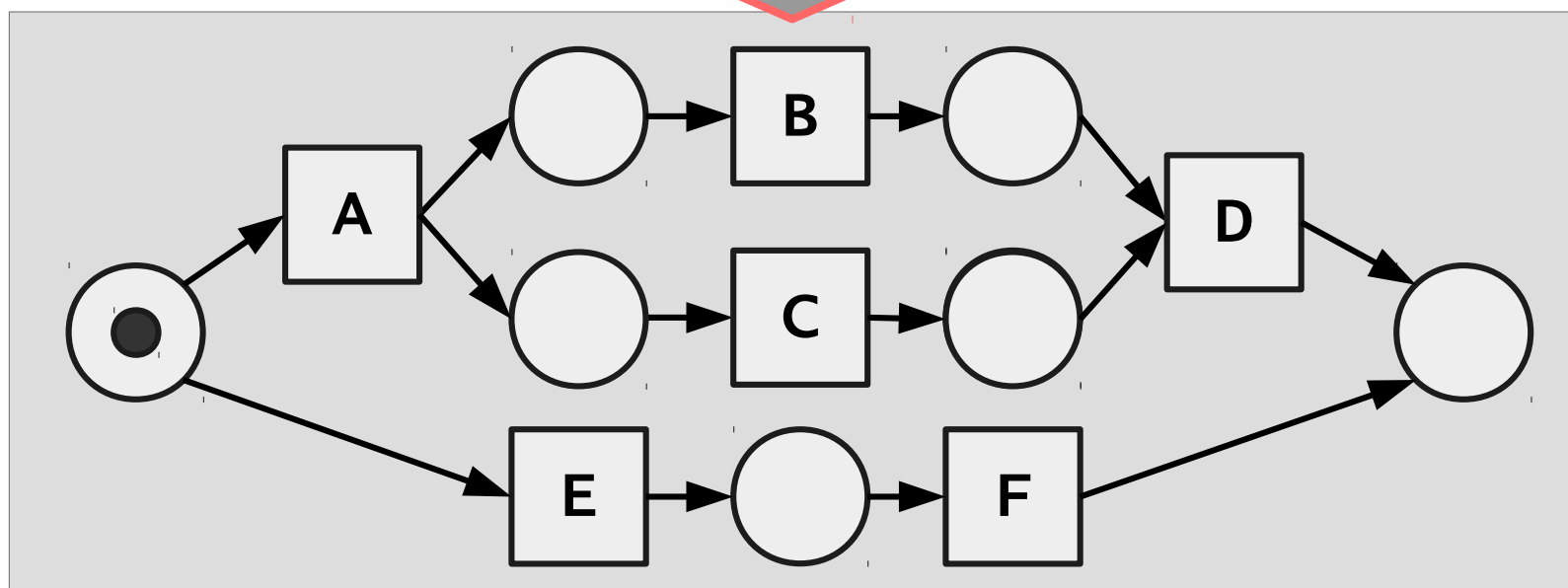
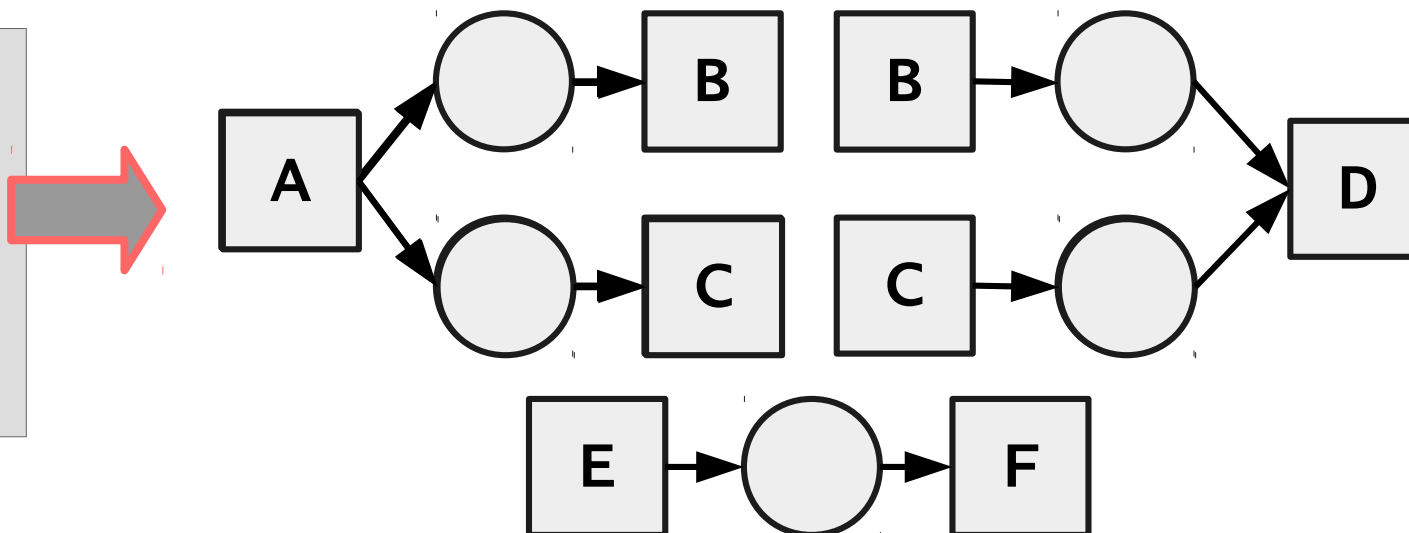
|         |          |                   |
|---------|----------|-------------------|
| $A > B$ | $B    C$ | $A \rightarrow B$ |
| $A > C$ | $C    B$ | $A \rightarrow C$ |
| $B > C$ |          | $B \rightarrow D$ |
| $B > D$ |          | $C \rightarrow D$ |
| $C > B$ |          | $E \rightarrow F$ |
| $C > D$ |          |                   |
| $E > F$ |          |                   |

# Альфа-алгоритм: переходим к сети Петри



# Альфа-алгоритм: переходим к сети Петри

|       |   |  |   |       |
|-------|---|--|---|-------|
| A > B | B |  | C | A → B |
| A > C | C |  | B | A → C |
| B > C |   |  |   | B → D |
| B > D |   |  |   | C → D |
| C > B |   |  |   | E → F |
| C > D |   |  |   |       |
| E > F |   |  |   |       |



# Направления исследований лаборатории

(подробнее: <http://pais.hse.ru/en/publications/>)

Алгоритмы извлечения из логов, синтеза, анализа и преобразования моделей процессов в виде систем переходов, моделей BPMN 2.0, высокоуровневых сетей Петри и др.

Инструмент проведения крупномасштабных экспериментов (анализ множества журналов событий одновременно)

Критерии совместимости веб-сервисов

Методы декомпозиции моделей процессов

Анализ многомерных журналов событий

Теоретические исследования в области методов анализа сетей потоков работ, вложенных сетей Петри

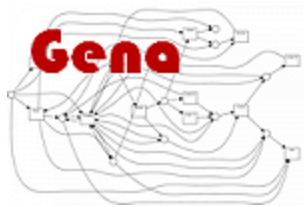
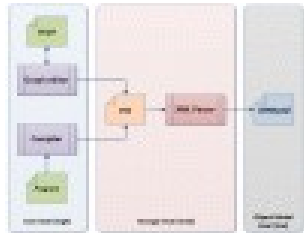
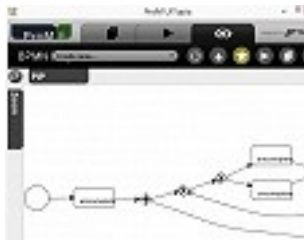
Симуляция и генерация тестовых журналов событий

Разработка программных инструментов для исследователей и инженеров

Применение методов анализа процессов в прикладных задачах

Исправление моделей процессов

# Наши инструменты



**BPMNDiffViz:** a Tool for Visual Comparison of Business Processes  
([link](#))

Discovering, Analyzing and Enhancing BPMN Models in **ProM**  
([link](#))

**DPMine:** Modeling and Process Mining Language and Tools  
([link](#))

**Gena:** Artificial Event Log Generator  
([link](#))





Спасибо за внимание!  
<http://pais.hse.ru>