Feature Aggregation in Process Mining
User Manual

January 14, 2022

1 Prerequisite

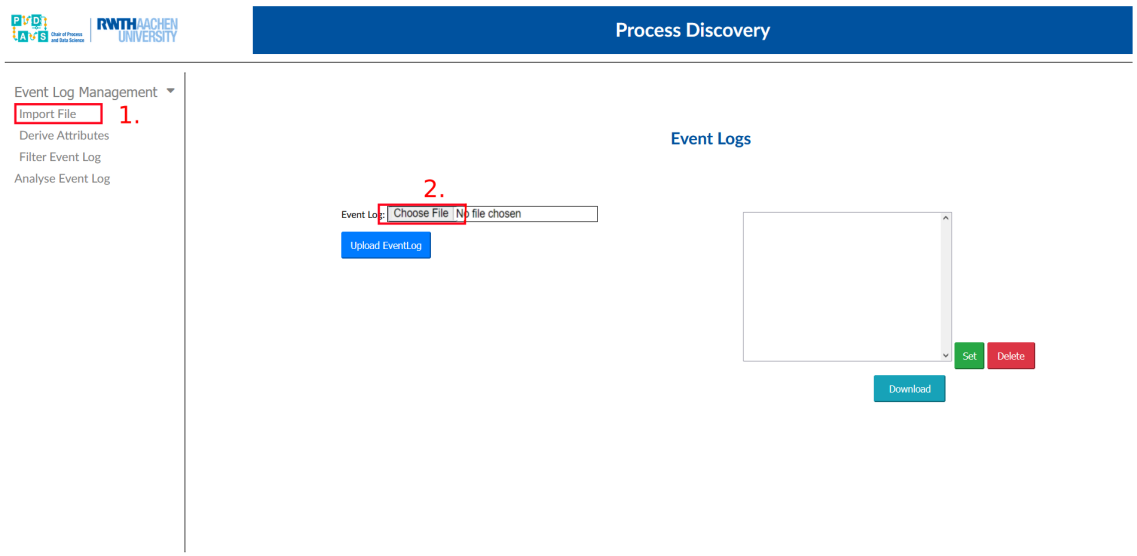
Make sure Docker is installed on your system. If not it can be installed from `https://docs.docker.com/desktop/windows/install/`.

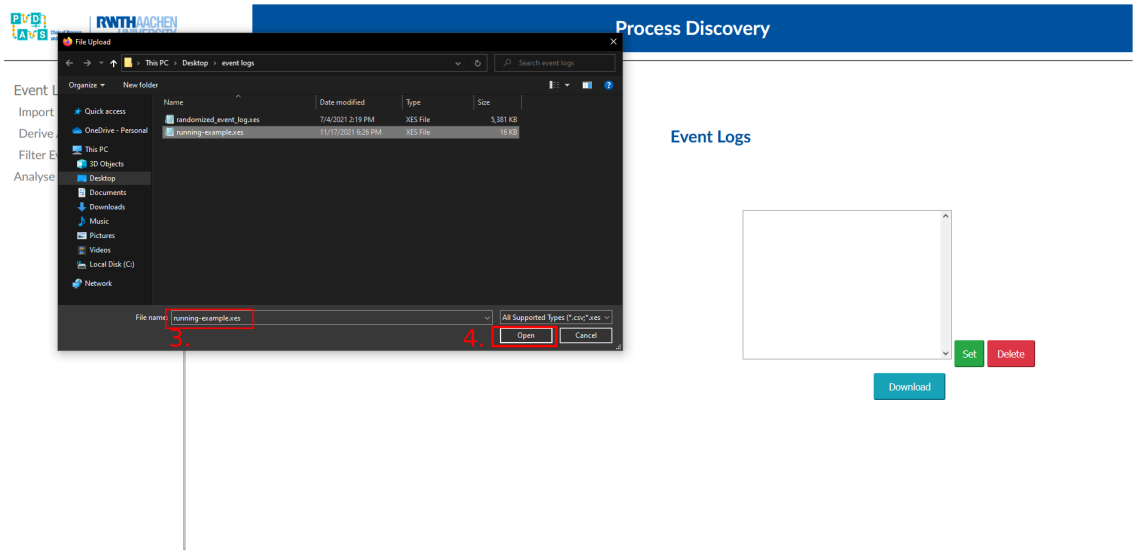
2 Start the Application

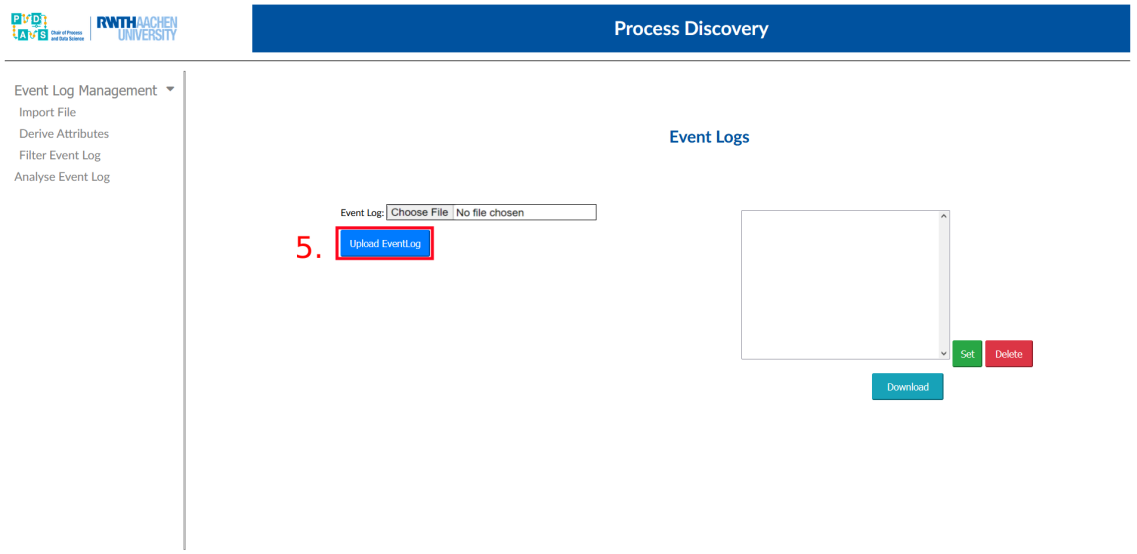
1. clone the application from:
   `https://github.com/KLiehr/WS2122---Feature-Aggregation-and-Clustering/tree/master3`
2. open the command prompt and navigate to the directory folder of the application
3. Build the docker image using command:
   docker build . -t feature_agg
4. Run the docker container using command:
   docker run -t -p 8000:8000 feature_agg
5. Open **http://127.0.0.1:8000/** or **http://localhost:8000/** in browser to start the application

3 Upload Event log

1. Press **Import File** listed under **Event Log Management**
2. Press **Choose File**
3. Select the event log you want
4. Press **Open** to upload the event log
5. Press **Upload EventLog**
6. The uploaded eventlog can now be selected


Process Discovery
Event Log Management
Import File
1.
Derive Attributes
Filter Event Log
Analyse Event Log
Event Logs
2.
Event Log:
Choose File
No file chosen
Upload EventLog
Set
Delete
Download


Process Discovery
File Upload
This PC
Desktop
event logs
Search event logs
Organize
New folder
Quick access
OneDrive - Personal
This PC
3D Objects
Desktop
Documents
Downloads
Music
Pictures
Videos
Local Disk (C:)
Network
Name
Date modified
Type
Size
randomized_event_log.xes
7/4/2021 2:19 PM
XES File
5,381 KB
running-example.xes
11/17/2021 6:26 PM
XES File
16 KB
File name:
running-example.xes
All Supported Types (*.csv;*.xes
3.
4.
Open
Cancel
Event Logs
Set
Delete
Download

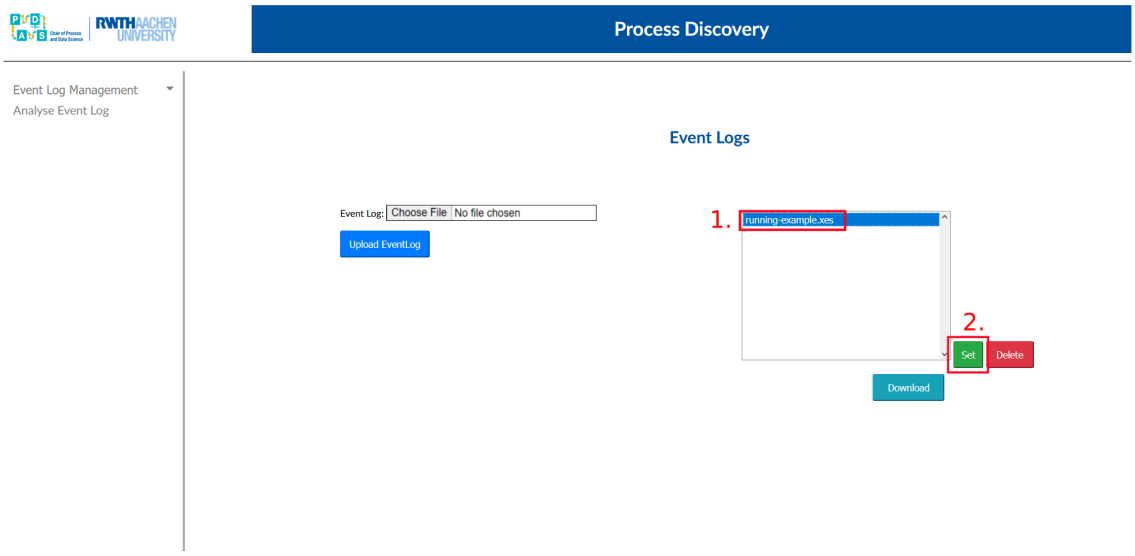

Process Discovery
Event Log Management
Import File
Derive Attributes
Filter Event Log
Analyse Event Log
Event Logs
Event Log:
Choose File
No file chosen
5.
Upload EventLog
Set
Delete
Download

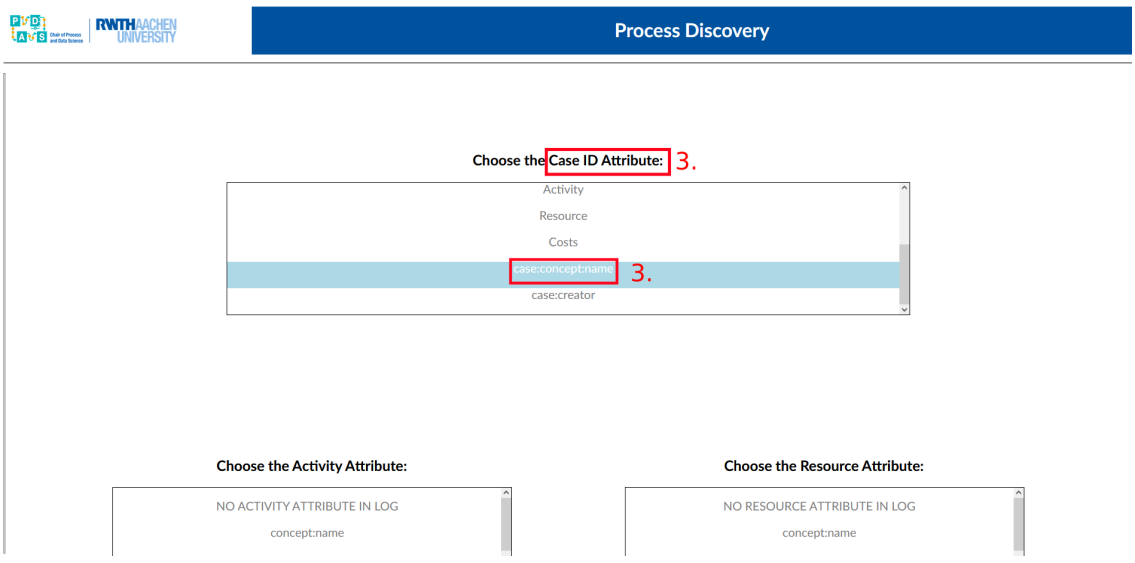
4 Set Event log

1. Press the Event log you want to select
2. Press **Set**
3. Press the name of the column in the event log corresponding to the Case ID Attribute
4. Repeat for the **Activity, Resource, Timestamp** and **Lifecycle Attribute**

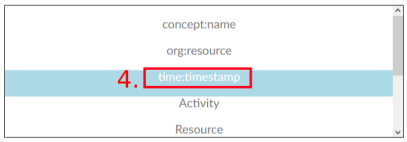
   The **timestamp** attribute should contain datetime objects, the **lifecycle** attribute should be one with the value *Start* or *Complete*. Otherwise set to NO LIFECYCLE ATTRIBUTE IN LOG.

   If an event contains two time stamp attributes, one denoting the start time of an event, the other the end time of an event, set these as start and end time respectively. Otherwise set to NO START/END TIME ATTRIBUTE IN LOG respectively.
5. Press **Save**

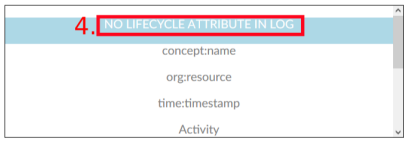


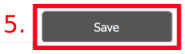


Choose the Timestamp Attribute:



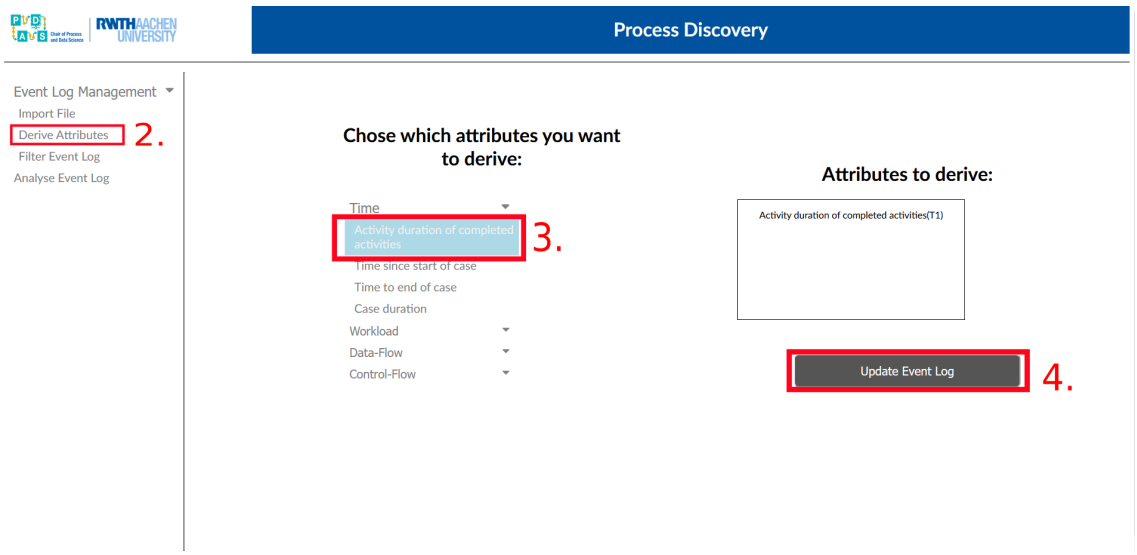
Choose the Lifecycle Attribute:





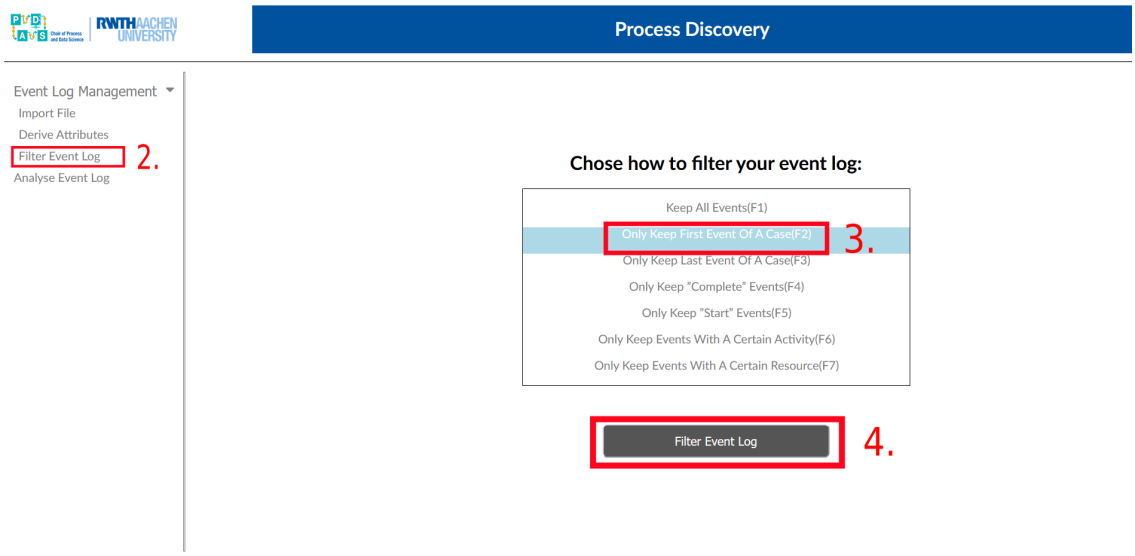
5 Derive Attributes

1. Select an event log of choice (see section 4) if not already set
2. Press **Derive Attributes** listed under **Event Log Management**
3. Choose the derived attribute by pressing the attribute category and then the attribute, multiple attributes are possible. For more info on the possible attributes that can be derived see section 9.
4. Press **Update Event Log**
5. The augmented event log can now be found in the **Import File** section



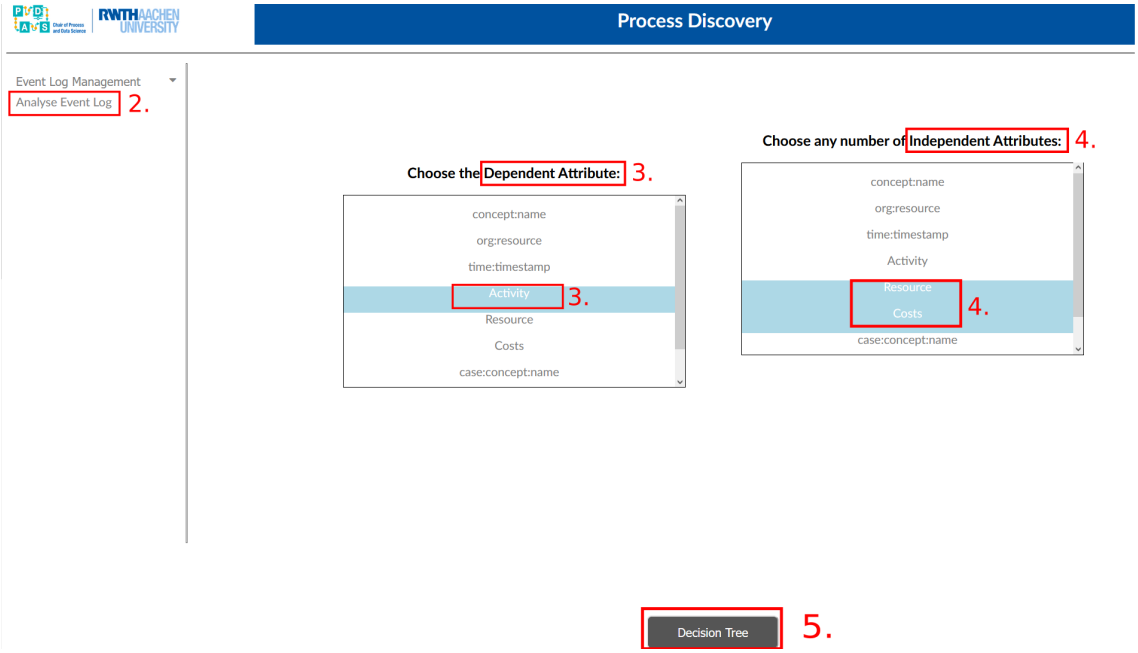
6 Filter Event Log

1. Select an event log of choice (see section 4) if not already set
2. Press **Filter Event log** listed under **Event Log Management** in the left corner.
3. Choose the filter option by pressing the filter of choice. For more info on the possible filters that can be applied see section 10.
4. Press **Filter Event Log**
5. The filtered event log can now be found in the **Import File** section

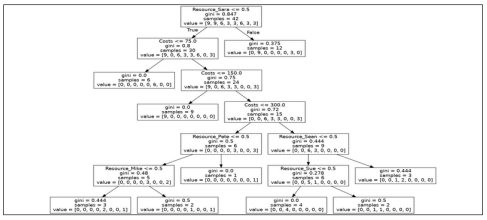


7 Use Case Analyses

1. Select an event log of choice (see section 4) if not already set
2. Press **Analyse Event Log**
3. Choose the target variable by selection it under **Dependent Attribute**
4. Choose the feature variables by selection them under **Independent Attributes**
5. Press **Decision Tree**
6. You can now see the decision tree. Press **Clustering** to continue
7. To see the different process models press **Process Model**
8. Shuffle between the process models by press **Back** or **Next**
9. The decision tree, the process model and the clusters can now all be found in the **Import File** section



Process Discovery



Back to Use Case

Clustering 6.

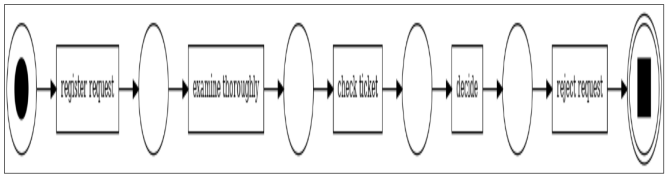
Process Discovery

Clustering completed, you can proceed to Process Mining!!!

Back to Use Case

Process Model 7.

Process Discovery



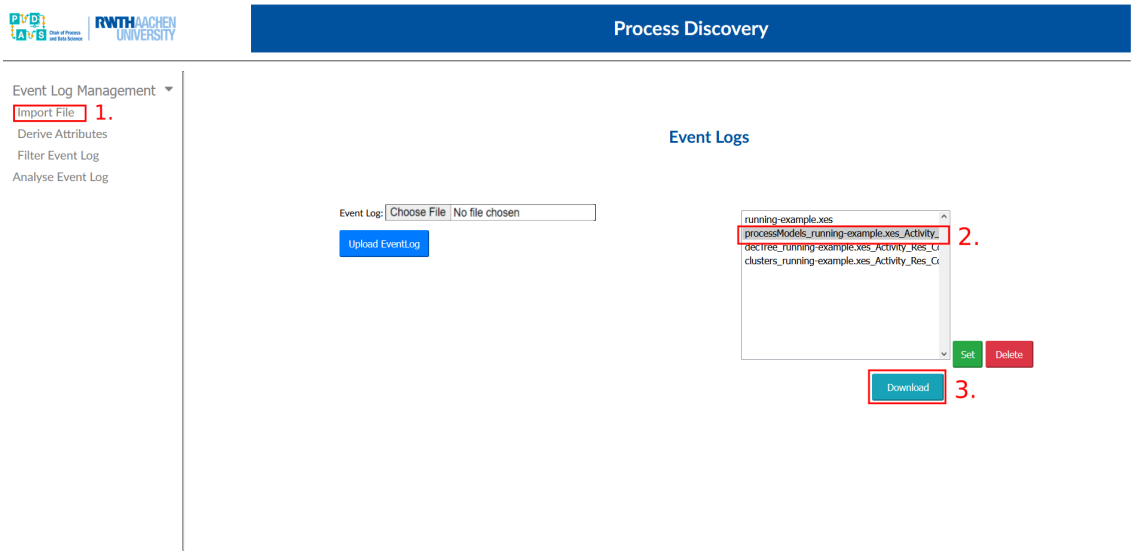
Back 8.

Back to Use Case

8. Next

8 Download Event Log

1. Press **Import File** listed under **Event Log Management**
2. Select the event log you want to download
3. Press **Download**



9 Attributes that can be derived

- Time related attributes:

  1. Activity duration of completed activities(T1):
     If a lifecycle attribute has been designated, check for the difference between start and complete event time. Should there be no start event or no lifecycle attribute, just take the event prior to the complete for the difference.
     If the start and end times were set, just subtract these.
  2. Time since start of case(T2):
     Checks for the difference of any event to the first event of the trace(ordered by timestamp).
     If start time is set, subtract start of event with start of first event of trace.
  3. Time to end of case(T3):
     Same as above, only with last event of trace.
     If end time is set, same as above with end time instead.
  4. Case duration(T4):
     Time since first event of trace to its last.
     If both start and end time are set, subtract last event of trace's end time with the trace's first event's start time.

- Workload related attributes:

  1. Total workload(R1):
     After a given event, check across the entire log to count the number of events happening after.
  2. Workload of event resource at time of event(R2):
     After a given event, check across the entire log to count the number of events happening after of the same resource as the event.

- Data flow attribute:

  The control flow attributes ALL require that the user to choose for which attribute to calculate the attributes, multiple are possible.

  Furthermore D3,4,5 and 6 only allow numerical attributes.

  1. Last assigned value of a certain attribute prior to event(D1):
     Denotes the given attributes last assigned value in the same trace prior to the current event.
  2. Last assigned value of a certain attribute after current event(D2):
     Denotes the given attributes last assigned value in the same trace at time of current event.
  3. Average value of a certain attribute after current event(D3):
     Denotes the given attributes average value in the same trace at time of current event.
  4. Max value of a certain attribute after current event(D4):
     Denotes the given attributes MAX value in the same trace at time of current event.
  5. Min value of a certain attribute after current event(D5):
     Denotes the given attributes MIN value in the same trace at time of current event.
  6. Sum value of a certain attribute after current event(D6):
     Denotes the given attributes summed up value in the same trace at time of current event.

- Control flow attributes:

  1. Number of times a certain activity is executed before an event(C1):
     Counts how often the given event's activity was executed prior in the same trace.
  2. Next activity after event(C2):
     Denotes the activity after a given event in the same trace.
  3. Activity prior to event(C3):
     Denotes the activity preceding the event in the same trace.

10 Filters that can be applied

1. Keep All Events(F1):
   Does nothing. (was used for testing)
2. Only Keep First Event Of A Case(F2):
   Creates a log which contains each trace with its first event only(ordered by timestamp).
3. Only Keep Last Event Of A Case(F3):
   Creates a log which contains each trace with its last event only(ordered by timestamp).
4. Only Keep "Complete" Events(F4):
   Keeps only events with the lifecycle attributes value being *Complete* or *complete*.
   Should not be visible if no lifecycle attribute was set.
5. Only Keep "Start" Events(F5):
   Keeps only events with the lifecycle attributes value being *Start* or *start*.
   Should not be visible if no lifecycle attribute was set.
6. Only Keep Events With A Certain Activity(F6):
   Requires user choice: Value for the activity attribute.
   Only keep events whose activity attribute is of the chosen value.
7. Only Keep Events With A Certain Resource(F7):
   Requires user choice: Value for the resource attribute.
   Only keep events whose resource attribute is of the chosen value.