Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar
Mesterséges Intelligencia és Rendszertervezés Tanszék

Logikai ágens, szabályalapú rendszerek

Előadó:
Dr. Hullám Gábor

Előadás anyaga:
Dr. Dobrowiecki Tadeusz
Dr. Hullám Gábor

Problémamegoldás

Keresés

Kényszer kielégítés

Keresés ellenséges térben

Heurisztikák

Tudásreprezentáció következtetés

Ontológia

Logika

Való-színűségi hálók

Bizony-talanság kezelése

Döntések

Hasznosság

Döntési hálók

Megerősítéses tanulás

Szekvenciális döntések

Markov véletlen folyamat

Regressziós modellek

Döntési fák

Ajánló rendszerek

Felügyelt tanulás

Neurális hálók

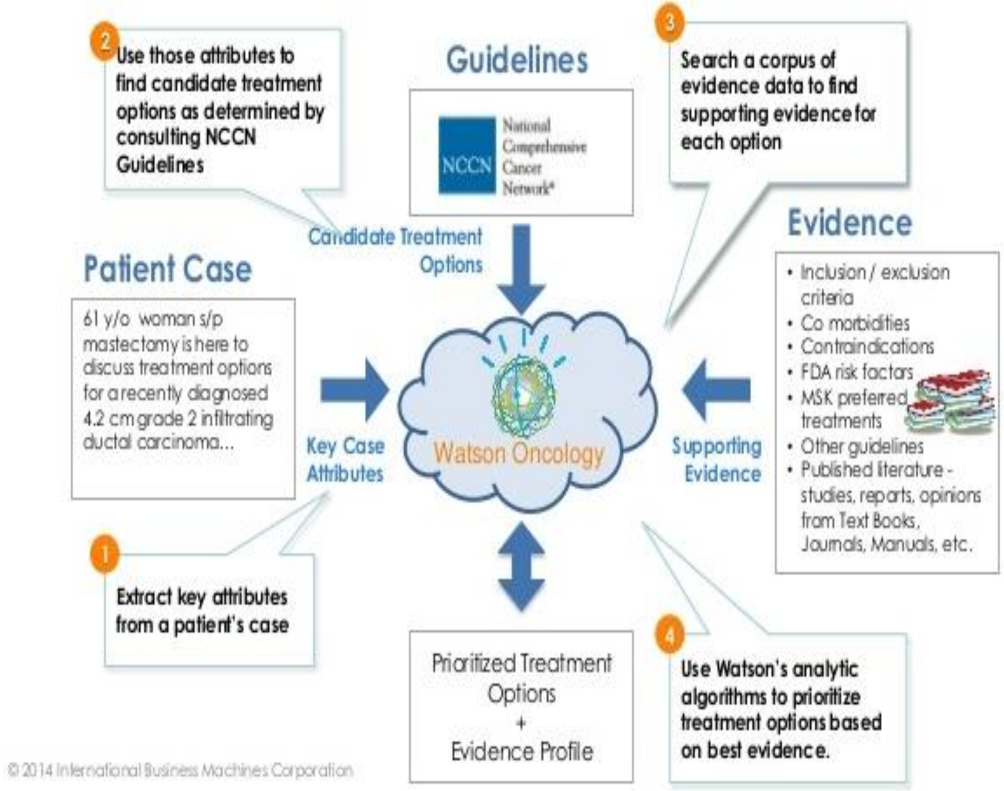
Együttes modellek

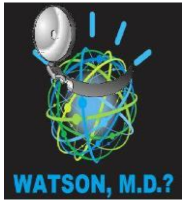
MŰEGYETEM 1782

IBM Watson (2011) Jeopardy kvízjáték

- **IBM** Grand Challenge
  - 1997: **Deep Blue** legyőzi a G. Kaszparov sakkvilágbajnokot.
  - 1999-2006<: **Blue Gene**, fehérje struktúra predikció
  - 2011: **Watson**

Orvosi döntéstámogató és szakértői rendszerek





Szakértői rendszerek az MI történetében

- ~1930 Univerzális számítási modell: Turing-gép (1936), Univerzális Turing-gép, Church-Turing hipotézis, Zuse, Neumann,..: „vezérlő program is adat”:
- 1943 McCulloch & Pitts: Bináris kapcsolati agymodell
- 1950 Turing: "Computing Machinery and Intelligence"
- 1956 Dartmouth találkozó: "Artificial Intelligence” megnevezés elfogadása
- 1950s Korai MI programok: sakk, tételbizonyítás
- **Számítás(keresés) alapú MI**
- - A Fizikai Szimbólumrendszer hipotézise: A.Newel&H.A.Simon (1976)
- 1966—73 Számítási komplexitási korlátok a keresésben
Elméleti korlátok a neurális hálózatokban
- **1969—79 Tudásalapú szakértői rendszerek**
(szabályalapú szakértői rendszerek)
- **Tudásalapú MI**

Szakértői rendszerek az MI történetében

- **1969—79 Tudásalapú szakértői rendszerek (szabály alapú szakértői rendszerek)**

Tudásalapú MI

- 1986-- Neurális hálózatok újbóli megjelenése
- 1988-- Valószínűségi szakértői rendszerek
- 1995-- Gépi tanulás gyors fejlődése

Adatvezérelt MI (‚data driven’)

- 2001-- ‚Big Data’

Autonóm tanulás alapú MI

- 2011 -- Mély tanulás (deep learning)

Szabály alapú rendszerek = Tudásalapú rendszerek = Szakértő rendszerek

Mycin, Stanford, ca. 1972, kb. 500 szabály, szakorvos szintje

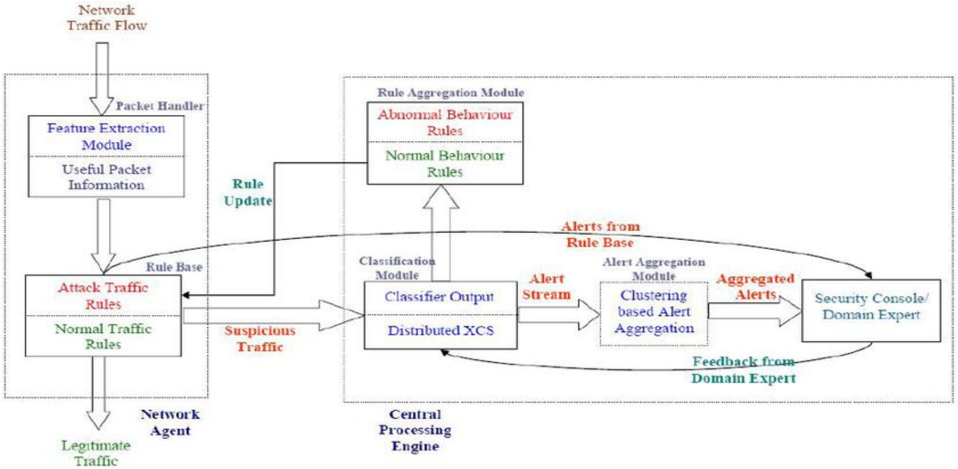
IF The site of the culture blood, and
The gram stain of the organism is gramneg, and
The morphology of the organism is rod, and
The portal of entry of the organism is urine, and
The patient has not had a genito-urinary manipulative procedure, and
Cystitis is not a problem for which the patient has been treated
THEN There is suggestive evidence (.6) - the identity of the organism is e.coli

IF The identity of the organism is bacteroides
THEN Recommend therapy chosen from the following drugs:
1. clindamycin (0.99) 2. chloramphenicol (0.99)
3. erthromycin (0.57) 4. tetracycline (0.28)
5. carbenicillin (0.27)

A logika szerepe a komplex MI alkalmazásokban

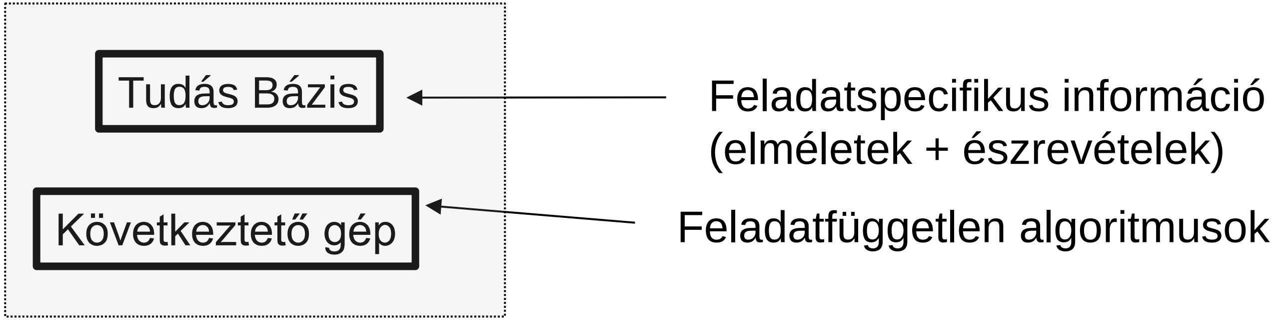
- Logikai szabályalapú következtetés és beavatkozás

- Egy vagy több feltétel logikai állításként történő megfogalmazása
- Reakciók, beavatkozások megadása logikai szabályok formájában



https://www.researchgate.net/figure/An-Adaptive-Rule-Based-Intrusion-Detection-Architecture_fig1_238621576

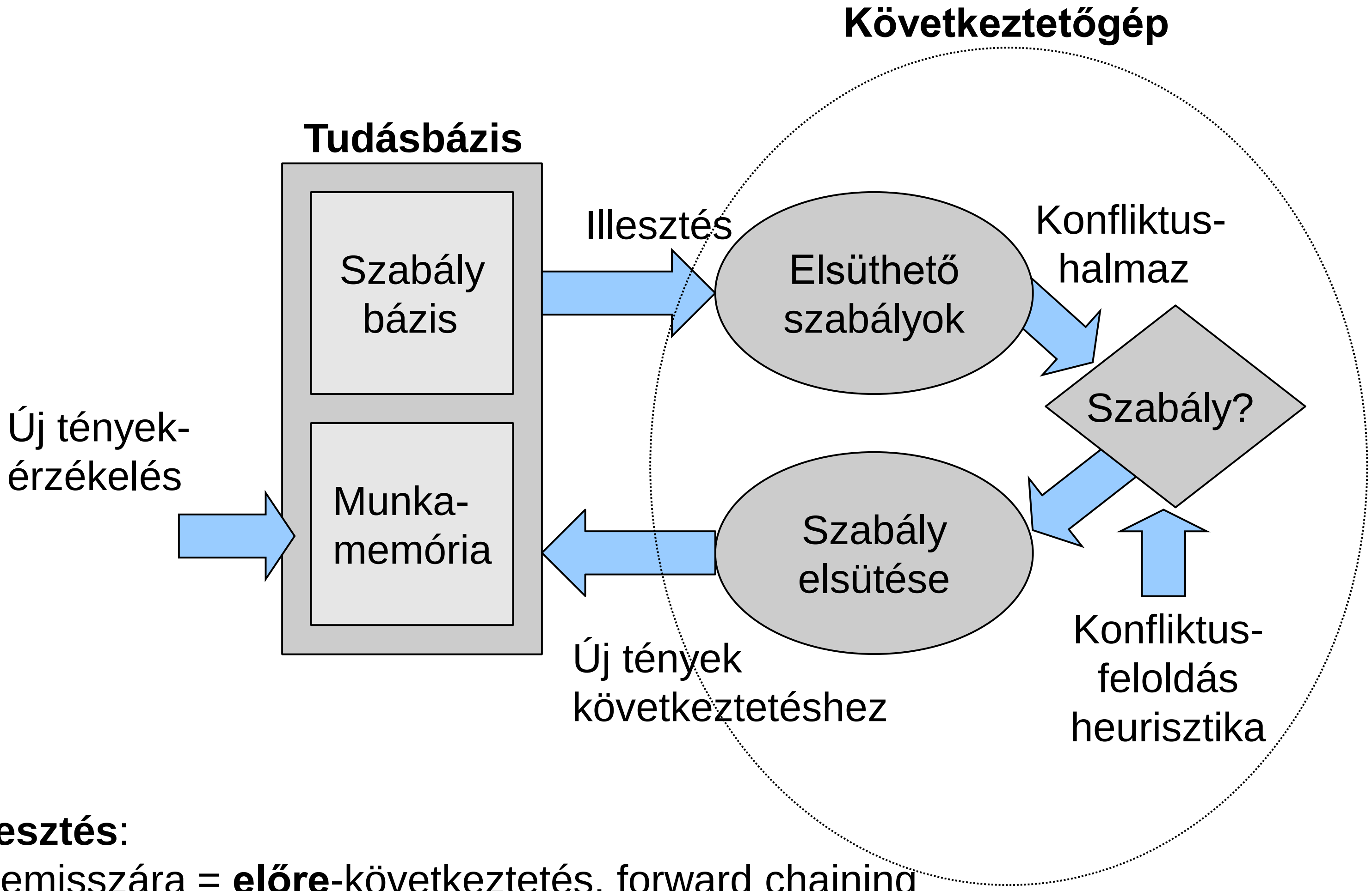
Logika



Mire használhatja az Ágens logikai képességeit?

Hogyan fejezze ki ehhez a tudását?

Milyen algoritmusok lesznek jók, ügyesek, …?



Illesztés:
Premisszára = **előre**-következtetés, forward chaining
Következményre = **hátra**-következtetés, backward chaining

Szabályalapú szakértői rendszerek LLM-ek előtt

- Jogi szakértői rendszerek
- Adó- és pénzügyi tanácsadó rendszerek
- Ügyfélszolgálat és technikai támogató rendszerek

- Természetes nyelvű interfészek
  - Digitális asszisztensek
  - Csetbotok

Szabályalapú szakértői rendszerek alkalmazásai

https://medium.com/@teamrework/the-promise-of-deep-learning-for-rule-based-tasks-e806fa688478

Logikai következtetések fajtái

Tudás átalakítás (manipulálás) modellje

Dedukció (már az ógörögök)

formálisan érvényes következtetés
olyan tények származtatása, amelyek a premisszákból mindenképpen következnek (un. 'igazságtartó' eljárás)

Ha kutya nagy, akkor sokat eszik (feltétel)
kutya nagy (tapasztalat, evidencia)
sokat eszik (következtetés)

Tömörítés, általánosítás: példákból tanulás modellje!

Indukció (középkori arabok, skolasztikusok)

Kutya1 nagy, Kutya2 nagy, …, Kutya1000 nagy (tapasztalat)
Minden kutya nagy (induktív általánosítás)

miért fontos? sok tény helyett egyetlen egy (új) tény
probléma mérete csökken,
exponenciális jelleg kevésbé zavaró,
de formálisan nem igaz! Miért?

Abdukció - 'belátás' folyamata (Charles Pierce, 1839–1914)

ha kutya nagy, akkor sokat eszik (feltétel)
_____________________ sokat eszik (tapasztalat)
kutya nagy (hipotézis)

Diagnózis, magyarázatadás modellje!

ha kutya nagy, akkor sokat eszik
ha kutya éhes, akkor sokat eszik
______________________ sokat eszik (tapasztalat)
éhes, vagy nagy? (melyik hipotézis esélyesebb)

formálisan miért nem igaz? de miért nagyon fontos?

Természetes rendszermodell
(a feladat megfogalmazása)

ha ‘egy rendszer X állapotban van',
akkor 'Y a rendszer megfigyelhető viselkedése'

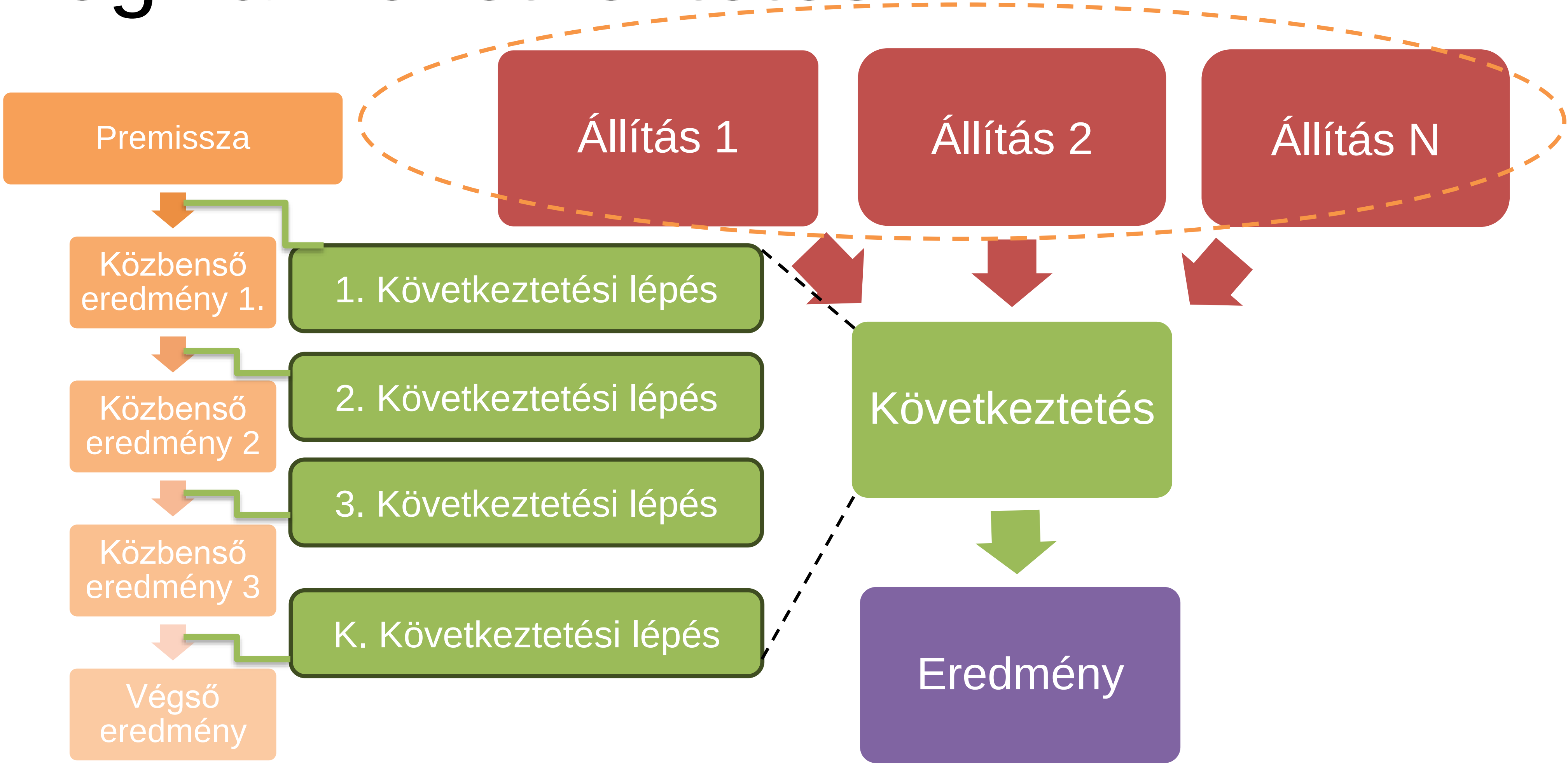
Természetes kérdés (feladat megoldása)
milyen állapotban van a rendszer, ha konkrét információt (Y) kapok a viselkedéséről?

Tudjuk, hogy a beteg lázas.
Dedukció, vagy abdukció?

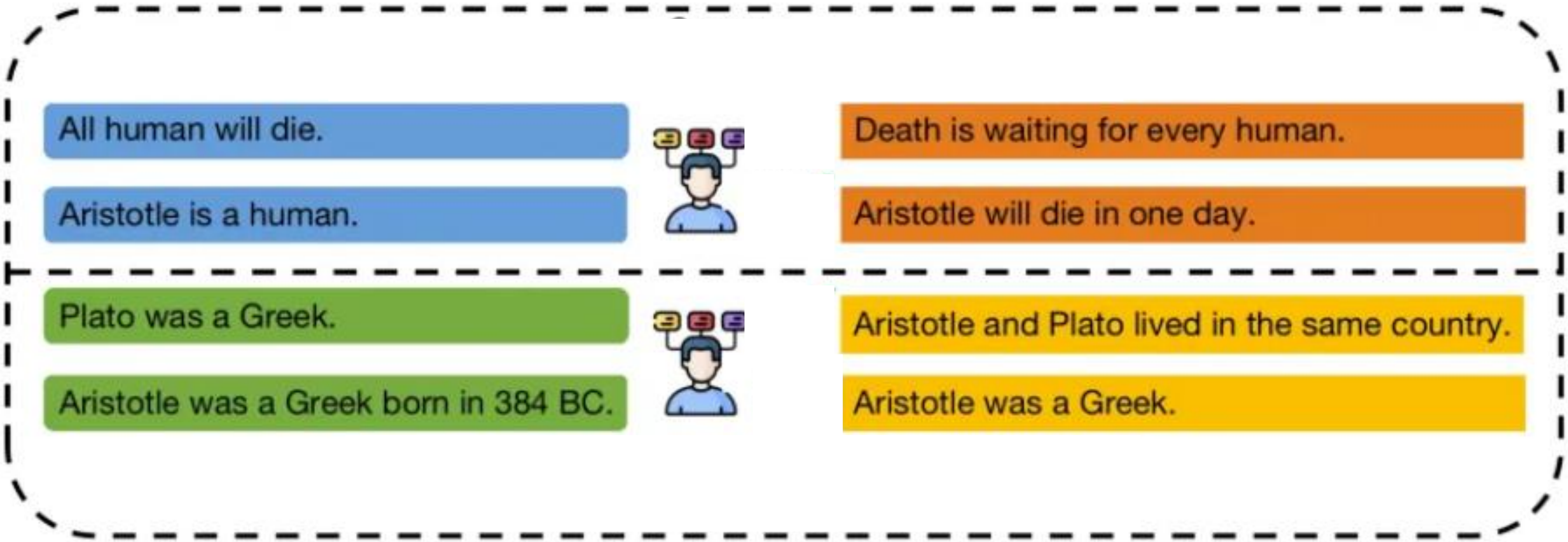
ha INFLUENZÁS, akkor LÁZAS
______________________ LÁZAS
INFLUENZÁS?

ha LÁZAS, akkor INFLUENZÁS
LÁZAS ______________________
INFLUENZÁS?

Logikai következtetés

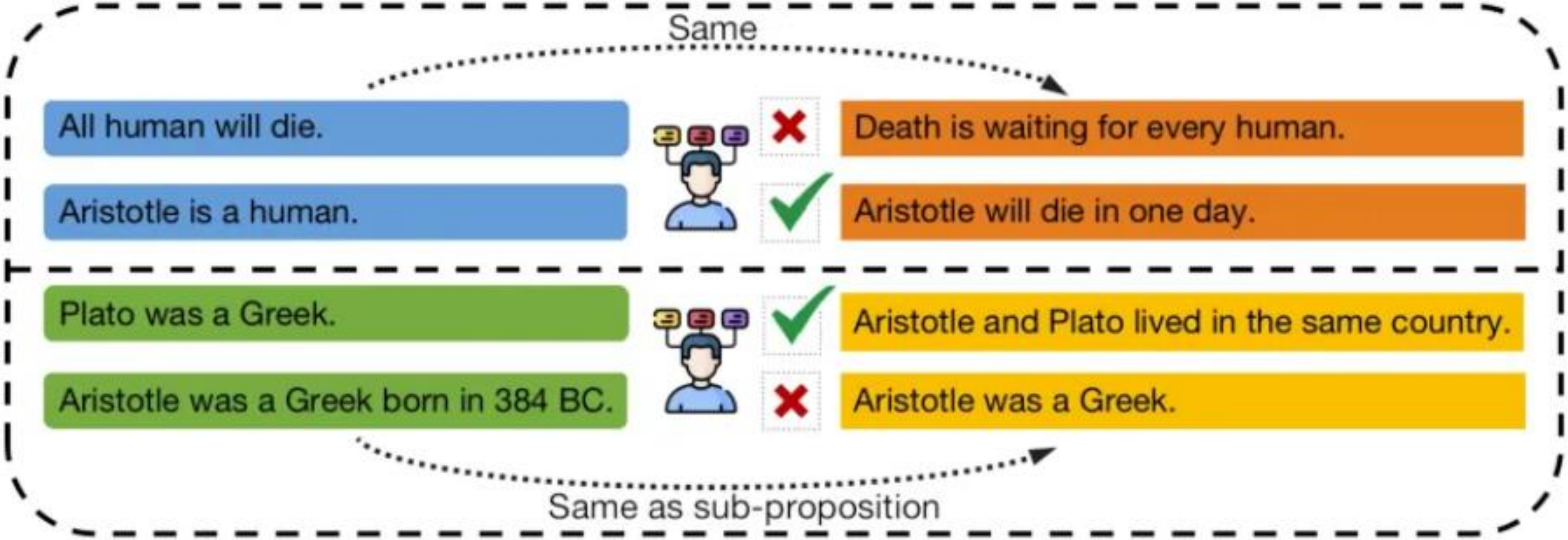


Mi tekinthető következtetésnek?



https://isamu-website.medium.com/understanding-the-current-state-of-reasoning-with-llms-dbd9fa3fc1a0

Mi tekinthető következtetésnek?



https://isamu-website.medium.com/understanding-the-current-state-of-reasoning-with-llms-dbd9fa3fc1a0

Logika – reprezentáció és manipuláció eszköze

Szintaktika - legálisan létrehozható szimbolikus mondatok
Szemantika - a világ tényei, amikre a szimbolikus mondatok vonatkoznak

Vonzatreláció és a következtetés

Tények neveznek valós dolgokat, amelyek egymással kapcsolatban vannak. Ha az egyik igaz a világban, és a másik szükségszerűen is az:

vonzat: logikai konzekvencia dolgok között.

Vonzatreláció tudásbázis TB és egy α mondat között:

α vonzata TB -nek: **TB |= α** , ha α minden olyan világban igaz, ahol TB is.

Következtetés: a vonzat „kiszámítása” a mondatok formális manipulálásával.
Bizonyítás = a következtetési algoritmus lépéssorozata

α bizonyítható TB -ből: **TB |-- α**

Logika – reprezentáció és manipuláció eszköze

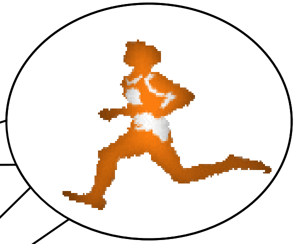
Szintaktika - legálisan létrehozható szimbolikus mondatok
Szemantika - a világ tényei, amikre a szimbolikus mondatok vonatkoznak

Mi egy mondat jelentése?

- A mondat írójának egy interpretációt kell adnia hozzá, kijelentve, hogy milyen tény tartozik hozzá.
- Egy mondat önmagában nem jelent semmit és lehet igaz vagy hamis.
- Egy mondat igaz egy bizonyos interpretációban, ha a dolgok állása ilyen.
- **Az igazság függ mind a mondat interpretációjától, mind a világ aktuális állapotától.**

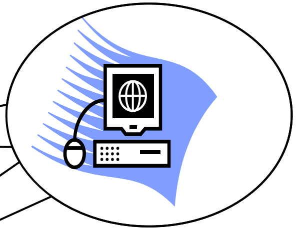
Szemantika és interpretáció

mi lehet pl. a **Fut_Ágens1** mondat jelentése?



Ha „fut” fizikai mozgás, „Ágens1” egy személy, és a világ most: akkor ez a mondat (most) **igaz**.

Ha „fut” nem terminált működés, „Ágens1” egy program, és a világ most ugyanilyen, akkor ez a mondat (most) **hamis**.



Ha „fut” fizikai mozgás, „Ágens1” egy személy és a világ most: akkor ez a mondat (most) **hamis**.

Ha „fut” nem terminált működés, „Ágens1” egy program, és a világ most ugyanilyen, akkor ez a mondat (most) **igaz**.

Modellek

Bármely (formális is) világ, ahol egy mondat igaz egy bizonyos interpretációban. Egy mondatnak számos modellje lehet.

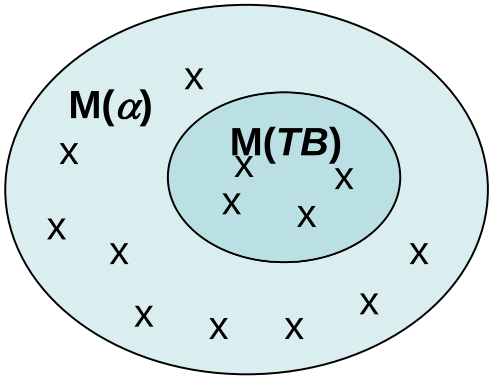
Minél többet állítunk (pl. minél több információt adunk hozzá a tudásbázishoz), annál kevesebb modellünk lesz.

Modellek nagyon fontosak (mert):

egy α mondat vonzata a TB tudásbázisnak, ha a TB modelljei mind modelljei az α-nak is.

Ha ez így van, akkor ha TB igaz, akkor α is igaz

TB $\models \alpha$ akkor és csak akkor, ha M(TB) $\subseteq$ M(α)



Ha adott egy TB, a **következtetés**:
(1) létrehozhat új α mondatot, amely vonzata a TB-nak
(2) ha adott α mondat, eldöntheti hogy α vonzata-e a TB-nak.

Igazságtartó következtetés (extenzionális), ha csak igazságfüggvényekkel dolgozik (igazságfüggvény logikai értéke csakis a részeinek és a logikai műveletek definíciójától függ, nem függ az interpretációtól).

Formális bizonyítás – igazságtartó következtetési eljárás lépései.

Következtetési eljárás teljes: ha minden vonzatmondathoz talál egy bizonyítást, avagy ami igaz, az bebizonyítható. **A |= B** -ból **A |- B**

Következtetési eljárás helyes: ha minden bizonyított mondat vonzatrelációban áll a felhasznált tényekkel, avagy ami bebizonyított, az igaz is. **A |- B** -ból **A |= B**

Érvényes (analitikus mondat, tautológia): ha minden világban minden lehetséges interpretációja igaz, függetlenül attól, hogy mit akar jelenteni és mi a világ állapota.

Pl. „Van bűz az [1,1]-ben, vagy nincs bűz az [1,1]-ben”

Kielégíthető: ha létezik olyan interpretáció, hogy valamely világban igaz.
Ami nem kielégíthető, az **kielégíthetetlen.**

A	B	$A \vee (B \vee \neg B)$	$A \wedge (B \wedge \neg B)$	$A \wedge B$
0	0	1	0	0
0	1	1	0	0
1	0	1	0	0
1	1	1	0	1

Melyik állítás érvényes, kielégíthető, ill. kielégíthetetlen?

Ítéletkalkulus (Arisztotelész, Euklidész, i.e. 300, Leibnitz, Boole, XVII sz.)

Szintaktika:

igaz és **hamis** logikai konstansok, P1, P2, ... ítélet szimbólumok

ítéletmondatok, **$\neg$ P**, **P1 $\wedge$ P2, P1 $\vee$ P2, P1 $\rightarrow$ P2, P1 $\Leftrightarrow$ P2** is ítéletmondatok.

precedencia: $\neg$, $\wedge$, $\vee$, $\rightarrow$ és $\Leftrightarrow$.

Szemantika

Minden modell igaz/hamis értéket rendel minden ítéletszimbólumhoz,
pl. P1 = igaz, P2 = hamis, P3 = igaz, ...

Tetszőleges mondat szemantikája - rekurzív elemzés, pl.:

$\neg$P1 $\wedge$ (P2 $\vee$ P3) = hamis $\wedge$ (hamis $\vee$ igaz) = hamis $\wedge$ igaz = hamis

Összekötő jelek igazságtáblái

P1	P2	$\neg$ P	P1 $\wedge$ P2	P1 $\vee$ P2	**P1 $\rightarrow$ P2**	P1 $\Leftrightarrow$ P2
0	0	1	0	0	**1**	1
0	1	1	0	1	**1**	0
1	0	0	0	1	**0**	0
1	1	0	1	1	**1**	1

N argumentumon értelmezett logikai függvény 2^{2^N}

Az ítéletlogika ekvivalencia szabályai

$(a \wedge b) \Leftrightarrow (b \wedge a)$	$\wedge$ kommutativitása	
$(a \vee b) \Leftrightarrow (b \vee a)$	$\vee$ kommutativitása	
$((a \wedge b) \wedge c) \Leftrightarrow (a \wedge (b \wedge c))$	$\wedge$ asszociativitása	
$((a \vee b) \vee c) \Leftrightarrow (a \vee (b \vee c))$	$\vee$ asszociativitása	
$\neg(\neg a) \Leftrightarrow a$	dupla negálás eliminálása	
$(a \rightarrow b) \Leftrightarrow (\neg b \rightarrow \neg a)$	kontrapozició	
$\mathbf{(a \rightarrow b) \Leftrightarrow (\neg a \vee b)}$	**implikáció elimináció**	
$(a \Leftrightarrow b) \Leftrightarrow ((a \rightarrow b) \wedge (b \rightarrow a))$	ekvivalencia eliminálás	
$\neg(a \wedge b) \Leftrightarrow (\neg a \vee \neg b)$	De Morgan	
$\neg(a \vee b) \Leftrightarrow (\neg a \wedge \neg b)$	De Morgan	
$(a \wedge (b \vee c)) \Leftrightarrow ((a \wedge b) \vee (a \wedge c))$	$\wedge$ disztributivitása $\vee$ felett	**DNF**
$(a \vee (b \wedge c)) \Leftrightarrow ((a \vee b) \wedge (a \vee c))$	$\vee$ disztributivitása $\wedge$ felett	**CNF**

két állítás logikailag ekvivalens, ha ugyanazokban a modellekben igaz:

$A \Leftrightarrow B$ a.cs.a, ha $A \models B$ és $B \models A$

TB $\models \alpha$ a.cs.a., ha (TB $\rightarrow \alpha$) érvényes

TB $\models \alpha$ a.cs.a., ha (TB $\wedge \neg \alpha$) kielégíthetetlen (reductio ad absurdum)

Következtetés (bizonyítások)

Modell-ellenőrzés – mondat kielégíthető-e (SAT satisfiability)
Igazságtábla listázás (exponenciális)
Visszalépéses keresés
Heurisztikus keresés modellek terében (helyes, de nem teljes)
(lokális kereséssel)

Következtetési szabályok alkalmazása
Új mondatok legális generálása régebbi mondatokból.

Bizonyítás: következtetési szabályok sorozata.
speciális operátorok **<u>megszokott</u>** keresési algoritmusokban.

- keresés általános operátorokkal (természetes dedukció)
- teljes keresés teljes operátorral általános logikában (rezolúció, exp.!)
- teljes keresés teljes operátorral redukált logikában
(Horn-klózok, Modus Ponens, redukált komplexitás!)

(más kombinációk nem garantálják a következtetés teljességét)

Az ítéletlogika következtetési mintái (operátorok)

Modus Ponens
(Implikáció eliminálása)
néha elég is

$$\frac{A \rightarrow B \quad A}{B}$$

AND eliminálása

$$\frac{A_1 \wedge A_2 \wedge \ldots \wedge A_n}{A_1, A_2, \ldots, A_n}$$

gépi világhoz

AND bevezetése

$$\frac{A_1, A_2, \ldots, A_n}{A_1 \wedge A_2 \wedge \ldots \wedge A_n}$$

TB = 1 állítás!

OR bevezetése

$$\frac{A_i}{A_1 \vee A_2 \vee \ldots \vee A_n}$$

Dupla negálás eliminálása

$$\frac{\neg \neg A}{A}$$

Elemi (egység)**rezolúció**

$$\frac{A \vee B \quad \neg B}{A}$$

Rezolúció

$$\frac{B \vee A \quad \neg B \vee G}{A \vee G}$$

Ítéletlogika eldönthető és teljes

Minden jól definiált mondat akár **igaz**, akár **hamis** volta belátható véges algoritmussal (a vonzat eldönthető) $\rightarrow$ a **következtetés igazságtábla módszere teljes**, mindig lehetséges kiszámolni a tábla 2^n sorát bármely n ítéletszimbólumot tartalmazó bizonyítás esetében.

A számítási idő azonban n-ben **exponenciális**.

Cook (1971): egy mondathalmaz kielégíthetőségi vizsgálata **NP-teljes**
(de nem minden ítéletlogikai következtetés $O(2^n)$ időt igényel! Ld. előbb)

Ítéletlogika monoton

Logika **monoton:** amikor új mondatokat adunk hozzá a TB-hoz, minden korábban maga után vonzott mondata az eredeti TB-nak továbbra is vonzata marad az új, nagyobb tudásbázisnak.

$$\textbf{ha } TB_1 \models a \text{ , akkor } (TB_1 \cup TB_2) \models a$$

Az igaz mondatok száma csak nőni tud! **Jó ez, vagy nem jó?!**

Jó: amit egyszer nehéz volt bebizonyítani, majd „ingyen van”.
Rossz: a változó világ logikai leírása is változik és ami igaz volt, nem biztos, hogy később annak megtartható.
Σ: amíg a probléma (a szükséges absztrakció szintjén) tökéletesen statikus, a monotonitással nyerünk, különben ráfizethetünk.

Horn klózok:

$P_1 \wedge P_2 \ldots \wedge \ldots P_n \rightarrow Q$

$\neg P_1 \vee \neg P_2 \ldots \vee \ldots \neg P_n \vee Q$

- Legfeljebb egy pozitív literál, pontosan egy pozitív literál = határozott klóz
- Egy hasznos mondattípus, amelyre létezik a TB méretében **lineáris idejű** következtetési eljárás.

Két fontos speciális eset:

$P_1 \wedge P_2 \wedge \ldots P_n \rightarrow$ Hamis, ua., mint $\neg P_i \vee \ldots \neg P_n$

Igaz $\rightarrow$ Q ua., mint Q.

Nem minden TB-beli állítás írható fel Horn klózként, de azoknál, amelyeknél ez megtehető: alkalmazzuk Modus Ponens-t, ameddig marad alkalmazható következtetés.

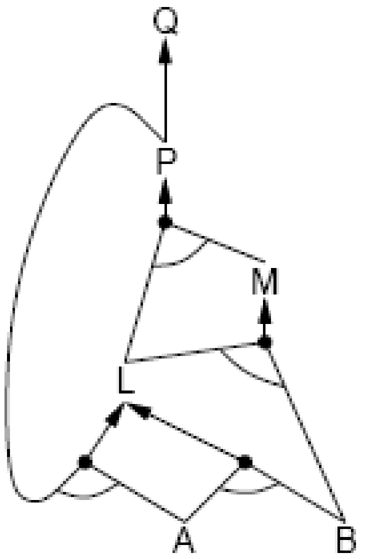
Modus Ponens **teljes** bizonyítási lépés a **Horn klózok** tudásbázisában!

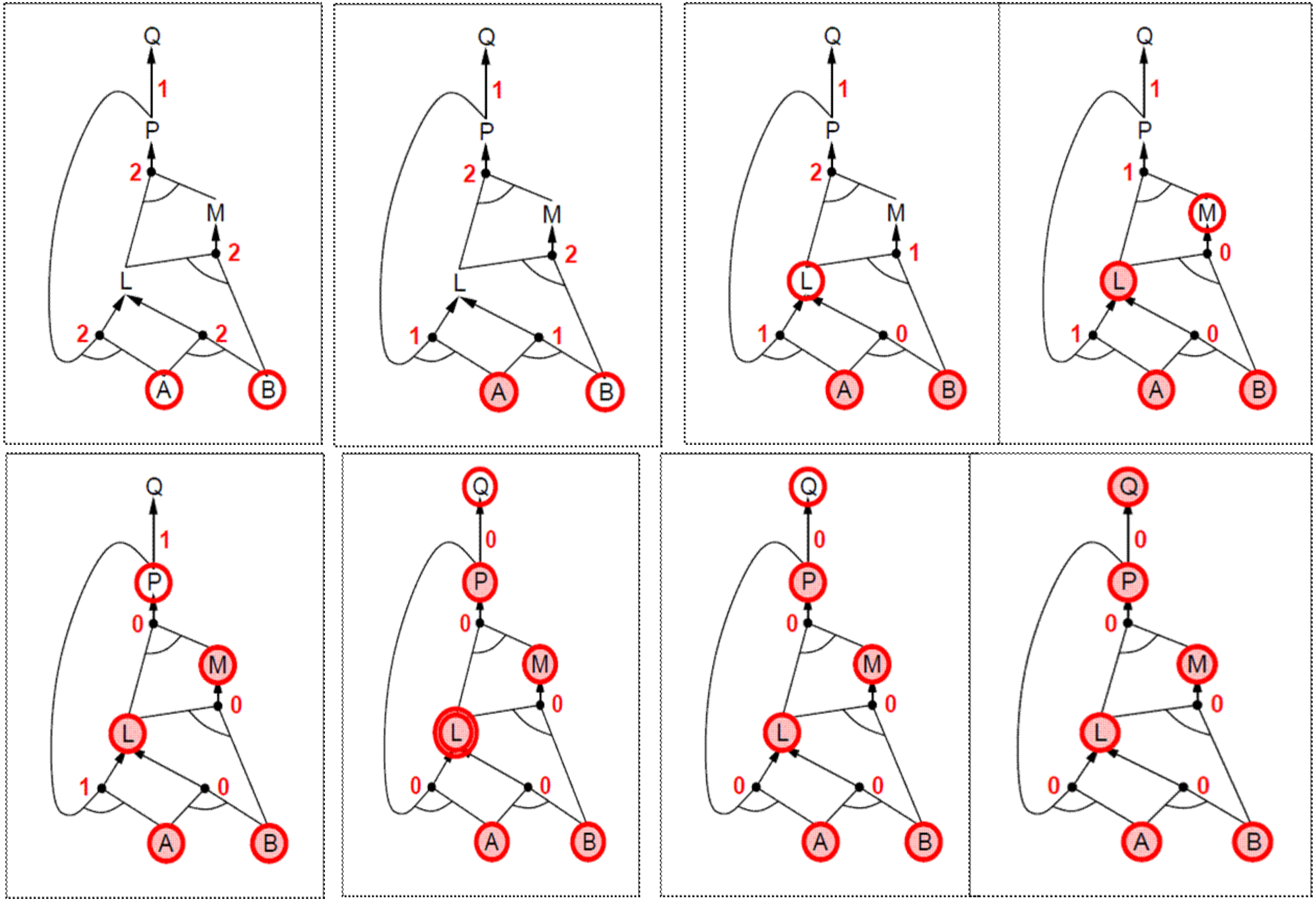
Előrecsatolt, hátracsatolt (Modus Ponens) következtetés

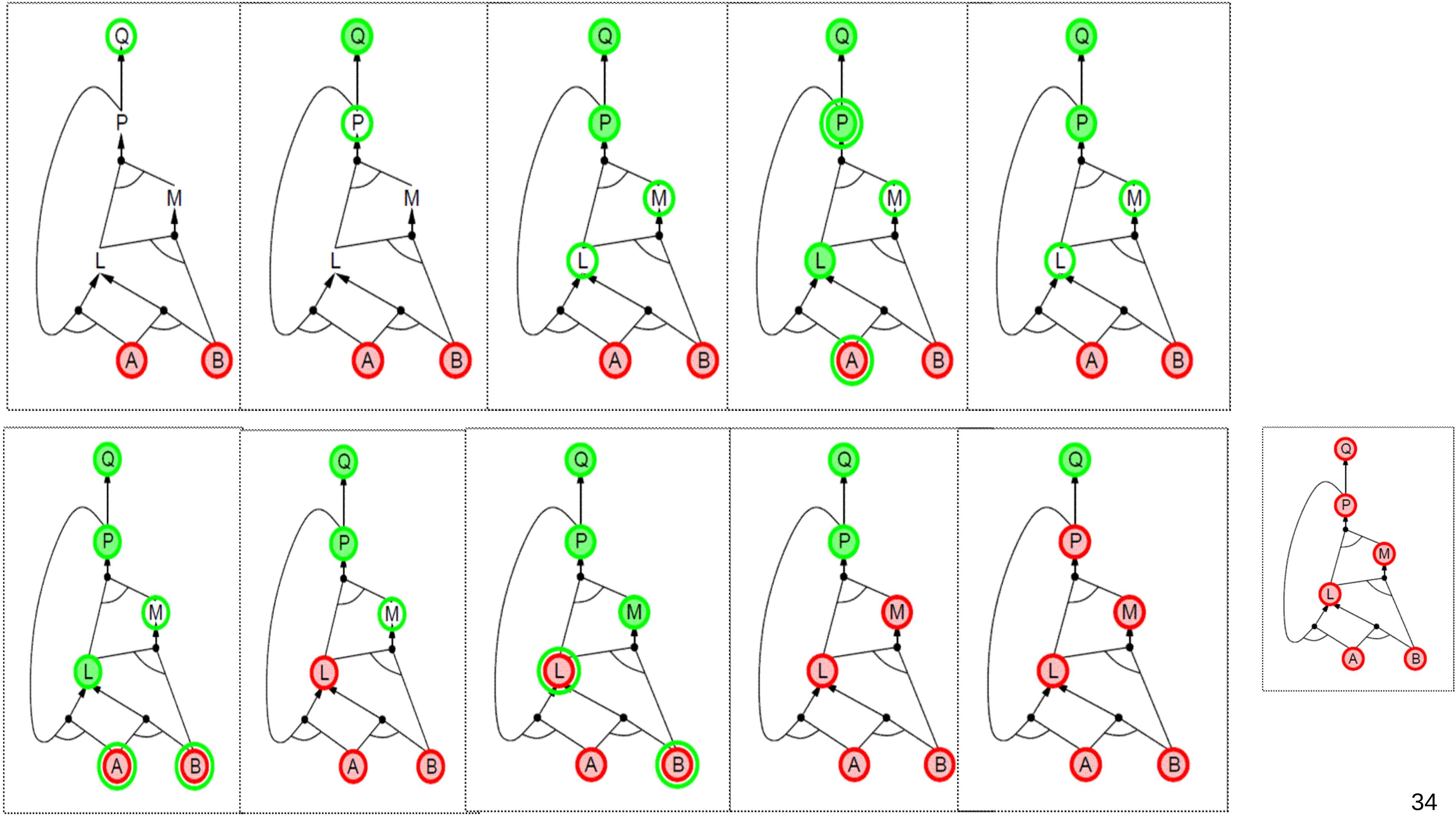
$$P \Rightarrow Q$$
$$L \wedge M \Rightarrow P$$
$$B \wedge L \Rightarrow M$$
$$A \wedge P \Rightarrow L$$
$$A \wedge B \Rightarrow L$$
$$A$$
$$B$$

ECs: Minden szabályt elsütni, melynek premisszája teljesül. Következmény hozzáadása a TB-hoz, amíg a lekérdezendő változó értéket nem kap (fixpont).

HCs: A lekérdezéstől visszafelé:
- ellenőrzés, netán a lekérdezés már igaz,
- elővenni egy szabályt, melynek következménye a lekérdezés és a premisszáit bizonyítani rekurzív Hcs-sal.




Q 1 P 2 M 2 L 2 2 A B
Q 1 P 2 M 2 L 1 1 A B
Q 1 P 2 M 1 L 1 0 A B
Q 1 P 1 M 0 L 1 0 A B
Q 1 P 0 M 0 L 1 0 A B
Q 0 P 0 M 0 L 0 0 A B
Q 0 P 0 M 0 L 0 0 A B
Q 0 P 0 M 0 L 0 0 A B


Q
P
M
L
A
B

Egy történet: „Ádám, Béla és Csaba közül valaki egy üvegház ablakát törte ki. Ádám azt állítja, hogy: 'Béla tette, Csaba nem bűnös.' Béla azt állítja, hogy: 'Ha Ádám bűnös, akkor Csaba is'. Csaba azt állítja, hogy: 'Nem én tettem; többiek közül valaki'."

Konzisztensek (egyszerre igazak) ezek az állítások?
Ha egyik sem bűnös, van-e, aki hazudik?
Ha mindenki igazat mond, akkor ki a bűnös? Stb.

A történet **ítélet szimbólumai**:
A: Ádám <u>nem</u> bűnös,
B: Béla <u>nem</u> bűnös,
C: Csaba <u>nem</u> bűnös és a mondanivalójuk:

SA $= \neg B \wedge C$ **SB** $= \neg A \rightarrow \neg C$ **SC** $= C \wedge (\neg B \vee \neg A)$

az un. „elmélet" pedig (az igaznak tartott mondatok összessége), ami egyben a TB tartalma: **TB** $= SA \wedge SB \wedge SC$

A	B	C	SA	SB	SC	TB = SA ∧ SB ∧ SC
H	H	H	H	I	H	H
H	H	I	I	H	I	H
H	I	H	H	I	H	H
H	I	I	H	H	I	H
I	H	H	H	I	H	H
I	H	I	I	I	I	I
I	I	H	H	I	H	H
I	I	I	H	I	H	H

felsorolásos bizonyítás,
2^N sor, $O(2^N)$ komplexitás (NP)

Visszalépéses keresés modellek Terében - Davis-Putnam, stb.

A	B	C	SA	SB	SC	TB
H	H	H	H	I	H	H
H	H	I	I	H	I	H
H	I	H	H	I	H	H
H	I	I	H	H	I	H
I	H	H	H	I	H	H
I	H	I	I	I	I	I
I	I	H	H	I	H	H
I	I	I	H	I	H	H

Keressük TB = SA ∧ SB ∧ SC =
¬B ∧ C ∧ (C ∧ (¬B ∨ ¬A)) ∧ (¬A → ¬C)
modelljét (ahol a TB kielégíthető, azaz igaz lesz)

¬B ∧ C ∧ (¬B ∨ ¬A) ∧ (¬A → ¬C)
igazához szükséges, hogy ¬B legyen igaz, azaz **B** legyen **hamis**

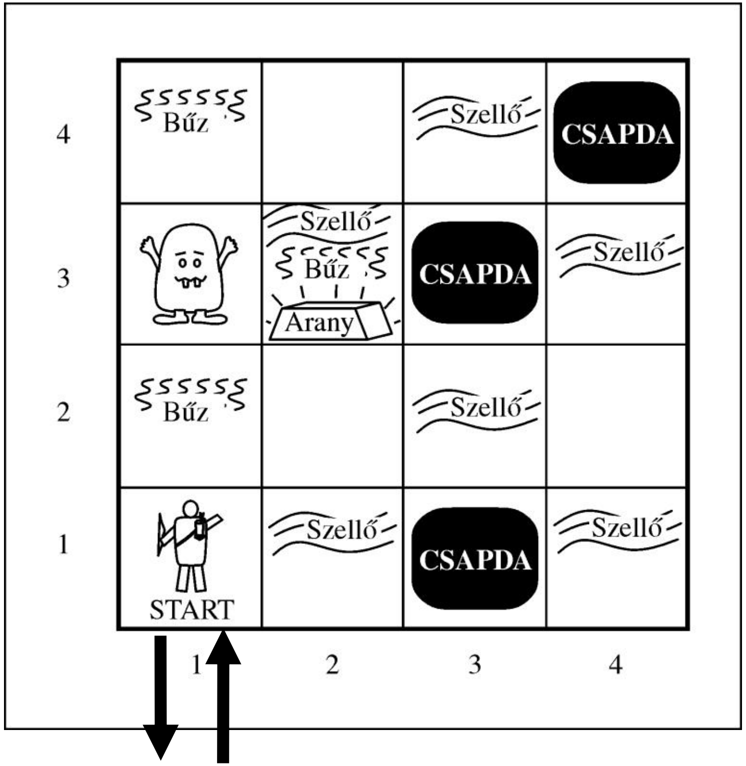
C ∧ (¬A → ¬C) igazához szükséges, hogy **C** legyen **igaz**

¬A → Hamis igazához szükséges, hogy ¬A legyen hamis, azaz **A** legyen **igaz**

(Vajon más kombinációkra is?)

Logikusan gondolkozó ágens - Mit jelentsen ez?

Esettanulmány - a „Wumpus világ” környezetében (jegyzet)



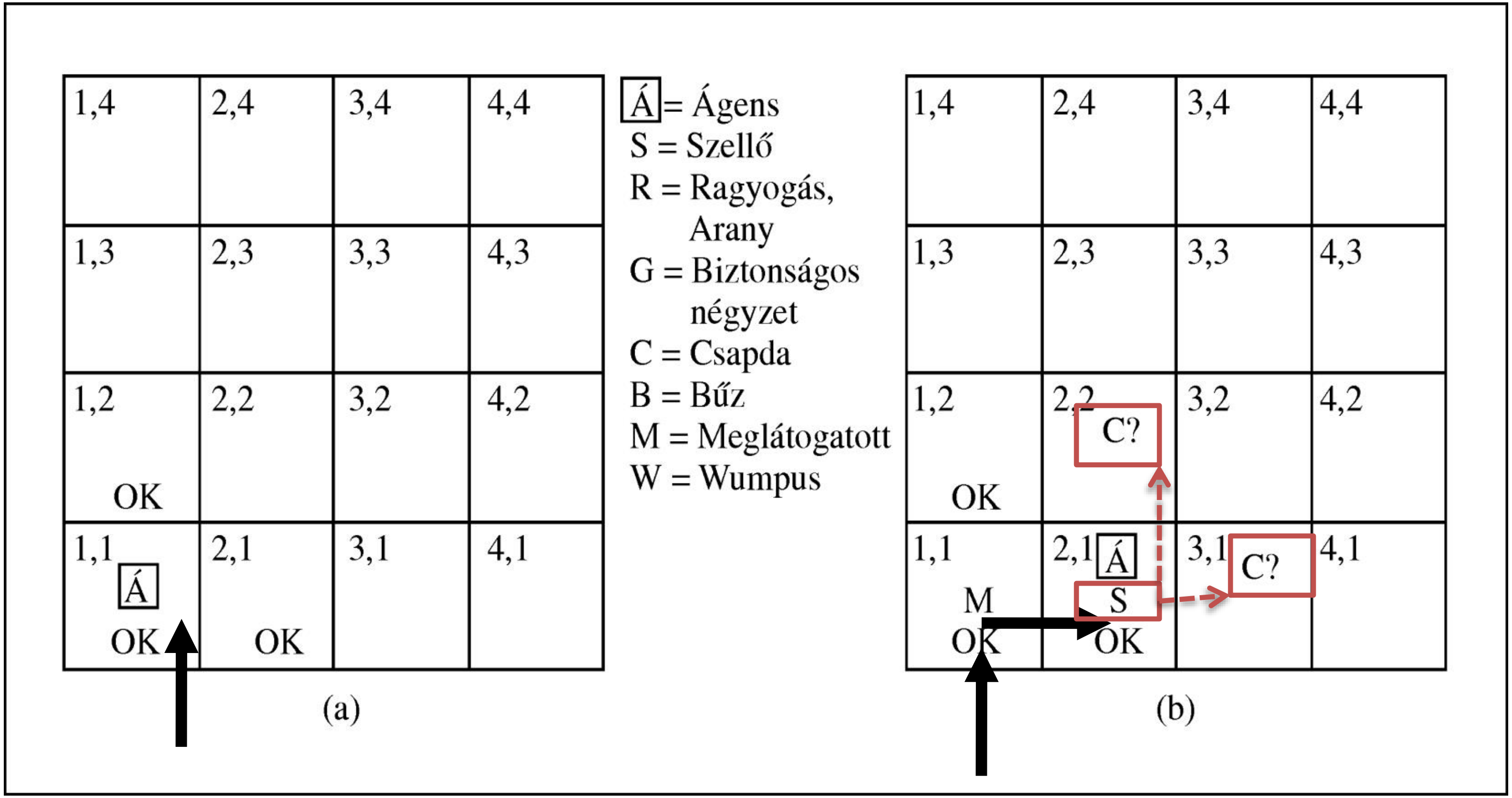
Ha ágens, akkor érzékelések, cselekvések és célok.

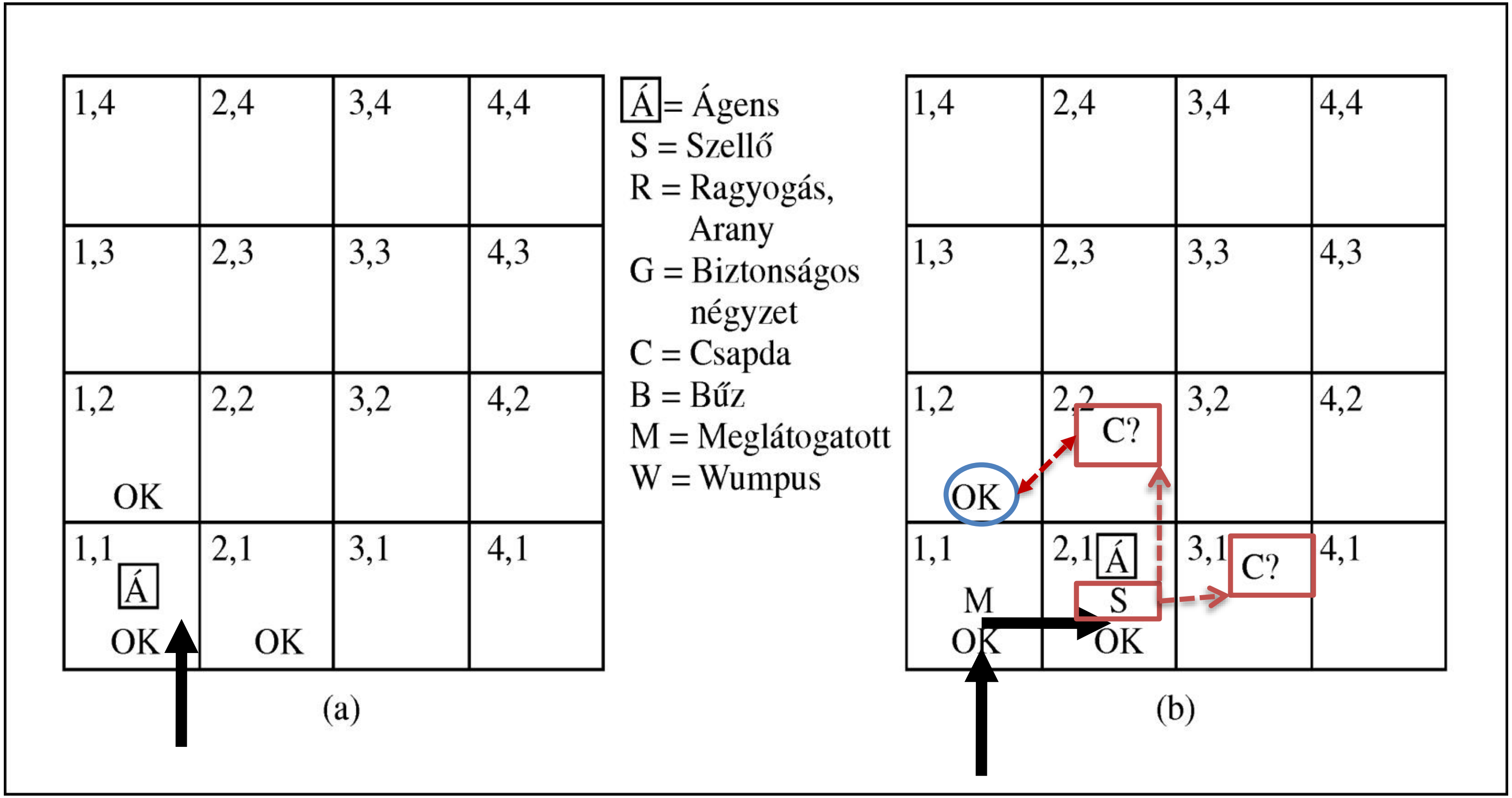
Érzet(ek):
[Bűz, Szellő, Csillog, Ütközés, Sikoly]
hol vagyunk, nem tudni.

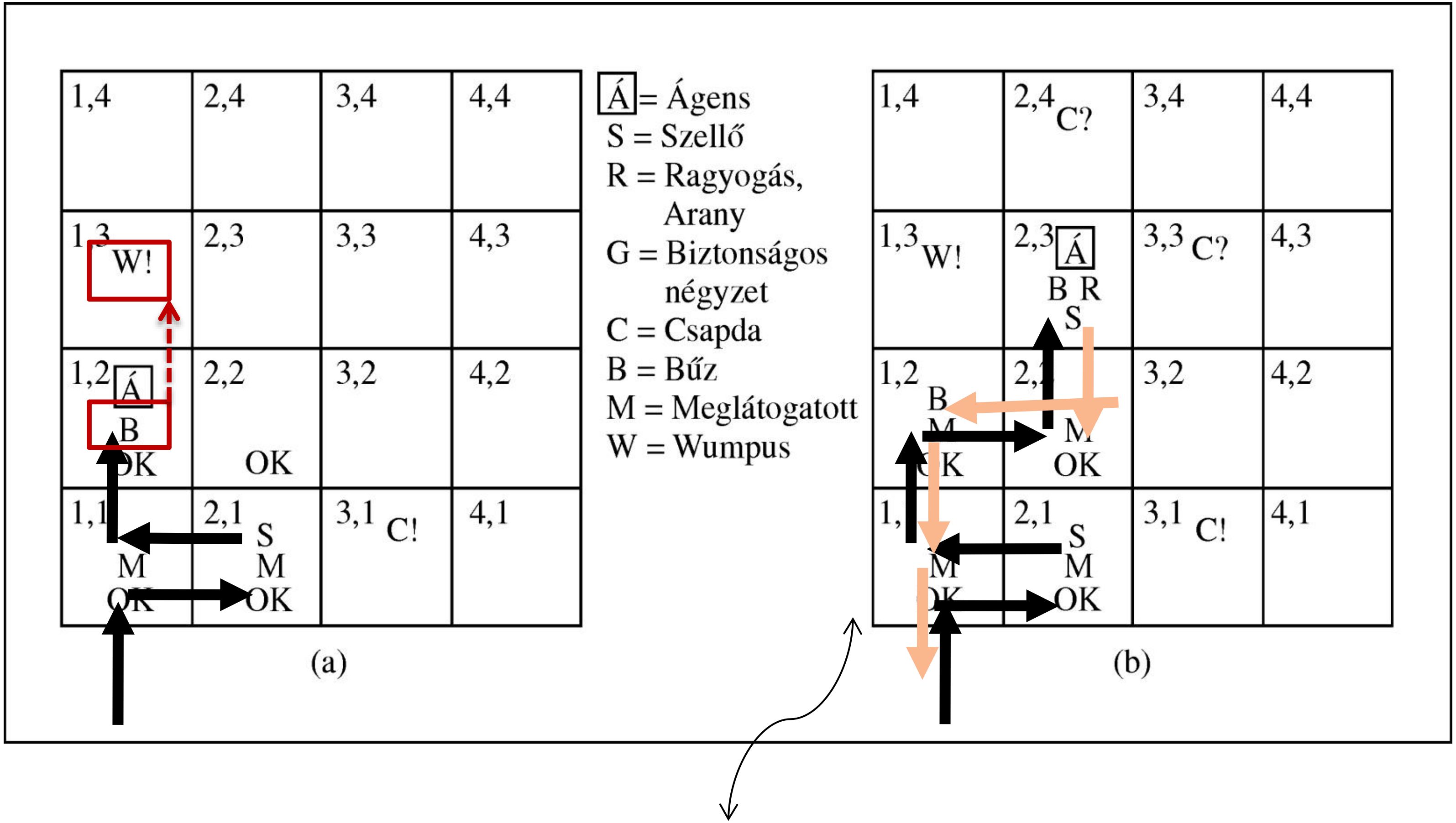
Cselekvések:
Előre, Jobbra/Balra, Megragad,
Lő, Mászik, … ágens meghal,
ha csapdára vagy Wumpusra rálép.

cél: arany - megtalálni és kihozni

Hogyan biztosítható a siker (logikai gondolkodással)?


1,4
2,4
3,4
4,4
1,3
2,3
3,3
4,3
1,2
OK
2,2
3,2
4,2
1,1
Á
OK
2,1
OK
3,1
4,1
(a)
Á = Ágens
S = Szellő
R = Ragyogás, Arany
G = Biztonságos négyzet
C = Csapda
B = Bűz
M = Meglátogatott
W = Wumpus
1,4
2,4
3,4
4,4
1,3
2,3
3,3
4,3
1,2
OK
2,2
C?
3,2
4,2
1,1
M
OK
2,1
Á
S
OK
3,1
C?
4,1
(b)


1,4
2,4
3,4
4,4
1,3
2,3
3,3
4,3
1,2
2,2
3,2
4,2
OK
1,1
Á
OK
2,1
OK
3,1
4,1
(a)
Á = Ágens
S = Szellő
R = Ragyogás, Arany
G = Biztonságos négyzet
C = Csapda
B = Bűz
M = Meglátogatott
W = Wumpus
1,4
2,4
3,4
4,4
1,3
2,3
3,3
4,3
1,2
2,2
C?
3,2
4,2
OK
1,1
M
OK
2,1
Á
S
OK
3,1
C?
4,1
(b)

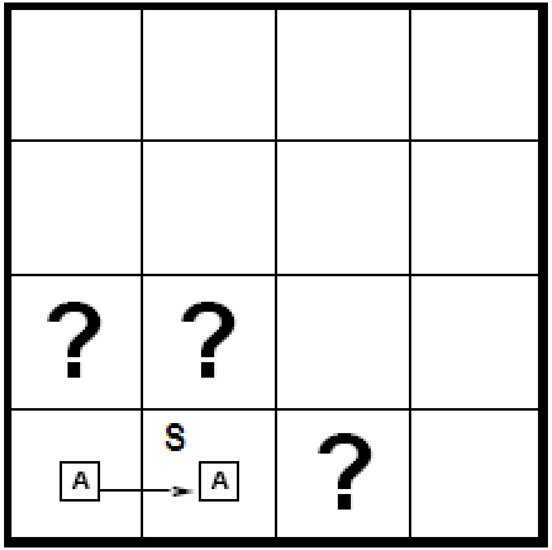


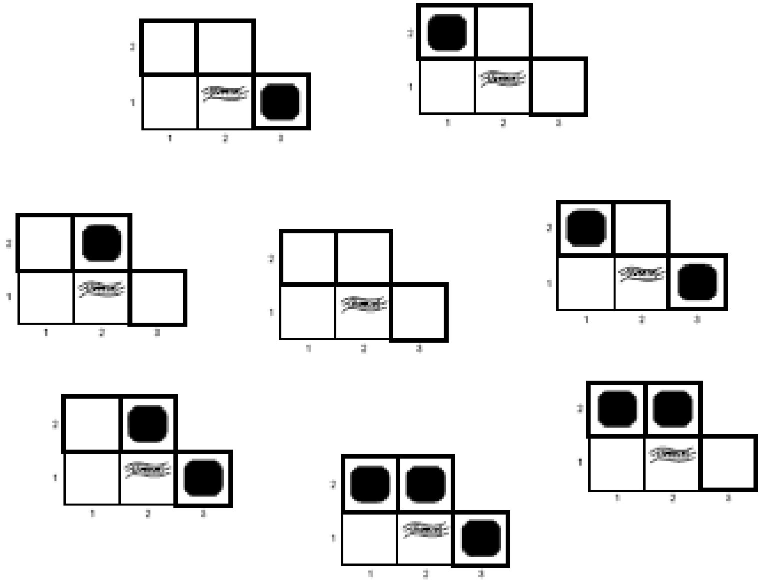
Környezeti titkok (nem hozzáférhető!) kiderítése logikai modellezéssel és gondolkodással, majd jó cselekvések eldöntése.

Vonzat és modellek Wumpus világban

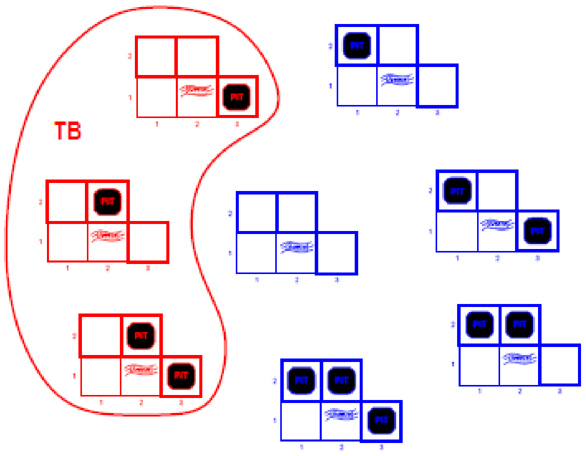
(1,1)-ben semmi érzet, lépés (2,1)-re, szellőt érezni, ...

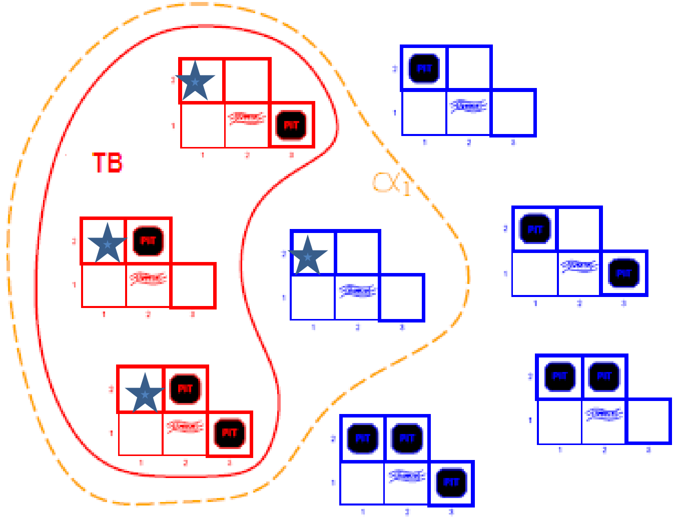
Lehetséges modellek a (?)-re, csakis csapdát feltételezve:
3 db bináris változó, 8 db lehetséges modell

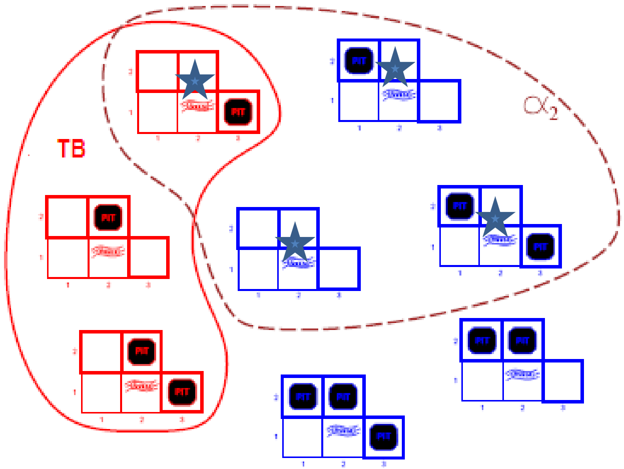




TB = Wumpus világ szabályai + megfigyelések







α1 = „[1,2] biztonságos",★

***TB* |= α1**, modellellenőrzéssel

α2 = „[2,2] biztonságos",★

***TB* |=/= α2**, modellellenőrzéssel

Egy ágens a Wumpus világ számára

Tudásbázis ciklus: ágens érzetei → mondatok → tudásbázis
további érvényes mondatok érzet mondatok vonzatai.

tegyük fel:
$B_{1,2}$ = "bűz van az [1,2]-ben" (de lehetne B12, vagy XYZ)
$S_{1,2}$ = „szellő van az [1,2]-ben", stb.

Tudásbázis: $\neg S_{1,1}$, $\neg B_{1,1}$, $S_{2,1}$, $\neg B_{2,1}$, $\neg S_{1,2}$, $B_{1,2}$, … stb.

amit az ágens tud a környezetéről általánosságban, pl.:

ha valahol nincs bűz, akkor sem ott, sem szomszédban nincs Wumpus.

$\mathbf{R_1}$: $\neg B_{1,1} \rightarrow \neg W_{1,1} \wedge \neg W_{1,2} \wedge \neg W_{2,1}$

$\mathbf{R_2}$: $\neg B_{2,1} \rightarrow \neg W_{1,1} \wedge \neg W_{2,1} \wedge \neg W_{2,2} \wedge \neg W_{3,1}$

$\mathbf{R_3}$: $\neg B_{1,2} \rightarrow \neg W_{1,1} \wedge \neg W_{1,2} \wedge \neg W_{2,2} \wedge \neg W_{1,3}$

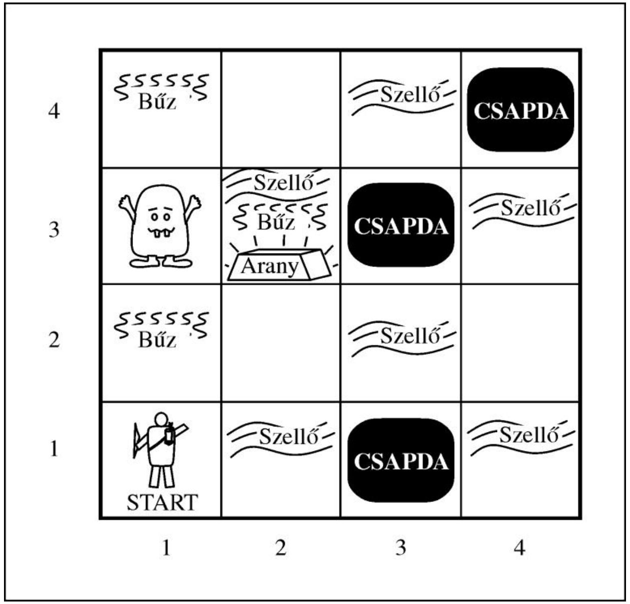
.....

ha bűz van az [1,2]-ben, akkor egy Wumpus van ott, vagy egy vagy több szomszédos négyzetben:

$\mathbf{R_4}$: $B_{1,2} \rightarrow W_{1,3} \vee W_{1,2} \vee W_{2,2} \vee W_{1,1}$

A Wumpus megtalálása - az ágensnek el kellene készítenie a **TB → $\mathbf{W_{1,3}}$** igazságtábláját, hogy megmutassa, hogy ez a mondat **érvényes** (mert akkor a Wumpus(1,3) jelenléte a TB (= ágens tudása) vonzata!).

kb. 12 ítélet szimbólum van: az igazságtáblának 2^{12} = 4096 sora lesz, minden sorban, amelyben a TB mondat igaz, a **$\mathbf{W_{1,3}}$** –nak is igaznak kellene lennie.
Alkalmazzunk inkább a következtetési mintákat! (de természetes dedukcióval)



$$\neg S_{1,1} \; \neg B_{1,1} \; S_{2,1} \; \neg B_{2,1} \; \neg S_{1,2} \; B_{1,2} \, , \ldots$$

$$\mathbf{W_{1,3}}$$

1. MP: $\neg B_{1,1}$ + R1-re: $\neg \mathbf{W}_{1,1} \wedge \neg \mathbf{W}_{1,2} \wedge \neg \mathbf{W}_{2,1}$

$\mathbf{R_1}$: $\neg \mathbf{B}_{1,1} \rightarrow \neg \mathbf{W}_{1,1} \wedge \neg \mathbf{W}_{1,2} \wedge \neg \mathbf{W}_{2,1}$

Modus Ponens
(Implikáció eliminálása)

$$\frac{A \rightarrow B \qquad A}{B}$$

$$\frac{\neg \mathbf{B}_{1,1} \rightarrow \neg \mathbf{W}_{1,1} \wedge \neg \mathbf{W}_{1,2} \wedge \neg \mathbf{W}_{2,1} \qquad \neg \mathbf{B}_{1,1}}{\neg \mathbf{W}_{1,1} \wedge \neg \mathbf{W}_{1,2} \wedge \neg \mathbf{W}_{2,1}}$$

2. És-eliminálás: $\neg \mathbf{W}_{1,1}$, $\neg \mathbf{W}_{1,2}$, $\neg \mathbf{W}_{2,1}$

3\. MP: $\neg B_{2,1}$ + R2-re, utána és-elimináció: $\neg \mathbf{W}_{2,2}$, $\neg \mathbf{W}_{2,1}$, $\neg \mathbf{W}_{3,1}$

R_2: $\neg \mathbf{B}_{2,1} \rightarrow \neg \mathbf{W}_{1,1} \wedge \neg \mathbf{W}_{2,1} \wedge \neg \mathbf{W}_{2,2} \wedge \neg \mathbf{W}_{3,1}$

Modus Ponens
(Implikáció eliminálása)

$$\frac{A \rightarrow B \qquad A}{B}$$

$$\frac{\neg \mathbf{B}_{2,1} \rightarrow \neg \mathbf{W}_{1,1} \wedge \neg \mathbf{W}_{2,1} \wedge \neg \mathbf{W}_{2,2} \wedge \neg \mathbf{W}_{3,1} \qquad \neg \mathbf{B}_{2,1}}{\neg \mathbf{W}_{1,1} \wedge \neg \mathbf{W}_{2,1} \wedge \neg \mathbf{W}_{2,2} \wedge \neg \mathbf{W}_{3,1}}$$

4. MP: $B_{1,2}$-re és R4-re: $\mathbf{W_{1,3} \vee W_{1,2} \vee W_{2,2} \vee W_{1,1}}$

$\mathbf{R_4}$: $\mathbf{B_{1,2} \rightarrow W_{1,3} \vee W_{1,2} \vee W_{2,2} \vee W_{1,1}}$

Modus Ponens
(Implikáció eliminálása)

$$\frac{A \rightarrow B \quad A}{B}$$

$$\frac{\mathbf{B_{1,2} \rightarrow W_{1,3} \vee W_{1,2} \vee W_{2,2} \vee W_{1,1}} \quad \mathbf{B_{1,2}}}{\mathbf{W_{1,3} \vee W_{1,2} \vee W_{2,2} \vee W_{1,1}}}$$

5. Egység rezolúció: $W_{1,3} \vee W_{1,2} \vee W_{2,2} \vee W_{1,1}$ és $\neg W_{1,1}$

Elemi (egység)**rezolúció**

$$\frac{A \vee B \qquad \neg B}{A}$$

$$\frac{W_{1,3} \vee W_{1,2} \vee W_{2,2} \vee W_{1,1} \qquad \neg W_{1,1}}{W_{1,3} \vee W_{1,2} \vee W_{2,2}}$$

az eredmény: $W_{1,3} \vee W_{1,2} \vee W_{2,2}$

6. Egység rezolúció: $W_{1,3} \vee W_{1,2} \vee W_{2,2}$ és $\neg W_{2,2}$
és az eredmény: $W_{1,3} \vee W_{1,2}$

7. Végül, mégy egy rezolúció: $W_{1,3} \vee W_{1,2}$ és $\neg W_{1,2}$
és az eredmény: $W_{1,3}$

1. MP: $\neg B_{1,1}$ + R1-re: $\neg \mathbf{W}_{1,1} \wedge \neg \mathbf{W}_{1,2} \wedge \neg \mathbf{W}_{2,1}$

2. És-eliminálás: $\neg \mathbf{W}_{1,1}$, $\neg \mathbf{W}_{1,2}$, $\neg \mathbf{W}_{2,1}$

3. MP: $\neg B_{2,1}$ + R2-re, utána és-elimináció: $\neg \mathbf{W}_{2,2}$, $\neg \mathbf{W}_{2,1}$, $\neg \mathbf{W}_{3,1}$

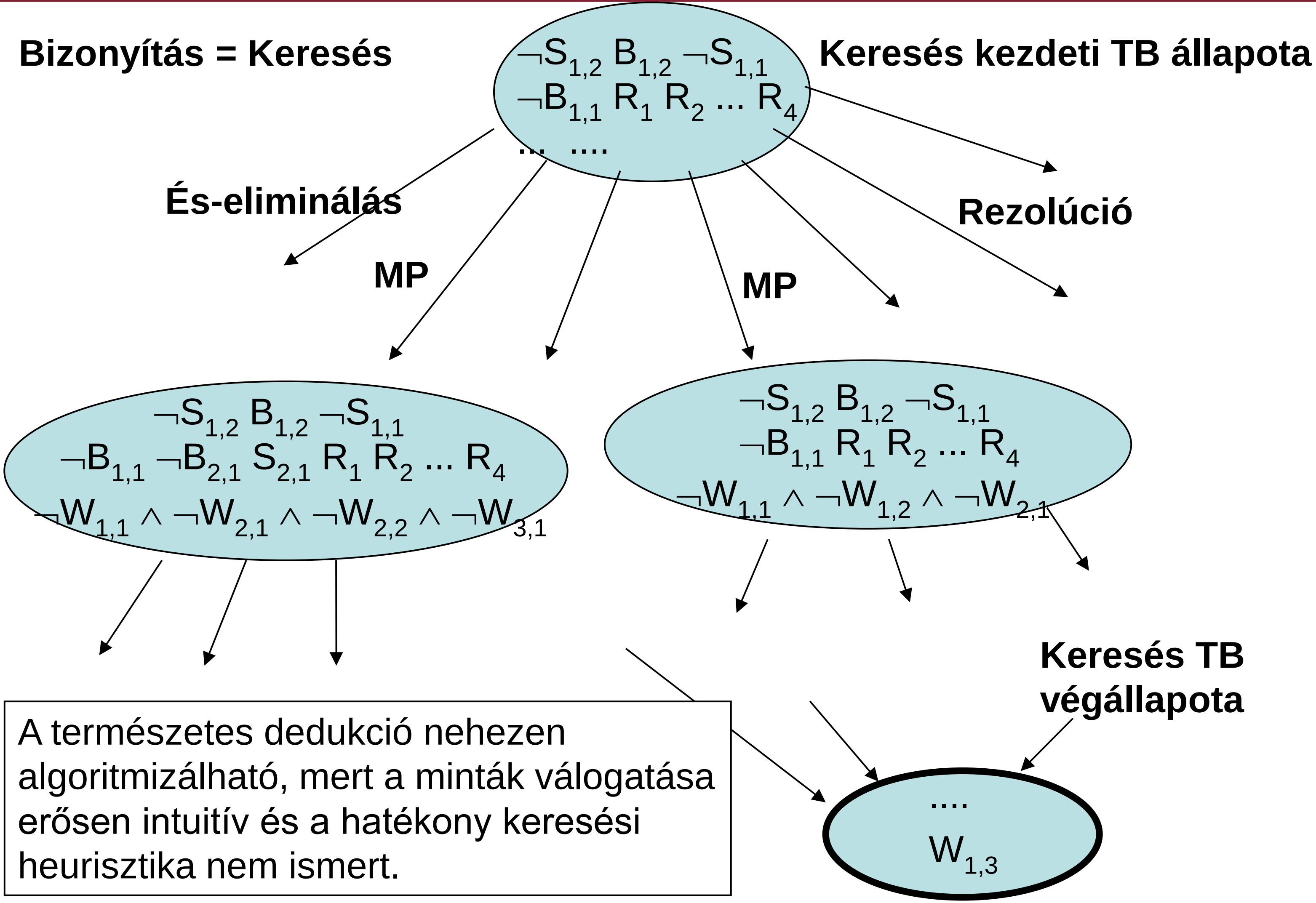
4. MP: $B_{1,2}$-re és R4-re: $\mathbf{W}_{1,3} \vee \mathbf{W}_{1,2} \vee \mathbf{W}_{2,2} \vee \mathbf{W}_{1,1}$

5. Egység rezolúció: $\mathbf{W}_{1,3} \vee \mathbf{W}_{1,2} \vee \mathbf{W}_{2,2} \vee \mathbf{W}_{1,1}$ és $\neg \mathbf{W}_{1,1}$
   az eredmény: $\mathbf{W}_{1,3} \vee \mathbf{W}_{1,2} \vee \mathbf{W}_{2,2}$

6. Egység rezolúció: $\mathbf{W}_{1,3} \vee \mathbf{W}_{1,2} \vee \mathbf{W}_{2,2}$ és $\neg \mathbf{W}_{2,2}$
   és az eredmény: $\mathbf{W}_{1,3} \vee \mathbf{W}_{1,2}$

7. Végül, mégy egy rezolúció: $\mathbf{W}_{1,3} \vee \mathbf{W}_{1,2}$ és $\neg \mathbf{W}_{1,2}$
   és megjelent a kívánt válasz: $\mathbf{W}_{1,3}$
   azaz a Wumpus tényleg az [1,3]-ban van.

Bizonyítás = Keresés



A természetes dedukció nehezen algoritmizálható, mert a minták válogatása erősen intuitív és a hatékony keresési heurisztika nem ismert.