Términos generales referentes a la minería de procesos:

* Evento: información relacionada con una actividad situada en una fecha de tiempo.
* Traza: sucesión de eventos para realizar un proceso. Sinónimo: caso o instancia o ejecución.
* Play-in: Inferencia de un modelo de procesos de forma automática a partir de datos de eventos sin procesar.
* Throughput: número de unidades por unidad de tiempo.
* Achieved throughput: máximo número de unidades por unidad de tiempo
* Waiting time: tiempo de espera de una actividad para poder realizarse.
* Processing time: tiempo que tarda en realizarse una actividad sin tener en cuenta el retraso del proceso (waiting time).
* Cycle time (Throughput time en Disco/Cellonis o como Sojurn time en ProM): tiempo total que tarda realizarse un proceso (processing time+Waiting time). Adicionalmente, en los artículos se menciona como transición a la suma de los cycle times de dos actividades consecutivas.
* Variante: trazas que siguen una misma secuencia de actividades y que por lo tanto siguen un patrón. Sinónimo: route (ruta).
* Process steps: etapas del proceso que engloban diferentes actividades (Submission, Judgement by various responsible roles, Payment). Estas etapas son definidas por la organización del concurso.
* Fitness score: valor de 0.0 a 1.0 que representa como de bien se ajusta un modelo a los datos de los logs.
* Happy path: variante más común de las trazas de un log.

Terminología general sobre filtrados en minería de procesos:

En la minería de procesos existen múltiples tipos de filtrados, basándonos en las técnicas disponibles en las distintas herramientas, hemos identificado en común los siguientes filtrados:

* “Variation filter”: filtrado de la variante más común. Equivalente a “Variation Filter” en Disco.
* “Time filter”: filtrado de trazas en un periodo de tiempo. Equivalente a “TimeFrame Filter” de Disco.
* “Activity filter”: filtrado de trazas que empiezan en una actividad X y acaban en una actividad Y, o que no contienen ciertas actividades, o que si contienen ciertas actividades. Equivalente a “Endpoints Filter” de Disco o “Activity selection” de Cellonis.
* “Process flow filter”: filtrado de trazas que cumplen un cierto patrón de actividades. Es equivalente a “Follower filter” en Disco o “Process flow selection” de Cellonis.
* “Attribute filter”: filtrado de trazas dependiendo de sus atributos. Es equivalente a “Attribute filter” de Disco o “Attribute selection” de Cellonis.
* “Rework filter”: filtrado de trazas en los que una actividad ocurre un número de veces superior o inferior a un límite. Equivalente a “Rework selection” de Cellonis.
* “Performance filter”: filtrado de trazas en las que el “cycle time” /”waiting time” /”processing time” entre dos actividades es superior o inferior a un umbral. Equivalente a “Throughput time selection” en Cellonis o “performance filter” de Disco.

Términos referentes a Disco:

* “Variation Filter”: sirve para identificar la variante más común. Para ello siguen estos pasos:
  + Definición de variantes
  + Cálculo del número de casos contemplado en cada variante.
  + Filtrado de variantes que no contengan un mínimo número de casos.
  + Filtrado de variantes que superen un umbral máximo de casos.
  + Representación de un histograma con las variantes restantes para decidir cuales se incluyen en la definición de la variante más común. Para ello, se revisan las frecuencias de los casos de cada variante y se establece un punto de corte a partir del cual se considerarán todas las variantes para definir la variante común.
* “Timeframe Filter”: filtrado para seleccionar trazas entre ciertas fechas.
* “Endpoints Filter” (sinónimo en Cellonis: “Activity selection”): filtrado para seleccionar casos que empiezan por una actividad X y que acaben por una actividad Z.
* “Attribute Filter” (sinónimo en Cellonis: “Attribute selection”): filtrado de casos dependiendo de un atributo, de la ausencia de una actividad, o de la presencia de una actividad.
* “Follower Filter” (sinónimo en Cellonis: “Process flow selection”): sirve para filtrar casos en los que se cumple alguna o varias de las siguientes condiciones (además de por valores de atributos y por restricciones temporales):
  + “Eventually followed”: un patrón de actividades ocurre en algún momento en el tiempo, filtrando incluso aunque no ocurren de forma seguida (por ejemplo, si queremos filtrar por las actividades X e Y da igual que Z suceda a la mitad de las otras).
  + “Directly followed”: un patrón de actividades o de valores debe ocurrir posteriormente de forma directa (es decir X debe ocurrir justo antes que Y, no se admite que ocurran en momentos distintos ordenes).
  + “Never eventually followed”: nunca pueden ocurrir la actividad X y la actividad Y en un caso,
  + “Never directly followed”: nunca pueden ocurrir las actividades X e Y de forma seguida.
* “Max repetitions mode”: Modo de visualización, que se usa para ver dónde se produce el mayor retrabajo. Los valores en el mapa de procesos muestran el número máximo de veces que se realizó una actividad para un solo caso.
* “Performance filter”: filtro de trazas que exceden “cycle time”, o cycle time de las trazas en las que exista un evento de inicio u otro final. También se puede filtrar por processing time o waiting time.

Términos referentes a Celonis:

Tipos de filtrados:

* “Attribute selection”: filtrado de trazas por atributos.
* “Activity selection”: filtrado de trazas que empiezan o acaban en ciertas actividades, o que contienen ciertas actividades.
* “Process flow selection”: filtrado de trazas en las que una actividad es seguida o no específicamente por otra actividad. Al igual que en Disco se puede seleccionar que ese patrón se siga de forma directa o no (es necesario que ese patrón de actividades se cumpla de forma inmediata), o que se siga de forma indirecta o no (es necesario que estén dos actividades, pero no de forma inmediata).
* “Throughput time selection” (Processing time selection): filtrado de trazas en las que el “cycle time” entre dos actividades es superior o inferior a un umbral.
* “Rework selection”: filtrado de trazas en los que una actividad ocurre un número de veces superior o inferior a un límite.

Otros métodos:

* “Conformance checker”: método para comparar un modelo de procesos ideal con el existente en un log para saber en qué grado se ajusta el modelo con los datos.

Términos referentes a ProM:

* ActiTrac:

Plugin de ProM, el cual es un algoritmo de “clustering” que agrupa trazas con comportamiento similar para realizar el descubrimiento de procesos (<https://ieeexplore.ieee.org/document/6507222>). Sigue a grandes rasgos los siguientes pasos:

* + Selección de las trazas que van a formar unos grupos iniciales dependiendo de su frecuencia, usando las que son más frecuentes.
  + Agrupación del resto de trazas en los grupos si mantiene la cohesión del grupo (se mide la cohesión midiendo el fitness del modelo de proceso que se crearía). Sino cumplen esa condición se crean otros grupos hasta alcanzar un máximo de grupos.
  + Las trazas restantes se agruparán con el grupo que menos empeoren (midiendo el fitness que tendría el modelo proceso si se añadiese).
  + Finalmente se obtienen procesos de los grupos generados, siendo el grupo más numeroso el “happy path” para esos datos.
* Replay a log on a Petri Net (<http://www.processmining.org/online/conformance_checker>):

Se vuelven a ejecutar las trazas de los logs en una red de Petri para comprobar si faltan tokens en el proceso o en que partes del proceso se producen retrasos (se puede entender token como la condición necesaria para que se cumpla una actividad). Finalmente se puede obtener un mapa de procesos indicando donde se producen más retrasos o faltan más tokens.

* Process Comparator:

Este plugin recibe dos grupos de eventos y sirve para comparar sus correspondientes mapas de procesos.

* Inductive miner:

Algoritmo para descubrir los procesos de forma inductiva separando las actividades desde un nivel más bajo.

* Social network mining:

Algoritmo para descubrir redes sociales en minería de procesos.

Más información en:

<http://www.processmining.org/_media/publications/aalst2005g.pdf>

<https://www.futurelearn.com/info/courses/process-mining/0/steps/15652>

* Doing similar task miner:

Plugin para detectar los roles de los recursos involucrados en un proceso.

Términos referentes a Python:

Librerías de Python:

* Librería Sberbank: <https://pypi.org/project/sberbank/>

Algoritmos de procesamiento de datos:

* Algoritmo t-sne: técnica no supervisada que sirve para reducir la dimensionalidad de los datos. Más información en: <https://sitiobigdata.com/2018/08/27/algoritmo-t-sne-con-python/>

Tests estadísticos:

* ANOVA: consiste en comparar las desviaciones de los valores medios de distintos grupos para encontrar diferencias significativas. Más información en: <https://www.cienciadedatos.net/documentos/19_anova>