

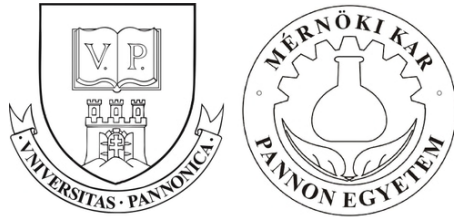
Doktori (PhD) értekezés tézisei

**Folyamatirányítási rendszerek  
gyakori mintázat keresés és  
folyamatbányászat alapú elemzése**

Bántay László

Pannon Egyetem  
Vegyészmérnöki- és Anyagtudományok  
Doktori Iskola

Témavezetők  
Dr. Abonyi János

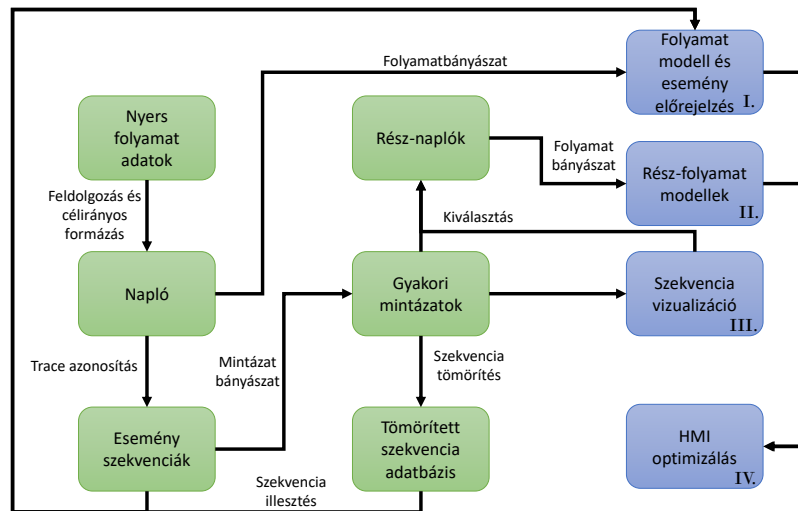


Folyamatmérnök Intézeti Tanszék  
Veszprém  
2025



# 1. Bevezetés és a dolgozat célja

Kutatásom az ipari irányítási rendszerek folyamataihoz kapcsolódó események elemzését és modellezését támogató adat- és folyamattudományi eszközök feltérképezésére irányult. Az Ipar 4.0 szemlélet megjelenése felértékelte a tárolt adatok jelentőségét a folyamatok mélyebb megismerésének igénye miatt. A folytonos adatok kezelésével kapcsolatban gazdag irodalom áll rendelkezésre, azonban a diszkrét adatok folyamatbányászat alapú elemzése és modellezése még tartalmaz kiaknázatlan lehetőségeket. Munkám célja a meglévő technikák feltérképezése mellett új, ipari napló fájlok feldolgozását támogató eszközök kifejlesztése volt, fókuszban a hibajelek és vészjelzések kezelésével.



1. ábra. A disszertáció felépítése. A folyamatbányászat alapja a tárolt nyers adatok, amiket szabványos formátumra kell hozni. Ez a szabványos formátum lehetővé teszi folyamatmodellek és esemény mintázatok feltárását, támogatva egy esemény előrejelző rendszer megalkotását. A gyakori esemény mintázatok al-folyamat modellek létrehozására is felhasználhatóak, a megfelelő mintázatok kiválasztását támogathatja azok hálózatszerű ábrázolása. A felsorolt technikák együttesen egy gépi tanulás alapú ember-gép interfész optimalizáló eszköztárat alkotnak.

Az ipari adatok gyűjtése és tárolása egyszerűvé és olcsóvá vált a huszadik század elején, ami a számítási sebesség exponenciális emelkedésével együtt áttörést okozott a folyamatbányászat területén. A folyamatbányászat bemeneti feltétele egy megfelelő formátumú napló fájl, ami rendelkezik a tárolt adatokkal kapcsolatos minden lényeges információval. A kötelező tulajdonságok mellett (esemény azonosító, időbélyeg, eset azonosító) az adatok gazdagíthatóak (erőforrás, szerkezeti szerep, költség, stb.), ezáltal célzott vagy hierarchikus folyamatbányászattal specifikus *folyamatmodellek* válnak elérhetővé. Egy megfelelő formátumú napló fájl további előnyöket hordoz magában, például gyakori mintázatok nyehetőek ki belőle amikkel *előrejelezhetünk eseményeket*. A gyakori szekvenciákat tömörített adatbázisba mentjük, majd az aktuális eseményláncokból alkotott szekvenciákat szekvencia illesztéssel összehasonlítjuk ezzel az adatbázissal létrehozva egy valószínűségi előrejelzést.

Az ipari naplófájlok egyik problémája a párhuzamosan futó folyamatok ugyanazon naplóba rögzítése, ami hibás és nehezen értelmezhető folyamat modelleket eredményezhet. Feltételezve, hogy az egyes folyamatok tipikus eseménycsoportokból állnak, gyakori szekvencia kereső technikákkal azonosíthatóak a rész-folyamatok rész-naplói, támogatva a *rész-folyamat modellek* előállítását. Az ilyen jellegű napló particionálási feladatoknál a megfelelő gyakori mintázatok kiválasztása okoz kihívást. Nagyméretű naplóknál hatalmas mennyiségű mintázat adódhat, amit nehéz értelmezni. Ennek kezelésére léteznek gyakori szekvencia utókezelési módszerek, az egyik lehetőség a vizualizáció. A szekvenciákat egy hálózat elemeiként értelmezve létrehozható egy információgazdag ábrázolás. A kifejlesztett módszerrel a rész-napló generálás mintázat választó lépése gyorsabb és egyszerűbb.

A rész-log alapú célzott folyamat feltárások mellett a kapott modellek hasznosak lehetnek kettes és hármas szintű ember-gép interfész (*HMI*) felületek fejlesztéséhez. Ezek a felületek jellemzően a

---

P&I diagrammok alapján kerülnek megrajzolásra, ami megfelelő az egyes szintű képernyőkhöz, de a magasabb szintűekhez más megközelítés szükséges. A dolgozók munkavégzésének támogatására, ami ember-központú HMI-k fejlesztését jelenti, a folyamat szemléletű elrendezések nyújthatnak megoldást. A folyamatbányászat eredményeit hálózatelmélettel kombinálva azonosíthatóvá válnak az egy művelethez tartozó beavatkozók, amiket egy képernyőre kell helyezni.

## 2. Felhasznált módszerek és eszközök

Kutatásom során, átfogó irodalmi áttekintésre támaszkodva összegyűjtöttem a létező adat- és folyamatbányászati technikákat, azonosítva hiányosságaikat DCS (distributed control system) adatokon végzett ipari folyamatvizelésekben alkalmazásuk során. A felmerülő hiányosságok kezelésére új folyamat elemzési eszközök kerültek kidolgozásra:

- egy célzott napló átalakítási módszer a naplók standard formátumra hozása érdekében,
- egy szekvencia illesztés-alapú esemény előrejelző eljárás,
- egy gyakori szekvencia keresésen alapuló napló fájl partícionálási megoldás,
- egy hasonlóság- és hálózat elmélet-alapú szekvencia ábrázolási módszer,
- és egy gépi tanlással támogatott HMI optimalizáló eljárás.

A napló adatok két forrásból származtak:

- DCS adatok a MOL százhalombattai finomítójának különböző egységeiből,
- és kattintási adatok egy publikus adattárból.

Az adatfeldolgozási és elemzési lépések részben ProM-ban zajlottak. Annak érdekében, hogy a kifejlesztett megoldások rugalmasan alkalmazhatóak legyenek bármilyen forrásadatra, a keretrendszer Python alapokra lett helyezve. A Python környezet széles körben biztosít megoldásokat, napló átalakítást XES (eXtensible Event Stream) formátumra, az SPMF könyvtár segítségével gyakori mintázatok keresését, a pm4py folyamatbányászati eszköztárat, és más könyvtárakat vizualizációs feladatokhoz (graphviz, networkx). A kifejlesztett módszertan és Python csomag az adatforrástól függetlenül felhasználható diszkrét adatok elemzéséhez.

---

### 3. Tézisek

#### Tézis I.

**Bizonyítottam, hogy célirányos trace azonosítással és napló formázással a folyamatbányászat alkalmas eszköz ipari folyamatok elemzésére, feltárására és előrejelzésére. [1,2]**

Az ipari folyamatok elemzése információ kinyerés a tárolt kapcsolódó eseményekből. A kapcsolódó események trace-eket alkotnak a naplóban. Ezen trace-ek meghatározása alapvető fontosságú az ipari folyamatok feltárásában. Mindazonáltal, ez nem egy triviális feladat, mivel a naplózott adatok rendszerint nincsenek címkézve ilyen szempontból. Vannak bizonyos technikák az események csoportosítására; az egyik módszer egy minimum időablak meghatározása két esemény között. A napló struktúrálása és címkézése is elkerülhetetlen. A megfelelő adat címkézés célirányos és hierarchikus folyamatbányászatot tesz lehetővé, így a folyamatok minden aspektusa értelmezhetővé válik. Alarm menedzsmentben az alarm-ok, az operátori beavatkozások, és a return-to-normal események együtt és külön is kiértékelhetőek, ugyanannak a folyamatnak a különböző rétegeit feltárva.

Egy jól megtervezett napló lehetővé teszi gyakori szekvenciák kinyerését, amik hasznos információt nyújtanak a folyamatok előrehaladásáról. A leggyakoribb esemény láncok tömörített szekvencia adatbázisba gyűjtésével és azok összehasonlításával az aktuális eseménysorral, a legvalószínűbb bekövetkező események előrejelezhetővé válnak.

A kifejlesztett trace azonosítási és napló szerkezet definiáló módszer egy hidrogén-fluorid alkiláló üzem alarm menedzsment racionalizáló projekt során került alkalmazásra. Azonosítottam a kulcs eseményeket és különböző naplókat állítottam elő a különböző célú folyamat elemzési feladatokhoz. A kapott naplók folyamatbányász eszközökkel kerültek feldolgozásra. A projekt bebizonyította, hogy a folyamatbányászat eszköztárával az alarmok racionalizálhatóak; ezáltal az operátorok munkája biztonságosabb és hatékonyabb lesz, valamint a munkaterhelés is csökkenthető. A szekvencia kompresszió és illesztő eljárás egy késleltetett kocszoló üzem valós adataival került alkalmazásra, alkalmazási korlátainak meghatározása mellett. Az eredmények alapján megállapítható, hogy a módszer nagyon hatékonyan gyűjti a gyakori mintázatokat, ami a szekvencia illesztéssel kiegészítve képes másodpercek alatt felismerni egy működési állapotot mindössze néhány tipikus alarm után és összehasonlítani azt a meglévő historikus mintázatokkal; így könnyedén nyerhetőek ki nagy valószínűségű előrejelzések.

#### Tézis II.

**Kifejlesztettem egy gyakori szekvencia alapú napló partícionáló módszert az egyazon naplóban tárolt részfolyamatok specifikus modelljeinek előállítására érdekében. [3]**

A javasolt módszer egyszerre szűri ki a felesleges eseményeket és hozza létre a rész-naplókat, kombinálva a gyakori mintázat keresést és a hagyományos process cube műveleteket. Az eredeti naplón gyakori elemhalmaz és szekvencia keresést végeztem, a kapott elemhalmazok és szekvenciák képezték a napló partícionálás alapját. Amennyiben a napló szerkezete megfelelő (Tézis I.), hierarchikus partícionálás is lehetővé válik.

A módszert egy ipari hidrogén-fluorid alkiláló üzem esemény naplóján alkalmaztam. A kapott modelleket a javasolt mérőszámok alapján értékeltem. Az eredmények igazolták a módszer hatékonyságát az események folyamat szemléletű csoportosítása terén, amivel célzott folyamat elemzés válik lehetővé kezelve a tárolt párhuzamos folyamatok problematikáját.

- Előnyök
  - A szekvencia kereső és folyamatbányász algoritmusok ugyanazt a bemeneti adatforrást igénylik, így nincsen szükség a napló további feldolgozására.
  - A módszer számítási kapacitás szempontjából kevésbé erőforrás igényes, mivel jelentősen csökkent méretű naplókat dolgoz fel.
  - Az előzetes ismeretek segíthetik a releváns gyakori szekvenciák azonosítását, egyszerűsítve az iteratív munkavégzést.
  - A meghatározott mérőszámok lehetővé teszik az eredmények objektív értékelését. A kiértékelési lépés elősegíti a módszer automatizálását.

- 
- Követelmények - Megkötések
    - Ugyan a módszer mintázat kereső eszközt igényel, ez nem probléma, mivel széles körben elérhetőek open-source eszközök.
    - Valamennyi rendszer ismeret szükséges a megfelelő mintázat kereső algoritmus kiválasztásához. Ezzel kapcsolatban ajánlások kerültek megfogalmazásra.
    - A folyamat előzetes ismerete megkerülhetetlen. Hasonlóan a *Knowledge Discovery Databases* (KDD) folyamat modellhez, a módszer iteratív és interaktív karaktere könnyít ezen a problémán.

### Tézis III.

**Kifejlesztettem egy hálózat-alapú gyakori szekvencia ábrázoló módszert (NBVFS) ami támogatja az eseményláncok gyors megértését. [4]**

A kifejlesztett módszer az esemény szekvenciákat ego-hálózatokká alakítja, amelyek a hálózat középpontjába helyezett közös kiindulási eseménnyel rendelkező szekvenciákból állnak. Mivel a különböző szekvenciák különböző esemény bővüléssel rendelkeznek, az elágazások fa struktúrárt eredményeznek. A módszer kulcsa, hogy miként helyezzük el a hálózat csomópontjait (az  $n$ -hosszúságú szekvenciákat). A szekvenciák hosszát, frekvenciáját (*support*), és a különböző eseményláncok közötti hasonlóságot használtam. Három vizualizációs eljárás került kidolgozásra:

- **NBVFS-WN (Network-Based Visualization of Frequent Sequences - Weighted Network):** ez a módszer egy konfidencia-alapú mátrixot és a szekvenciák számított mérőszámait használja. Eredménye egy súlyozott hálózat, amiben azon szekvenciák szerepelnek, amelyek egymás bővítései. A súlyérték az átmenetek konfidenciája. Ez a fajta vizualizáció elősegíti az egyes esemény előfordulások feltételeit és következményeit.
- **NBVFS-CM (Network-Based Visualization of Frequent Sequences - Confidence-based MDS projection):** ez a módszer hasonlóság számításán alapul, az átmenetek konfidencia értékeit felhasználva a hasonlóság mértékének megállapításához. A szomszédsági mátrixot a nem direktben kapcsolódó szekvenciák hasonlóság értékeivel gazdagíthatjuk, tranzitív távolság kalkuláció alkalmazásával. A csomópontok pozícióját Multidimensional scaling (MDS) segítségével határozhatjuk meg, minél hasonlóbb két szekvencia, annál közelebb kerülnek a hálózaton belül. Ez a típusú vizualizáció több információval szolgál a szekvenciák kapcsolatáról.
- **NBVFS-TM (Network-Based Visualization of Frequent Sequences - Transaction-based MDS projection):** a csomópontok ebben esetben is MDS segítségével számítnak, a különbség a hasonlóság számításban van az NBVFS-CM módszerhez képest. Itt a hasonlóság alapja a közös támogató tranzakciók száma. Ez a megközelítés közelebb áll a folyamatbányászathoz, mivel a tranzakciók megfeleltethetőek a trace-eknek, így visszajelzést kaphatunk a bányászati folyamatról.

Az idősor jellegű adatbázisok hálózat jellegű ábrázolása sok területen alkalmazható. A bemutatott eljárás egy célzott elemző eszköz az eseményszekvenciák vizualizációja által esemény adatbázisokból történő hasznos tudás kinyerésére. Interaktív jellege által alkalmas iteratív adat-elemző feladatokra, például lehetővé teszi a gyakori szekvencia kereső algoritmusok paraméter hangolását. A javasolt megjelenítési módszer a fontos eseményláncok gyors felismerését biztosítja azok legfontosabb paramétereivel - támogatottság és konfidencia - együtt. Az ilyen jellegű információ értelmezés az információ feldolgozási eljárások támogatása mellett felgyorsíthatja a gépi tanulási algoritmusok paraméterezését is. Ezekon kívül validálhat trace generálási szabályokat folyamatbányászati tevékenységek során, működését az adott feladat függvényében kell személyre szabni.

---

## Tézis IV.

### Javaslatot tettem egy gépi tanuláson alapuló ember-központú és folyamat-tudatos HMI képernyő fejlesztési módszerre. [5]

A kifejlesztett módszer az ISA 101 szabvány *Navigation and Layout* elvének alkalmazását támogatja, egészen pontosan a *grouping related elements together* megközelítést. A szóban forgó iteratív eljárás a megfelelő gyakori mintázat- és folyamatbányászati technikák kombinációját alkalmazza. A javasolt HMI képernyők a P&I diagrammok, a tipikus esemény/alarm szekvenciákat alkalmazó folyamat deviancia modellek, valamint a jellemző operátori beavatkozásokból előállított folyamatirányítási modellek kombinált értelmezésén alapulnak.

Egy hidrogén-fluorid alkiláló üzem alarmjai, operátori beavatkozásai és képernyőn végzett akciói kerültek elemzésre. A napló a tárolt események gyakori szekvenciáinak keresésével került partíciónálásra. A folyamatok modelljeinek felderítése folyamatbányászati eszközökkel történt. A tipikus folyamatok hálózat szerű ábrázolása lehetővé tette a munkafolyamatok számának és az azokat tartalmazó képernyők számának összehasonlítását. A HMI képernyők meglévő felépítése és a kapott modellek alapján javaslatot tettem egy magasabb rendű képernyő felépítésre.

Annak ellenére, hogy nem állt rendelkezésre a rendszerre vonatkozó széleskörű ismeret, a módszer hatékonyan bizonyult. Mindazonáltal manuális munkára is szükség van a képernyők validálásához, ami szubjektív döntésekhez vezethet. A módszer továbbfejleszthető objektív teljesítmény mutatók definiálásával, amik támogatják a minimalizációs probléma megoldását.

## 4. Az eredmények jövőbeli alkalmazási lehetőségei

Ugyan a kifejlesztett módszereket sikeresen alkalmaztam alarm és folyamat elemzési feladatokban, függetlenül az adatforrástól más kutatási területeken - például EDU-miningban - is alkalmazhatóak, valamint adottak potenciális fejlesztési irányok.

Az aktuális folyamatadatok összehasonlítása előzőleg meghatározott eseménysorokkal lehetővé teszi egy részben vagy teljesen automatizált HAZOP elemző eszköz kifejlesztését. A kapott folyamat mintázatok referencia modellel történő összehasonlításával az operátorok munkája minősíthető, ami elősegítheti az Operator Training System (OTS) fejlesztését.

Egy másik fejlesztési irány lehet a gépi tanulás integrálása SCADA rendszerekbe, ami az iparban egy aktuális topik ahogy egyre több megoldás veszi igénybe a mesterséges intelligencia segítségét. A javasolt módszertan képes feldolgozni a megszerzett adatokat és mintázatokat találni azon eseményláncok között amik hibákhoz vagy bizonyos operátori beavatkozásokhoz vezetnek. A gépi tanulás alkalmazása lehetővé teszi az ipari irányítási rendszerek dinamikus és adaptív fejlesztését.

A trend chart-ok témaköre is érdekes lehet. Gépi tanulással támogatott megoldások segítségével folyamatváltozók összefüggéseit nyerhetjük ki a historikus adatokból, képessé téve a rendszert bizonyos predikciókra. Például, amennyiben az operátor előhív egy trend chart-ot, egy intelligens rendszer javaslatot tehet más trend chart-ok előhívására is, amelyek hasznosak lehetnek az adott munkafolyamathoz.

## Publikációk

1. L. Bántay, G. Dörgő, F. Tandari and J. Abonyi, „Simultaneous Process Mining of Process Events and Operator Actions for Alarm Management”, *Complexity, Frontiers in Data-Driven Methods for Understanding, Prediction, and Control of Complex Systems*, Vol. 2022, 2022.
2. L. Bántay and J. Abonyi, „Frequent pattern mining-based log file partition for process mining”, *Engineering Applications of Artificial Intelligence*, Vol 123., Part A, 2023.
3. L. Bántay, N. Sas, G. Dörgő and J. Abonyi, „Sequence Compression and Alignment-Based Process Alarm Prediction”, *Industrial & Engineering Chemistry Research*, Vol. 62(27), 10577-10586, 2023.
4. L. Bántay and J. Abonyi, „Machine Learning-Supported Designing of Human–Machine Interfaces”, *Applied Sciences, New Insights into Human-Computer Interaction*, Vol. 14(4), 2024.
5. L. Bántay and J. Abonyi, „Network-based visualisation of frequent sequences”, *PLOS ONE*, Vol. 19(5), 2024.