

Folyamatbányászat 1

Big Data, Event Data



Az adat az új sütőolaj: tudni fogják, mit eszel a nyáron

A turisztikai szervezet vezérigazgató-helyettese szerint „csak akkor lehetünk sikeresek az európai versenytársakkal szemben, ha megőrizzük és növeljük a digitális adatokon alapuló tervezési és stratégiai előnyünket”.

Az előadás anyagai megtalálhatóak:

virt.uni-pannon.hu oldalon

Tantárgyak menüpont

Folyamatbányászat tárgy

Moodle: dolgozatíráshoz, feltöltéshez

Számonkérés:

- **2 félévközi dolgozat** (60 pont) külön-külön min. 40%-os teljesítése,
- **1 beadandó – csoport munka elkészítése** (40 pont)
- Mindösszesen 100 pontot lehet gyűjteni
- **Aláíráshoz szükséges: 40 pont**
- Ponthatárok: 0-45=1 46-58=2 59-71=3 72-84=4
85-100=5

Beadandó feladat

- Feladat: **4 fős csoportok létrehozása**, minden tag választ egy **szerepkört** (manager, ügyfél, technológus, projekt manager) és annak szerepében látja el a feladatait a csoporton belül.
November 24-én egy 15 perces prezentációban be kell mutatni a csoport tagjai által elvégzett feladatokat, milyen log került elemzésre, milyen KPI-ok kerültek megfogalmazásra, milyen eszközök kerültek felhasználásra, mi lett az elvégzett munka eredménye és a kapott eredményeket, következtetéseket, javaslatokat milyen célra lehet felhasználni.
- **A csoportokat legkésőbb szeptember 17-ig kell létrehozni.** Ha a csoporton belül megvannak a szerepkörök, akkor a projekt manager megkeresi az oktatót és megkapja azt a **logot**, amellyel foglalkozni kell.
- A feldolgozáshoz fel lehet használni a ProM rendszer pluginjeit, illetve egyedileg megtervezett, implementált modulokat is létre lehet hozni az eredmények előállításához.
- **Feltöltés a Moodle-be, bemutatás határideje: 2025. november 24. 12:00**

Szerepkörök a csoportban

- A **manager** a cég tulajdonosait, a befektetőket képviseli. Az ő feladata az, hogy értéket teremtsen. Mivel egy cégtulajdonos elszámoltatható törvényileg is, ezért neki arra is ügyelnie kell, hogy ne sértsék meg a törvényt. Azonban ha a törvény betartása költséges, akkor akár azt is választhatja, hogy nem tartják be. Például a Google és Facebook gigantikus bírságot kapnak az adatokkal való visszaélés miatt, de erre épül az üzleti modelljük, hogy az adataikkal kereskednek. Tehát a management-et a profit érdekli elsősorban, és hogy simán menjen az üzletmenet. Ők határozzák meg egy cég hosszú távú stratégiáját, projekteket indítanak vagy leállítanak.
- Az **ügyfelet** az érdekli, hogy jó minőségű szolgáltatást kapjon, sok szolgáltatás legyen elérhető, és rövid idő alatt szolgálják ki.
- A **technológust** ezzel szemben az érdekli, hogy olajozottan működjön a technika. Sokszor egy csapaton belül több technológia is van, ami ütközik. Például a fejlesztő és tesztelő szerepkör, vagy a rendszergazda és a biztonsági szakember. Utóbbi esetben a rendszergazda ad hozzáférést rendszerekhez, a biztonsági szakember pedig ezt korlátozza.
- A **projekt manager** feladata egyensúlyt teremteni. A felső vezetéstől kapja a fizetést, ezért az ő érdekeiket is kell képviselnie, valamint feléjük képviseli a projektet. Egy projekt akkor sikeres, ha az ügyfél elégedett, ezért az ügyfél kéréseit szem előtt kell tartania. A csapaton belül a technológiai területen a konfliktusokat is ő oldhatja fel. A projekt manager tehát egyfajta híd szerepét tölti be a többi szereplő között.

Témakörök

Cél: Folyamatokból származó adatok felhasználása a döntéstámogatáshoz

1. A folyamatbányászat különböző típusai, hogyan kapcsolódik a folyamatbányászat az adatbányászathoz,
3. Munkafolyamat hálók és stabilitás, Alpha algoritmus – Folyamat felfedező algoritmus;
4. Minőségi kritériumok a folyamat feltáráshoz, Üzleti folyamat modell;
5. Függőségi gráfok és oksági hálók;
6. Tanulás függőségi gráfokkal, tanulás oksági hálókkal, tanulás átmenet rendszerekkel;
7. További folyamat feltáró technikák;
8. Konformancia ellenőrzés;
9. Esemény adatok feltárása;
10. Döntési pontok feltárása, feltárási akadályok;
11. Szociális háló feltárása, szervezeti háló feltárása, különböző perspektívák kombinálása;
12. Működés támogatás – detektálás, előrejelzés és javaslat megadása;
13. Folyamatbányászati szoftverek;

Sok-sok adat "Az adat az új olaj!"

Az elmúlt 10 percben több adatot generáltunk mint 2003 előtt összesen.

Adatokat generálunk:

- jegyet vásárolunk egy repülőjára
- veszünk egy kávét
- feltöltjük üzemanyaggal az autónkat
- kezdeményezünk egy telefonhívást
- küldünk egy emailt
- megnézünk egy klippet a youtube-on
- stb.



Honnan?

Az iPhone-on pl. nagyon sok féle szenzor található:

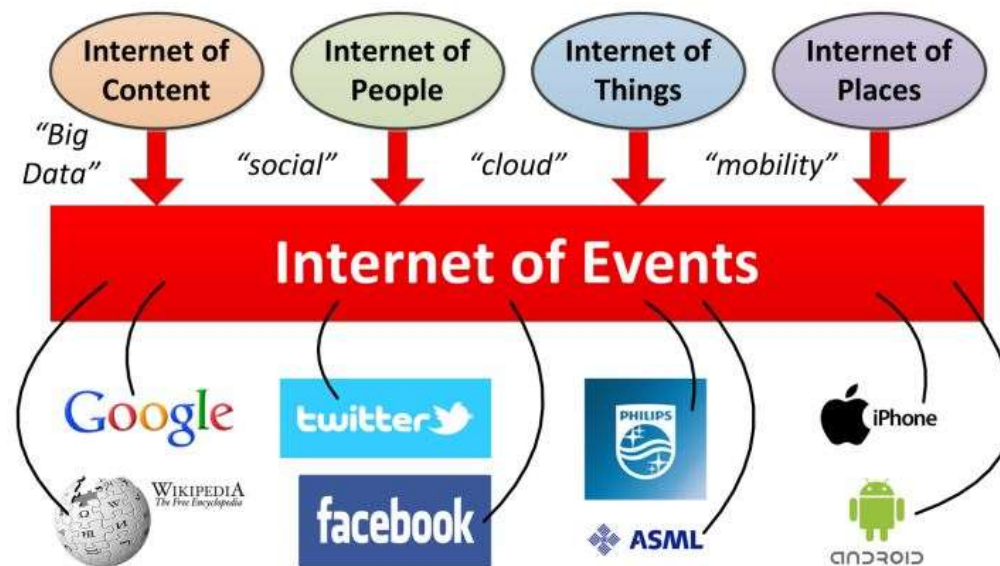
- camera (elől, hátul),
- mágnestermérő (magnetometer),
- érintés érzékelő,
- gyorsulásmérő (accelerometer),
- Bluetooth,
- GPS,
- ujjlenyomat vizsgáló,
- környezeti fényerősség érzékelő (ambient light sensor)



Honnan származhatnak pl. az adatok

Az esemény adatoknak (event data) 4 forrását különböztethetjük meg:

- Tartalmak internetje: Wikipedia, Google stb.
- Személyek internetje: twitter, facebook stb.
- Dolgok, tárgyak internetje: hűtőszekrény, mosógép gyártók stb.
- Helyek internetje: mobil alkalmazások stb.



Növekedés

- számítási kapacitás
- tároló kapacitás (exponenciális növekedés)
- pl. 40 évvel ezelőtt Párizsból New Yorkba 7 óra alatt juthattunk el repülővel, ma ezt megtehetjük Moore törvényét figyelembe véve 24 ms alatt



Diagram by W. van der Aalst

Hogyan tovább? Mi várhat ránk?

Kérdések, felvetések

- Az MI és a robotika nem fog egész iparágakat törölni
- Egészségügy:
 - Orvos: információ feldolgozásra koncentrálni → elemzi az orvosi adatokat, ez alapján diagnózist állít fel → **MI háziorvos az okostelefonon (FOLYAMATBÁNYÁSZAT)**
 - Ápoló: jó motoros és érzelmi képességekkel rendelkezik → bead egy injekciót, kicserél egy kötést, megfékez egy erőszakos páciens → **nem helyettesítheti egy ápolórobot**
- Az emberek egyre tovább élnek, kevesebb gyerek születik → valószínű **az idős gondozás az egyik leggyorsabban növekvő szektor az emberi munkaerőpiacon (MI-s segítő alkalmazások megjelenhetnek (FOLYAMATBÁNYÁSZAT))**

Beethoven a gépben

- Egyetlenegy munkahely sem lehet biztonságban az automatizálástól.
- A művészetről általában az emberi érzésekre asszociálunk
 - azt gondoljuk, hogy belső pszichológiai erőket csatornáznak be a művészek
 - a művészet célja, hogy összekössön bennünket az érzelmeinkkel, vagy újakat keltsen bennünk
- Mi történik akkor, ha az algoritmusok képesek lesznek jobban megérteni és manipulálni az emberi érzelmeket, mint Shakespeare vagy Beyoncé?
- Az érzelmek biokémiai folyamatok! → Egy tanuló algoritmus képes lehet elemezni a szenzorai által a testünkből és testünkről közvetített biometrikus adatokat (ADATELEMZÉS, FOLYAMATBÁNYÁSZAT) → meghatározzák személyiség típusunkat és hangulatunkat → kiszámítja milyen hatással lesz ránk egy- egy dal

Beethoven a gépben

- A zene alkalmas arra, hogy a big data elemezze, mivel az input és az output precízen felírható matematikailag:
 - Input = hanghullámok matematikai mintázatai
 - Output = a neurális kisülések elektrokémiai mintázatai
- Csak néhány évtized szükséges ahhoz, hogy több millió zenei élményt végigtanulmányozó algoritmus képes legyen megjósolni, hogy adott input milyen outputot fog kiváltani (ALGORITMUSOK)
- Nincs az DJ, aki vetekedhetne egy ilyen képességgel!?

Munkák, amelyek eddig nem voltak

- Általános orvosok (legismertebb betegségek diagnosztizálása) helyett → MI orvosok (ADATELEMZÉS, FOLYAMATBÁNYÁSZAT) → több pénz juthat pl. sebészeti eljárások kifejlesztésére
- Emberi pilóták drónokkal való helyettesítése → új munkahelyek karbantartás, távirányítás, adataelemzés, kiberbiztonság területén
- USA 30 ember szükséges pl. a Szíria felett pilóta nélküli Predatorok és Reaperok működtetéséhez, min. 110 ember kell az adatok elemzéshez → ember hiány az ember nélküli repülőgépekhez?!

Munkák, amelyek eddig nem voltak

- A 2050-es munkaerőpiacot az **ember-MI együttműködés** kell, hogy jellemezze, nem a versengés
- Hatékonyabbak lehetnek a vegyes ember-MI csapatok pl. rendfenntartás, banki szféra
- Deep Blue 1997 legyőzte Garri Kaszparovot → emberek tovább sakkoztak → MI-edzőknek köszönhetően a sakk mesterek jobbak lettek, DE!
- A „kentaur” becenévre hallgató ember-MI csapatok túltettek az emberen és a számítógépen külön-külön → kinevelhetők a legjobb nyomozók, bankárok, katonák stb.

(ADATELEMZÉS, FOLYAMATBÁNYÁSZAT)

Problémák lehetnek

- Az új munkahelyek nagy szakértelmet igényelnek → a nagyszámú képzetlen munkaerő nem vagy nehezen talál munkát
- **Átképzések, továbbképzések szükségesek folyamatosan!**
- Pl. ha 2050-ben egy bolti pénztáros vagy gyári munkás elveszíti az állását a robotizálás miatt, nem lesz képes rákkutatóként, drónkezelőként vagy egy ember-MI bankárcsapat tagjaként dolgozni → magas munkanélküliség, képzett munkaerő hiánya → **már most sem ülhetünk a babérjainkon!**

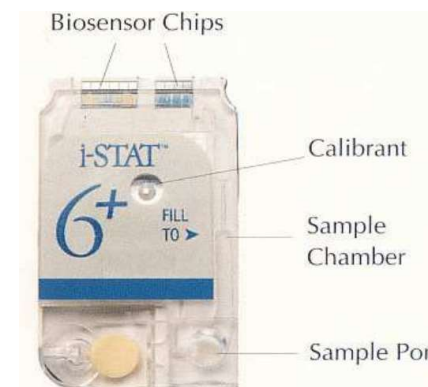
Nem lesz „nyugdíjas szakma”

- Folyamatos átképzés → érzelmi állóképesség?! → már most stresszjárvány, mi lesz később?!
- Info- és biotechnológia (külön-külön és összekapcsolódva) kihívásai az elkövetkezendő évtizedekben
- Jön a „Hallgass az algoritmusokra!” korszak?!
- Két óriási forradalom összefolyását éljük meg: biológusok fejtik meg az emberi test, az agy és az érzések rejtelseit + komputertudósok biztosítanak számunkra korábban nem látott adatfeldolgozási kapacitást → érzéseinket megértő algoritmusok (ADATELEMZÉS, FOLYAMATBÁNYÁSZAT)

Info- és biotechnológia

- Néhány évtizeden belül 24 órában monitorozhatják egyézségi állapotunkat olyan algoritmusok, amelyek folyamatosan testünk biometrikus adataihoz férnek hozzá → pl. influenza, Alzheimer-kór felfedezése az észlelt tünetek előtt → személyiségünkre, DNS-ünkre, fizikumunkra szabott kezelések, életmód és étrend javaslatok → **folyton „betegek” leszünk?!** → folyamatosan javaslatok **(ADATELEMZÉS, FOLYAMATBÁNYÁSZAT)**
- **Az egészségügyi algoritmusok, elemzések jelzéseit figyelembe vesszük vagy sem?**

Info- és biotechnológia

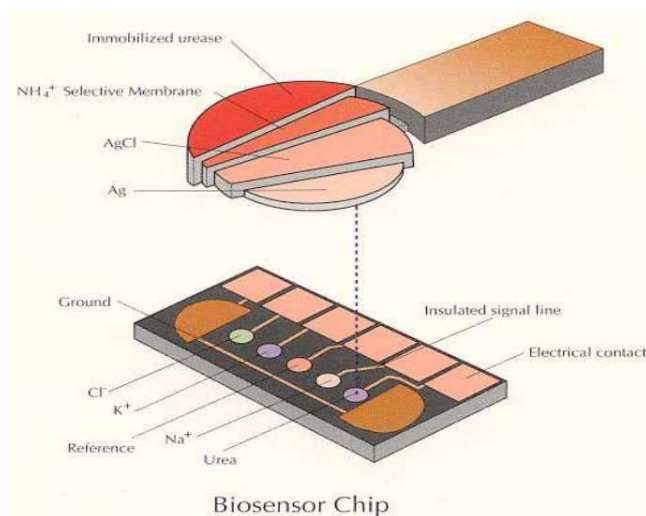


- „Video on demand” – **ajánlórendszerek sokasága**
- Megadjuk milyen filmeket szeretünk, óriási statisztikai adatbázis segítségével az algoritmus megtalálja a tökéletes választást → **önbevallás megbízhatatlan!**
- Valós időben adatok gyűjtése, amikor filmet nézünk (pl. tudni fogja, hogy jónak hirdettük, de fél óránál tovább soha nem tudtuk megnézni)
- A szem és arcizmok mozgása alapján az ember érzelmeinek felismerése (kamera + algoritmus → nevetés, sírás stb.)
- Algoritmus összekapcsolása bioszenzorokkal → információ a pulzusunkról, vérnyomásunkról, agytevékenységünkéről → különbség az erőltetett nevetés és a szívből jövő nevetés között

(ADATELEMZÉS, FOLYAMATBÁNYÁSZAT)

Bioszenzorok

- TV (televízió = messze látni) **hamarosan látni fog bennünket?**
→ lehet, hogy ha végig nézünk egy filmet, nem is emlékszünk rá, de pl. a Netflix (aki működteti a tévéalgoritmust) ismerni fogja a személyiségtípusunkat → tudni fogja hogyan hasson ránk → tökéletes film választás → de tovább is léphetünk, akár megmondható mit tanuljunk, hol dolgozzunk stb. (ADATELEMZÉS, FOLYAMATBÁNYÁSZAT)



Collect just 2-3 drops of fresh whole blood and fill cartridge.



Insert cartridge into the portable, battery-powered analyzer.



View quantitative test results in less than 2 minutes.

Konklúzió

- Ha ma nem is tudjuk, hogyan váltsuk az adatokat készpénzre, akkor is érdemes megtartani azokat, mert az élet irányításának és alakításának a kulcsát rejthetik a jövőben.
- Érdemes foglalkozni az adatfeldolgozás kérdésével, hogy a jövőben olyan szakmánk, tudásunk legyen, ami számunkra megélhetést biztosít.
- De találjuk meg a helyes irányt!!!

(Megjegyzés: A hétköznapi ember ezt támogatja is minden oldalról, ahol csak lehet adatot közöl, ad át pl. ingyenes szolgáltatásért vagy kutyusos videóért cserébe!?)

Legfontosabb kérdés lehet: Hogyan szabályozzuk az adatok tulajdonjogát?)

Hogyan tudunk használható adatokat kinyerni?

Ma ez egy nagy kihívás. 4 dolgot szükséges feltétlenül figyelembe venni:

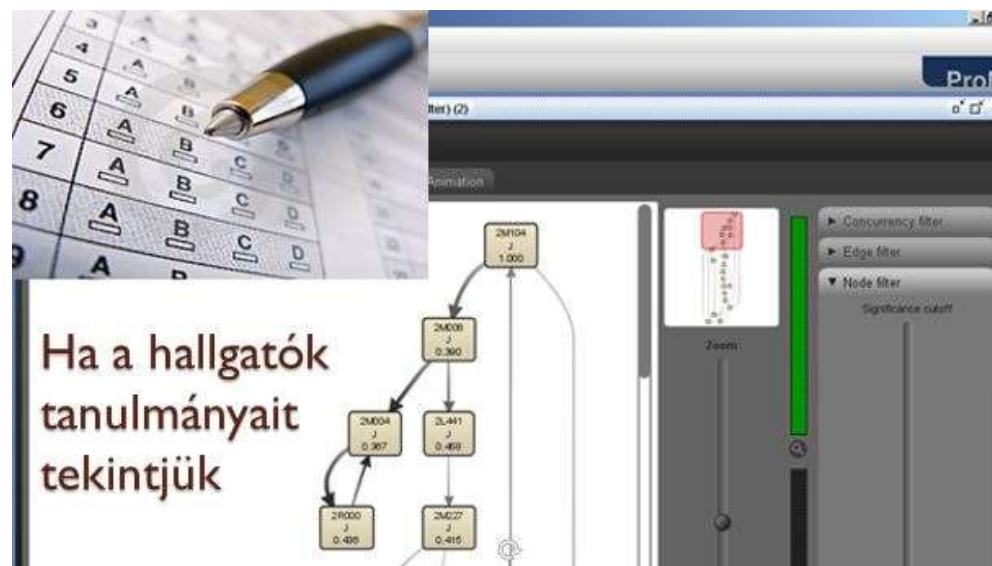
- **Mennyiség** (Volume) - adat mennyiség, adathalmazok mérete
- **Sebesség** (Velocity) – az adathalmaz méretének folyamatos növekedése
- **Változatosság** (Variety) - az adatok különböző formában való megjelenése, adattípusok különbözősége
- **Igazságtartalom** (Veracity) - megfelel-e az igazságnak a tárolt adatok valóságrepresentáló képessége
(**Hiteles forrásokból tájékozódjunk???**!!!)

Vizsgálható kérdések 1

Az igazi kihívás manapság már nem az adatok generálása, hanem az adathalmazokból az értékes adatok kinyerése.

4 alap kérdést tehetünk fel:

- Mi történt?
- Miért történik mindez?
- Mi fog történni?
- Mi a legjobb ami történhet?



Vizsgálható kérdések 2

Vegyük példaként az egészségügyi ellátást:

- Minden orvos követi a protokollt?
- Miért kell a betegnek olyan sokáig várnia?
- Mi okozza a késedelmet?
- Meg tudjuk jósolni a páciensek várakozási idejét?
- Meg tudjuk előre határozni, hogy a következő napon az adott osztályon mekkora személyzetre lesz szükség?
- Hogyan lehet csökkenteni a költségeket a minőségi ellátás megtartása mellett?



Vizsgálható kérdések 3

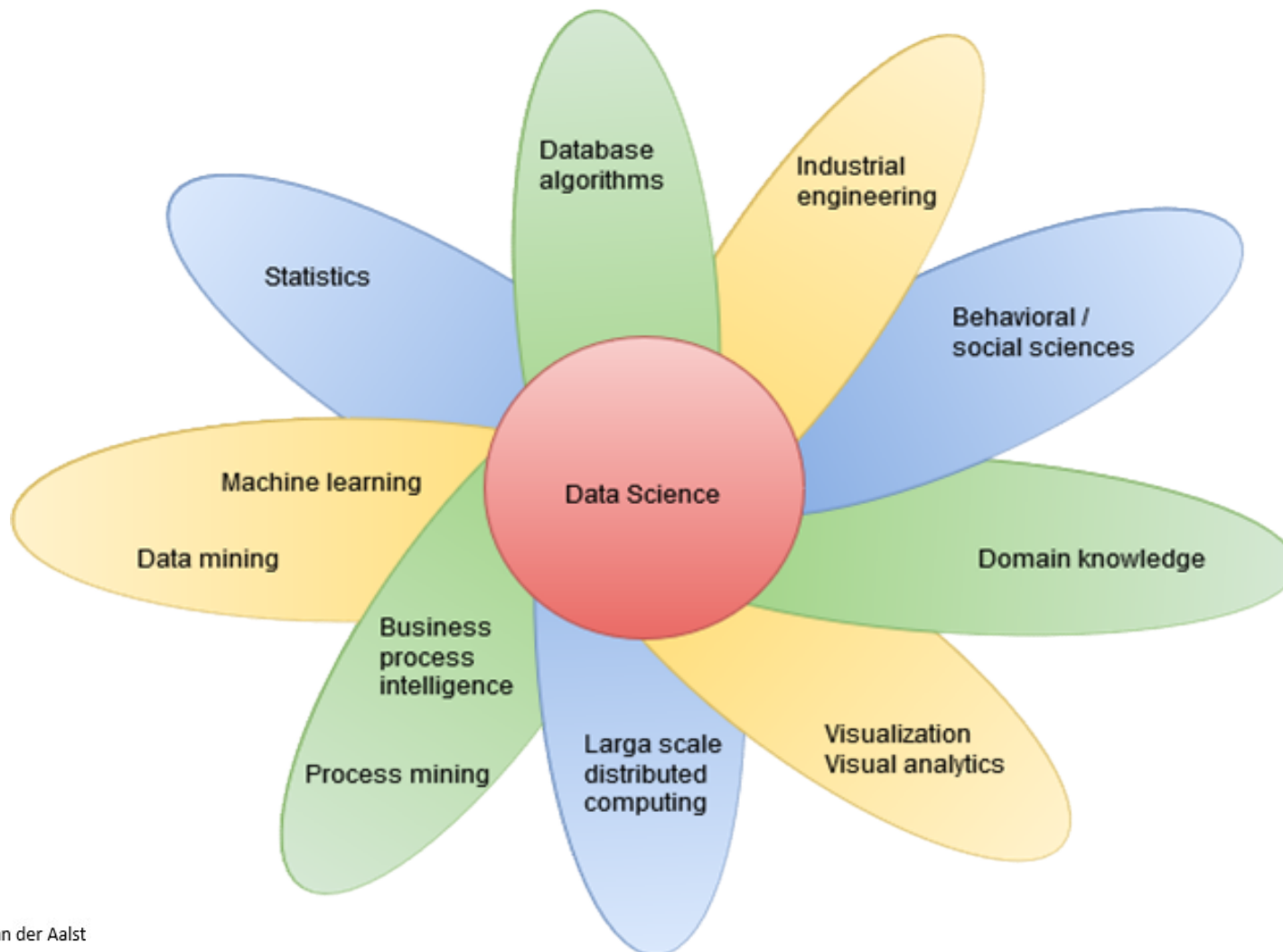
Vegyünk példaként egy gyártási folyamatot:

- Melyik komponens törött el, hibásodott meg?
- Mikor tört el/hibásodott meg és miért?
- Melyik komponenst kell kicserélni?
- Meg tudjuk jósolni, hogy melyik komponenssel fog történni valami?
- Tudunk-e tanulni a létező problémákból, hogy mely részeket kell javítani?
- Mi okozza a késedelmet a folyamatban?
- Meg tudjuk előre határozni, hogy a következő napon az adott osztályon mekkora személyzetre lesz szükség?
- Hogyan lehet csökkenteni a költségeket a minőségi ellátás megtartása mellett?



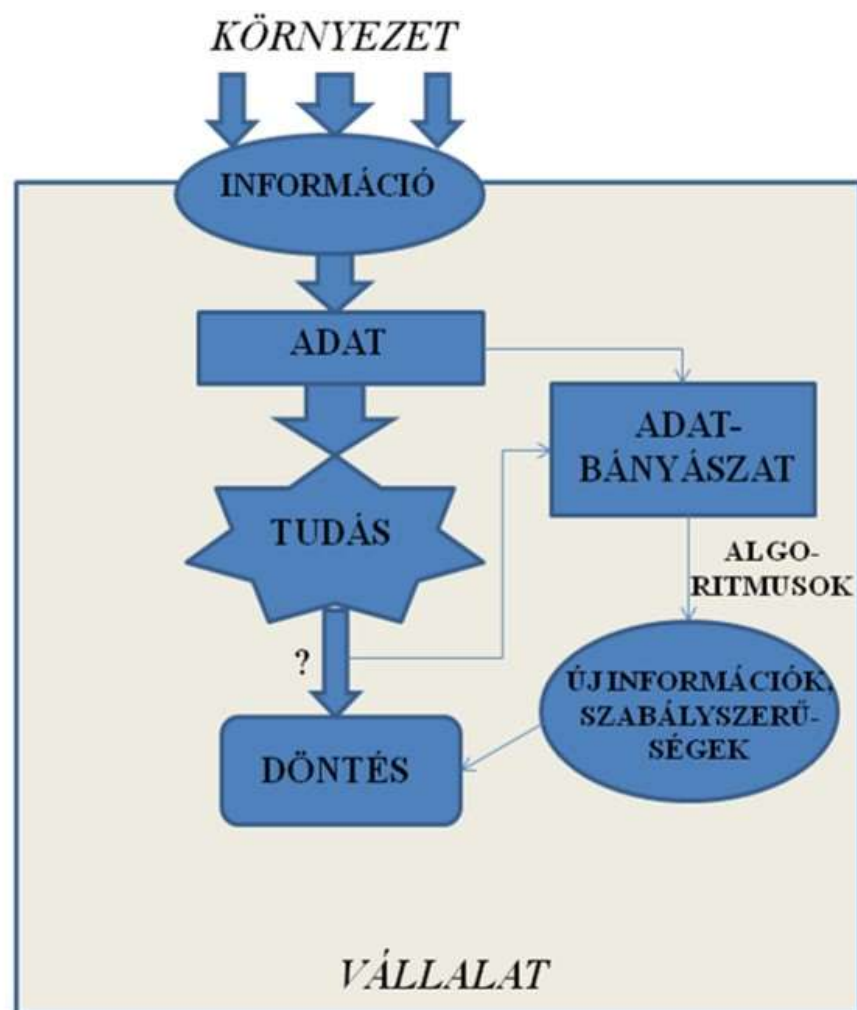
Kérdések megválaszolása

Számos tudományterületre van szükség:



Adatok felhasználása

A folyamatbányászat az adatbányászat egy speciális területének is tekinthető



Folyamatbányászat vs. adatbányászat

- mindkettő adatokkal dolgozik
- az adatbányászati technikák nem folyamat centrikusak
- a tradicionális adatbányászati technikákkal a folyamat felfedezés, összehasonlítás, torlódás elemzés nem elvégezhető

A két módszer együttes használatával **komplex kérdésekre adhatunk választ.**

Mi is a folyamatbányászat?

- A folyamatbányászat célja: **a folyamatokból valódi hasznosítható tudásanyag kinyerése.**
- Felderíthetők és megjeleníthetők azok az összefüggések (pl. az ügyfélcsoportok, régiók, a termékek stb. között), amelyek
 - a hosszú átfutási időért,
 - a magas költségekért,
 - a rossz minőségért felelősek.
- Felderíthetőek a gyártási folyamatban megjelenő hibák, amelyek
 - a leállásokért,
 - a hosszabb gyártási időért felelősek.
- Megnézni:
 1. What is Process Mining? <https://www.youtube.com/watch?v=-u5y2dST7ng>
 2. What is Celonis Process Mining? Analyze and optimize your processes https://www.youtube.com/watch?v=Do2r15_epcl

Jellemzők

A folyamatbányászat egy híd:

- a **folyamat modell elemzés** (szimuláció, ellenőrzés, optimalizálás stb.) és
- az **adat orientált elemzés** (adatbányászat, gépi tanulás, üzleti intelligencia) között.

A folyamatbányászat **2 fő elemzési területtel** foglalkozik:

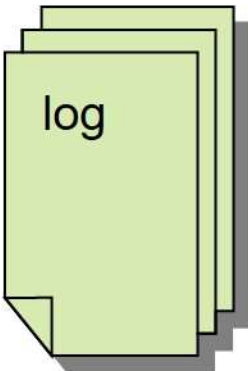
- hatékonyság orientált kérdések, problémák és megoldásaik
- megfelelés orientált kérdések, problémák és megoldásaik

Kiemelt területek

- Folyamat felfedezése , feltérképezése (Process discovery): **mi történik valójában?**
- Megfelelőség elemzés (Conformance checking): **az történik amit elvárunk, amit előre meghatároztunk?**
- Teljesítményelemzés (Performance analysis): **hol lehet javítani, változtatni?**
- Folyamat előrejelzés (Process prediction): **vajon egy eset be fog következni?**
- Folyamat fejlesztés, tökéletesítés (Process improvement): **hogyan alakítsuk át a folyamatot?**
- Stb.

A folyamatbányászat perspektívái

Minden folyamatot megvizsgálhatunk más-más szempont szerint is, ezeket perspektíváknak nevezzük.



Tartalma	Perspektíva
A feladatok végrehajtásának körülményei és azok kapcsolatai	Ellenőrzött folyam (control flow) perspektíva
A végrehajtó személyek / rendszerek	Szervezeti (organizational) perspektíva
További információk (adatmezők értékei) a feladatokról	Eset szerinti (case) perspektíva

Hol alkalmazhatjuk a folyamatbányászatot?

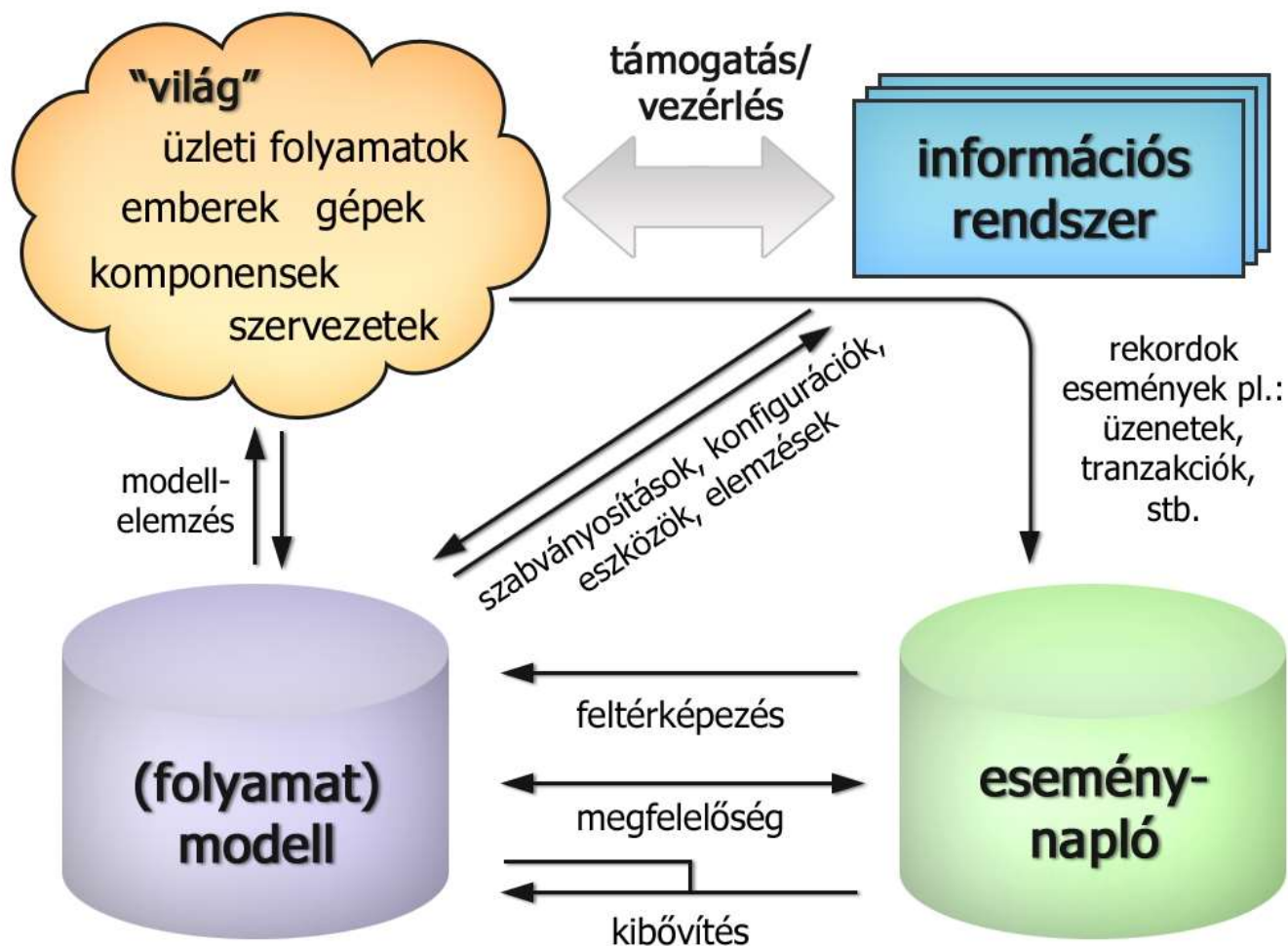
- Önkormányzatoknál (pl.: Alkmaar, Heusden, Harderwijk)
- Kormányzati ügynökségeknél (pl.: Rijkswaterstaat, Centraal Justitieel Incasso Bureau, Justice department)
- Biztosításokhoz kapcsolódó társaságoknál (pl.: UWV)
- Bankoknál (pl.: ING Bank)
- Kórházakban (pl.: AMC kórház, Catharina kórház)
- Multi cégeknél (pl.: DSM, Deloitte)
- Média vállalatoknál (pl.: Winkwaves)
- High-tech rendszerek gyártóinál és fogyasztóiknál (pl.: Philips Healthcare, ASML, Thales)
- stb.

Kiindulási pont: esemény adatok

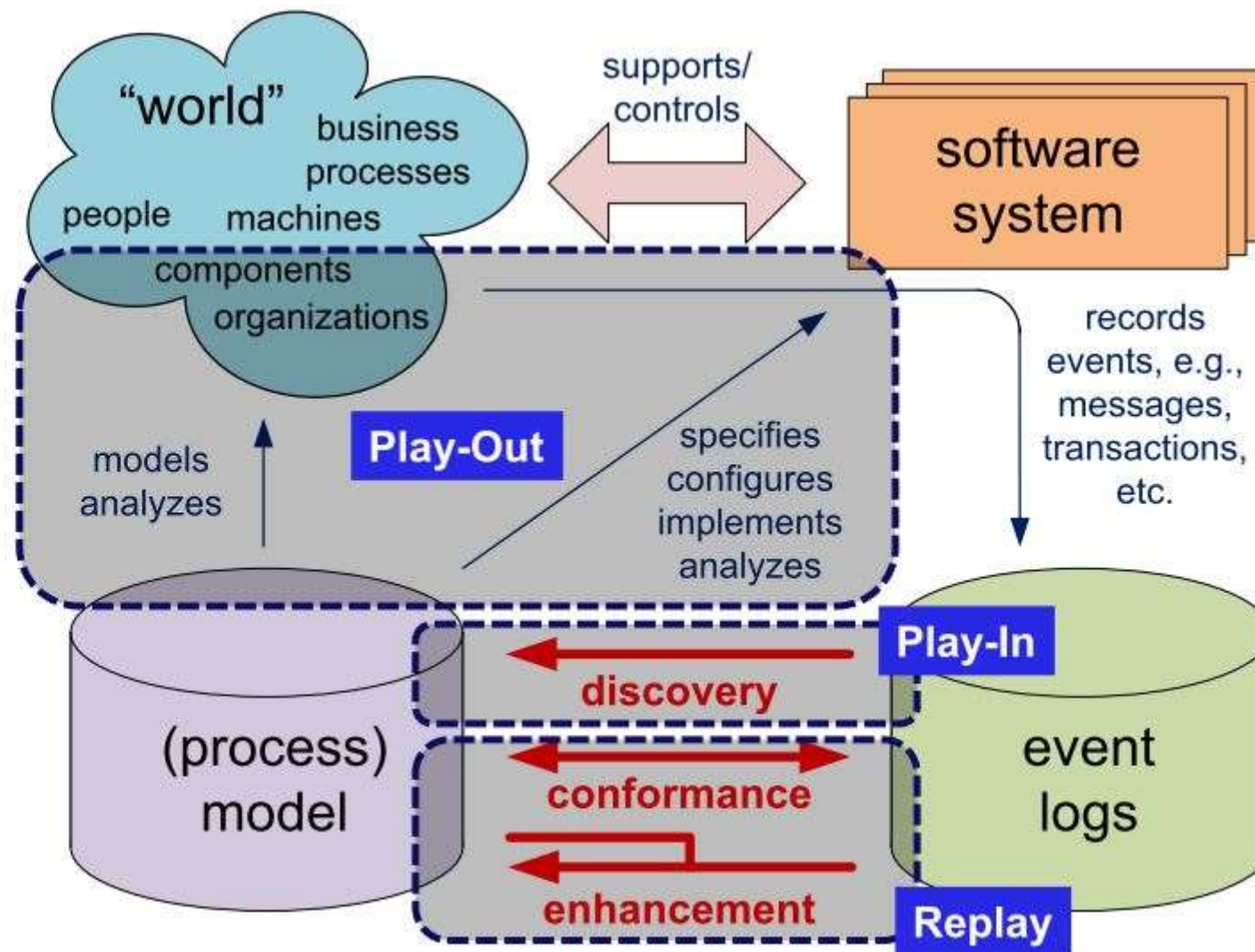
Minden sor egy eseményt azonosít (eset azonosító, esemény neve, időbélyeg, egyéb).

Hallgató neve	Kurzus	Vizsga időpontja	Jegy
Kovács Péter	Programozás I.	2016.01.15	4
Molnár Hajnalka	Programozás I.	2016.01.15	5
Fehér Klára	Programozás I.	2016.01.15	2
Fekete Péter	Programozás I.	2016.01.15	3
Kovács Péter	Programozás II.	2016.01.18	4
Molnár Hajnalka	Programozás II.	2016.01.18	5
Fehér Klára	Mesterséges intelligencia	2016.01.22	3
Kovács Péter	Mesterséges intelligencia	2016.01.22	5
Molnár Hajnalka	Mesterséges intelligencia	2016.01.22	4

A folyamatbányászat magas szintű modell diagramja



Folyamatbányászati feladatkörök



Kapcsolatok: folyamat modellek - esemény logok

- **Play-out** (szimuláció, munkafolyam automatizálás, modell ellenőrzés stb.): A folyamatból kell lefutásokat generálni. Ezeknek a lefutásoknak a száma ugrásszerűen növekedik a folyamat bonyolultságától függően. Amennyiben a folyamat tartalmaz visszacsatolást, akkor a lehetséges lefutások halmaza akár végtelen nagy is lehet.
- **Play-in** (logokból modell felépítése (folyamat felfedezés)): Minél nagyobb a lefutások halmaza, annál jobb folyamat generálható hozzá. Korlátot szab a folyamat bonyolultsága, minél összefüggőbb, annál nehezebb ütemezett folyamatot pontosan visszaadni, hiszen csak a valós lefutások álnak rendelkezésre inputnak.
- **Replay** (adott a modell és a viselkedés, megfelelés vizsgálat, teljesítmény elemzés, előrejelzés, torlódás vizsgálat stb.): Egy meglévő folyamathoz kell a valós lefutásokat viszonyítani

Példa1: A folyamat felfedezés hasonlít a nyelvtanuláshoz

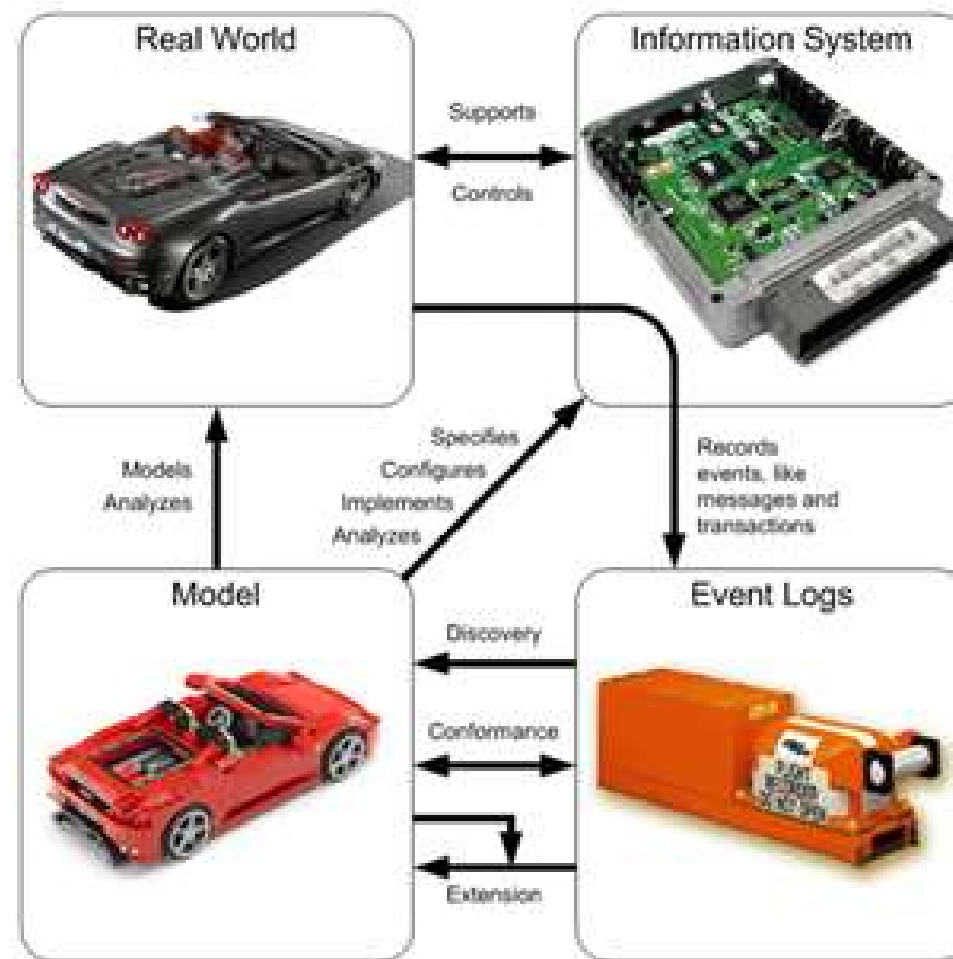
A mondatok olyanok mint a trace-ek az esemény logban, a nyelv pedig a folyamat modell.

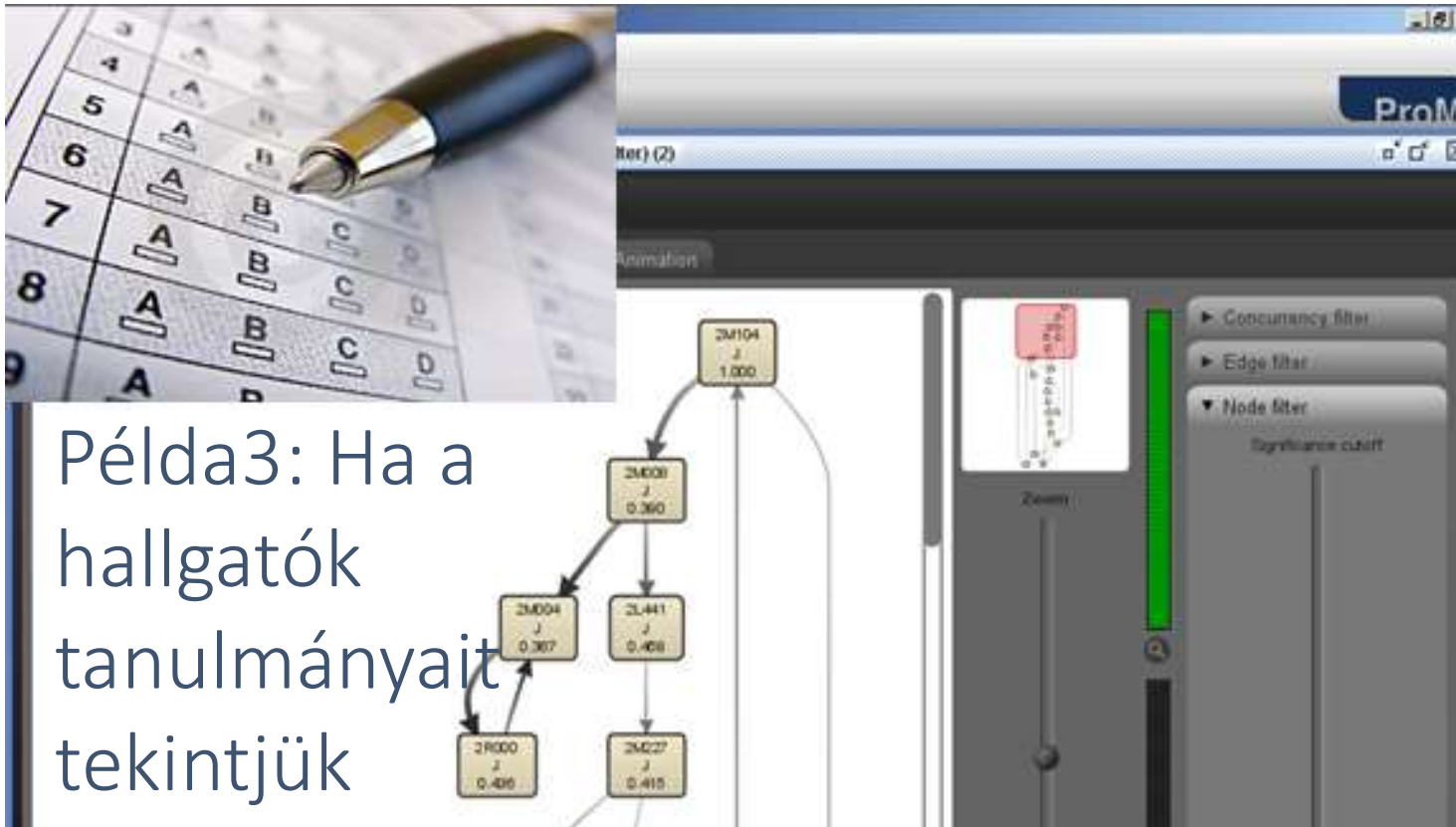


Diagram by W. van der Aalst

Az összehasonlítás elemzés hasonlít a helyesírás ellenőrzéshez: a nyelv modellje a helyesírás ellenőrző.

Példa2: járművek fejlesztése



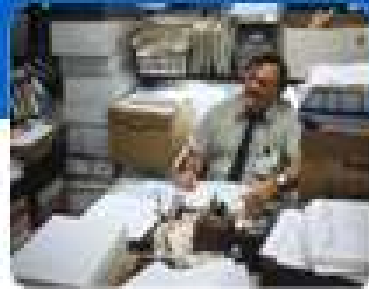


- Folyamat felfedezése: **milyen a valós tanterv?**
- Hasonlóság elemzés: **a hallgatók találkoznak-e az előfeltételekkel?**
- Teljesítményelemzés: **hol lehet változtatni?**
- Folyamat előrejelzés: **be tudja fejezni a hallgató a tanulmányait?**
- Folyamat fejlesztés, tökéletesítés: **hogyan lehetne újratervezni a tantervet?**

Honnan érkeznek az esemény adatok?

- adatbázis rendszerekből (pl. páciensek adatai egy kórházban)
- kórházi információs rendszerekből (pl. Chipsoft)
- tranzakciós logokból (pl. kereskedelmi rendszer)
- üzleti rendszerekből, vállalati erőforrás tervező rendszerekből (pl. SAP, Oracle)
- üzenet logokból (pl. IBM middleware)
- folyamatirányító rendszerekből (pl. WebSphere software, Staffware, Flower, Eastman software, People Soft, Adept)
- web szerverek (pl. Apache)
- Szoftver Konfiguráció Menedzsment rendszerekből (pl. Subversion)
- ügyfél kapcsolatokat menedzselő rendszerekből (pl.: Microsoft Dynamics CRM)
- vesszővel elválasztott értékek fájljaiból (CSV) vagy táblázatokból
- stb.

Példák:



Egy másik példa esemény log-ra, páciensek adatai

páciens ID	művelet	időpont	orvos	kor	költség
23456	CT scan	2016.01.17 17:10	dr. Kovács	23	40.000
25673	MR	2016.01.17 12:00	dr. Molnár	36	60.000
27654	MR	2016.01.18 11:20	dr. Molnár	45	60.000
21117	radiológiai terápia	2016.01.18 11:30	dr. Kiss	65	10.000
24567	vérvizsgálat	2016.01.18 11:30	Mérő Henrietta	34	25.000
27673	CT scan	2016.01.18 12:30	dr. Kovács	37	40.000
26765	CT scan	2016.01.18 12:30	dr. Kovács	51	40.000
23654	vérvizsgálat	2016.01.18 13:00	Mérő Henrietta	21	25.000
27134	vérvizsgálat	2016.01.18 13:10	Mérő Henrietta	70	25.000

Minden sor egy eseményt azonosít (eset azonosító, esemény neve, időbélyeg, erőforrás, költség).

Az esemény adatok ilyen és hasonló listáit tudjuk elemezni.

Nem mindig ilyen egyértelmű!

Tekintsük a mail box-unkat, egy email egy esemény.

Egy emailben van:

- küldő ("Honnan") → erőforrás, tevékenység név
- fogadó ("Kinek") → egyéb adat
- tárgy → eset azonosító
- időbélyeg → időbélyeg
- szövegtörzs → egyéb adat
- stb. →

Kérdés, hogy mi lesz az eset azonosító, a tevékenység, az időbélyeg stb.

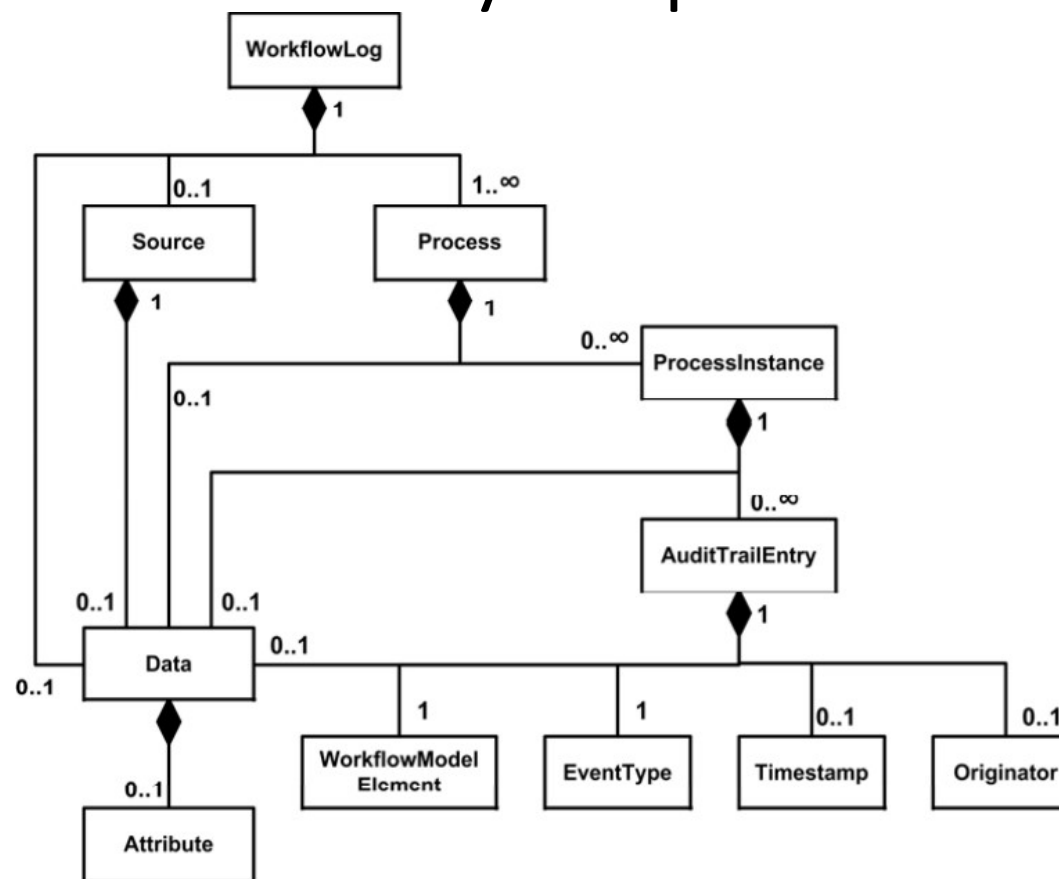
Több féle válasz is lehetséges.

[Adott kontextusban kell vizsgálni.](#)

Kiindulási pont: esemény napló



esemény logok, ellenőrzési
naplók, adatbázisok, üzenet
logok stb.



egyértelmű esemény log
(MXML formátum)

A Staffware rendszer saját formátumú naplózásának MXML formátumra való fordításából kapott eseménynapló egy részlete:

Esemény log:

- folyamatok
 - folyamat előfordulások

Egy esemény:

- művelet neve
- (esemény típusa)
- (végrehajtó)
- (időbélyeg)
- (adat)

```
<Source program="Staffware">
  <Data>
    <Attribute name="version">7.0</Attribute>
  </Data>
</Source>
<Process id="main_process">
  <Data>
    <Attribute name="description">complaints handling</Attribute>
  </Data>
  <ProcessInstance id="Case 1">
    <AuditTrailEntry>
      <WorkflowModelElement>Case start</WorkflowModelElement>
      <EventType unknowntype="case_event">unkown</EventType>
      <Timestamp>2011-04-16T11:06:00:000+001:00</Timestamp>
    </AuditTrailEntry>
    <AuditTrailEntry>
      <WorkflowModelElement>Register
complaint</WorkflowModelElement>
      <EventType>schedule</EventType>
      <Timestamp>2011-04-16T11:16:00:000+001:00</Timestamp>
```

A log előállítás alapjai

- Az első lépés egy log felhasználása előtt, annak áttekintése, az esetleges **formai hibák kijavítása**.
- Az egyik leggyakrabban használt forma, melyet szinte minden program fel tud dolgozni az ún. **CSV fájl**.
- Ez egy olyan egyszerű fájlformátum, amelyet nagy hatékonysággal lehet **táblázati adatok elhelyezésére** használni.
- Fontos része a **fejléce**, azaz az első sor, ahol a CSV fájlban szereplő adatok sorrendje és tartalma van felsorolva.
- **Sorokból** tevődik össze, melyben **az aktuális adatok valamilyen írásjellel vannak elválasztva**, általában vesszővel, vagy pontosvesszővel.
- Mivel nagyon egyszerű, ezért kiválóan használható különböző programok közötti információk cseréjére.
- Legkönnyebben Microsoft **Excel segítségével állítható elő**.

CSV fájl - részlet

```
Fájl Szerkesztés Beállítások Kikódolás Súgó 10 %
["Case ID; Operation; Timestamp; Via; Customer ID; Complexity; Servicetype; Originator;"
"Case 1; Incoming task;07-11-2011 12:26:44;other; Customer 2;complex;backup;ASA;"
"Case 1; Handling task;10-11-2011 09:01:10;other; Customer 2;complex;backup;KR;"
"Case 1; Closing task;12-11-2011 04:55:11;other; Customer 2;complex;backup;KR;"
"Case 2; Incoming task;07-11-2011 05:31:13;telephone; Customer 2;simplex;update;ASA;"
"Case 2; Handling task;09-11-2011 12:51:52;telephone; Customer 2;simplex;update;EK;"
"Case 2; Closing task;12-11-2011 06:34:00;telephone; Customer 2;simplex;update;EK;"
"Case 3; Incoming task;08-11-2011 01:55:55;telephone; Customer 2;medium;upload;ASA;"
"Case 3; Handling task;12-11-2011 01:35:16;telephone; Customer 2;medium;upload;APN;"
"Case 3; Closing task;14-11-2011 09:22:29;telephone; Customer 2;medium;upload;APN;"
"Case 4; Incoming task;08-11-2011 12:15:07;via mail; Customer 2;simplex;database;ASA;"
"Case 4; Handling task;11-11-2011 21:25:59;via mail; Customer 2;simplex;database;AM;"
"Case 4; Closing task;13-11-2011 14:56:59;via mail; Customer 2;simplex;database;AM;"
"Case 5; Incoming task;09-11-2011 23:41:48;via fax; Customer 2;medium;scan;ASA;"
"Case 5; Handling task;13-11-2011 08:42:04;via fax; Customer 2;medium;scan;DB;"
"Case 5; Closing task;14-11-2011 17:54:08;via fax; Customer 2;medium;scan;DB;"
"Case 6; Incoming task;08-11-2011 03:52:38;via mail; Customer 2;medium;scan;ASA;"
"Case 6; Handling task;10-11-2011 11:48:58;via mail; Customer 2;medium;scan;AM;"
"Case 6; Closing task;12-11-2011 13:51:30;via mail; Customer 2;medium;scan;AM;"
"Case 7; Incoming task;07-11-2011 22:11:16;other; Customer 1;complex;audit;TRSZ;"
"Case 7; Handling task;10-11-2011 03:37:54;other; Customer 1;complex;audit;UH;"
"Case 7; Closing task;12-11-2011 16:17:28;other; Customer 1;complex;audit;UH;"
"Case 8; Incoming task;09-11-2011 13:28:14;via fax; Customer 1;complex;audit;TRSZ;"
"Case 8; Handling task;10-11-2011 19:32:15;via fax; Customer 1;complex;audit;UH;"
"Case 8; Closing task;14-11-2011 17:16:42;via fax; Customer 1;complex;audit;UH;"
"Case 9; Incoming task;08-11-2011 06:12:09;via fax; Customer 1;simplex;Server boarding;TRSZ;"
"Case 9; Handling task;12-11-2011 00:44:22;via fax; Customer 1;simplex;Server boarding;UT;"
"Case 9; Closing task;15-11-2011 18:27:47;via fax; Customer 1;simplex;Server boarding;UT;"
"Case 10; Incoming task;09-11-2011 22:08:09;via mail; Customer 2;medium;validation;ASA;"
"Case 10; Handling task;11-11-2011 20:30:54;via mail; Customer 2;medium;validation;APN;"
"Case 10; Closing task;13-11-2011 03:02:45;via mail; Customer 2;medium;validation;APN;"
"Case 11; Incoming task;07-11-2011 16:14:57;via fax; Customer 3;medium;incident;ICSN;"
"Case 11; Handling task;10-11-2011 14:53:46;via fax; Customer 3;medium;incident;EL;"
"Case 11; Closing task;11-11-2011 16:06:48;via fax; Customer 3;medium;incident;EL;"
```

CSV fájl → MXML formátum

- Ahhoz, hogy a CSV fájlt fel tudjuk használni a ProM-ben, **szükség van annak átalakítására** egy speciális, ún. MXML vagy XES formátumba.
- Ennek végrehajtásához **szükség van egy köztes program**, a Nitro használatára. A Nitro Fluxicon cég terméke, egy gyors és hatékony eszköz a folyamatbányászathoz, a <http://www.fluxicon.com/> oldalon érhető el.
- Ebbe kell betölteni a CSV fájlt, és a megfelelő oszlopokhoz hozzá kell rendelni a megfelelő értékeket, pl.:
 - Case = Case ID,
 - Activity = Operation,
 - Timestamp = Timestamp,
 - Via, CustomerID, Complexity, Servicetype = Other,
 - Originator = Resource.

Folyamatbányászati szolgáltatás konkrét cég által (DataPortum)

Mik ennek a lépései?

- Kiválasztjuk a folyamatot
- Meghatározzuk a folyamat vizsgálatához szükséges adatok körét
- Azonosítjuk, hogy a vizsgálatához szükséges adatok mely forrásrendszerekben vannak elmentve és hogyan lehet azokat kinyerni onnan
- Egyszeri adatexport keretében átvesszük az adatokat
- Feldolgozzuk a kapott adatokat a saját Celonis környezetünkben
- Itt létrehozuk a folyamat digitális röntgenképét
- Azonosítjuk a kritikus pontokat, folyamat hiányosságait, veszteségforrásokat
- Létrehozuk azokat a felületeket, amelyek segítségével vizualizáljuk és elemizzük a folyamatot
- Megmutatjuk a mérések, elemzések eredményét és ezeket együtt kiértékeljük
- Az eredményeket prezentáljuk és átadjuk

Mik a várható eredmények?

- Folyamat költségeinek a csökkenése
- Folyamat átfutási idejének a csökkenése
- Folyamat átláthatóságának a biztosítása
- Folyamat előkészítése szoftveres robotizációra
- Folyamat mentén KPI analízis (Key Performance Indicator = A KPI olyan **mérhető érték** takar, amelyek összessége segít jobban megérteni a különböző üzleti folyamatok hozzájárulását a kitűzött céljaid eléréséhez, ezáltal mérhetővé teszi az üzleti sikert.)

Folyamatbányászati eszközök



ABBYY Timeline

Tool Name: ABBYY Timeline

Vendor: ABBYY (Milpitas, CA, USA)

Company Size: 1001-5000 employees



Tool Name: Celonis Process Mining

Vendor: Celonis SE (Munich, Germany & New York, NY, USA)

Company Size: 1001-5000 employees



apromore

Tool Name: Apromore

Vendor: Apromore Pty Ltd (Melbourne, Australia)

Company Size: 11-50 employees



Disco
by Fluxicon

Tool Name: Disco

Vendor: Fluxicon BV (Eindhoven, Netherlands)

Company Size: 1-10 employees



Tool Name: ARIS Process Mining

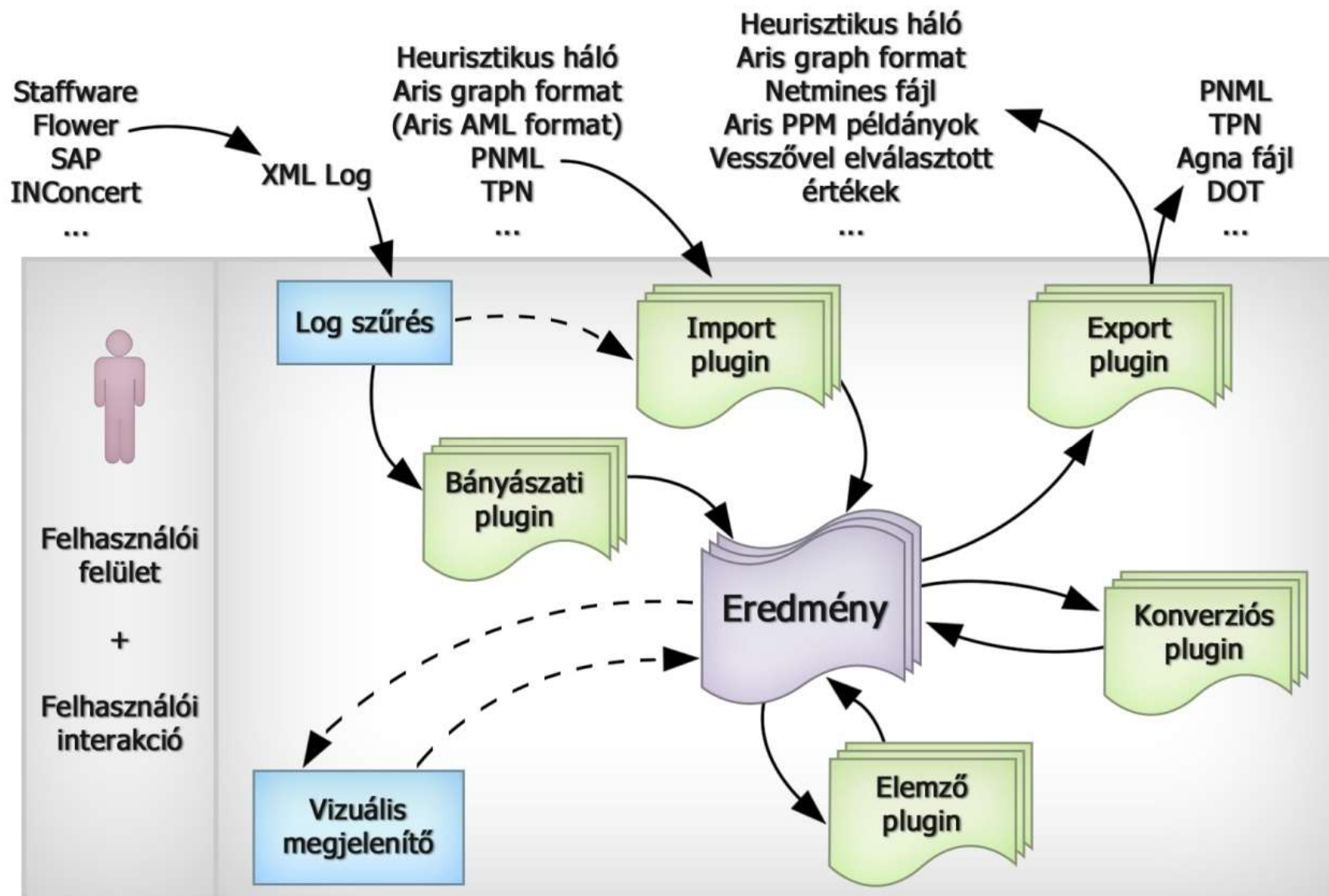
Vendor: Software AG (Darmstadt, Germany)

Company Size: 1001-5000 employees

Figure 1: Magic Quadrant for Process Mining



A ProM keretrendszer architektúrája



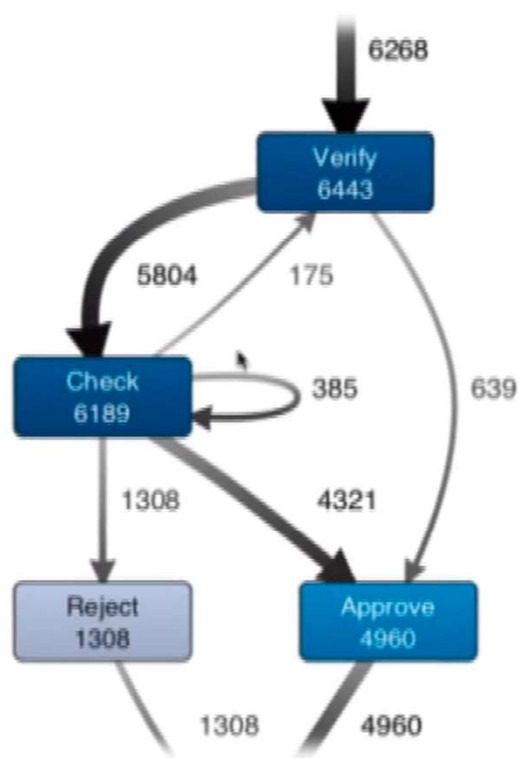
A ProM fontosabb plugin-jai

a folyamatbányászat perspektíváinak tükrében

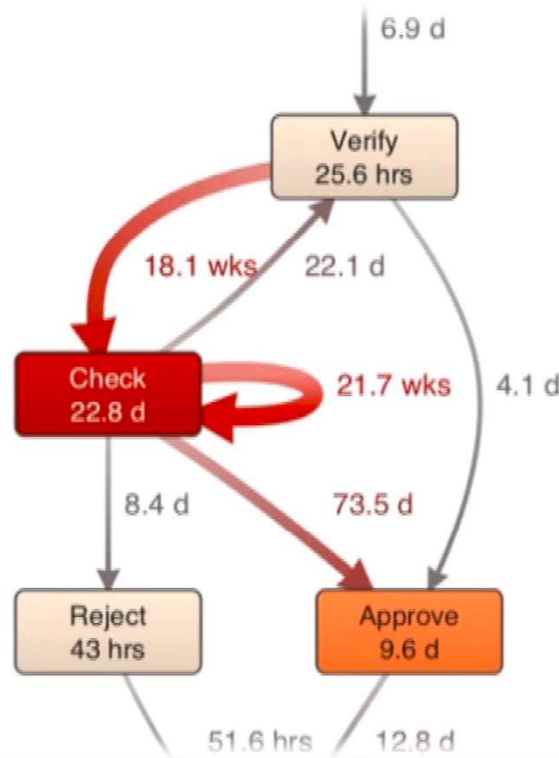
Perspektíva	ProM plugin
Ellenőrzött folyam perspektíva	Alpha algorithm Multi-phase Macro plugin Heuristic miner Genetic algorithm plugin
Szervezeti perspektíva	Social Network Miner Organizational Miner Analyze Social Network Originator by Task Matrix
Eset szerinti perspektíva	Performance Sequence Diagram Analysis
Egyéb	Semantic LTL Checker Conformance Checker Performance Analysis with Petri net Decision Point Analysis

Process Mining and Data Science Using ProM

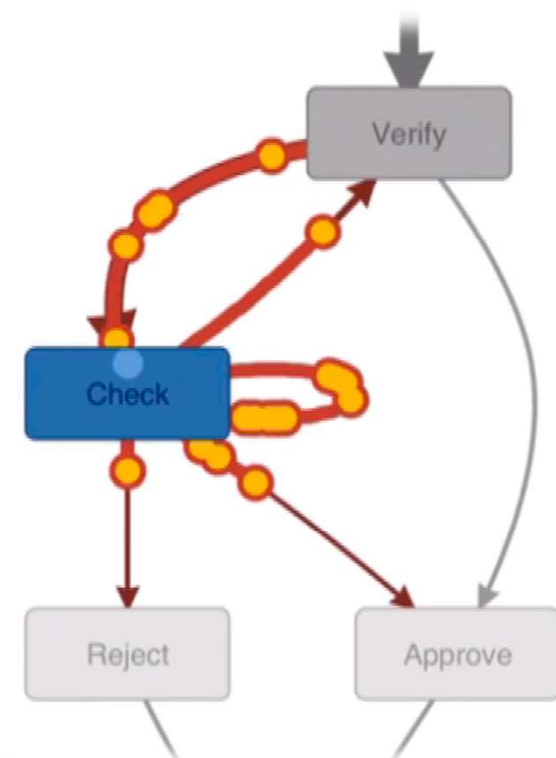
<https://youtu.be/azrwGm7BC3s>



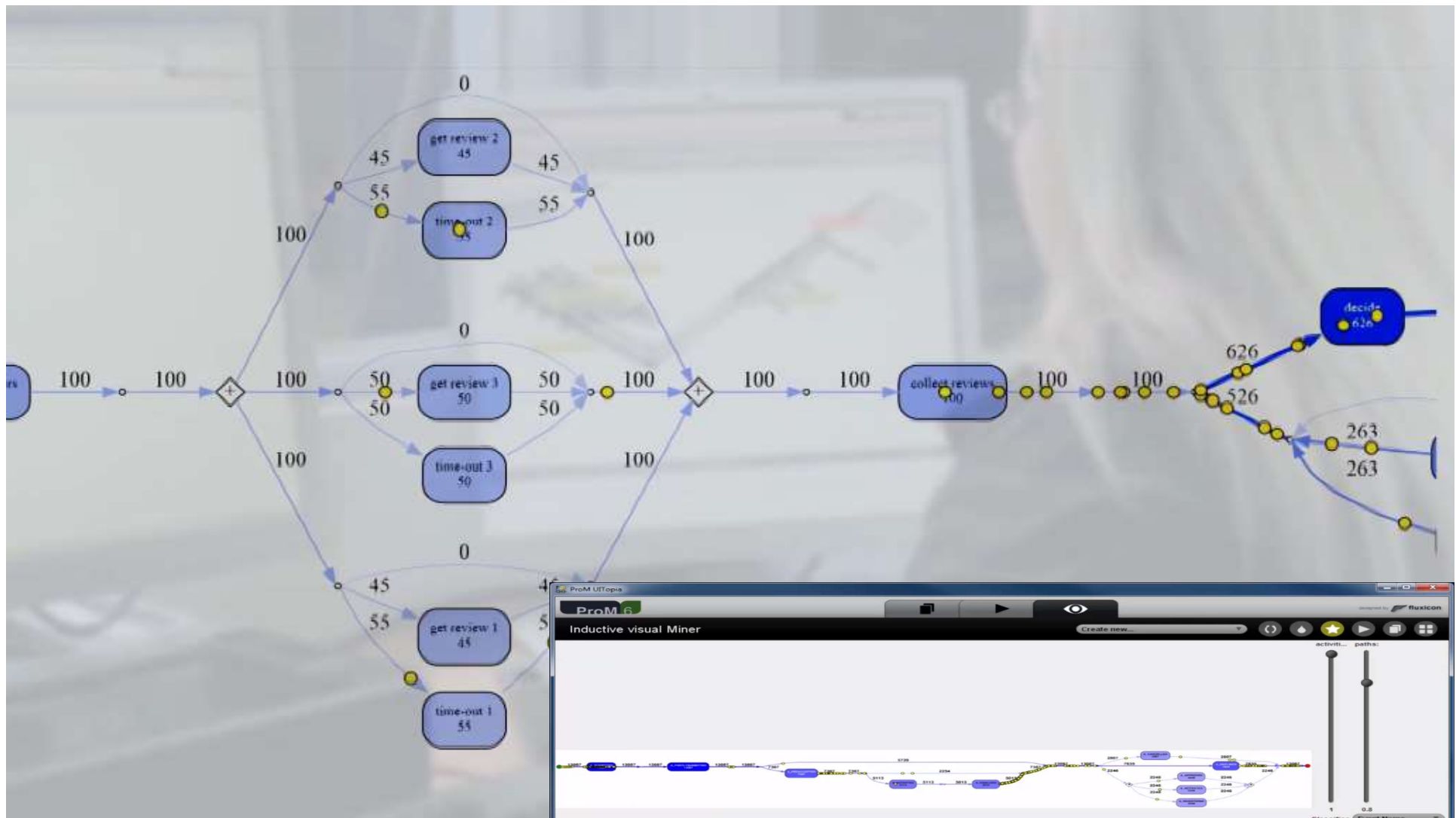
DEVIATIONS



REWORK



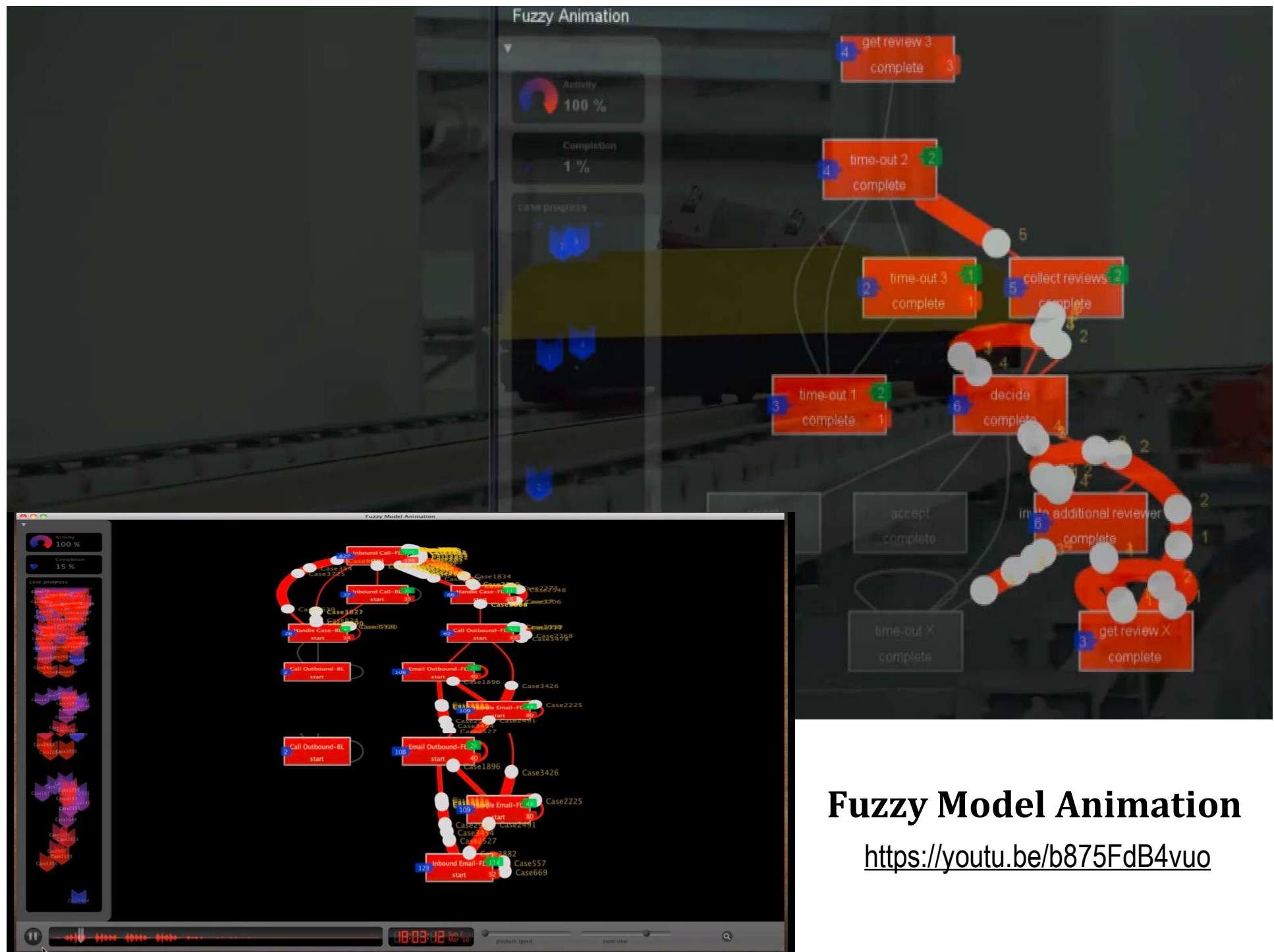
BOTTLENECKS



Inductive Visual Miner

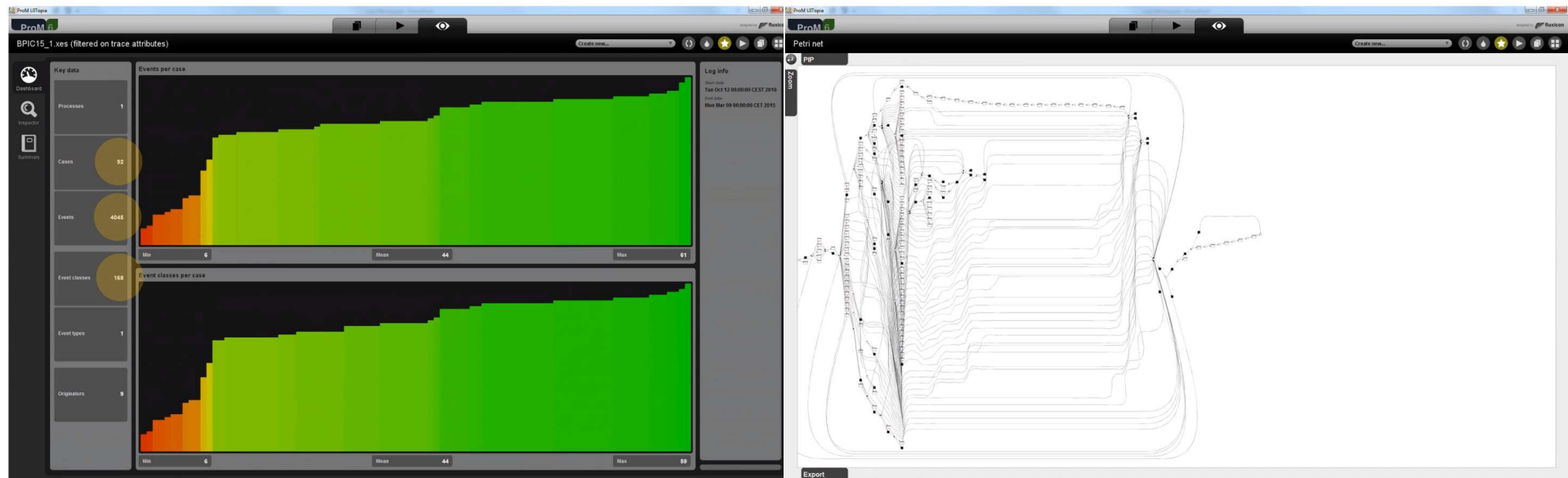
<https://youtu.be/dO4pIHIGeFY>





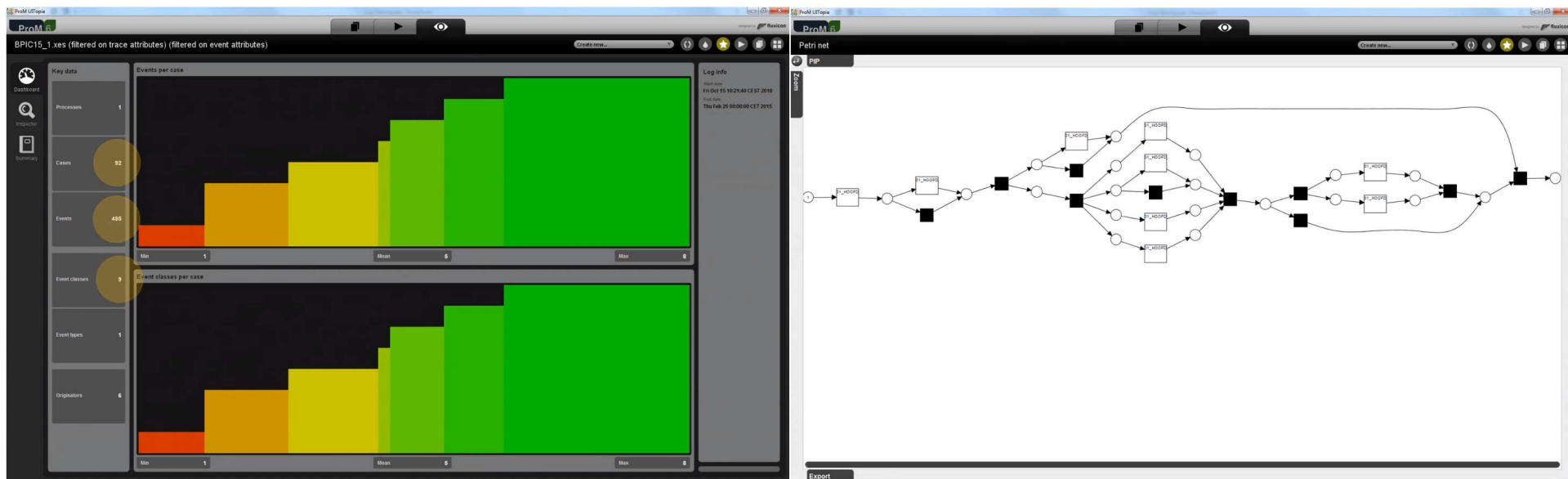
Fuzzy Model Animation

<https://youtu.be/b875FdB4vuo>



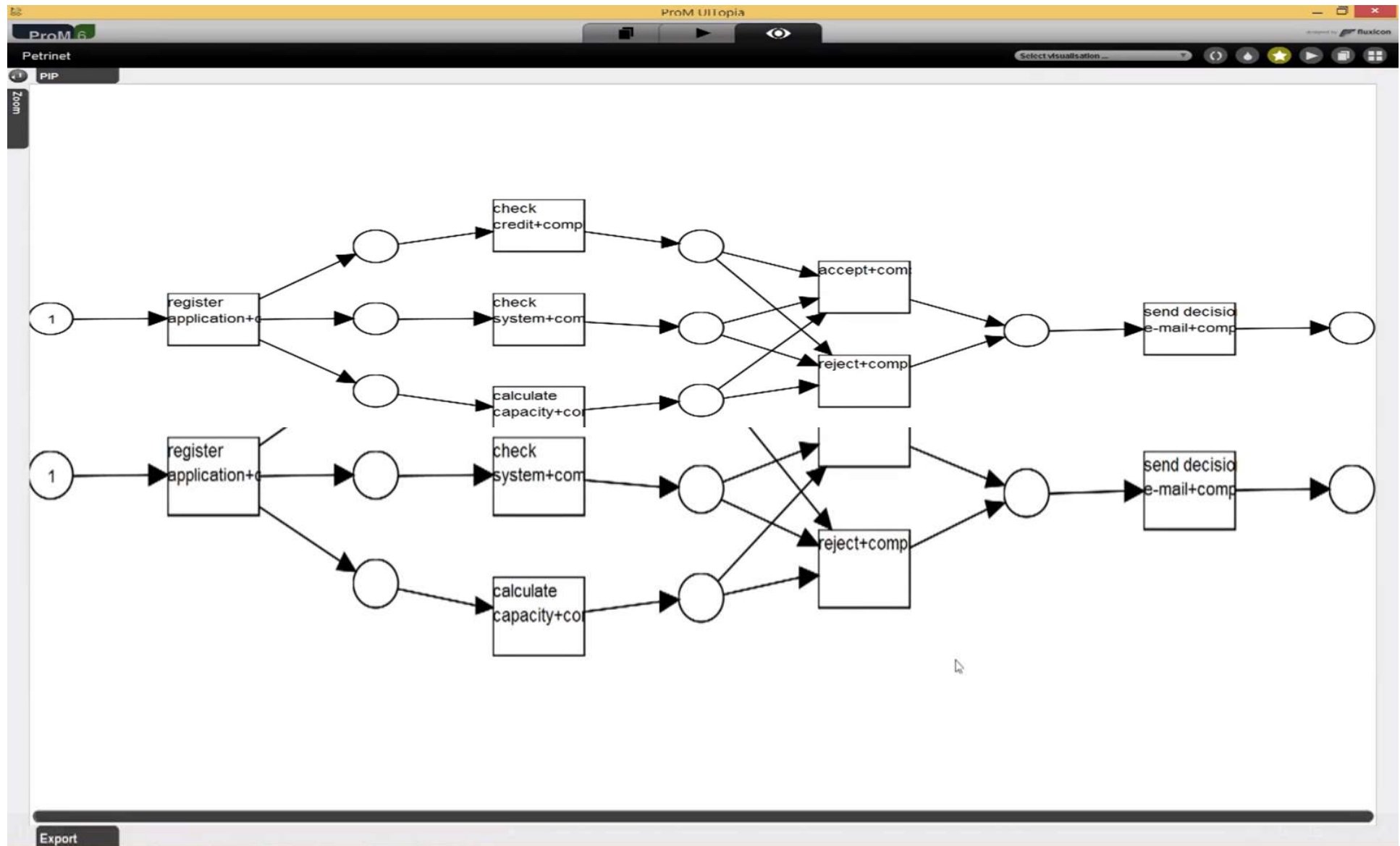
Log Filtering

<https://youtu.be/vCpxWycroSM>



Alpha miner in ProM

<https://www.futurelearn.com/courses/process-mining/0/steps/15637>



oktató anyagok



FutureLearn

Esemény logok:

- Event logs: <https://www.futurelearn.com/courses/process-mining/0/steps/15621>
- Event logs in ProM: <https://www.futurelearn.com/courses/process-mining/0/steps/15622>

ProM lite telepítése és használata:

- Installing ProM lite: <https://www.futurelearn.com/courses/process-mining/0/steps/15623>
- Using ProM lite: <https://www.futurelearn.com/courses/process-mining/0/steps/15624>
- ProM lite information: <https://www.futurelearn.com/courses/process-mining/0/steps/15625>
- Converting a CSV file to an event log: <https://www.futurelearn.com/courses/process-mining/0/steps/15626>
- Exploring event logs with the dotted chart: <https://www.futurelearn.com/courses/process-mining/0/steps/15627>
- Filtering event logs: <https://www.futurelearn.com/courses/process-mining/0/steps/15628>

Összefoglaló videó az esemény logokról és ProM lite használatáról:

<https://www.futurelearn.com/courses/process-mining/0/steps/15629>

Alfa algoritmus:

- Alpha miner: <https://www.futurelearn.com/courses/process-mining/0/steps/15636>
- Background information on the Alpha miner: <https://www.futurelearn.com/courses/process-mining/0/steps/15638>
- Alpha miner in ProM: <https://www.futurelearn.com/courses/process-mining/0/steps/15637>