

## Folyamatbányászat – beadandó

### Bevezetés – folyamatbányászat, a ProM rendszer

Mai világunkban mindennapi életünket folyamatok hálózata be, legyen szó akár munkáról, oktatásról, vásárlásról, szórakozásról. A folyamatbányászat az ezen folyamatok lefutása által generált naplófájlokból való információ, tudás kinyerésével foglalkozó tudományág. Az információ pedig különösen nagy értéket képvisel manapság, akinek sok, különféle információ áll rendelkezésére, az egyfajta hatalmat is jelent számára.

A ProM egy ingyenes letölthető, bővíthető keretrendszer, amely számtalan különböző bányászati algoritmust bocsát rendelkezésünkre pluginok formájában, ezzel támogatva a folyamatbányászati munkát. A következőkben egy rendelkezésünkre álló logfájlt fogunk ezen program segítségével feldolgozni és az eredményt jobban megvizsgálni.

### A rendelkezésünkre álló logfájl

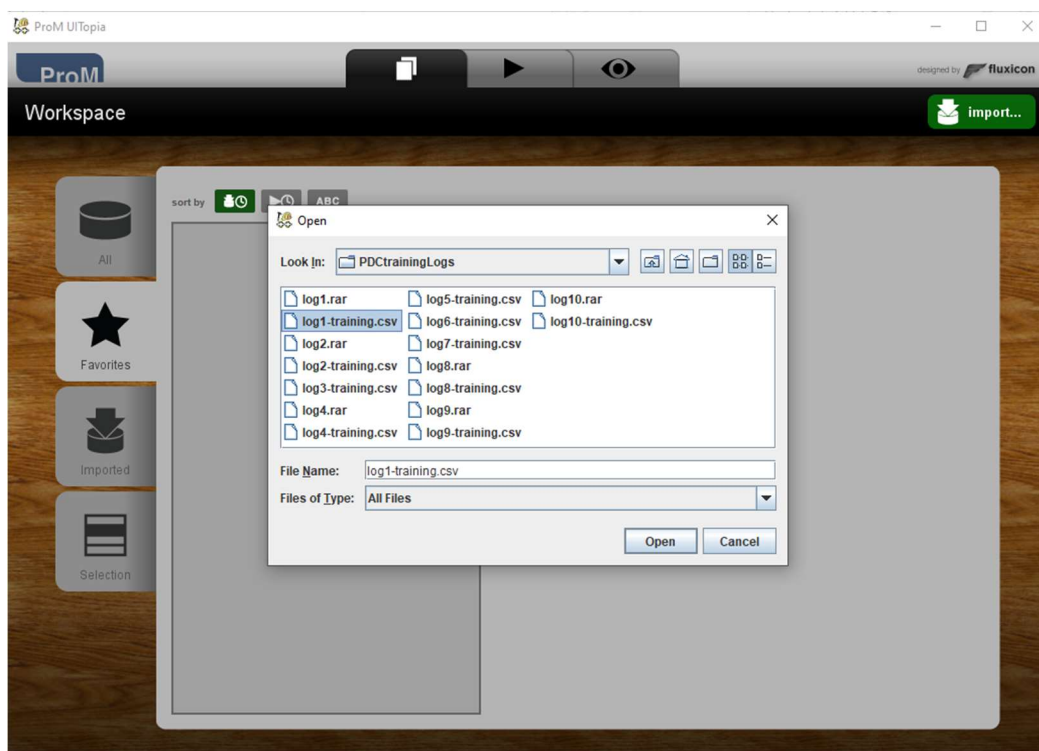
	A
1	case,event,z_2,z_1,z_4,z_3,z_6,z_5,z_8,z_7,z_9,z_10,z_11
2	0,c,0.848164873,0.79712514,0.680985114,0.195903765,FALSE,0.229026421,0.612067226,0.384932612,TRUE,0.821351412,FALSE
3	0,h,0.848164873,0.79712514,0.680985114,0.195903765,FALSE,0.229026421,0.612067226,0.384932612,TRUE,0.821351412,FALSE
4	0,j,0.848164873,0.79712514,0.680985114,0.195903765,FALSE,0.229026421,0.612067226,0.384932612,TRUE,0.821351412,FALSE
5	0,af,0.848164873,0.79712514,0.680985114,0.195903765,FALSE,0.229026421,0.612067226,0.384932612,TRUE,0.821351412,FALSE
6	0,ao,0.848164873,0.79712514,0.680985114,0.195903765,FALSE,0.229026421,0.612067226,0.384932612,TRUE,0.821351412,FALSE
7	0,m,0.848164873,0.79712514,0.680985114,0.195903765,FALSE,0.229026421,0.612067226,0.384932612,TRUE,0.821351412,FALSE
8	0,d,0.848164873,0.79712514,0.680985114,0.195903765,FALSE,0.229026421,0.612067226,0.384932612,TRUE,0.821351412,FALSE
9	0,a,0.848164873,0.79712514,0.680985114,0.195903765,FALSE,0.229026421,0.612067226,0.384932612,TRUE,0.821351412,FALSE
10	0,g,0.848164873,0.79712514,0.680985114,0.195903765,FALSE,0.229026421,0.612067226,0.384932612,TRUE,0.821351412,FALSE
11	0,ag,0.848164873,0.79712514,0.680985114,0.195903765,FALSE,0.229026421,0.612067226,0.384932612,TRUE,0.821351412,FALSE
12	0,an,0.848164873,0.79712514,0.680985114,0.195903765,FALSE,0.229026421,0.612067226,0.384932612,TRUE,0.821351412,FALSE
13	0,b,0.848164873,0.79712514,0.680985114,0.195903765,FALSE,0.229026421,0.612067226,0.384932612,TRUE,0.821351412,FALSE
14	0,y,0.848164873,0.79712514,0.680985114,0.195903765,FALSE,0.229026421,0.612067226,0.384932612,TRUE,0.821351412,FALSE
15	0,ae,0.848164873,0.79712514,0.680985114,0.195903765,FALSE,0.229026421,0.612067226,0.384932612,TRUE,0.821351412,FALSE
16	0,ac,0.848164873,0.79712514,0.680985114,0.195903765,FALSE,0.229026421,0.612067226,0.384932612,TRUE,0.821351412,FALSE
17	0,f,0.848164873,0.79712514,0.680985114,0.195903765,FALSE,0.229026421,0.612067226,0.384932612,TRUE,0.821351412,FALSE
18	1,as,0.868846197,0.538704679,0.071917416,0.411046626,FALSE,0.257845111,0.68149208,0.199458918,TRUE,0.932251847,TRUE
19	1,b,0.868846197,0.538704679,0.071917416,0.411046626,FALSE,0.257845111,0.68149208,0.199458918,TRUE,0.932251847,TRUE
20	1,c,0.868846197,0.538704679,0.071917416,0.411046626,FALSE,0.257845111,0.68149208,0.199458918,TRUE,0.932251847,TRUE
21	1,q,0.868846197,0.538704679,0.071917416,0.411046626,FALSE,0.257845111,0.68149208,0.199458918,TRUE,0.932251847,TRUE

A rendelkezésünkre ezúttal csv (comma-separated values) fájlok állnak, amelyekben az értékek egymástól egy „ , ” karakterrel vannak elválasztva. A 10 lehetséges training log közül én az 1-eset fogom feldolgozni. A fájl felső sorában láthatjuk az attribútumok neveit, amelyek

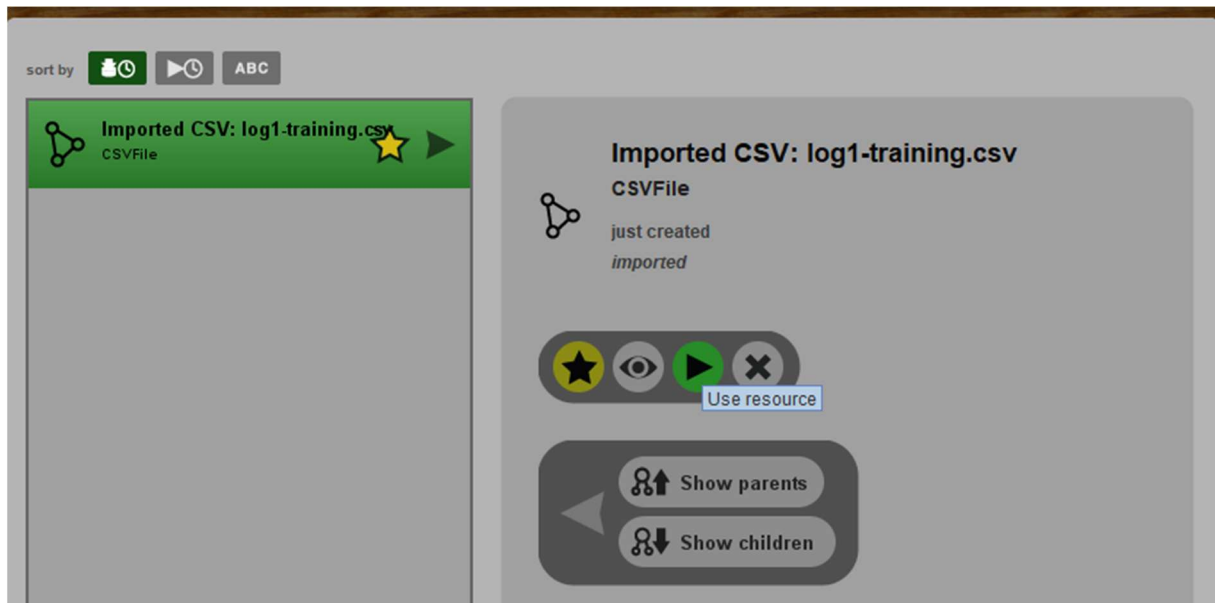
jelen esetben a case, azaz a lefutás azonosítója, az event, azaz a futások során megtörtént események, valamint z\_1, ..., z\_11 egyéb, különböző típusú attribútumok. A ProM rendszerben most ezzel a fájlal fogunk dolgozni.

## A ProM rendszer használata – logfájl importálása

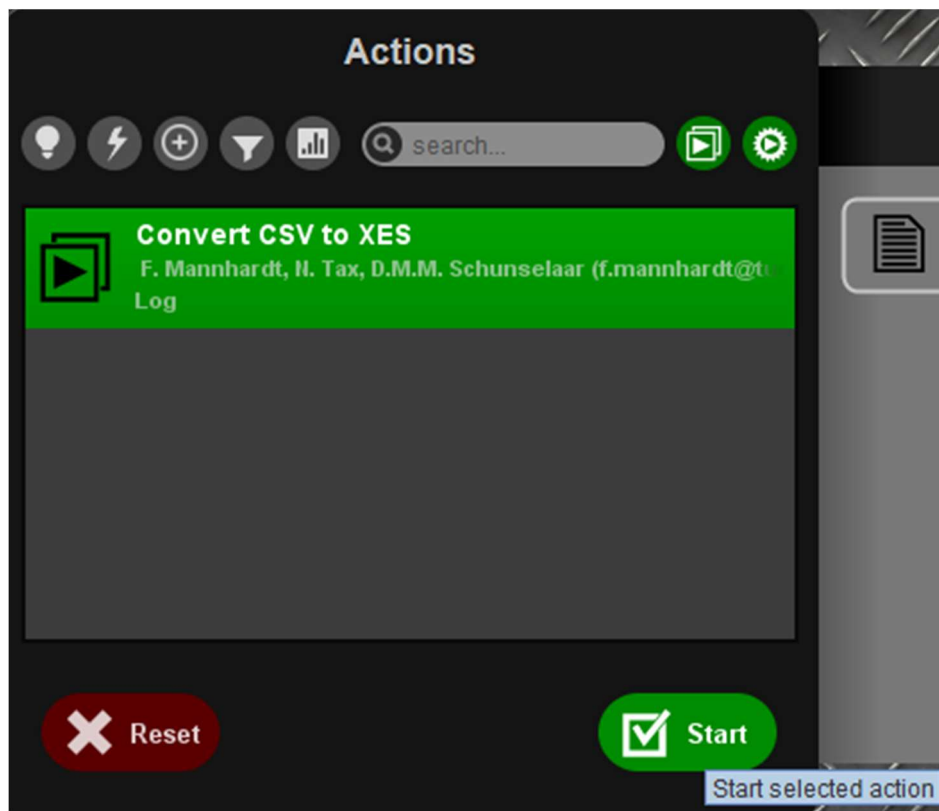
A ProM programot megnyitva az alábbi felületen az import gomb megnyomásával, majd a böngészőablakban a megfelelő fájlt kiválasztva tudjuk a kívánt logfájlt importálni a rendszerbe.



Ha ez megtörtént, az alábbi listában láthatjuk az importált logfájlt. Ezt kiválasztva a Use Resource gomb megnyomásával tudunk ezen további műveleteket végezni.



Mivel az elemzés XES formátumú fájlokon történik, ezért CSV fájlunkat a beépített plugin segítségével át kell erre a formátumra konvertálnunk. A listából ezt kiválasztjuk, majd a Start gombra kattintunk. Az ezt követő beállításokat szerencsére a ProM automatikusan felismeri és jól beállítja, tehát csak a Next gombot nyomogatva helyesen el tudjuk végezni az átalakítást.



## A ProM rendszer használata – az XES logfájl elemzése

Mindezek után az XES event log fájlunkról a következő ábrán már rengeteg hasznos információt kaphatunk, mint például az összes eset- és eseményszám: 700, illetve 12048, az esetenkénti események minimális, átlagos és maximális száma: 5, 17 és 27, valamint az eseményosztályok, az eseménytípusok és a kezdeményezők száma: 45, 1 és 1.



Az inspector menüpontra áttérve az egyes lefutásokat vizsgálhatjuk meg több nézetben is, azaz hogy az adott azonosítójú trace-ekben mely események milyen sorrendben futottak le. Az alábbi, 117-es azonosítójú lefutásban például láthatjuk, hogy 12 esemény következett be az alábbi sorrendben:

The screenshot displays the Log Inspector interface with three main panels:

- Instances:** A list of instance IDs from 0 to 121. Instance 117 is highlighted in red.
- 117 12 events:** A detailed view of the events for instance 117, showing a sequence of 12 events with their respective labels and completion status.
- Attributes for case 117:** A panel showing the attributes for the selected instance, including the text "LITERAL TYPED" and "conceptname: 117".

Event #	Label	Status
	g	
#1	complete	- no timestamp -
	a	
#2	complete	- no timestamp -
	j	
#3	complete	- no timestamp -
	c	
#4	complete	- no timestamp -
	h	
#5	complete	- no timestamp -
	ae	
#6	complete	- no timestamp -
	an	
#7	complete	- no timestamp -
	f	
#8	complete	- no timestamp -

A summary menüpontban egy összeggést kapunk a logunkról és a benne található események abszolút és relatív előfordulásáról, legyen az akár csak a kezdeti- vagy csak a végeseményekre szűrve. Mint láthatjuk, a leggyakrabban előforduló esemény a c esemény, amely mind a 700 lefutásban előfordult, a relatív előfordulása pedig 5,81%, azaz a logfájlban megtalálható összes esemény 5,81%-a c esemény volt. A c ezen kívül a leggyakrabban előforduló kezdeti esemény is, lentebb pedig a leggyakrabban előforduló végeseményeket láthatjuk.

**XES Event Log**

Select visualisation ...

Log Summary save HTML...

**Log Summary**

Total number of process instances: **700**  
 Total number of events: **12048**

**Event Name**

Event classes defined by Event Name  
**All events**

Total number of classes: **45**

Class	Occurrences (absolute)	Occurrences (relative)
c	700	5,81%
h	698	5,793%
ag	689	5,719%
u	397	3,295%
b	368	3,054%
d	368	3,054%
af	368	3,054%

**Start events**

Total number of classes: 17

Class	Occurrences (absolute)	Occurrences (relative)
c	113	16,143%
a	83	11,857%
e	69	9,857%

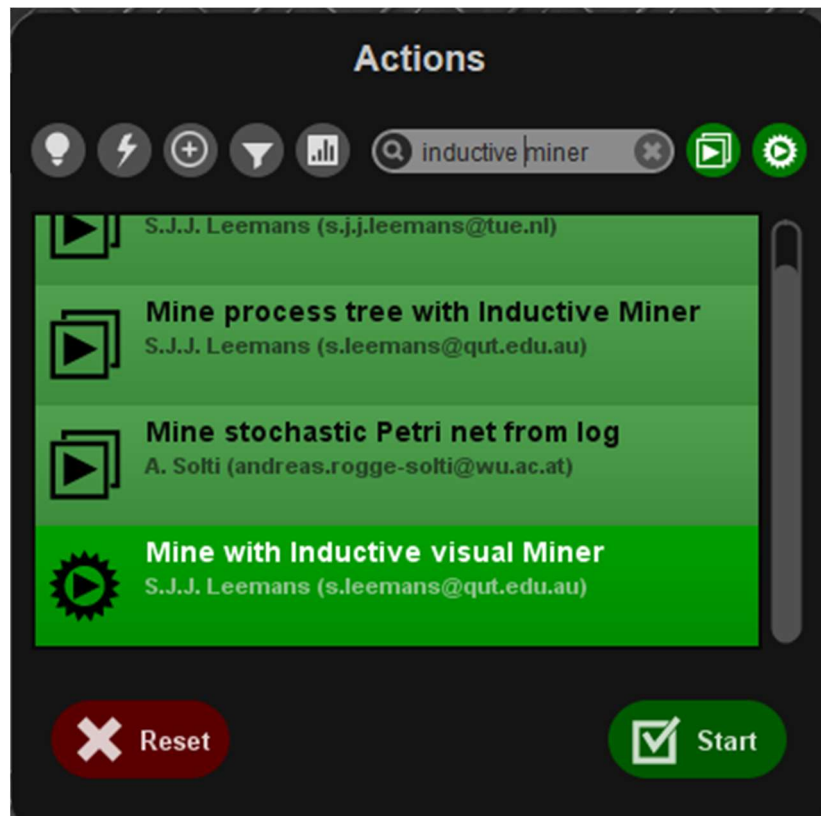
**End events**

Total number of classes: 25

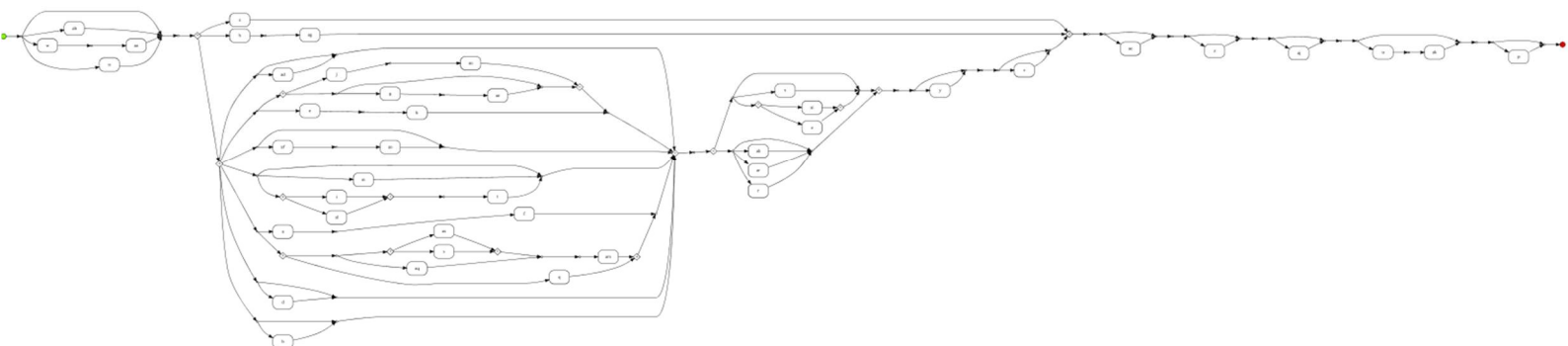
Class	Occurrences (absolute)	Occurrences (relative)
p	209	29,857%
aj	138	19,714%
ak	112	16,0%
v	63	9,0%

## A ProM rendszer használata – Inductive Visual Miner

Ezután, immár az XMS fájlunkat felhasználva az előbbi listában már rengeteg különböző plugin között válogathatunk. Választásom az Inductive Visual Minerrel való bányászatra esett, amely hasznos információkat gyűjt számunkra a folyamat teljesítményéről. Ezt a következő néven találjuk meg:

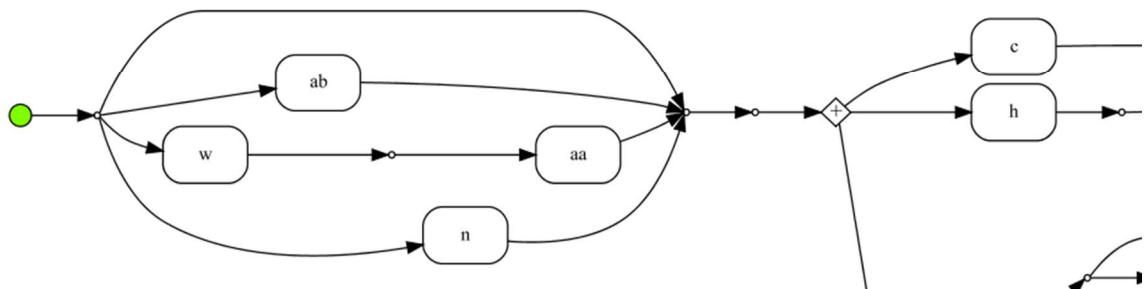


A plugint futtatva eredményül a következő modellt kapjuk a log alapján:

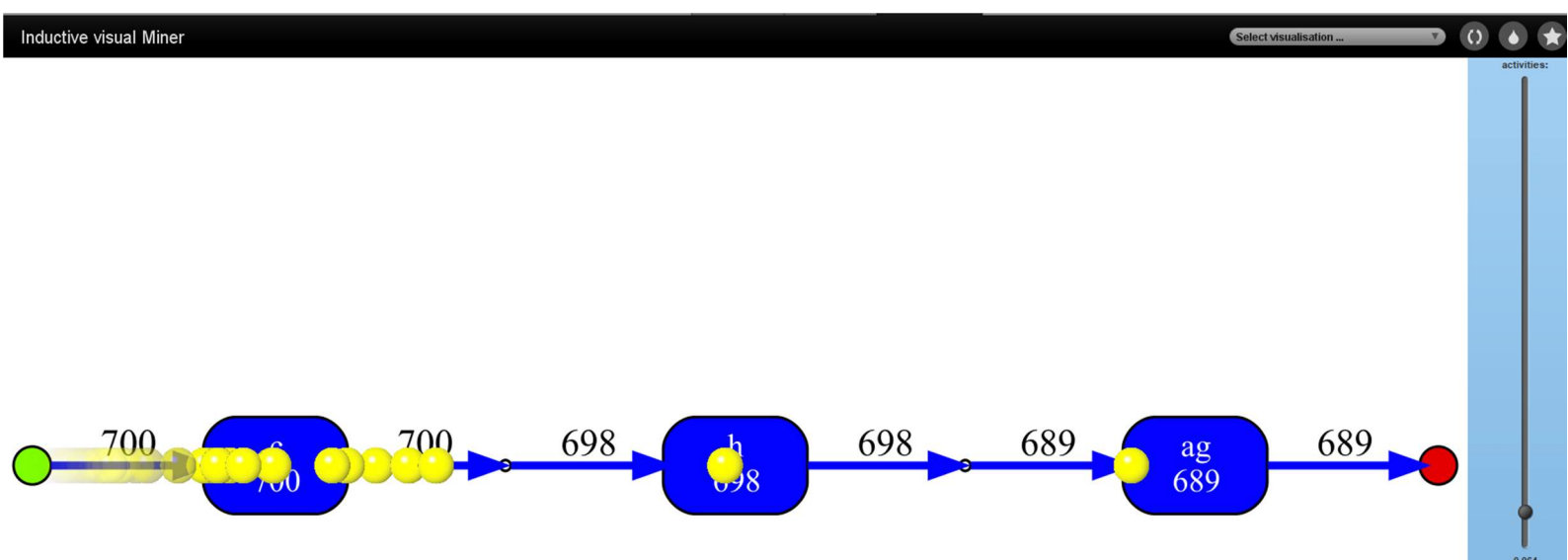


Minden egyes lefutás a zöld pontból indul ki és a piros pontban végződik.

Az ábrán is látható a többféle lehetséges kezdeti esemény, melyek a Summary menüben is szerepeltek, mint pl. a c, valamint lentebbi ágakban az a és az e esemény is. Ebből a részletből például azt is láthatjuk még, hogy az aa esemény mindig a w eseményt követi, vagy hogy az ab, w->aa, n ágak közül mindig csak az egyik vagy egyik sem fut le, tehát ezek mind opcionális események.

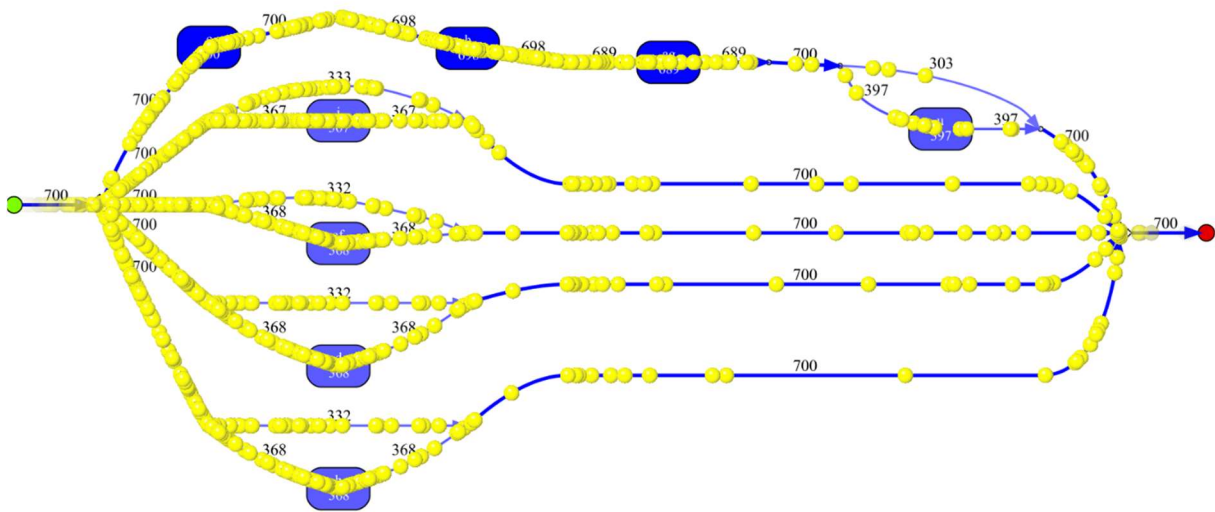


Az activities csúszka módosításával az eseményeket előfordulási gyakoriságuk szerint tudjuk előzetesen szűrni. Mint láthatjuk, egy igen alacsony értéknél az események csak egy nagyon kicsi töredékéből, a leggyakrabban előfordulókból fog a modell felépülni. Az egyes eseményeknél látható szám azt jelenti, hogy az adott esemény hány trace-ben fordult elő, jelen esetben az ag esemény pl. 689-szer. Az Inductive Visual Miner még animáció segítségével is vizualizálja a lefutásokat.





Az animáció jóval szemléletesebb egy kevésbé szűrt modell esetén, azonban a programnak ezt jóval több időbe is tart, hogy legenerálja:



Az Inductive Visual Miner lehetőséget nyújt továbbá az optimális lefutástól való eltérések feltárására, tehát konformancia-ellenőrzésre, valamint az egyes események között eltelt időtartamok megjelenítésére is, azonban a jelenlegi logfájlunkban nem volt példa ilyenre és nincsenek meg az ehhez szükséges időbélyegek.

## Konklúzió

Láthatjuk, hogy a ProM egy nagyon hasznos alkalmazás a folyamatbányászat területén mind modellek feltárására, elemzésére, a trace-ek vizualizációjára, stb.. Rengeteg lehetőség rejlik a számtalan elérhető pluginban, amit én bemutattam, az csak egy a sok közül.