

Folyamatbányászat

Konformancia elemzés – megfelelőség
elemzés I.

A vizsgálat szükségessége

- Ipari rendszerekben a munkafolyamatok lefutása sok esetben eltérhet a tervezett (ütemezett) lefutástól.
- Ezen eltérések okainak felkutatásának első lépése, hogy megtaláljuk, hogy a valós lefutás hol, milyen mértékben tér el a tervezettől.
- Egy ipari folyamat modellje tartalmazza az adott műveletsor lépéseit, átmeneteit, erőforrás (emberi, anyagi) felhasználását, időigényeit, viszont nem tartalmazhat minden apró részletet a folyamattal és pl. az emberekkel kapcsolatban.
- Ezért van szükség a konformancia vizsgálatra, mely segítségével megkereshetjük a modell és a tényleges lefutások közötti különbségeket.
- Mivel a modell az optimális lefutást hivatott leírni, ezért bármilyen eltérés a modelltől negatív hatással van a lefutásra és eltávolítja azt az optimálistól.

Példák

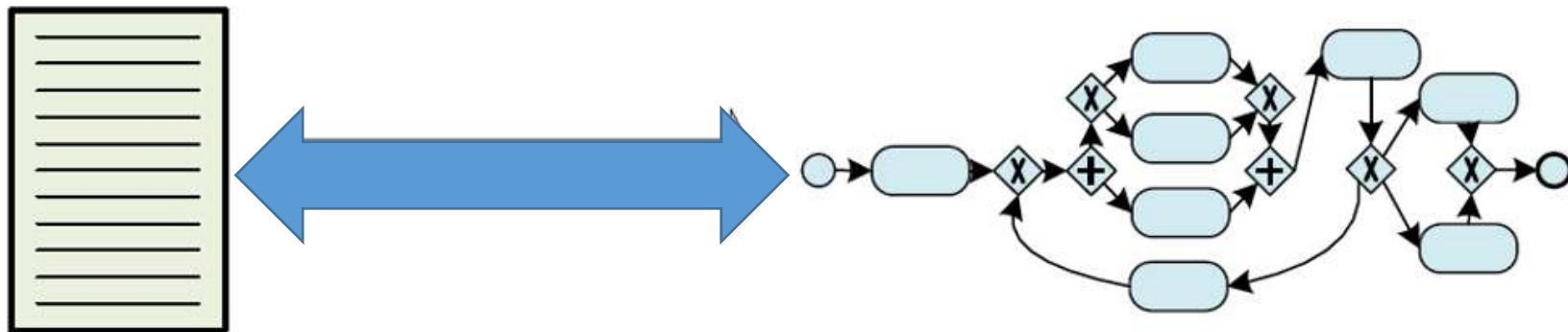
- **Első példa**, amikor **egy gépi erőforrás meghibásodik** és azt javítani vagy másikkal helyettesíteni kell. Ebben az esetben **a valós lefutás jelentősen el fog térni a modeltől**, mivel jelentős időbeli és erőforrásbeli eltéréseket mutathat.
- **Másik példa**, nagyon sok ipari folyamat során szükség van emberi erőforrásra, vagyis munkásokra, legyen ez szalag melletti összeszerelés vagy éppen könyvelés vagy más irodai munkák.

A gépesített munkavégzés egyik előnye a kézi munkavégzéssel szemben a gyorsaság és a pontosság, ellenben **amikor valamilyen hiba merül fel, akkor az ember képes azt felismerni és a legtöbb esetben gyorsan korrigálni**, míg egy meghibásodott gép sokkal drasztikusabb kiesést eredményez a termelésben.

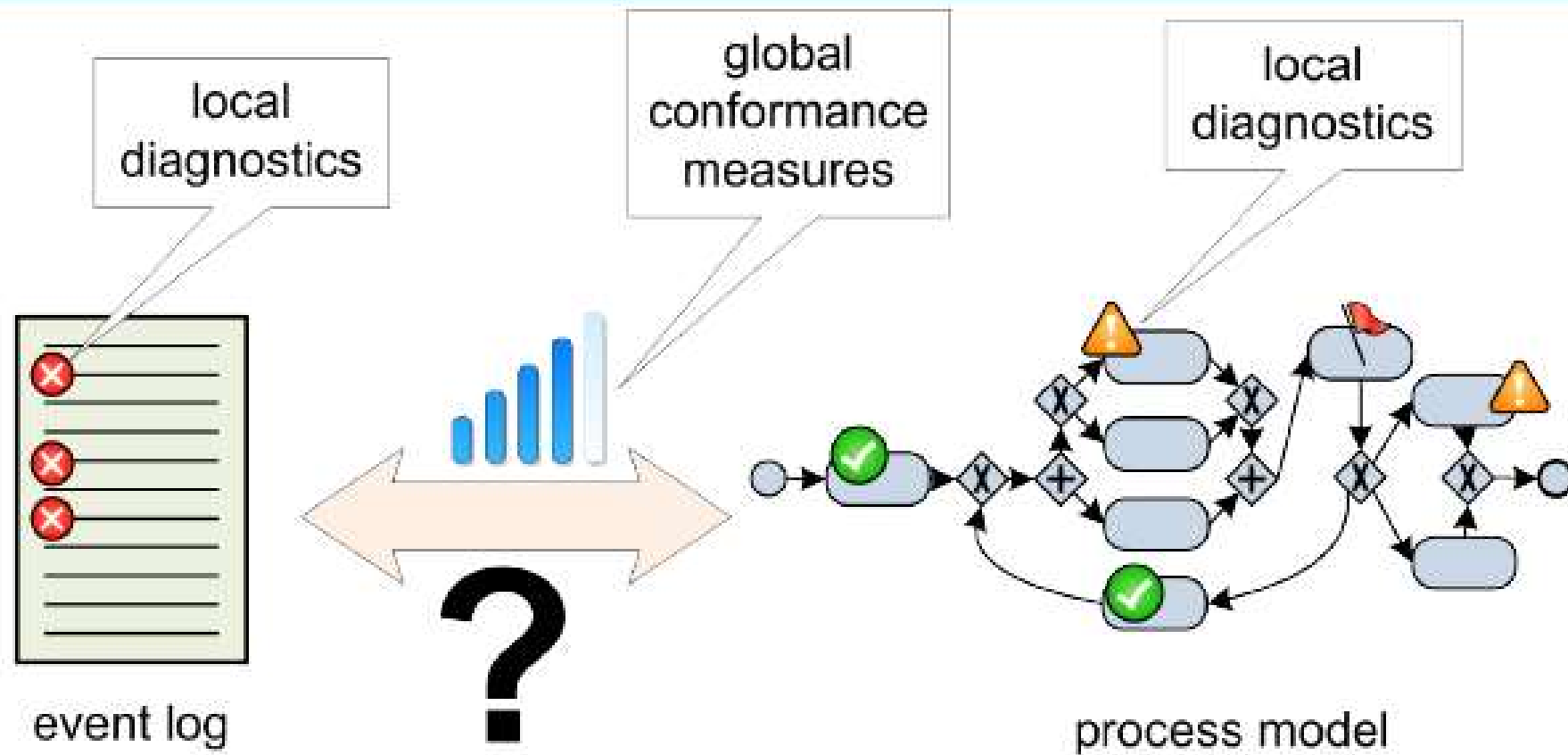
Még ezek tudatában is sokkal **nehezebb modellt készíteni egy emberek által elvégzett munkára**, mert az egyes emberek közötti különbségeken felül az idő és az aktuális hangulat is nagyon befolyásolja a munkavégzést.

Konformancia analízis (Conformance checking)

- Akkor használatos, ha van előzetes modell.
- A létező modellt hasonlítjuk össze a folyamat eseménynaplójával.
- Analizálja az eltéréseket az események és a modell között.
- Segítségével meghatározzuk az egyezés mértékét a modell által leírt működés és a tényleges esemény sorok között.



Konformancia, diagnosztika és mértékek



Megfelelően elkészített modell

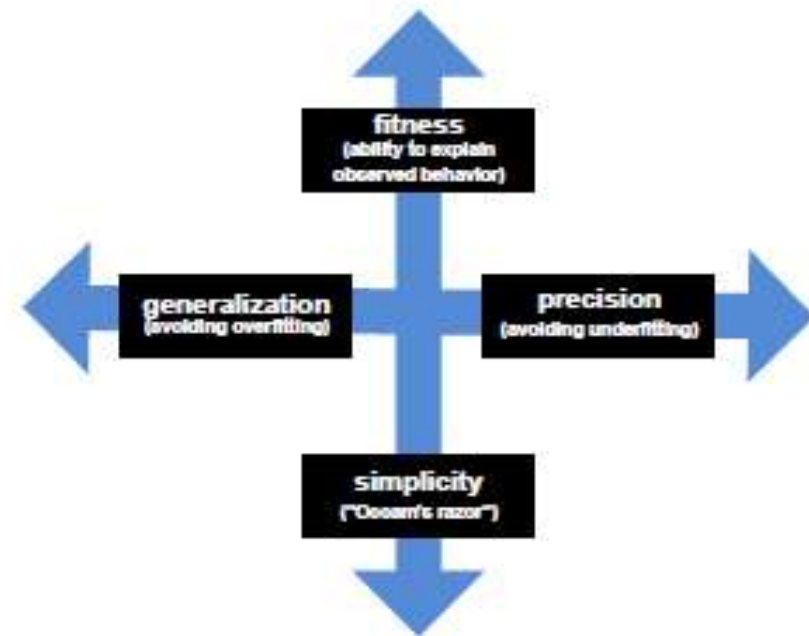
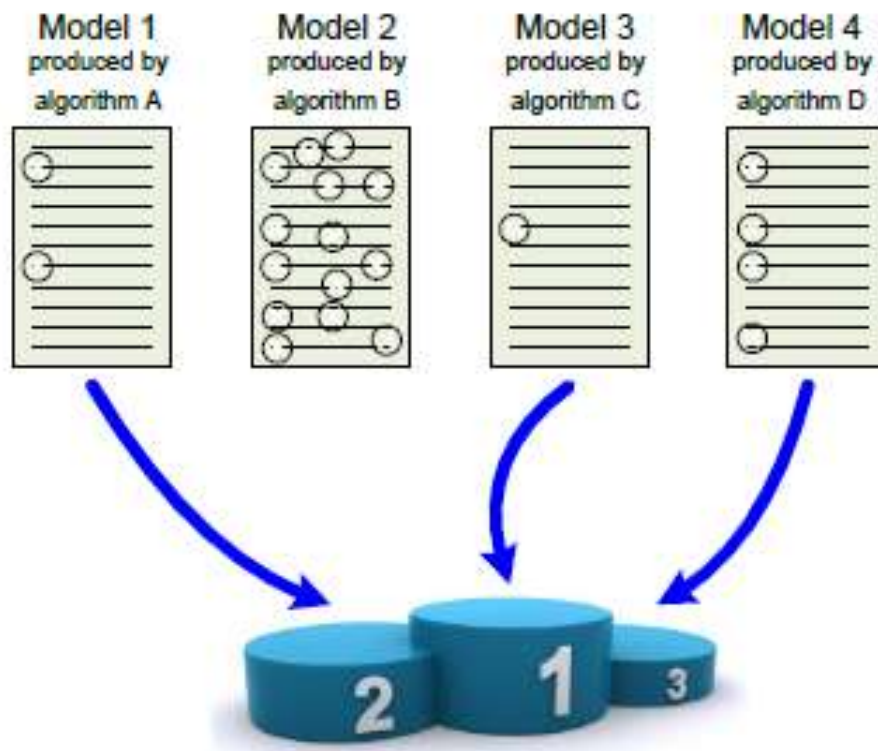
A modell négy fő tulajdonsággal rendelkezik:

- **megegyezőség** (fitness): az eseménynaplóban felsorolt események mennyire jól illeszkednek, azaz az általuk leírt útvonallal milyen mértékig lehet végig menni a modellen,
- **egyszerűség** (simplicity): a lehetséges legegyszerűbb modell mely leírja az események viselkedését a legjobb,

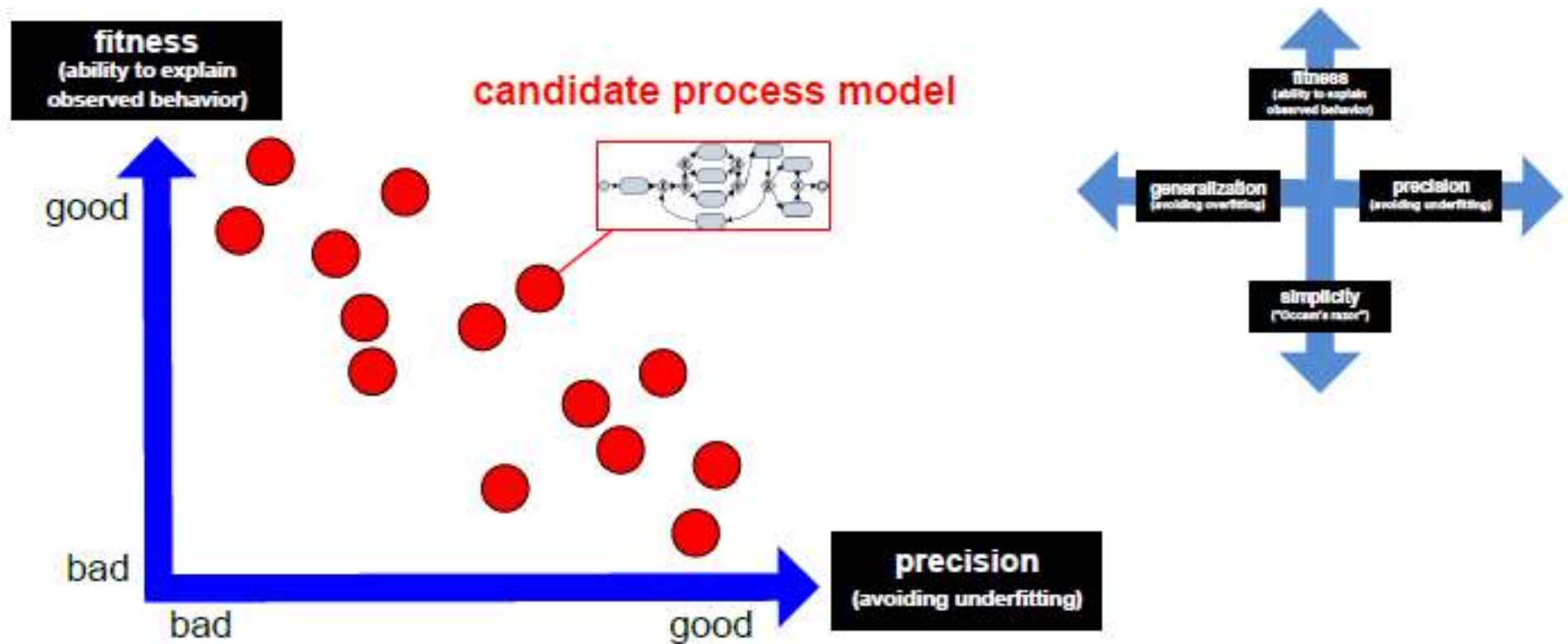
(Megj.: A megegyezőség és az egyszerűség önmagukban nem elegendőek a modell minőségének megállapításához.)

- **általánosítás** (generalization): törekedni kell minél általánosabb modell készítésére, azaz a modell minél több viselkedést ír le annál általánosabb,
- **pontosság** (precision): arra törekszünk, hogy a lehető legkevesebb viselkedést írjuk le (Megj.: Ha túlságosan specifikusan állítjuk össze a modellt, akkor az eseménynaplóban leírt események csak egy kisebb csoportját írja le jól, a többit hiányosan vagy egyáltalán nem.)

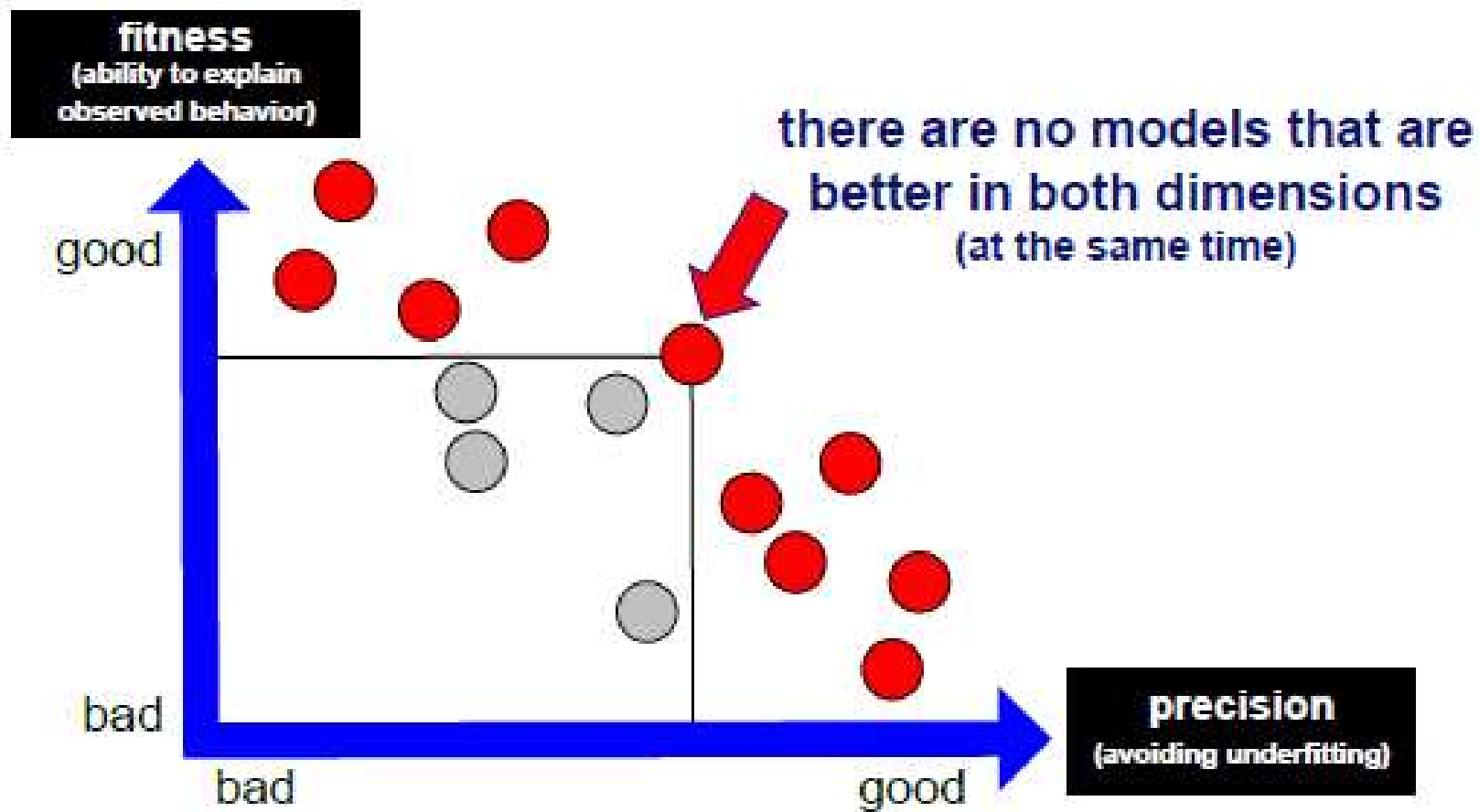
Folyamat feltáró algoritmusok vizsgálata



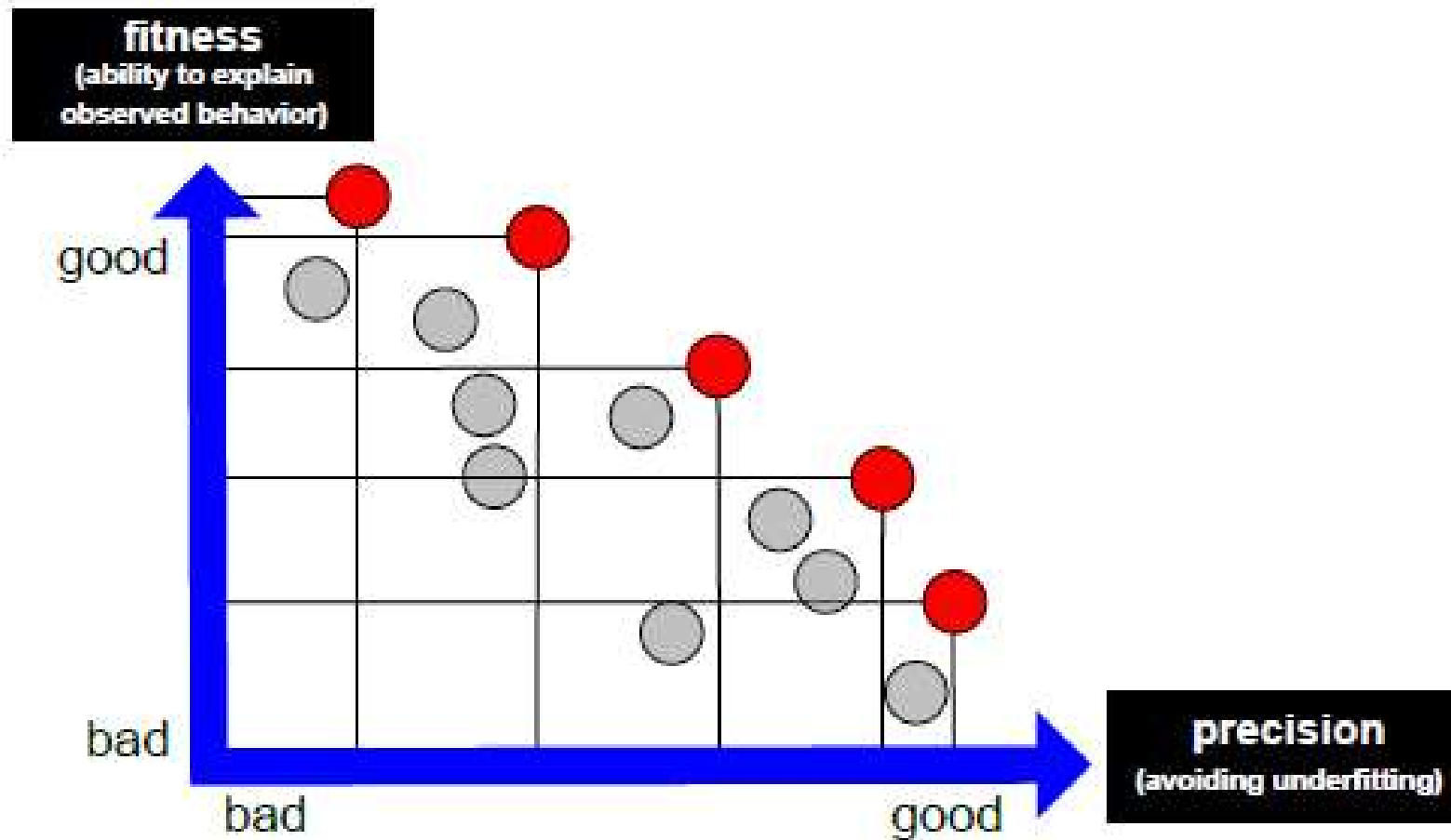
Pareto front 2 dimenziót figyelembe véve



Pareto front 2 dimenziót figyelembe véve



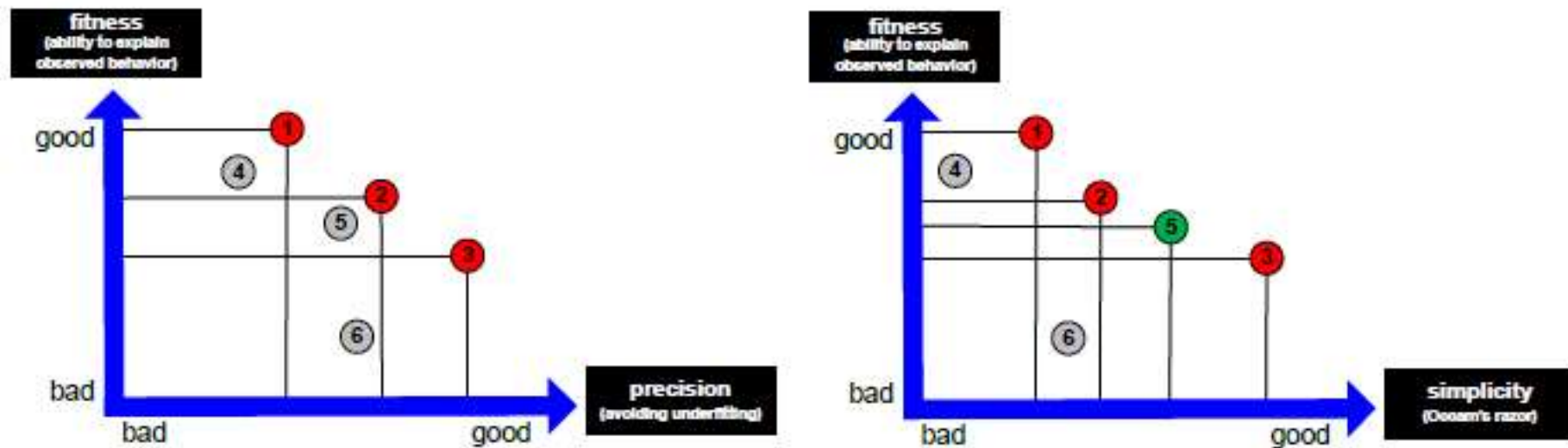
Pareto front 2 dimenziót figyelembe véve



Egy további dimenziót is megvizsgálunk

Nincs olyan modell, amely egy időben a legjobb fitness, pontosság és egyszerűség értékkel rendelkezik.

Nehéz a feltáró modelleket összehasonlítani!

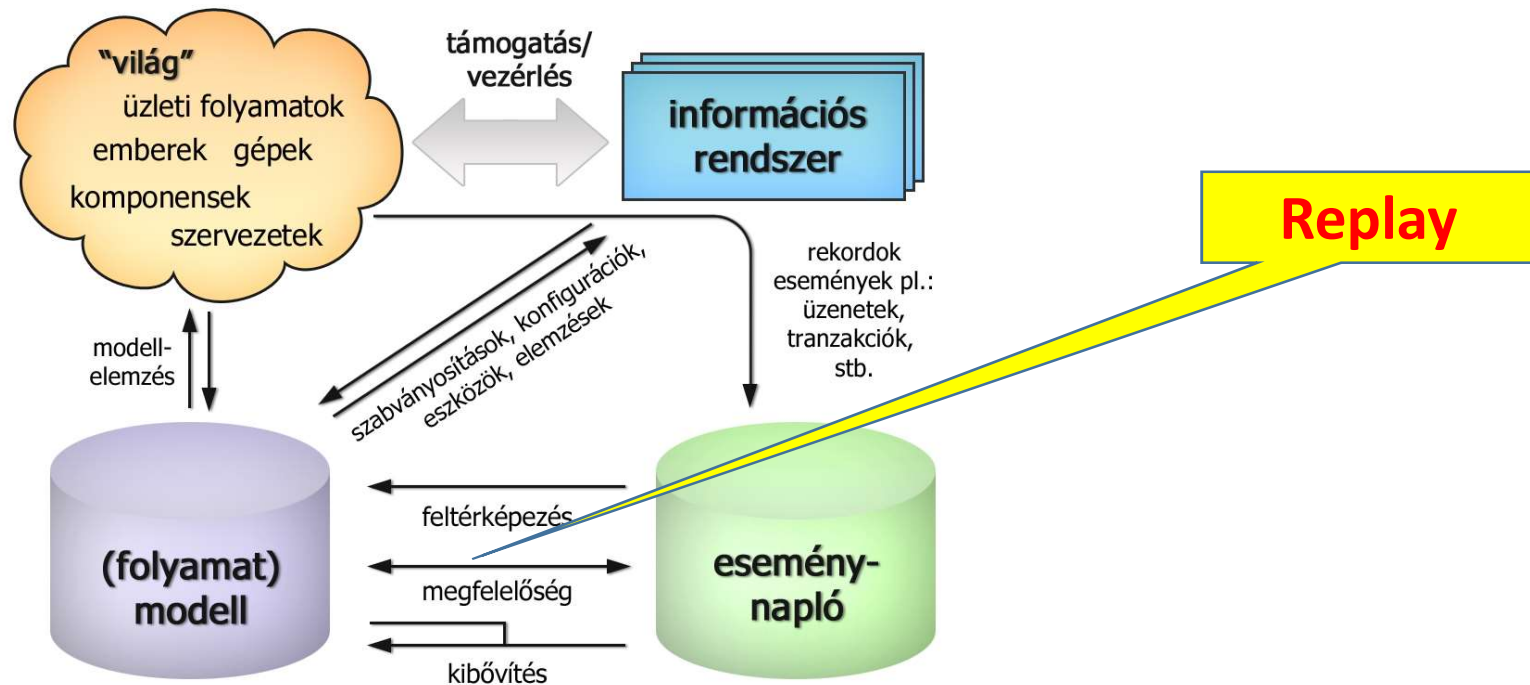


Miért nehéz a folyamat pontos feltárása?

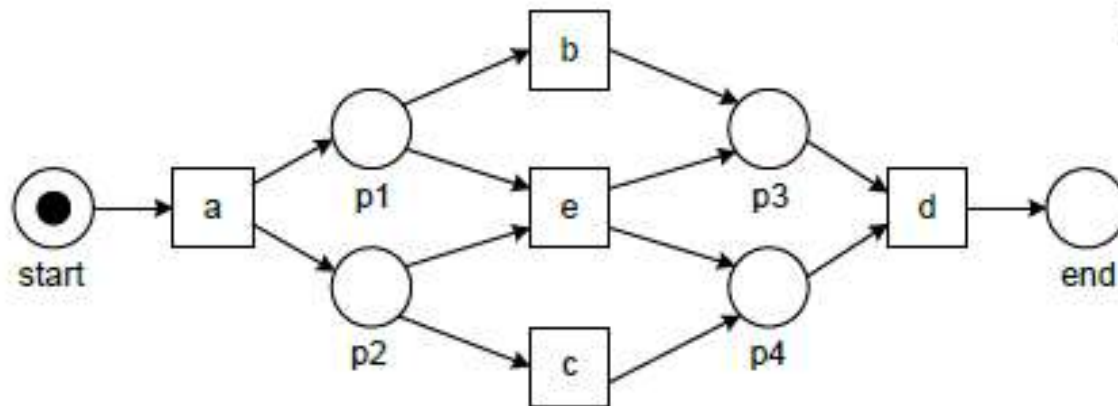
- **Nincsenek negatív példák** (azaz a logból láthatjuk, hogy mi történt, de azt nem láthatjuk, hogy mi nem történt)
- A párhuzamosság, a ciklusok és választások a keresési térnek egy komplex struktúráját adják meg, de **a logok** jellemzően csak **a lehetséges viselkedések egy töredék részét tartalmazzák**.
- **Nincs egyértelmű kapcsolat a modell mérete és viselkedése között.**
- Nagyon fontos a reprezentációból adódó **torzítás mértékének** meghatározása, figyelembevétele.
- A **modellező nyelvek adta lehetőségektől** is függ, hogy mit tudunk megjeleníteni, mi fér bele a modell leírásba. (UML, BPMN, Petri háló, stb.)

A konformancia ellenőrzés módszerei

1. Konformancia ellenőrzés **ok-okozati lábnyomokkal**,
2. Konformancia ellenőrzés **token alapú visszajátszással**,
3. **Illesztés alapú** megfeleléség **ellenőrzés**.



Konformancia ellenőrzés ok-okozati lábnyomokkal



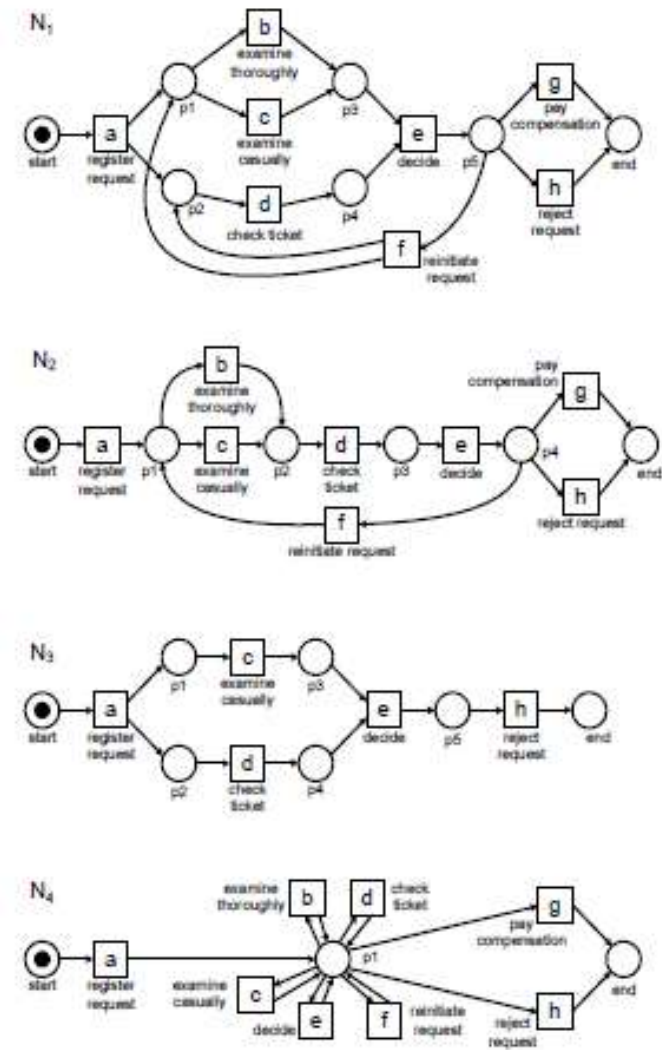
$$L_1 = [\langle a, b, c, d \rangle^3, \langle a, c, b, d \rangle^2, \langle a, e, d \rangle]$$

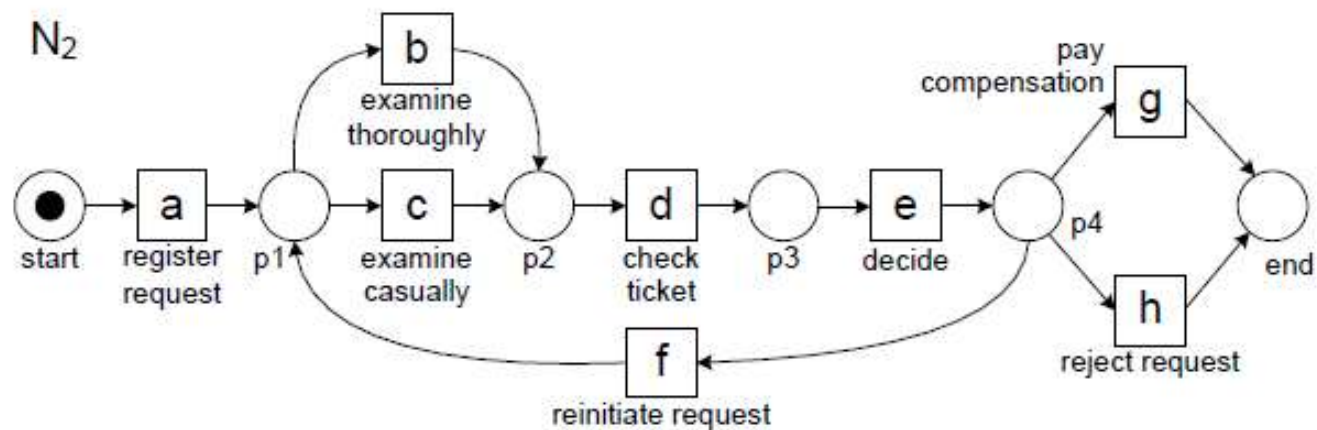
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
<i>a</i>	$\#_{L_1}$	$\rightarrow_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$
<i>b</i>	$\leftarrow_{L_1}$	$\#_{L_1}$	$\parallel_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
<i>c</i>	$\leftarrow_{L_1}$	$\parallel_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
<i>d</i>	$\#_{L_1}$	$\leftarrow_{L_1}$	$\leftarrow_{L_1}$	$\#_{L_1}$	$\leftarrow_{L_1}$
<i>e</i>	$\leftarrow_{L_1}$	$\#_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$

lábnyom alapú megfelelés = 1 (1 = tökéletes egyezés, 0 = nincs egyezés)

A lábnyomok megadják a log és a modell egybevágóságát.

#	trace
455	acdeh
191	abdeg
177	adceh
144	abdeh
111	acdeg
82	adceg
56	adbeh
47	acdefdbeg
38	adbeg
33	acdefdbeg
14	acdefdbeg
11	acdefdbeg
9	adcefcdeh
8	adcefdbeg
5	adcefbdeg
3	acdefbdefdbeg
2	adcefdbeg
2	adcefbdefdbeg
1	adcefbefdbeg
1	adbefbdefdbeg
1	adcefbefcdefdbeg
1391	





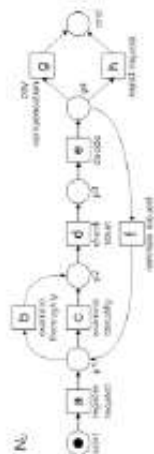
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
<i>a</i>	#	→	→	#	#	#	#	#
<i>b</i>	←	#	#	→	#	←	#	#
<i>c</i>	←	#	#	→	#	←	#	#
<i>d</i>	#	←	←	#	→	#	#	#
<i>e</i>	#	#	#	←	#	→	→	→
<i>f</i>	#	→	→	#	←	#	#	#
<i>g</i>	#	#	#	#	←	#	#	#
<i>h</i>	#	#	#	#	←	#	#	#

L_{full}

#	trace
455	acdeh
191	abdeg
177	acdeh
144	abdeh
111	acdeg
82	addeg
56	adbeh
47	acdehbeh
38	addeg
33	acdehdeh
14	acdehdeg
11	acdehdeg
9	acdehdeh
8	acdehdeh
5	acdehdeg
3	acdehdegdeh
2	acdehdeg
2	acdehdegdeh
1	acdehdehdeh
1	acdehdehdeg
1	acdehdehdegdeh

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
<i>a</i>	#	→	→	→	#	#	#	#
<i>b</i>	←	#	#		→	←	#	#
<i>c</i>	←	#	#		→	←	#	#
<i>d</i>	←			#	→	←	#	#
<i>e</i>	#	←	←	←	#	→	→	→
<i>f</i>	#	→	→	→	←	#	#	#
<i>g</i>	#	#	#	#	←	#	#	#
<i>h</i>	#	#	#	#	←	#	#	#

N_2



	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
<i>a</i>	#	→	→	#	#	#	#	#
<i>b</i>	←	#	#	→	#	←	#	#
<i>c</i>	←	#	#	→	#	←	#	#
<i>d</i>	#	←	←	#	→	#	#	#
<i>e</i>	#	#	#	←	#	→	→	→
<i>f</i>	#	→	→	#	←	#	#	#
<i>g</i>	#	#	#	#	←	#	#	#
<i>h</i>	#	#	#	#	←	#	#	#

A különbségek számszerűsítése

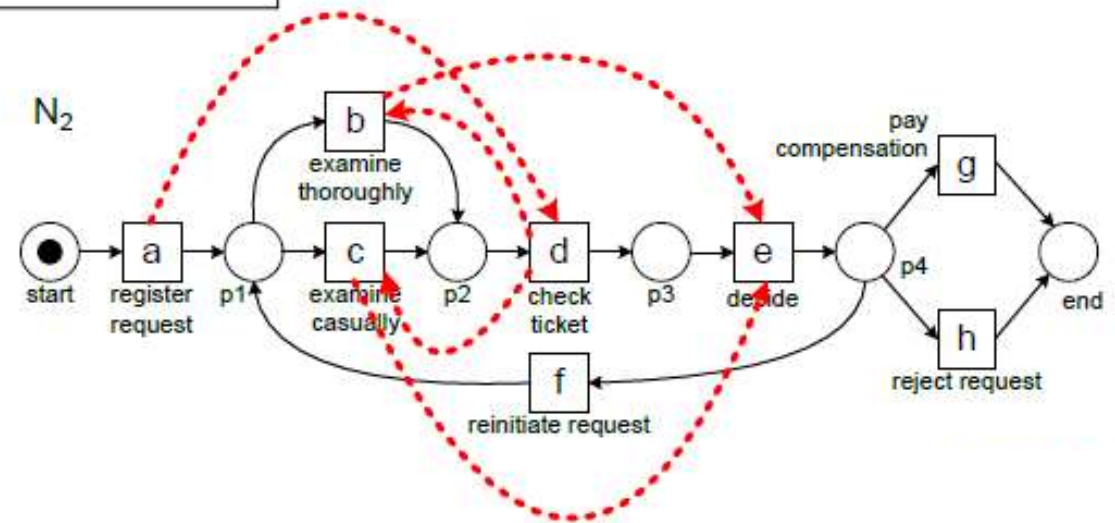
1 - x/y, ahol **x** a log-beli eltérések száma a modellbeli bejegyzésekhez képest, **y** az összes lábnyom száma a táblázatban

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
<i>a</i>				→: #				
<i>b</i>				:→	→: #			
<i>c</i>				:→	→: #			
<i>d</i>	←: #	:←	:←			←: #		
<i>e</i>		←: #	←: #					
<i>f</i>				→: #				
<i>g</i>								
<i>h</i>								

$$1 - \frac{12}{64} = 0.8125$$

A kapott eredmény felhasználása diagnosztikához

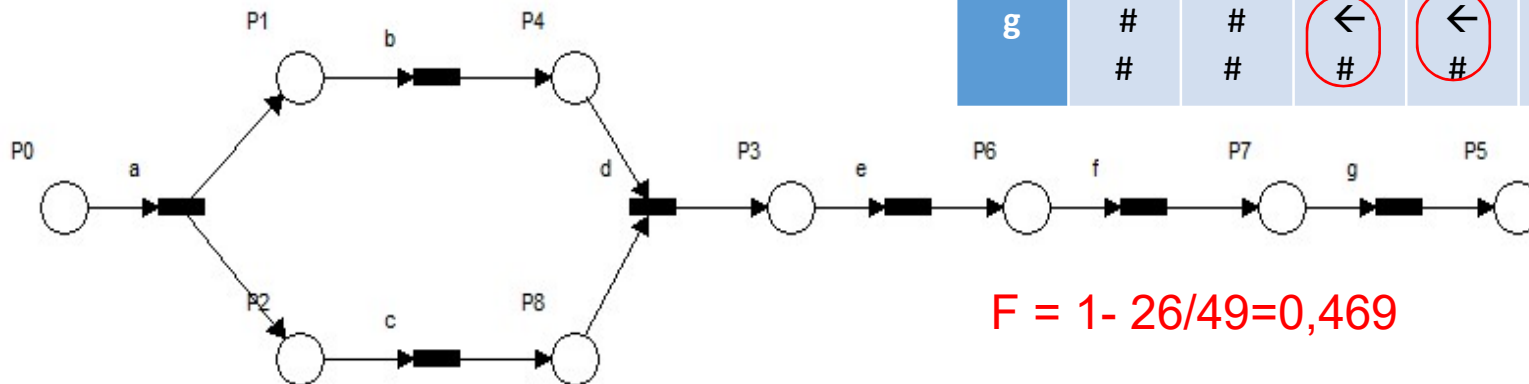
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
<i>a</i>				$\rightarrow; \#$				
<i>b</i>				$\parallel; \rightarrow$	$\rightarrow; \#$			
<i>c</i>				$\parallel; \rightarrow$	$\rightarrow; \#$			
<i>d</i>	$\leftarrow; \#$	$\parallel; \leftarrow$	$\parallel; \leftarrow$			$\leftarrow; \#$		
<i>e</i>		$\leftarrow; \#$	$\leftarrow; \#$					
<i>f</i>				$\rightarrow; \#$				
<i>g</i>								
<i>h</i>								



Számolja ki az egyezőség mértékét az alábbi log és a modell alapján!
Milyen következtetéseket lehet levonni a folyamat végrehajtásával kapcsolatban?

- a = lemez ketté vágása
- b = préselés
- c = esztergálás
- d = hegesztés
- e = csiszolás
- f = méretre szabás
- g = festés

$L=[\langle abcdg \rangle, \langle abdfeg \rangle, \langle abdfbeg \rangle, \langle aeg \rangle, \langle abfdcg \rangle]$



	a	b	c	d	e	f	g
a	# #	→ →	# →	# #	→ #	# #	# #
b	← ←	# #	→ 	→ →	→ #	 #	# #
c	# ←	← 	# #	 →	# #	# #	→ #
d	# #	← ←	 ←	# #	# →	 #	→ #
e	← #	← #	# #	# ←	# #	# →	→ #
f	# #	 #	# #	 #	# ←	# #	# →
g	# #	# #	← #	← #	← #	# ←	# #

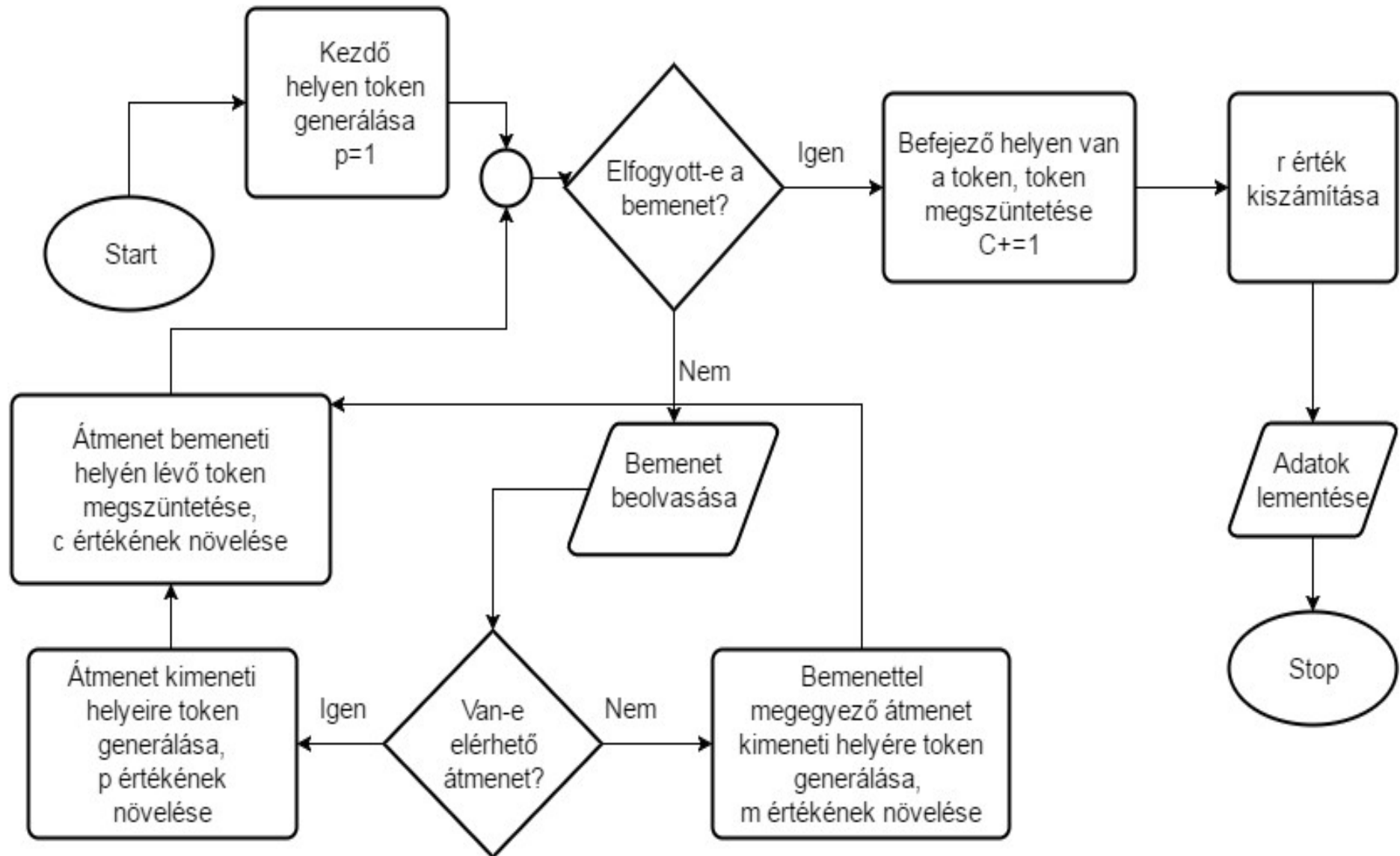
$$F = 1 - 26/49 = 0,469$$

- Javaslatok pl.:**
1. Folyamat újratervezése a valós tapasztalatok alapján.
 2. e és f művelet kihagyása, amely kötelező volt. Vizsgálat!
 3. Előbb végezték el az f munkát mint az e-t. Vizsgálat!
 4. A b művelet kétszer került végrehajtásra. Miért? Stb.

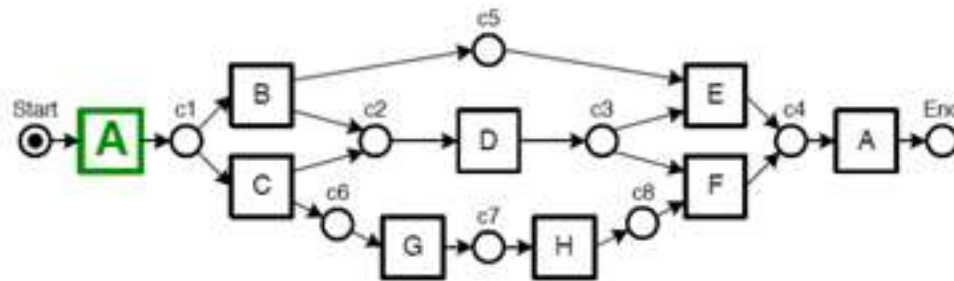
Konformancia token alapú egyezőség-vizsgálata

- Ehhez szükség van a **follyamat modell**jére, pl. Petri háló formájában
- A **napló fájl**ok rengeteg információt tartalmaznak, melyek sok esetben nem feltétlen szükségesek a vizsgálat elvégzéséhez. Ezért az adatoknak előzetes átalakításon kell átesniük, mely folyamán csak **azon esemény tulajdonságokat tartjuk meg, melyek számunkra szükségesek.**
- Bevezetünk négy különböző mérőszámot:
 1. ***p*** létrehozott token (produced token),
 2. ***c*** elhasznált token (consumed token),
 3. ***m*** hiányzó token (missing token),
 4. ***r*** megmaradt token (remaining token).

Konformancia token alapú egyezőség-vizsgálatának folyamatábrája

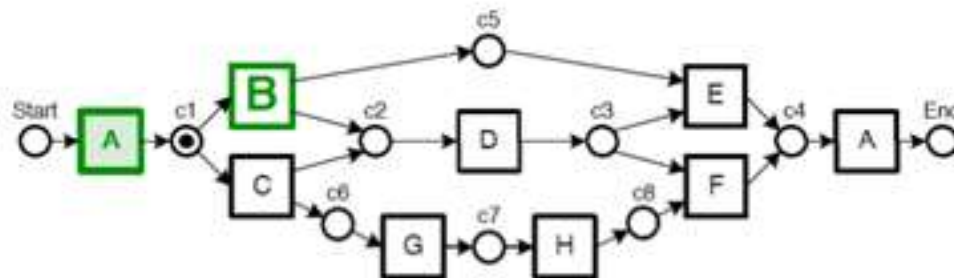


Fitnessz érték figyelése az események végig követésével egy trace-ben



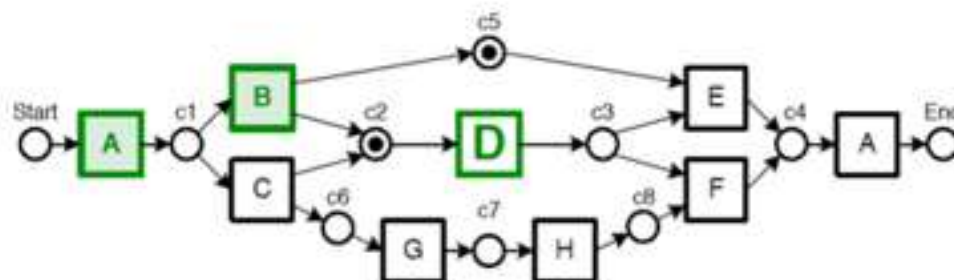
$m = 0$
 $r = 0$
 $c = 0$
 $p = 1$

No. of Instances	Log Traces
1207	→ A BDEA
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA



$m = 0$
 $r = 0$
 $c = 1$
 $p = 2$

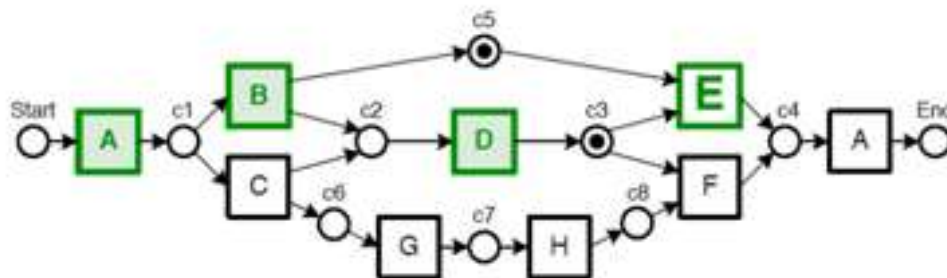
No. of Instances	Log Traces
1207	→ A B DEA
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA



$m = 0$
 $r = 0$
 $c = 2$
 $p = 4$

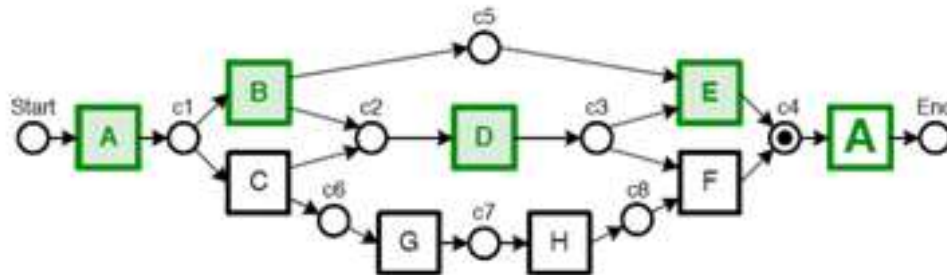
No. of Instances	Log Traces
1207	→ A B D EA
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA

Ha nincs probléma ($m=0, r=0$):



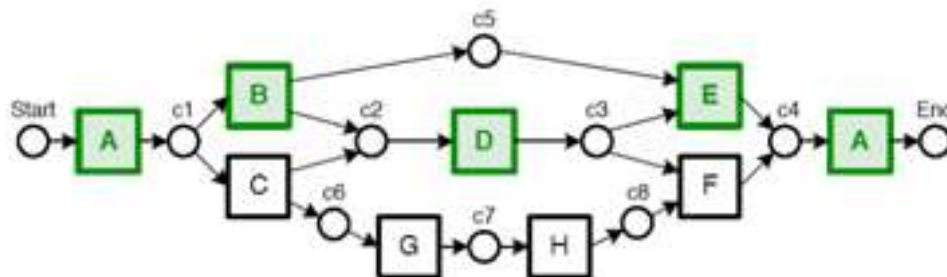
$m = 0$
 $r = 0$
 $c = 3$
 $p = 5$

No. of Instances	Log Traces
1207	→ ABDEA
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA



$m = 0$
 $r = 0$
 $c = 5$
 $p = 6$

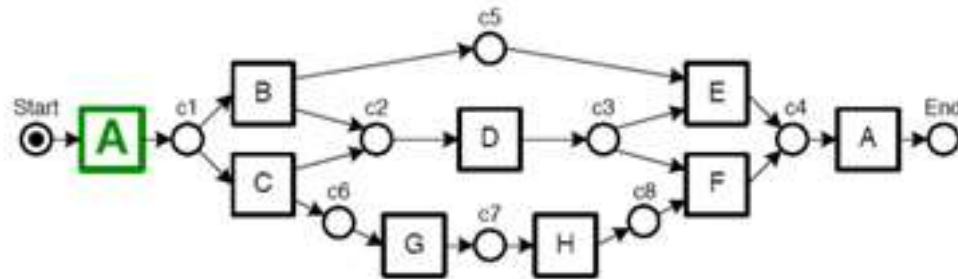
No. of Instances	Log Traces
1207	→ ABDEA
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA



$m = 0$
 $r = 0$
 $c = 7$
 $p = 7$

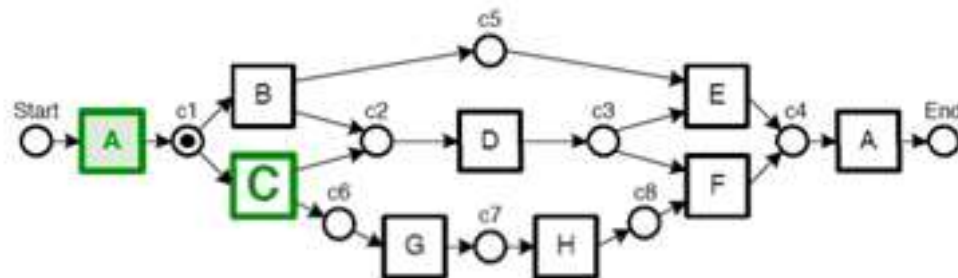
No. of Instances	Log Traces
1207	→ ABDEA
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA

Egy másik (megvalósíthatatlan) trace



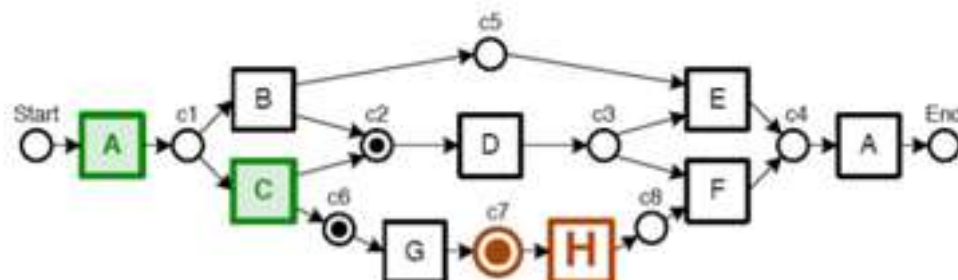
m = 0
r = 0
c = 0
p = 1

No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ ACHDFA
28	ACDHFA



m = 0
r = 0
c = 1
p = 2

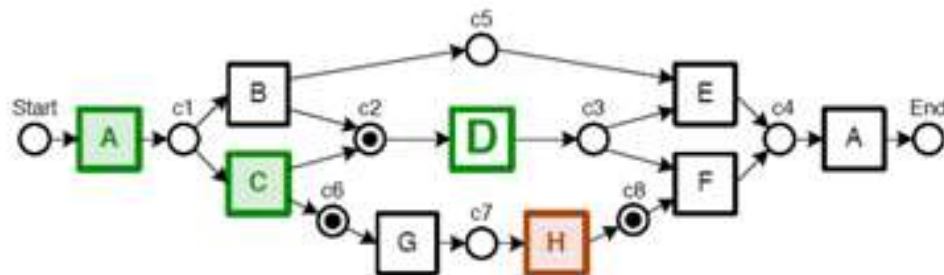
No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ AC ^C HDFA
28	ACDHFA



m = 1
r = 0
c = 2
p = 4

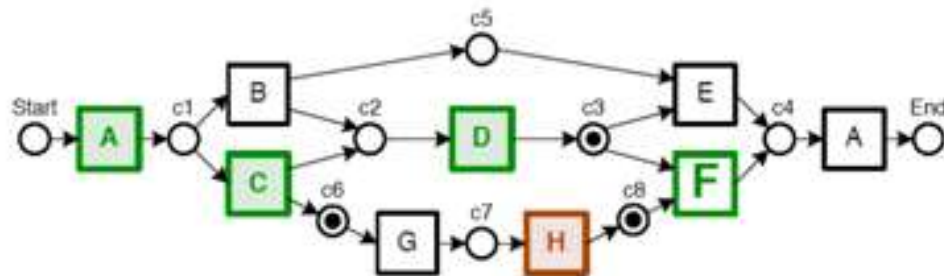
No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ AC ^C ^H DFA
28	ACDHFA

Folytatás



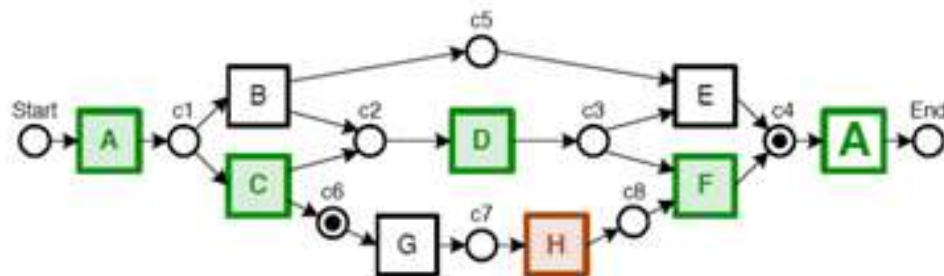
$m = 1$
 $r = 0$
 $c = 3$
 $p = 5$

No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ ACHDFA
28	ACDHFA



$m = 1$
 $r = 0$
 $c = 4$
 $p = 6$

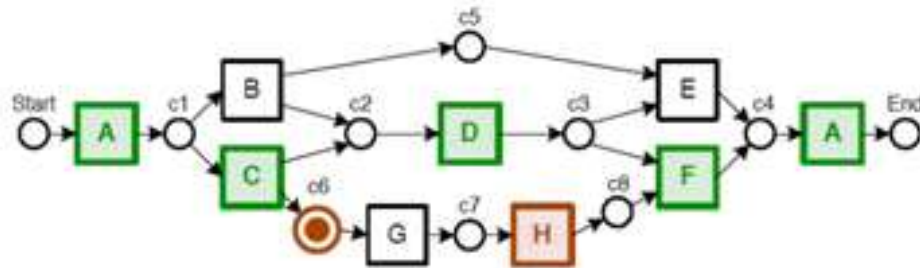
No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ ACHDFA
28	ACDHFA



$m = 1$
 $r = 0$
 $c = 6$
 $p = 7$

No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ ACHDFA
28	ACDHFA

Fitnessz érték számolása a teljes logra



$m = 1$
 $r = 1$
 $c = 8$
 $p = 8$

No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ ACHDFA
28	ACDHFA

különböző
 tracek száma

hány azonos
 trace van

hiányzó token

megmaradt token

$$f = \frac{1}{2} \left(1 - \frac{\sum_{i=1}^k n_i m_i}{\sum_{i=1}^k n_i c_i} \right) + \frac{1}{2} \left(1 - \frac{\sum_{i=1}^k n_i r_i}{\sum_{i=1}^k n_i p_i} \right)$$

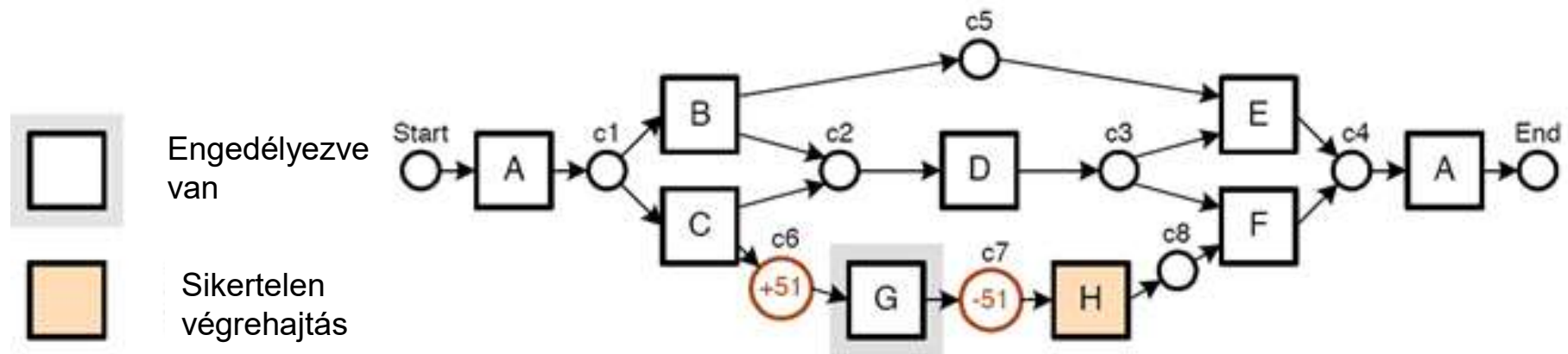
elhasznált token

létrehozott token

A fenti log esetében:

$$f(M1, L2) = \frac{1}{2} \left(1 - \frac{51}{10666} \right) + \frac{1}{2} \left(1 - \frac{51}{10666} \right) \approx 0.995$$

Hogyan használhatjuk a módszert diagnosztikához



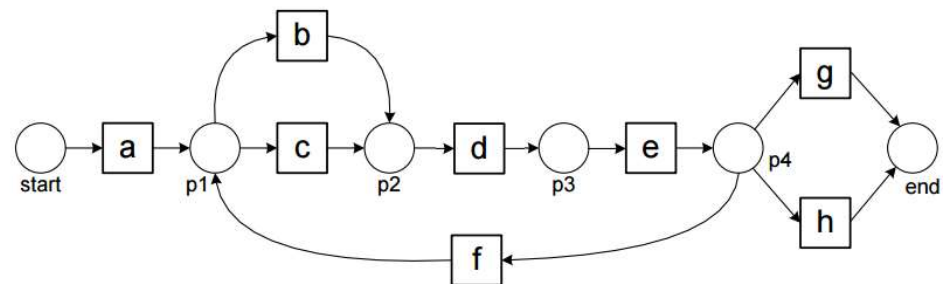
A H művelet valamilyen oknál fogva nem került végrehajtásra.
Ki kell deríteni, hogy mi volt az oka!

Másik példa

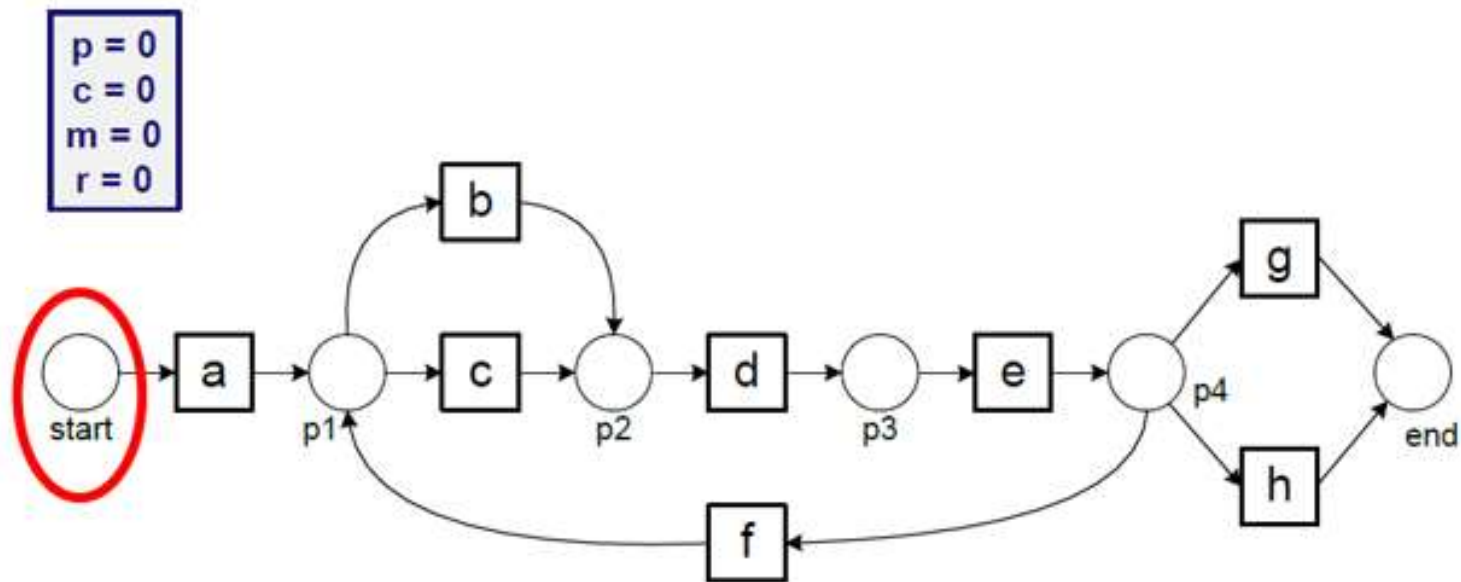
- A napló fájl átalakítása után csak azon esemény tulajdonságokat tartjuk meg melyek számunkra szükségesek.
- Ezután előfordulási gyakoriság szerint összerendezzük őket.
- A vizsgálat elvégzéséhez szükséges adatok maradnak meg: az azonosító, a gyakoriság és az útvonal

455	σ_1	$\langle a, c, d, e, h \rangle$
191	σ_2	$\langle a, b, d, e, g \rangle$
177	σ_3	$\langle a, d, c, e, h \rangle$
144	σ_4	$\langle a, b, d, e, h \rangle$
111	σ_5	$\langle a, c, d, e, g \rangle$
82	σ_6	$\langle a, d, c, e, g \rangle$
56	σ_7	$\langle a, d, b, e, h \rangle$
47	σ_8	$\langle a, c, d, e, f, d, b, e, h \rangle$
38	σ_9	$\langle a, d, b, e, g \rangle$
33	σ_{10}	$\langle a, c, d, e, f, b, d, e, h \rangle$
14	σ_{11}	$\langle a, c, d, e, f, b, d, e, g \rangle$
11	σ_{12}	$\langle a, c, d, e, f, d, b, e, g \rangle$
9	σ_{13}	$\langle a, d, c, e, f, c, d, e, h \rangle$

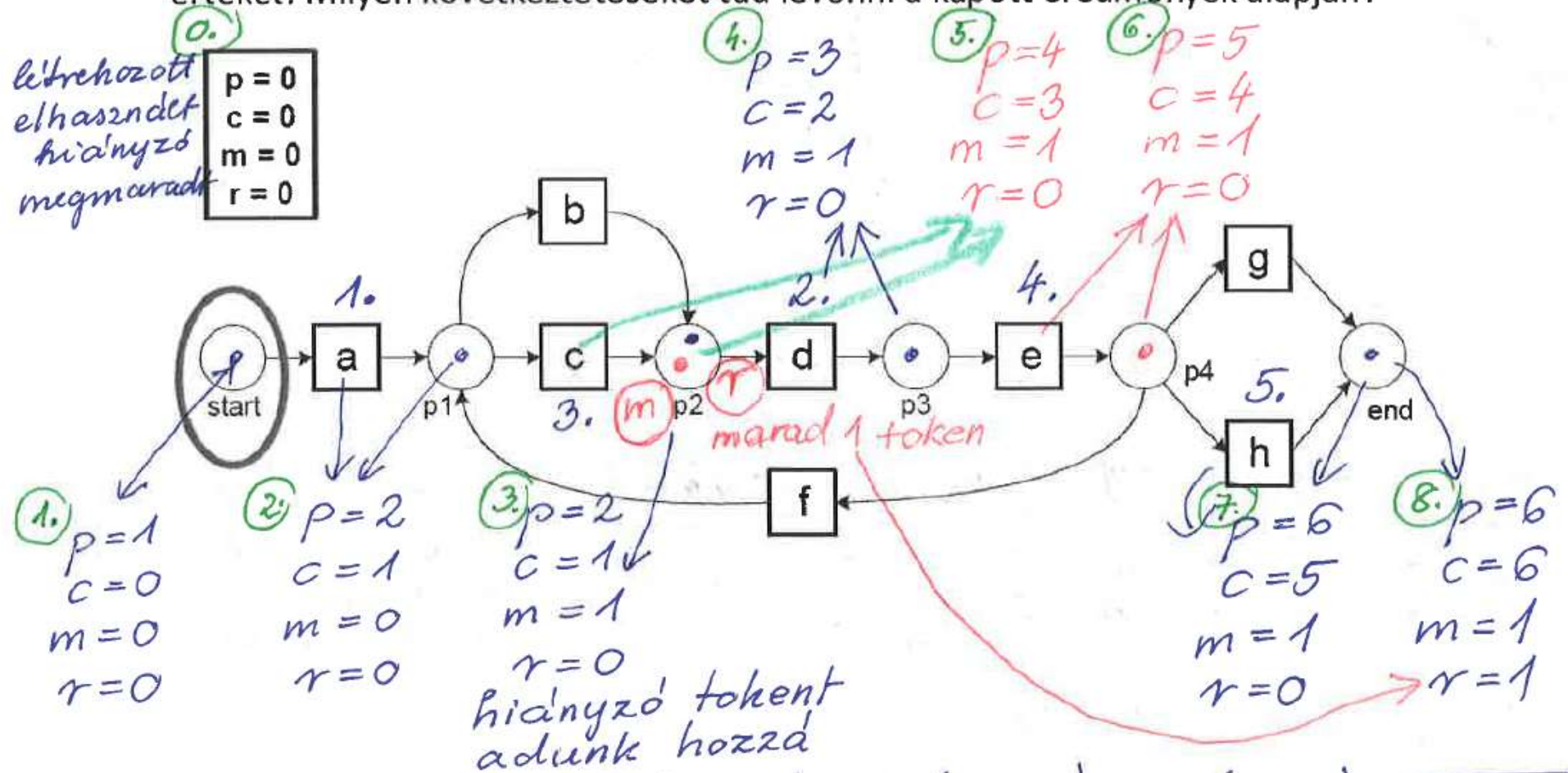
Elemezzük a megjelölt eseménysorozatot a rendelkezésre álló modell alapján!



Számolja ki a $t = \langle a, d, c, e, h \rangle$ trace alapján a megfelelőséget meghatározó fitness értéket! Milyen következtetéseket tud levonni a kapott eredmények alapján?

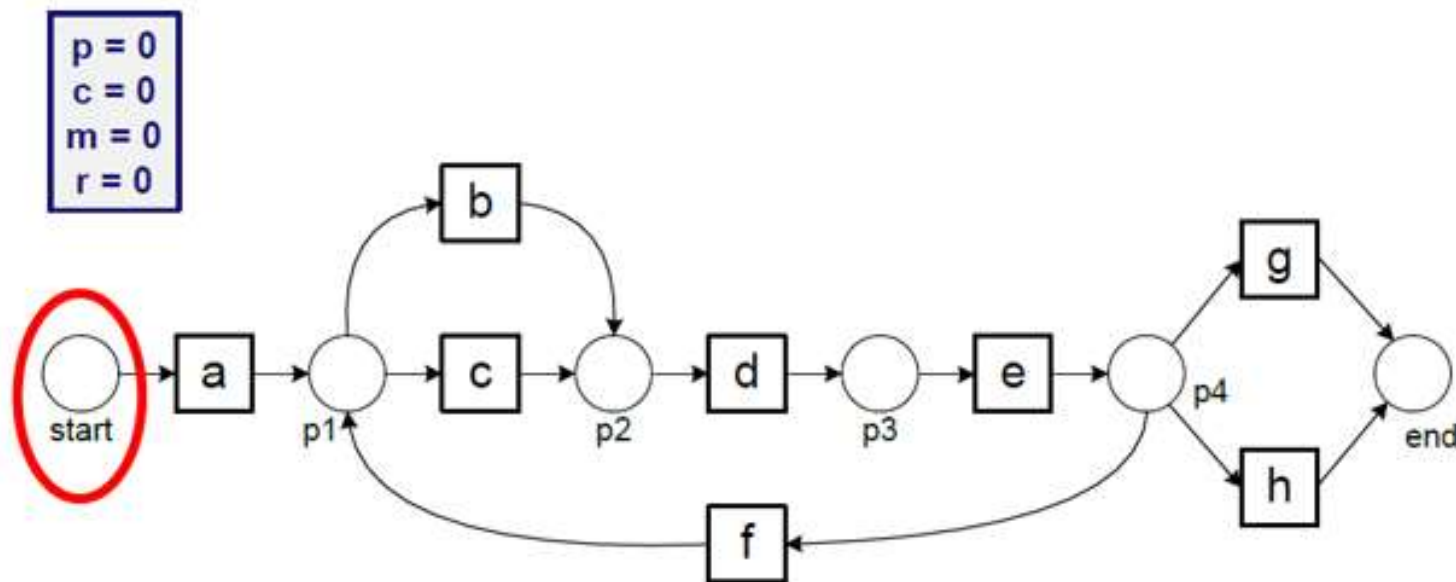


Számolja ki a $t=\langle a,d,c,e,h \rangle$ trace alapján a megfelelőséget meghatározó fitness értéket! Milyen következtetéseket tud levonni a kapott eredmények alapján?



$$\phi = \frac{1}{2} \left(1 - \frac{m}{c} \right) + \frac{1}{2} \left(1 - \frac{r}{p} \right) = \frac{1}{2} \left(1 - \frac{1}{6} \right) + \frac{1}{2} \left(1 - \frac{1}{6} \right) = \boxed{0,833}$$

Számolja ki a $t = \langle a, d, c, e, f, c, d, e, h \rangle$ trace alapján a megfelelőséget meghatározó fitness értéket! Milyen következtetéseket tud levonni a kapott eredmények alapján?



$t = \langle a, d, c, e, f, c, d, e, h \rangle$
 1. 2. 3. 4. 5. 6. 7. 8. 9.

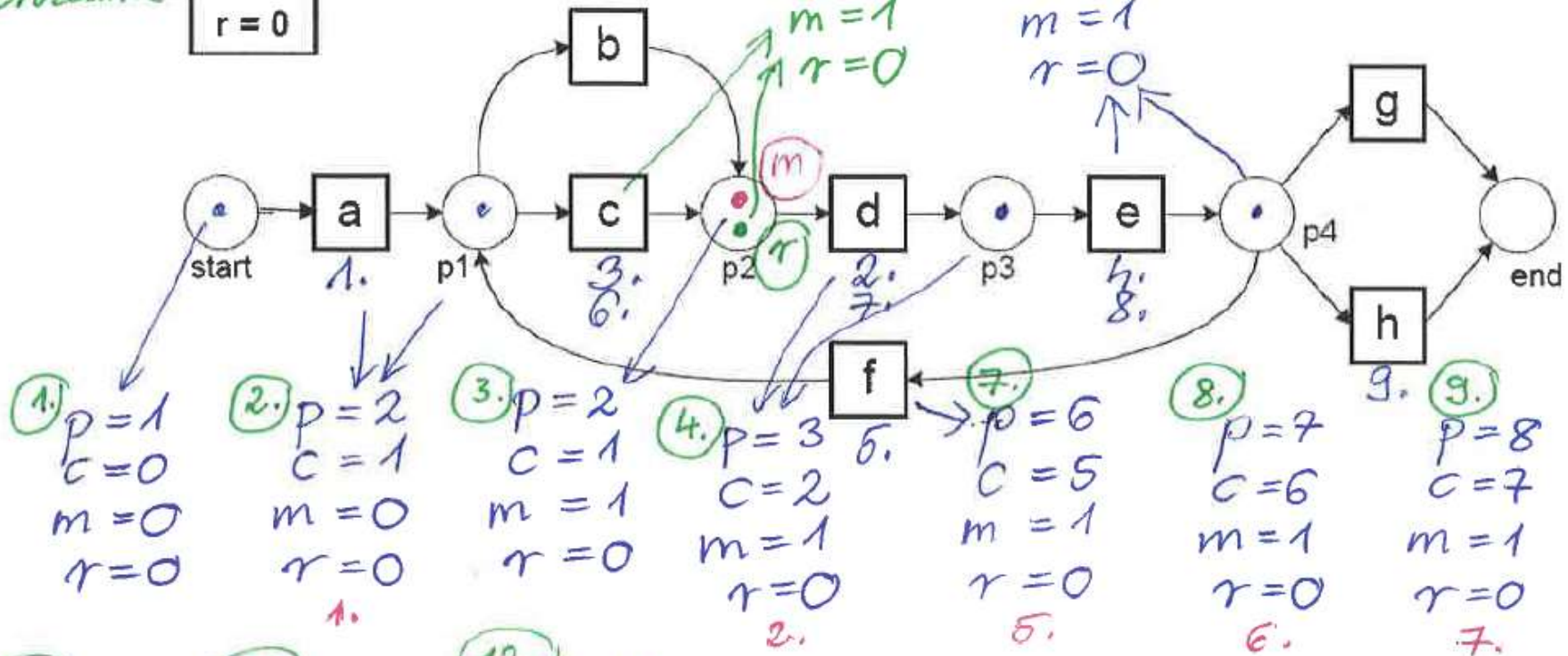
lepe's
sorozatma

$p = 0$
 $c = 0$
 $m = 0$
 $r = 0$

5. 3.
 $p = 4$
 $c = 3$
 $m = 1$
 $r = 0$

6. 4.
 $p = 5$
 $c = 4$
 $m = 1$
 $r = 0$

művelet sorozatma



10. $p = 9$
 $c = 8$
 $m = 1$
 $r = 0$
 8.

11. $p = 10$
 $c = 9$
 $m = 1$
 $r = 0$
 9.

12. $p = 10$
 $c = 10$
 $m = 1$
 $r = 1$

$$\phi = \frac{1}{2} \left(1 - \frac{m}{c} \right) + \frac{1}{2} \left(1 - \frac{r}{p} \right) = \frac{1}{2} \left(1 - \frac{1}{10} \right) + \frac{1}{2} \left(1 - \frac{1}{10} \right) = \underline{\underline{0.9}}$$

Vegyük figyelembe egyszerre a 2 logot

különböző tracek száma hány azonos trace van hiányzó token megmaradt token

$$f = \frac{1}{2} \left(1 - \frac{\sum_{i=1}^k n_i m_i}{\sum_{i=1}^k n_i c_i} \right) + \frac{1}{2} \left(1 - \frac{\sum_{i=1}^k n_i r_i}{\sum_{i=1}^k n_i p_i} \right)$$

elhasznált token létrehozott token

$\langle a, d, c, e, h \rangle^{177}$

$\langle a, d, c, e, f, c, d, e, h \rangle^9$

$$f = \frac{1}{2} \left(1 - \frac{177*1+9*1}{177*6+9*10} \right) + \frac{1}{2} \left(1 - \frac{177*1+9*1}{177*6+9*10} \right)$$

$$= 1 - \frac{177 + 9}{177 * 6 + 9 * 10} = \mathbf{0,8385}$$

Az elemzést elvégző alkalmazás felhasználói felülete, eredmények

File

File Name: D:\My Backups_20161219
\\a_tantargyak 2010-to\Process mining
\\konformancia szakdoga\Programkod\Teszt
\\Eseménynapló\teszt_események.txt
455 q1 <a,c,d,e,h>
191 q2 <a,b,d,e,g>
177 q3 <a,d,c,e,h>
144 q4 <a,b,d,e,h>

Modell file Name: D:\My Backups_20161219\\a_tantargyak 2010-to\Process mining
\\konformancia szakdoga\Programkod\Teszt\Modell\teszt_modell.pnml

Place id: 1 Place name: start
Place id: 2 Place name: p1
Place id: 3 Place name: p2
Place id: 4 Place name: p3
Place id: 5 Place name: p4
Place id: 6 Place name: end
Transition id: 7 Transition name: a
Transition id: 8 Transition name: b
Transition id: 9 Transition name: c
Transition id: 10 Transition name: d
Transition id: 11 Transition name: e
Transition id: 12 Transition name: g
Transition id: 13 Transition name: h
Transition id: 14 Transition name: f

Analysis

.....
Frekvency: 191
Event_name: q2
Produce: 6
Consumed: 6
Missing: 0
Remain: 0
Fitness: 1,000
.....
Frekvency: 177
Event_name: q3
Produce: 6
Consumed: 6
Missing: 1
Remain: 1
Missing_place_id: 3 Name: p2
Remain_place_id: 3 Name: p2
Fitness: 0,833
.....

Eredmények

- o1, o2, o4: Tökéletes illeszkedés.
 - Egyezőség mértéke: 1
 - Létrehozott token-ek száma: 6
 - Megszüntetett token-ek száma: 6
 - Hiányzó token-ek száma: 0
 - Megmaradt token-ek száma: 0
 - Egyéb hiba: Nincs
- o3: Hiányos illeszkedés.
 - Egyezőség mértéke: 0.833
 - Létrehozott token-ek száma: 6
 - Megszüntetett token-ek száma: 6
 - Hiányzó token-ek száma: 1
 - Megmaradt token-ek száma: 1
 - Egyéb hiba:
 - Hiányzó token hiba: Hely neve p2
 - Megmaradt token hiba: Hely neve p2
- A eseménynaplóban lévő összes eseményre kivetített egyezőség mértéke (teljes fitness): 0.97.