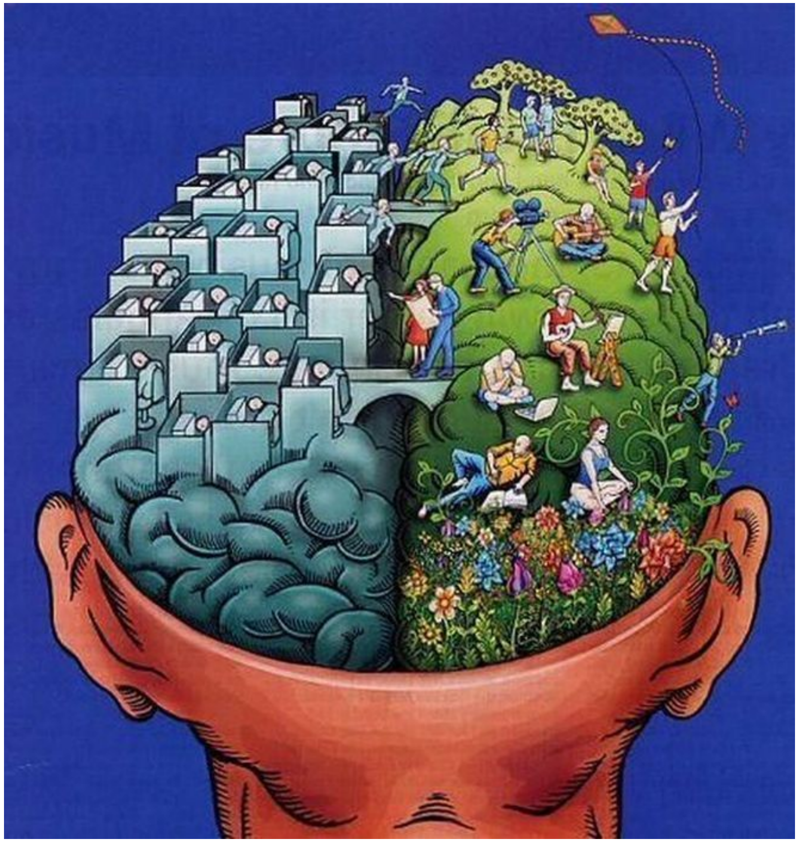
Mesterséges Intelligencia MI

Logikai ágens és problémái



Dobrowiecki Tadeusz
Eredics Péter, és mások

BME I.E. 437, 463-28-99
dobrowiecki@mit.bme.hu,
http://www.mit.bme.hu/general/staff/tade

Ugye volt már?

Boole-algebra: $-(A + B) = (-A)(-B)$

$(A + B) C = AC + BC$, …, stb.

Függvények: NOT, AND, OR, XOR, …, stb.

Függvények definíciói igazságtáblákkal

Ekvivalens átalakítások, redukciós formulák, pl.

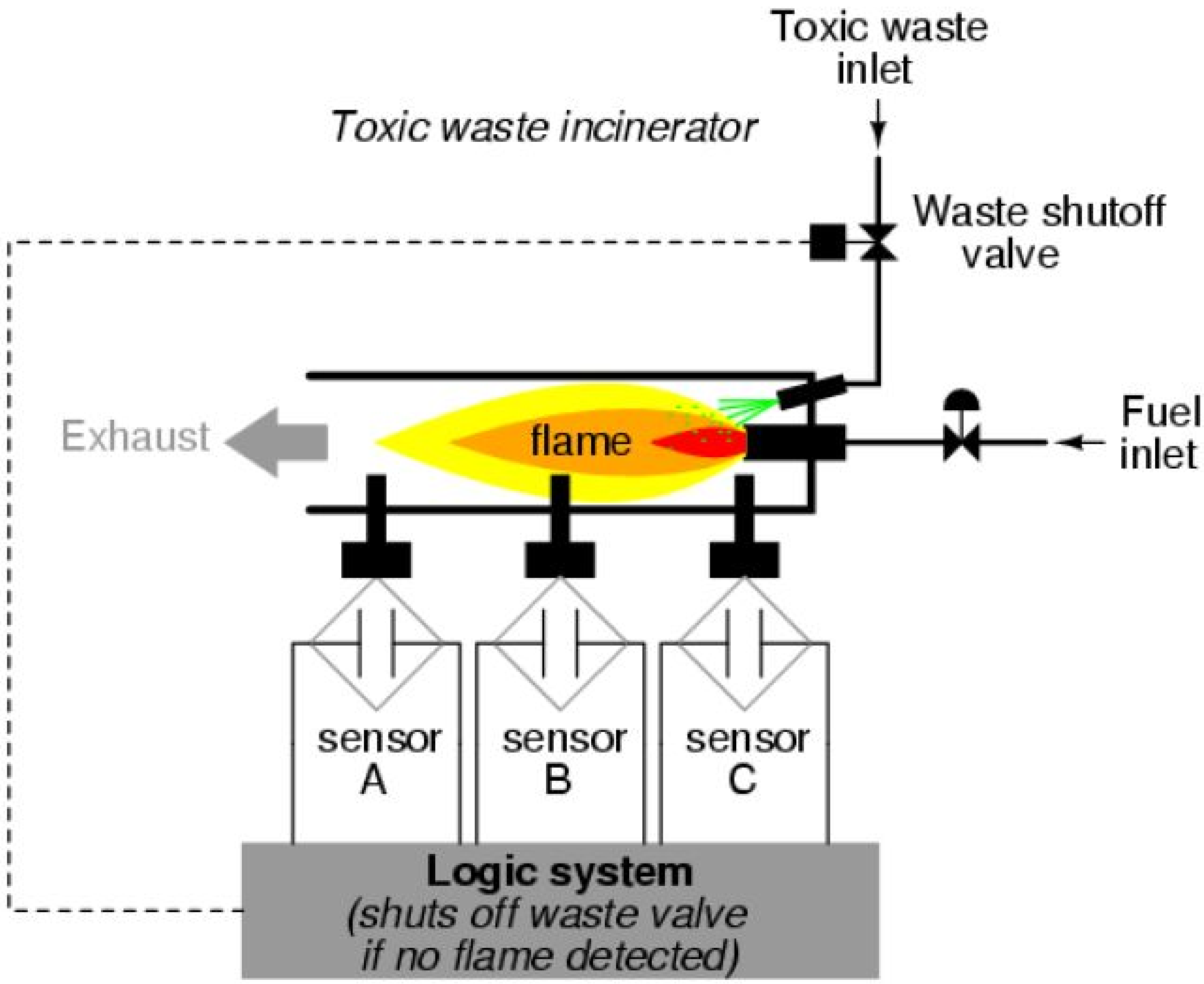
$A + C \text{ NOT } A = A + C$, …, stb.

Carnough táblák

Venn diagramok

Logikai függvények optimalizálása (digitális tervezés)

...


Toxic waste
inlet
Toxic waste incinerator
Waste shutoff
valve
Exhaust
flame
Fuel
inlet
sensor
A
sensor
B
sensor
C
Logic system
(shuts off waste valve
if no flame detected)

sensor inputs

A	B	C	Output
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	1

Output = 0 (close valve)

Output = 1 (open valve)

sensor inputs

A	B	C	Output
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

sensor inputs

A	B	C	Output	
0	0	0	0	
0	0	1	0	
0	1	0	0	
0	1	1	1	$\overline{A}BC = 1$
1	0	0	0	
1	0	1	1	$A\overline{B}C = 1$
1	1	0	1	$AB\overline{C} = 1$
1	1	1	1	$ABC = 1$

$$\text{Output} = \overline{A}BC + A\overline{B}C + AB\overline{C} + ABC$$

$$\text{Output} = AB + BC + AC$$

Milyen feladatai lehetnek itt logikának?

Rolle tétele

A tétel

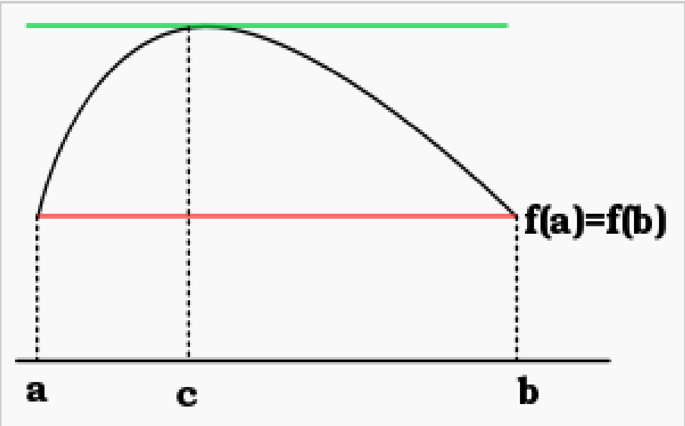
Ha az f függvény folytonos az $[a, b]$ intervallumban, differenciálható az intervallum belső pontjaiban és

$$f(a) = f(b),$$

akkor van olyan $a < c < b$ szám, hogy

$$f'(c) = 0$$

teljesül.



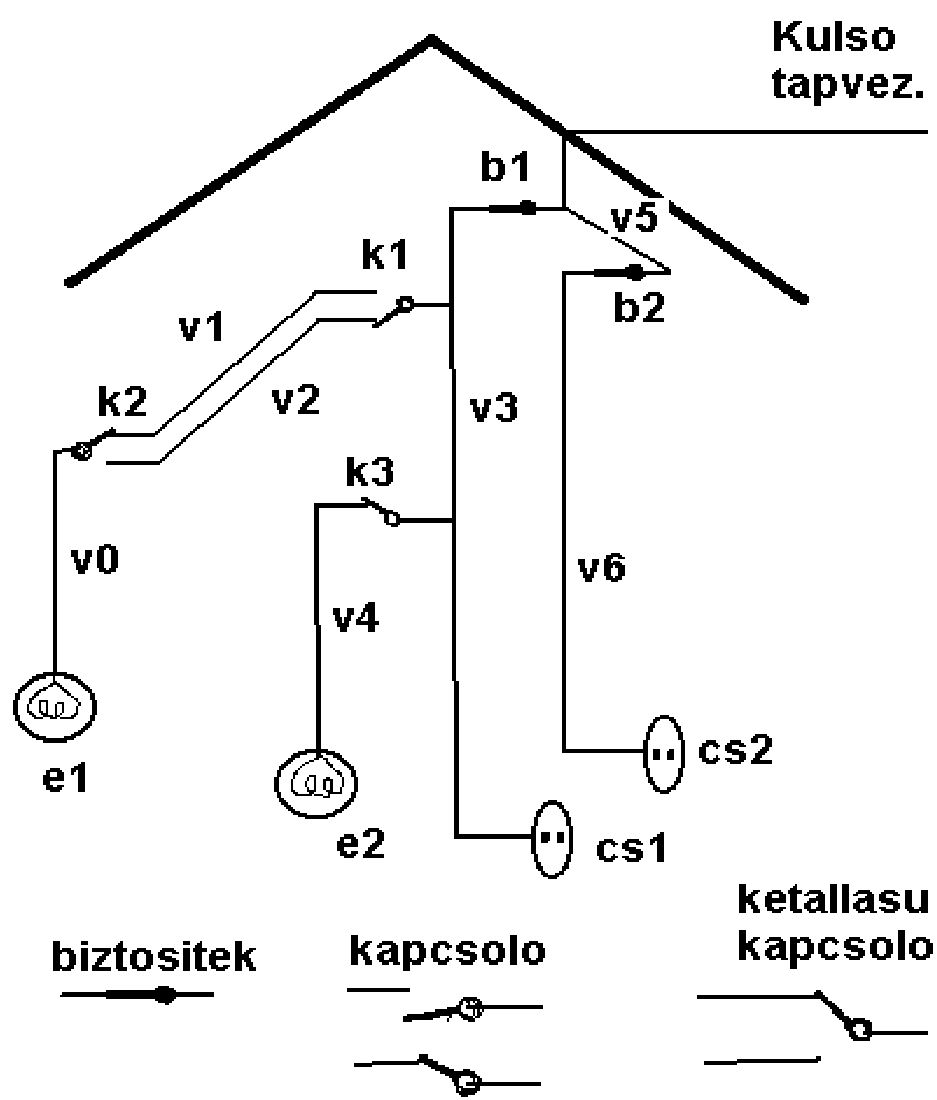
Ábra Rolle tételéhez

Bizonyítása

Ha az f függvény az (a, b) intervallumon végig az $f(a) = f(b)$ értéket veszi fel, akkor konstans, tehát deriváltja mindenütt 0.

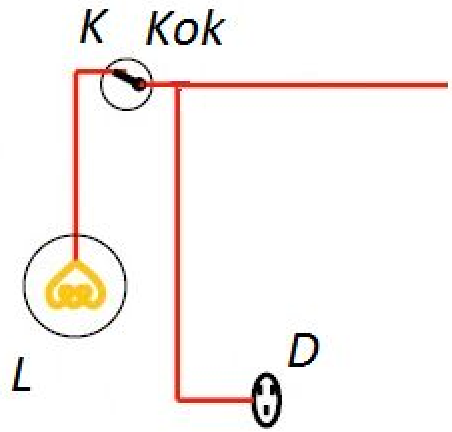
Tegyük fel, hogy egy pontban f értéke ettől eltér, mondjuk. Az általánosság megszorítása nélkül feltehető, hogy ez az érték nagyobb $f(a) = f(b)$-nél (ellenkező esetben ugyanezt a gondolatmenetet a $-f$ függvényre kell alkalmaznunk). Weierstrass tétele szerint a függvény az $[a, b]$ intervallumban valahol felveszi maximumát. Legyen c egy ilyen pont. c nem lehet a-val vagy b-vel egyenlő, mert akkor lenne nála nagyobb értékű hely, ami ellentmond $f(c)$ maximális tulajdonságának. Mivel f a c-ben (mely az értelemezési tartomány belső pontjában van) differenciálható és ott maximuma van, ezért a szélsőértékekre vonatkozó Fermat-tétel miatt ott a deriváltja 0. ■

Milyen feladatai lehetnek itt logikának?



PL.
(k2 v1) + (-k2 v2) = v0
v3 k3 = v4

Milyen feladatai lehetnek itt logikának?



D	K	Kok	L	Elm.
0	0	0	0	1
0	0	0	1	0
0	0	1	0	1
0	0	1	1	0
0	1	0	0	1
0	1	0	1	0
0	1	1	0	1
0	1	1	1	0
1	0	0	0	1
1	0	0	1	0
1	0	1	0	1
1	0	1	1	0
1	1	0	0	1
1	1	0	1	0
1	1	1	0	0
1	1	1	1	1

Elmélet: Konnektorban van áram AND
Kapcsoló ok AND
Kapcsoló állása fel =
Lámpa ég

Kapcsoló állása fel, de a lámpa nem ég!
Áram a konnektorban rendben van. És?

D	K	Kok	L	Elm.
0	0	0	0	1
0	0	0	1	0
0	0	1	0	1
0	0	1	1	0
0	1	0	0	1
0	1	0	1	0
0	1	1	0	1
0	1	1	1	0
1	0	0	0	1
1	0	0	1	0
1	0	1	0	1
1	0	1	1	0
1	1	0	0	1
1	1	0	1	0
1	1	1	0	0
1	1	1	1	1

D	K	Kok	L	Elm.
0	0	0	0	1
0	0	1	0	1
0	1	0	0	1
0	1	1	0	1
1	0	0	0	1
1	0	1	0	1
1	1	0	0	1
1	1	1	0	0

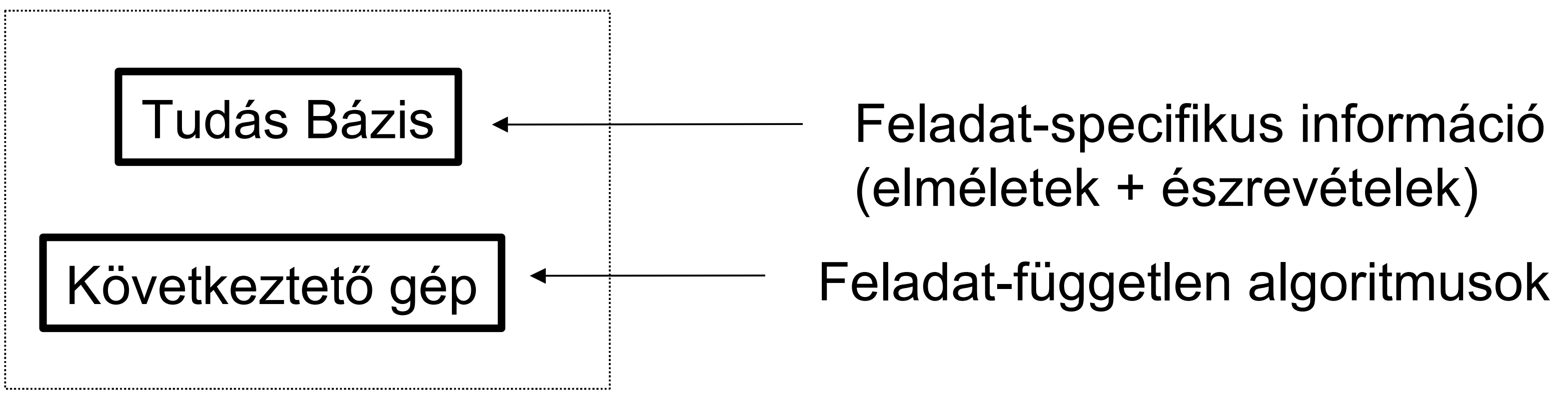
D	K	Kok	L	Elm.
1	0	0	0	1
1	0	1	0	1
1	**1**	**0**	**0**	**1**
1	1	1	0	0

Lámpa nem ég > Áram van > Kapcsoló fel > Ha elmélet igaz, akkor Kapcsoló NEM ok!

Nagyobb hálózatnál: (itt) 17 változó + berendezések állapotai …

2^{17} > 100 000 tábla bejegyzés, valami jobbat kell kitalálni!

Logika



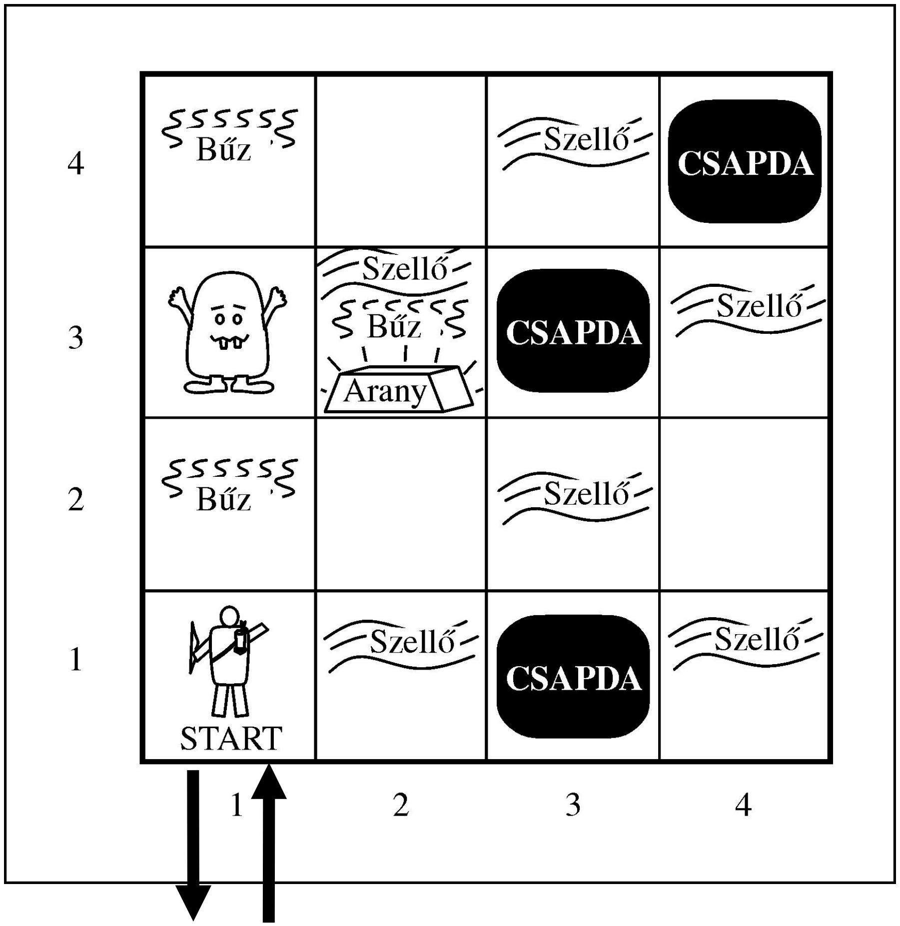
Ágens mire használhatja a logikai képességeit?

Hogyan fejezze ki ehhez a tudását,

Milyen algoritmusok lesznek jók, ügyesek, …?

Logikusan gondolkozó ágens - Mit jelentsen ez?

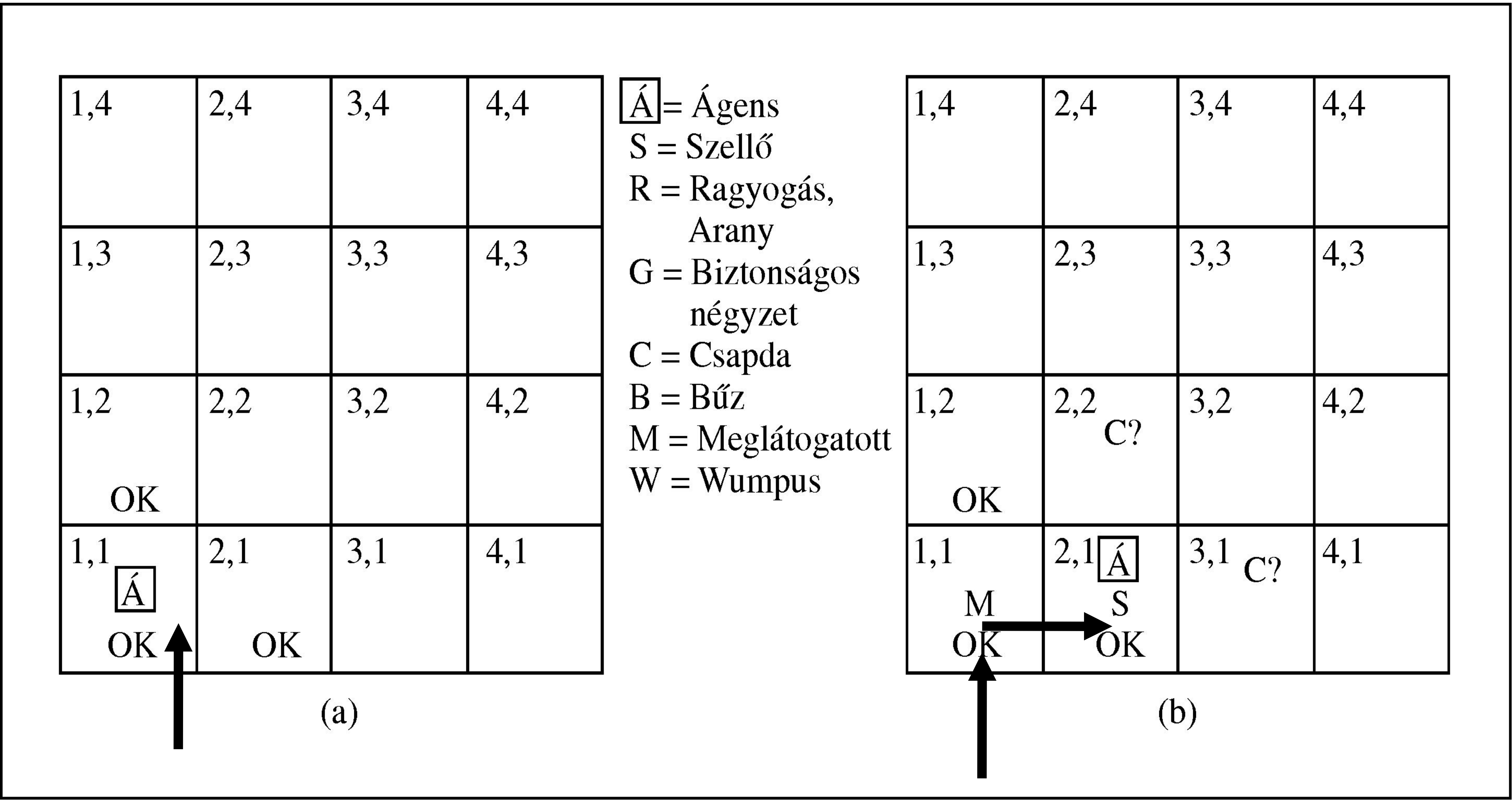
Esettanulmány - a „Wumpus világ” környezetében (jegyzet)

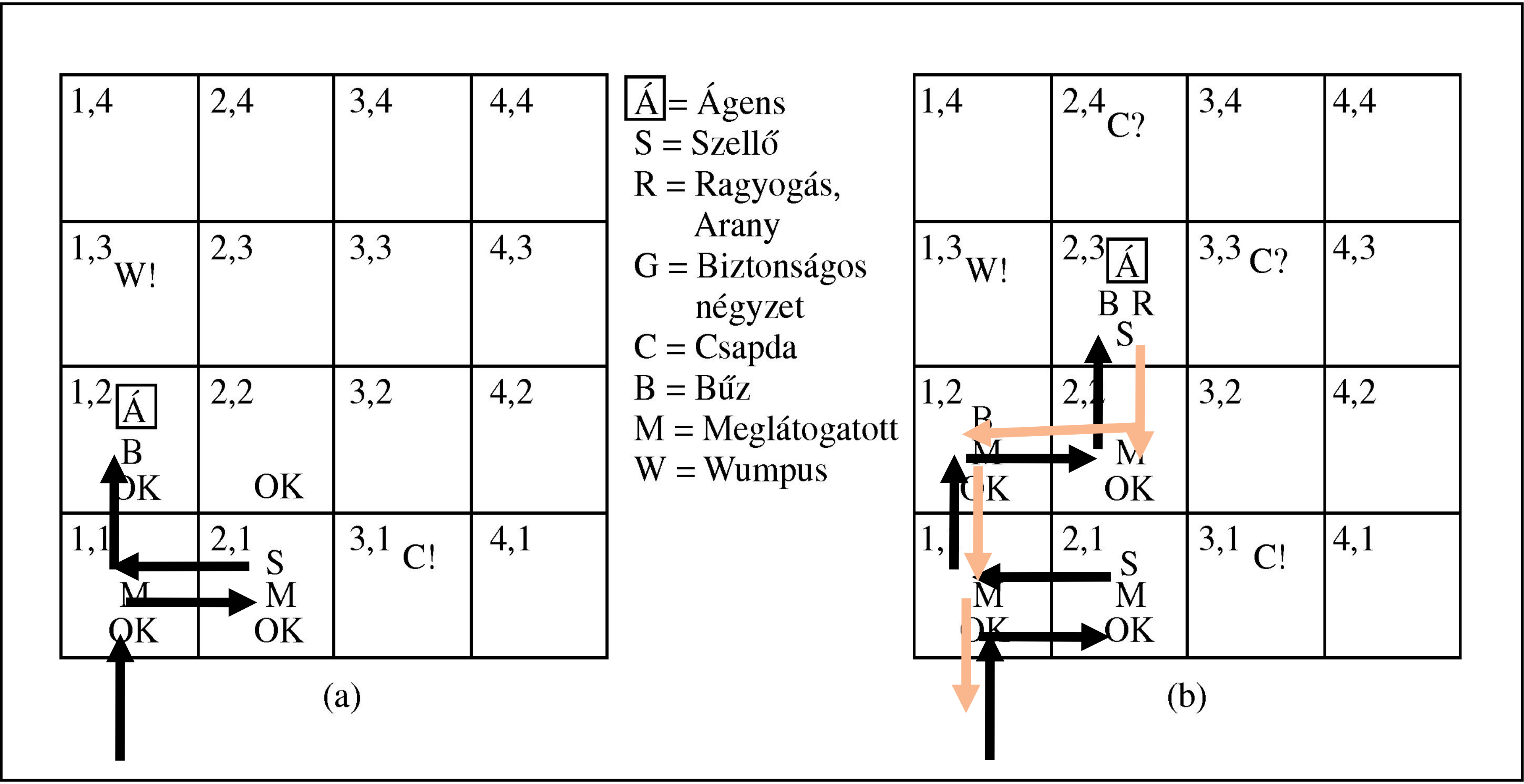


Ha ágens, akkor érzékelések, cselekvések és célok.

Bűz, szellő, az arany csillogása,
falak és a sikoltozó szörny,
hol vagyunk, nem tudni,
az érzet(ek):
[Bűz, Szellő, Csillog, Ütközés, Sikoly]
a cselekvések:
Előre, Jobbra/Balra, Megragad,
Lő, Mászik,
ágens meghal, ha csapdára vagy
egy élő Wumpusra rálép .
cél: arany - megtalálni, és
gyorsan kihozni

Hogyan biztosítható a siker?


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
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 W!
2,3
3,3
4,3
1,2 Á B OK
2,2 OK
3,2
4,2
1,1 M OK
2,1 S M OK
3,1 C!
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 C?
3,4
4,4
1,3 W!
2,3 Á B R S
3,3 C?
4,3
1,2 B M OK
2,2 M OK
3,2
4,2
1,1 M OK
2,1 S M OK
3,1 C!
4,1
(b)

Ez komoly teljesítmény volt, egy viszonylag nehéz következtetés.

Felhasználtunk ismereteket, amelyek különböző időpontokban és különböző helyeken lettek begyűjtve.

Érzetek hiányára is támaszkodunk.

Ezt kell most hatékonyan gépesíteni!

Logika – a reprezentáció és a manipuláció eszköze

Szintaktika - legálisan létrehozható szimbolikus mondatok

Szemantika - mik a világ tényei, amikre a szimbolikus mondatok vonatkoznak

Minden szintaktikailag helyes mondat állít valamit a világról.
Ha a mondat az ágens valamely fizikai konfigurációja által reprezentált, akkor az ágens hiszi a hozzátartozó mondatot.

Vonzat reláció és a következtetés

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

Vonzat: logikai konzekvencia dolgok között.

Vonzat reláció a tudásbázis TB és egy α mondat között

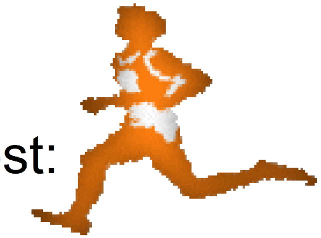
α vonzata TB -nek: **TB |= $\boldsymbol{\alpha}$** ha α minden olyan világban igaz, ahol TB is.

Következtetés: a vonzat „kiszámítása” a reprezentáción belül a mondatok formális manipulálásával.

Bizonyítás = a következtetési algoritmus lépéssorozata

α bizonyítható TB -ből: **TB |-- $\boldsymbol{\alpha}$**

Szemantika, interpretáció, szándékolt interpretáció



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

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

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

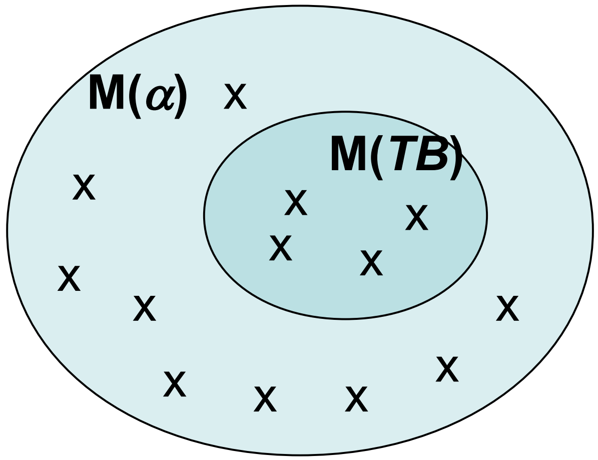
Ha „fut” fizikai mozgás, és „Ágens1” egy személy és a világ most: akkor ez a mondat (most) **hamis**.

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

Mi tehát 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.

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.

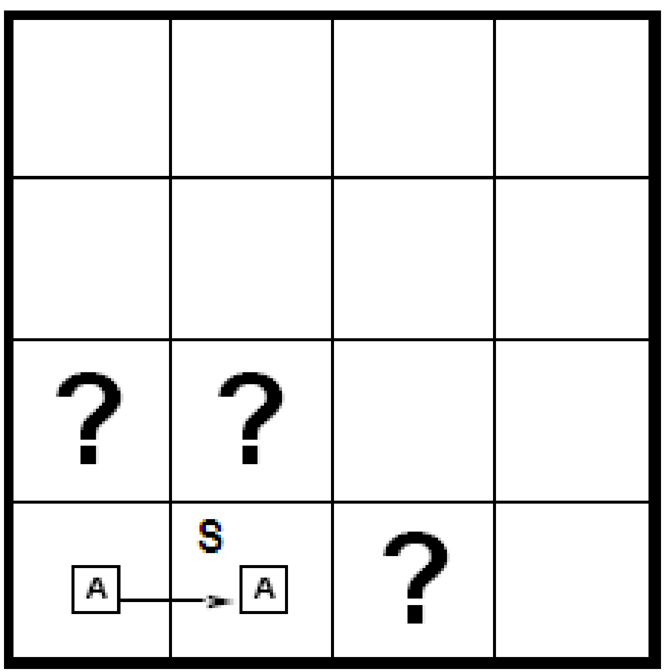
Ha ez így van, akkor ha TB igaz, akkor α is igaz

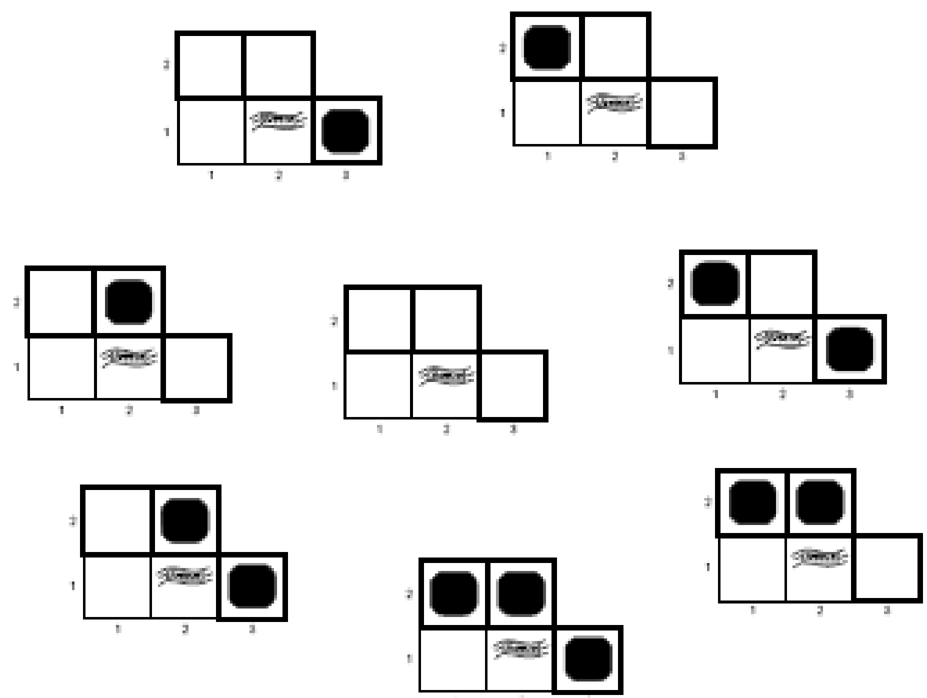
TB $\models \alpha$ akkor és csak akkor, ha M(TB) $\subseteq$ M(α)

Vonzat és modellek Wumpus világban

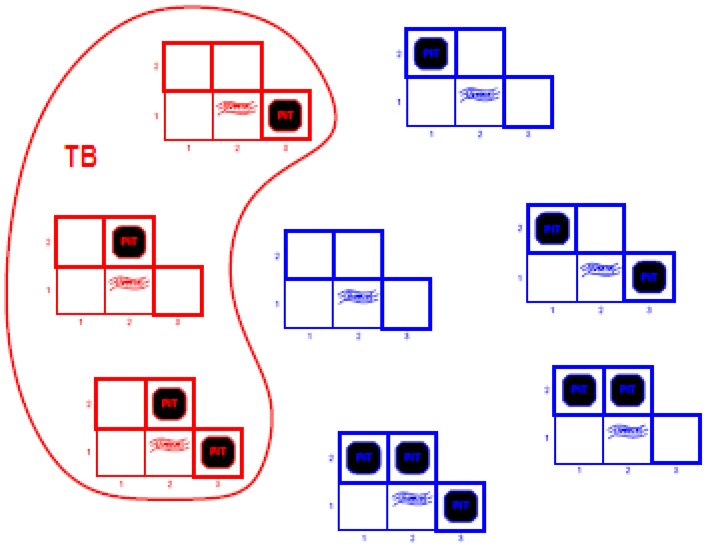
(1,1)-ben semmi, lépés át (2,1)-re, szellőt érezni, ...

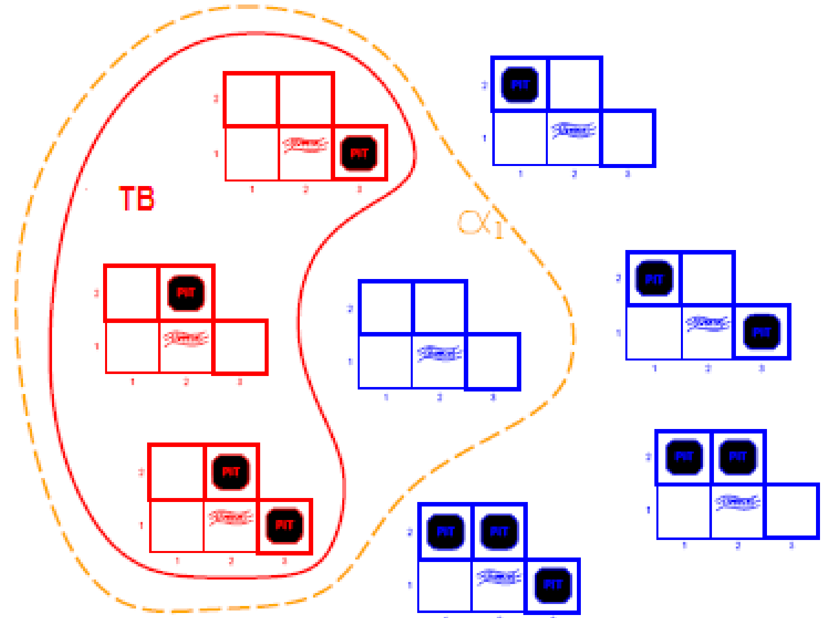
Lehetséges modellek a (?)-re, csakis csapdát feltételezve: 3 db bináris változó, 8 db lehetséges modell

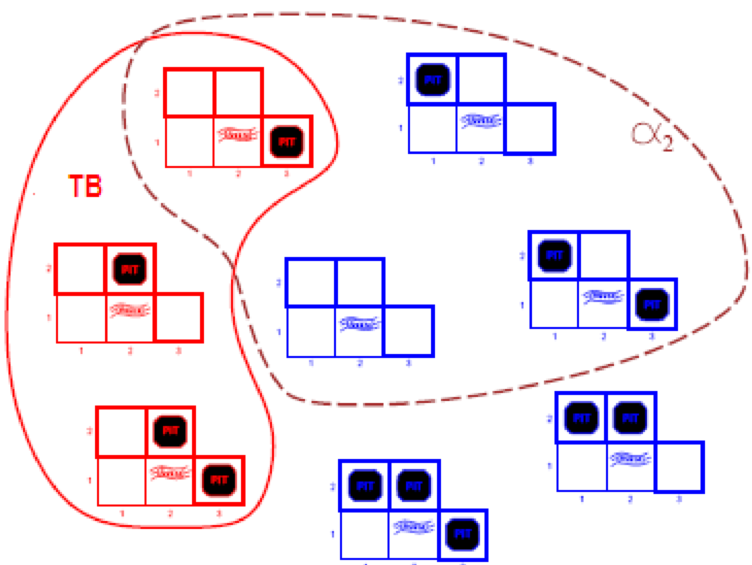




TB = Wumpus világ szabályai + megfigyelések







$\alpha 1$ = „[1,2] biztonságos",
TB |= α1, modellellenőrzéssel

$\alpha 2$ = „[2,2] biztonságos",
TB |=/= α2, modellellenőrzéssel

Vonzat reláció és a következtetés

Következtetési eljárás két dolgot tehet:
(1) ha adott egy TB, létrehozhat új α mondatot, amely vonzata a TB-nak
(2) ha adott egy TB és egy α mondat, megállapíthatja hogy α vonzata-e a TB-nak.

Igazságtartó következtetési eljárás (extenzionális), ha igazság függvényekkel dolgozik,
Igazság függvény logikai értéke csakis a részeinek és a logikai műveletek definíciójától függ.

Formális bizonyítás – igazságtartó következtetési eljárás lépéssorozata.

Egy **következtetési eljárás teljes:** ha minden vonzat mondathoz képes találni egy bizonyítást = avagy ami igaz, az bebizonyítható.
A |= B -ból **A |-- B**

Egy **következtetési eljárás helyes:**
ha minden bizonyított mondat tényleg vonzat relációban áll a felhasznált tényekkel = avagy ami bebizonyítható, az igaz is. **A |-- B** -ból **A |= B**

Szemantika miatt logikai mondatoknak több fajtája van

Érvényes (analitikus mondat, tautológia):
ha minden világban minden lehetséges interpretációja igaz, függetlenül attól, hogy mit akart 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.

Amely 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?

Az eddigiek minden logikára vonatkoznak, amit a matematika ismer, pontosabban matematika minden formális nyelvére, amilyen a matematikai logika is.

Most az informatikában (és MI-ben) használatos alaplogikákkal fogunk foglalkozni, meglátjuk, hogyan néznek belülről, mire tudjuk őket használni, milyen algoritmusokkal?

Ítéletlogika (ítéletkalkulus, propozicionális logika,
    0-ad rendű logika,kombinációs hálózatok, digitális technika logikája)
Predikátum kalkulus (első rendű logika, FOL)
    (programozási nyelvek logikája, intelligens ágens logikája)
Leíró logikák (description logic, DL)
    (Web alkalmazások (ontológiák) logikája)
Temporális logikák (modális logikák)
    (elosztott szoftvertesztelés logikája, ágenskommunikáció)
Episztemikus logikák (modális logikák)
    (intelligens ágensek gondolkodása és kommunikációja)
Több-értékű logikák
    (hardvertesztelés, hibatűrő tervezés)
...

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

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

ha P ítéletmondat, akkor $\neg$ **P** is (negálás).
ha P1 és P2, akkor **P1** $\wedge$ **P2** is (konjunkció).
ha P1 és P2, akkor **P1** $\vee$ **P2** is (diszjunkció).
ha P1 és P2, akkor **P1** $\rightarrow$ **P2** is (implikáció).
ha P1 és P2, akkor **P1** $\Leftrightarrow$ **P2** is (ekvivalencia).

literál: atomi mondat
v. negált atomi mondat

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

Szemantika

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

$\neg$ **P** igaz a.cs.a, ha P hamis.
P1 $\wedge$ **P2** igaz a.cs.a, ha P1 igaz **és** P2 igaz.
P1 $\vee$ **P2** igaz a.cs.a, ha P1 igaz, **vagy** P2 igaz.
P1 $\rightarrow$ **P2** igaz a.cs.a, ha P1 hamis **vagy** P2 igaz, azaz
hamis a.cs.a, ha P1 igaz **és** P2 hamis.
P1 $\Leftrightarrow$ **P2** igaz a.cs.a, ha P1 $\rightarrow$ P2 igaz **és** P2 $\rightarrow$ P1 igaz.

Szemantika

Tetszőleges mondat szemantikája - rekurzív elemzés, pl.:

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

Összekötő jelek igazságtáblái

P	Q	$\neg P$	$P1 \wedge P2$	$P1 \vee P2$	**$P1 \rightarrow P2$**	$P1 \Leftrightarrow P2$
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}

http://en.wikipedia.org/wiki/Truth_table

Mire jó egy igazságtábla?

„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 (azaz lehetnek-e egyszerre igazak) ezek az állítások?
Ha egyik sem bűnös, akkor ki hazudik?
Ha mindenki csak igazat mond, akkor ki a bűnös? Stb.

A történet **ítélet szimbólumai**:
A: Ádám nem bűnös, **B**: Béla nem bűnös, **C**: Csaba nem 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 (1)(3)
I	I	H	H	I	H	H
I	I	I	H	I	H	H (2)

(1) Létezik olyan A, B, C érték kombináció, hogy mindhárman igazat mondanak.
(2) Ha egyik sem bűnös, akkor Ádám és Csaba hazudtak.
(3) Ha igazat mondtak, akkor Béla a bűnös.

Vonzat, érvényesség és következtetés

TB |= α akkor és csak akkor, ha (TB $\rightarrow$ α) érvényes
TB |= α akkor és csak akkor, ha (TB $\wedge \neg$ α) kielégíthetetlen
(reductio ad absurdum)

TB	B	$\neg$ B	**TB $\rightarrow \neg$ B**	TB $\wedge$ B
0	0	1	**1**	0
0	1	0	**1**	0
1	0	1	**1**	0
1	1	0	0	1

Béla bűnös ($\neg$B) következik TB-ból.

Ádám bűnös ($\neg$A) NEM következik TB-ból.

TB	A	$\neg$ A	**TB $\rightarrow \neg$ A**	TB $\wedge$ A
0	0	1	**1**	0
0	1	0	**1**	0
1	1	0	**0**	1
1	0	1	1	0

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

Az ítéletlogika ekvivalencia szabályai

$(a \wedge b) \equiv (b \wedge a)$	$\wedge$ kommutativitása
$(a \vee b) \equiv (b \vee a)$	$\vee$ kommutativitása
$((a \wedge b) \wedge c) \equiv (a \wedge (b \wedge c))$	$\wedge$ asszociativitása
$((a \vee b) \vee c) \equiv (a \vee (b \vee c))$	$\vee$ asszociativitása
$\neg(\neg a) \equiv a$	dupla negálás eliminálása
$(a \to b) \equiv (\neg b \to \neg a)$	kontrapozició
$(a \to b) \equiv (\neg a \vee b)$	**implikáció elimináció**
$(a \Leftrightarrow b) \equiv ((a \to b) \wedge (b \to a))$	ekvivalencia eliminálás
$\neg(a \wedge b) \equiv (\neg a \vee \neg b)$	De Morgan
$\neg(a \vee b) \equiv (\neg a \wedge \neg b)$	De Morgan
$(a \wedge (b \vee c)) \equiv ((a \wedge b) \vee (a \wedge c))$	$\wedge$ disztributivitása $\vee$ felett
$(a \vee (b \wedge c)) \equiv ((a \vee b) \wedge (a \vee c))$	$\vee$ $\wedge$ disztributivitása $\wedge$ felett

két állítás logikailag ekvivalens, ha ugyanazokban a modellekben igaz:

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

CNF = diszjunkciók konjunkciója, **DNF** = konjunkciók diszjunkciója

Bizonyítások

Modell-ellenőrzés – 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 **megszokott** keresési algoritmusokban.

- keresés általános operátorokkal (természetes dedukció, algoritmusa?)
- 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)

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

Keresés modellek terében - visszalépéses keresés Davis-Putnam, stb.

Keressük pl. a TB = SA $\wedge$ SB $\wedge$ SC

= $\neg$B $\wedge$ C $\wedge$ (C $\wedge$ ($\neg$B $\vee$ $\neg$A)) $\wedge$ ($\neg$A $\rightarrow$ $\neg$C)

modelljét (ahol a TB kielégíthető, azaz igaz lesz)

$\neg$B $\wedge$ C $\wedge$ ($\neg$B $\vee$ $\neg$A) $\wedge$ ($\neg$A $\rightarrow$ $\neg$C) igazához szükséges, hogy $\neg$B legyen igaz, azaz **B** legyen **hamis**

C $\wedge$ ($\neg$A $\rightarrow$ $\neg$C) igazához szükséges, hogy **C** legyen **igaz**

$\neg$A $\rightarrow$ Hamis igazához szükséges, hogy $\neg$A legyen hamis, azaz **A** legyen **igaz**

(Vajon más kombinációkra is?)

$C1 : a \vee b$
$C2 : c \vee d$
$C3 : a \vee e \vee f$
$C4 : \neg b \vee \neg f \vee g$
$C5 : \neg f \vee h$
$C6 : \neg b \vee \neg h \vee i$
$C7 : \neg g \vee \neg i$

$\underline{a}$, a = 0

$C1 : a \vee b$
$C2 : c \vee d$
$C3 : a \vee e \vee f$
$C4 : \neg b \vee \neg f \vee g$
$C5 : \neg f \vee h$
$C6 : \neg b \vee \neg h \vee i$
$C7 : \neg g \vee \neg i$

b, b = 1

$C1 : a \vee b$
$C2 : c \vee d$
$C3 : a \vee e \vee f$
$C4 : \neg b \vee \neg f \vee g$
$C5 : \neg f \vee h$
$C6 : \neg b \vee \neg h \vee i$
$C7 : \neg g \vee \neg i$

$\underline{c}$, c = 0

$C1 : a \vee b$
$C2 : c \vee d$
$C3 : a \vee e \vee f$
$C4 : \neg b \vee \neg f \vee g$
$C5 : \neg f \vee h$
$C6 : \neg b \vee \neg h \vee i$
$C7 : \neg g \vee \neg i$

d, d = 1

$C1 : a \vee b$
$C2 : c \vee d$
$C3 : a \vee e \vee f$
$C4 : \neg b \vee \neg f \vee g$
$C5 : \neg f \vee h$
$C6 : \neg b \vee \neg h \vee i$
$C7 : \neg g \vee \neg i$

$\underline{e}$, e = 0

$C1 : a \vee b$
$C2 : c \vee d$
$C3 : a \vee e \vee f$
$C4 : \neg b \vee \neg f \vee g$
$C5 : \neg f \vee h$
$C6 : \neg b \vee \neg h \vee i$
$C7 : \neg g \vee \neg i$

f, f = 1

$C1 : a \vee b$
$C2 : c \vee d$
$C3 : a \vee e \vee f$
$C4 : \neg b \vee \neg f \vee g$
$C5 : \neg f \vee h$
$C6 : \neg b \vee \neg h \vee i$
$C7 : \neg g \vee \neg i$

g, g = 1

$C1 : a \vee b$
$C2 : c \vee d$
$C3 : a \vee e \vee f$
$C4 : \neg b \vee \neg f \vee g$
$C5 : \neg f \vee h$
$C6 : \neg b \vee \neg h \vee i$
$C7 : \neg g \vee \neg i$

TB = **klózok halmaza**, klóz egy CNF

i, i = 0

$C1: a \vee b$
$C2: c \vee d$
$C3: a \vee e \vee f$
$C4: \neg b \vee \neg f \vee g$
$C5: \neg f \vee h$
$C6: \neg b \vee \neg h \vee i$
$C7: \neg g \vee \neg i$

h, h = 1

$C1: a \vee b$
$C2: c \vee d$
$C3: a \vee e \vee f$
$C4: \neg b \vee \neg f \vee g$
$C5: \neg f \vee h$
$C6: \neg b \vee \neg h \vee i$
$C7: \neg g \vee \neg i$

ellentmondás

Vissza, e = 1, legközelebbi szabad elágazás

Keresés modellek terében - lokális keresés, WalkSAT stb.

Változók véletlen 0/1 hozzárendelése (egy modell).
Egy véletlen klóz a modellben hamis klózokból.
Egy véletlen változó a klózból.
P valószínűséggel a változó átbillenése a modellben (új modell, de az eddig teljesített klózokból néhányan kieshetnek).
Különben a klóz annak a változójának az átbillenése, ami a teljesített klózok számát maximálja.

… (szimulált lehűtés jellegű)

Korlátkielégítési probléma

Korlátok*:*

$C1 : a \vee b = 1$

$C2 : c \vee d = 1$

$C3 : a \vee e \vee f = 1$

$C4 : \neg b \vee \neg f \vee g = 1$

$C5 : \neg f \vee h = 1$

$C6 : \neg b \vee \neg h \vee i = 1$

$C7 : \neg g \vee \neg i = 1$

Változók*:*

$a = ?,\ b = ?,\ c = ?,\ d = ?,\ e = ?,\ f = ?,\ g = ?,\ h = ?,\ i = ?$

Értéktartomány = $\{0, 1\}$

…

Az ítéletlogika következtetési mintái (operátorok)

Modus Ponens
(Implikáció eliminálása)
néha elég is

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

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

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

stb.

Az ítéletlogika következtetési mintái (operátorok)

Minden a gépnek, semmit az embernek

Elemi (egység)**rezolúció**

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

Rezolúció

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

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)
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 - a '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 a legesélyesebb)

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 (a feladatmegoldáskor)

milyen állapotban van valójában a rendszer,
ha konkrét információt kapok a viselkedéséről?

ha János influenzás, akkor János lázas
János lázas
?

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

ha INFLUENZÁS, akkor LÁZAS

LÁZAS

INFLUENZÁS?

ha LÁZAS, akkor INFLUENZÁS

LÁZAS

INFLUENZÁS?

Analógia (már a középkori arabok, józan ész, tudomány, ...)

Néhány objektum néhány közös tulajdonsága alapján egyéb tulajdonságok közös meglétére következtetni.

$$\frac{\begin{array}{l}\text{a-tulajdonság}(P) \wedge \text{a-tulajdonság}(Q)\\ \text{b-tulajdonság}(P) \wedge \text{b-tulajdonság}(Q)\\ \text{c-tulajdonság}(P) \wedge \text{c-tulajdonság}(Q)\\ \text{x-tulajdonság}(P)\end{array}}{\text{x-tulajdonság}(Q)}$$

P olyan, mint Q, biztos? valószínű?

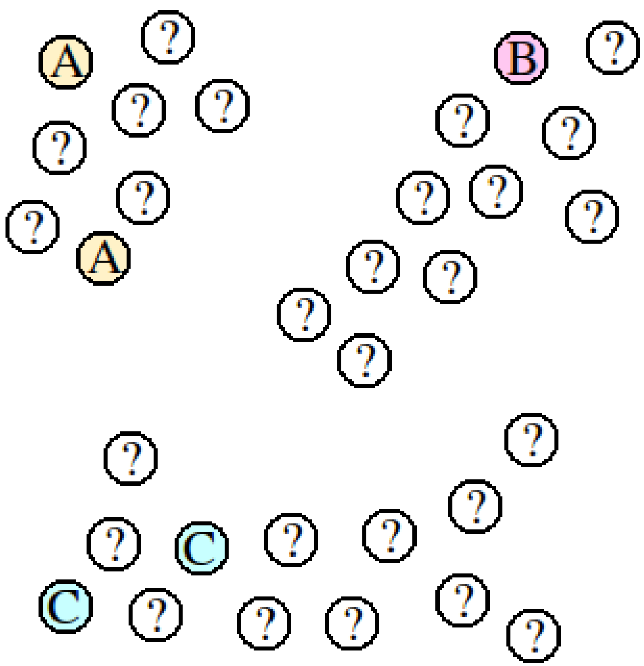
Jancsi olyan, mint egy medve (nagy, lassú, szeret mézet, …, téli álmot alszik?)

Transzdukció (Vladimir Vapnik, 1990-es évek)

Megfigyelt esetekről következtetés más esetekre

Mi a kérdőjeles objektumok címkéje?
Indukcióhoz kevés a példa
Legközelebbi szomszéd?

Transzdukció nem csak példákat veszi alapul, a kérdéses pontokat is, de a predikcióra nem képes



Particiós transzdukció

(a) Az egész ponthalmaz egy nagy partíció.

(b) Ha egy partíció tartalmaz nem azonos címkéjű pontokat: bontsuk fel kisebb partíciókra

(c) Minden partícióra: ugyanaz a címke minden pontra.

Az ítéletlogika következtetési mintái – mind dedukciók!

(pl. Modus Ponens)

$(A \wedge (A \rightarrow B)) \rightarrow B$

$\neg(A \wedge (A \rightarrow B)) \vee B$

$\neg A \vee \neg(A \rightarrow B) \vee B$

$\neg A \vee \neg(\neg A \vee B) \vee B$

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

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

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

$\text{Igaz} \wedge \text{Igaz} = \text{Igaz}$

És igazságtáblával?

És mit kapnánk egy abduktív séma esetében?

Az ítéletlogikai következtetés komplexitása

Ítéletlogika eldönthető és teljes!

Minden jól definiált mondat akár **igaz**, akár **hamis** volta belátható véges lépésszámú 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 **exponenciális** n-ben és így gyakorlatilag kivitelezhetetlen.
Cook (1971):

egy mondathalmaz kielégíthetőségi vizsgálata **NP-teljes**

(de nem minden példánya az ítélet logikai következtetésnek
2^n-nel arányos időt igényel! Ld. előbbi példa.)

A logika monotonitása

Tegyük fel: a tudásbázis maga után vonz egy mondathalmazt.

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.

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

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

Jó: amit egyszer nehéz volt bebizonyítani, majd „ingyen van”.
Rossz: a változó világ logikai leírása (axiómák) 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 ráfizetünk.

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)

mondatok egy hasznos osztálya, amelyre létezik a TB méretében **lineáris idejű** következtetési eljárás.

két fontos speciális eset:

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

(2) Igaz $\rightarrow Q$ ua., mint Q.

Nem minden TB-beli állítás írható fel Horn mondatként (**példa**?), de azoknál, amelyeknél ez megtehető: alkalmazzuk Modus Ponens-t, ameddig marad alkalmazható következtetés.

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

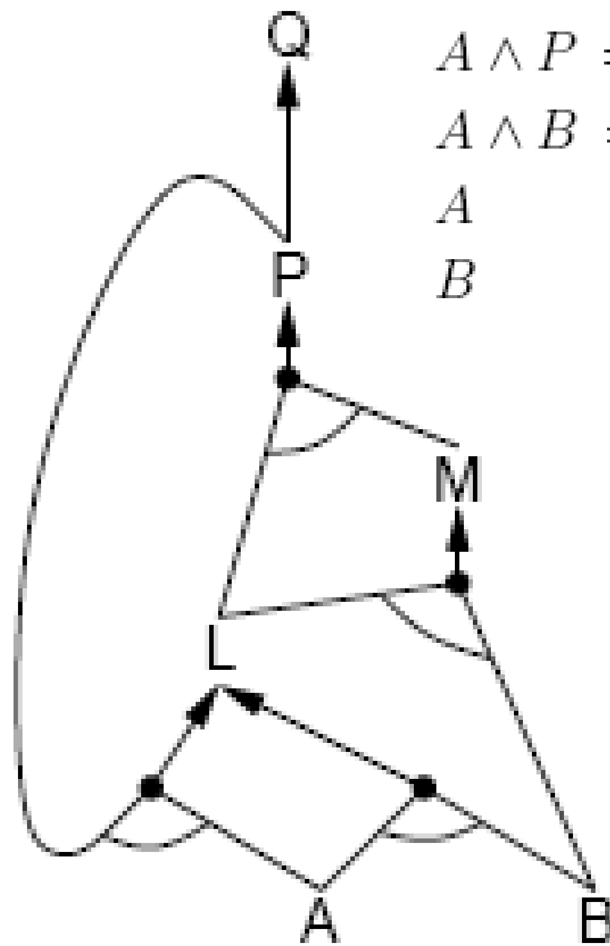
Előrecsatolt, hátracsatolt (Modus Ponens) következtetés

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

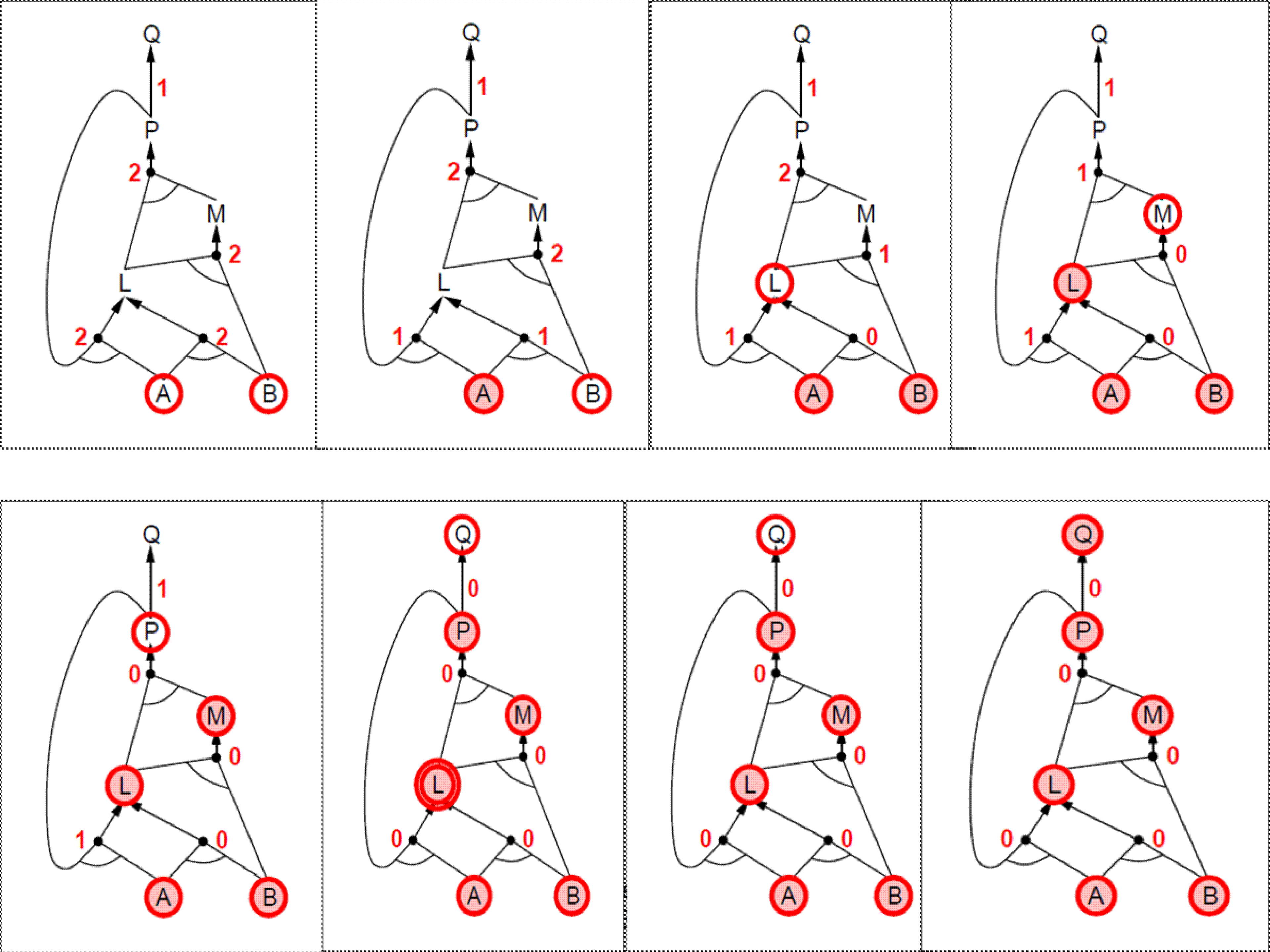
$$P \Rightarrow Q$$
$$L \wedge M \Rightarrow P$$
$$B \wedge L \Rightarrow M$$
$$A \wedge P \Rightarrow L$$
$$A \wedge B \Rightarrow L$$
$$A$$
$$B$$

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

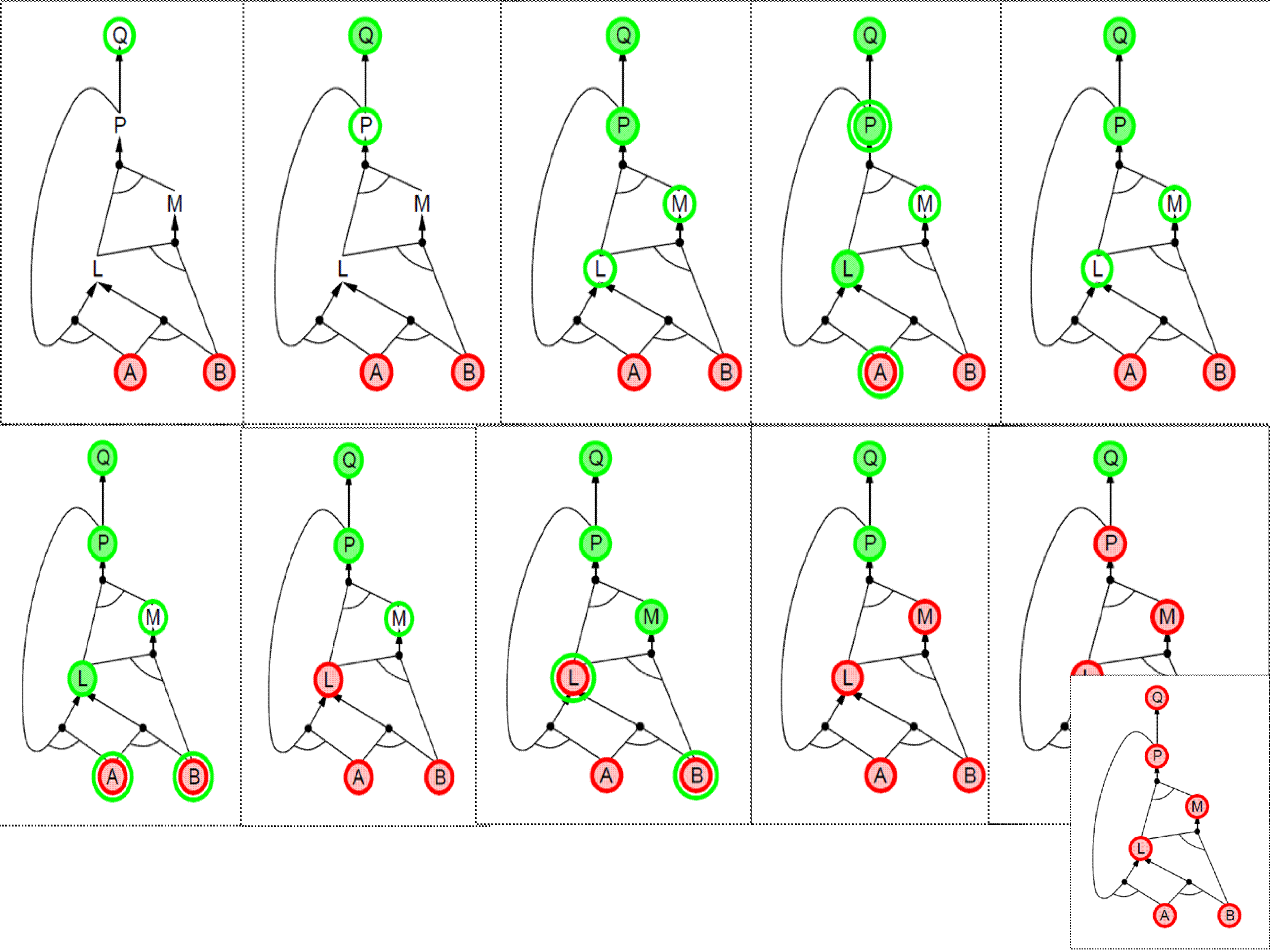
Hurkok kerülése: ellenőrzés, vajon új részcél már benne van a veremben.
Ismételt munka kerülése: ellenőrzés, vajon az új részcél
1) már bizonyított igaznak, vagy
2) kudarcba fulladt.



És-Vagy gráf



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 ágens a Wumpus világ számára

Tudásbázis ciklus: ágens érzetei $\rightarrow$ mondatok $\rightarrow$ 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.

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

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

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

.....

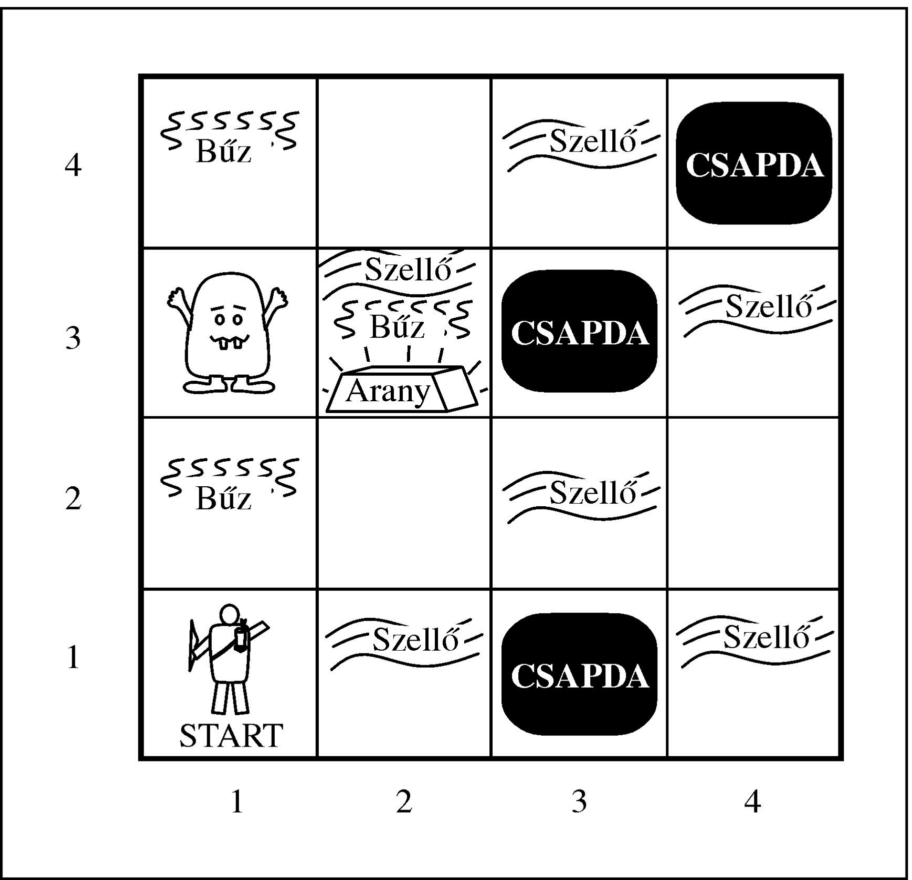
ha bűz van az [1,2]-ben, akkor egy Wumpus van ott, vagy egy vagy több szomszédos négyzetben:

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

A Wumpus megtalálása - az ágensnek el kellene készítenie a **TB → $W_{1,3}$** 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)



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

$$\Downarrow$$

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

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

2. És-eliminálás: $\neg \mathbf{W}_{1,1}$, $\neg \mathbf{W}_{1,2}$, $\neg \mathbf{W}_{2,1}$ (miért kell?)

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: $\mathbf{W}_{1,3} \vee \mathbf{W}_{1,2} \vee \mathbf{W}_{2,2} \vee \mathbf{W}_{1,1}$

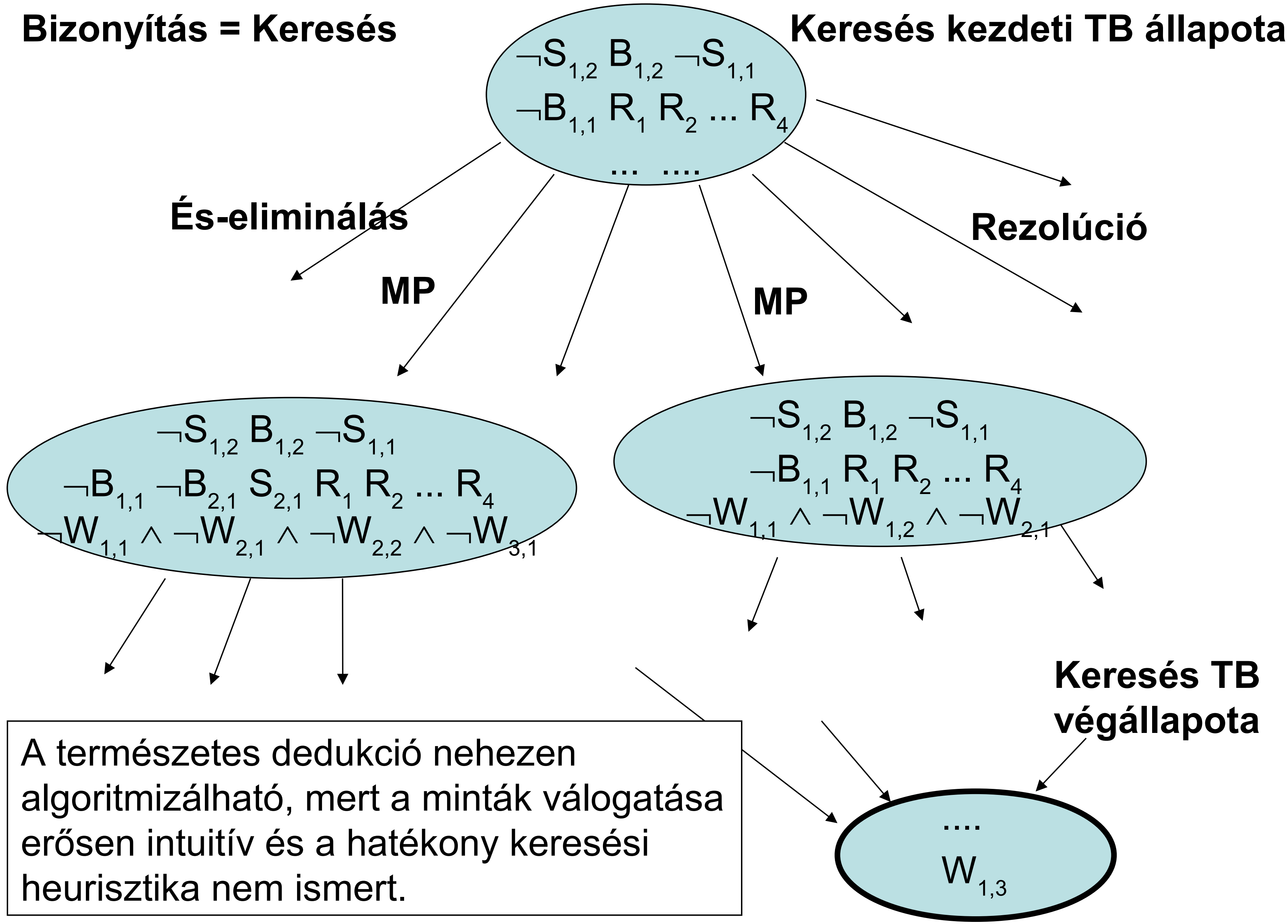
5. Egység rezolúció: $\mathbf{W}_{1,3} \vee \mathbf{W}_{1,2} \vee \mathbf{W}_{2,2} \vee \mathbf{W}_{1,1}$ és $\neg \mathbf{W}_{1,1}$
   az eredmény: $\mathbf{W}_{1,3} \vee \mathbf{W}_{1,2} \vee \mathbf{W}_{2,2}$

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.

Rezolúció (1963)

Orvoslás Gödel teljességi tételére és ad hoc term. dedukcióra.
Klózokkal dolgozik (vigyázz, ezek nem Horn-klózok!).
Átalakítás klóz formára (CNF + $\wedge$ eliminálás = csak $\neg$ és $\vee$ marad)
Rezolúció „cáfolat-teljes” lépés.
Teljes kereséssel párosítva teljes bizonyítás.

$$\frac{\begin{array}{c} B \vee A \\ \neg B \vee G \end{array}}{A \vee G}$$

Rezolúciós bizonyítás:

1. TB átírása az un. klóz formába.
2. Q kérdés negálása.
3. Negált Q kérdés átírása klóz formára.
4. Kiterjesztett TB’ = TB $\cup$ $\neg$Q tudásbázis létesítése.
5. Abban ciklikus rezolúciós lépés elvégzése:
   - üres rezolvens – kilép, TB’ ellentmondás, TB igaz, akkor Q igaz
   - a rezolvens nem üres, vissza 5.

(ha minden párosítás megvizsgálva, kilépés)

Transzformáció klóz formára:

1. Ekvivalencia el: $A \Leftrightarrow B = (A \to B) \wedge (B \to A)$
2. Implikáció el: $A \to B = \neg A \vee B$
3. Negálás atomi formulák szintjére (de Morgan).

$$\neg (A \vee B) = \neg A \wedge \neg B$$

4. Diszjunkciók literálok szintjére (disztributivítás).

$$(A \wedge B) \vee C = (A \vee C) \wedge (B \vee C)$$

konjunktív normal forma (CNF, Davis, 1960)

5. Konjunkciók el. Bontás diszjunktív klózokra
   (csak negálás és diszjunkció marad)

Példa:
$\mathbf{S \vee T \to Q}$
$\neg (S \vee T) \vee Q$
$(\neg S \wedge \neg T) \vee Q$
$(\neg S \vee Q) \wedge (\neg T \vee Q)$
$\mathbf{\neg S \vee Q}$
$\mathbf{\neg T \vee Q}$

Az eredeti (redundáns) állításforma és a (redundancia mentes) klózforma logikailag ekvivalensek!

TB=axiómák TB’

eredeti állítások	klózok
P	P
$P \wedge Q \rightarrow R$	$\neg P \vee \neg Q \vee R$
$S \vee T \rightarrow Q$	$\neg S \vee Q$
	$\neg T \vee Q$
T	T
R igaz-e ?	
TB |- R ?	$\neg R$
TB’ |- $\varnothing$?	

rezolúció menete

$\neg P \vee \neg Q \vee R$ $\quad \neg R$

$\neg P \vee \neg Q$ $\quad P$

$\neg S \vee Q$ $\quad \neg Q$ $\quad \neg T \vee Q$

$\neg S$ $\quad \neg T$ $\quad T$

üres klóz $\varnothing$

Egyszerű rezolúciós bizonyítás ítéletkalkulusban

(axiómarendszer!?)

A lényeg:

Egy feltehetően igaz állítást negálunk.

Konzisztens, hibátlannak tartott tudásbázishoz hozzáadjuk.
(a tudásbázis az **elvárásunk szerint most már nem konzisztens**!).

Rezolúcióval az **üres rezolvens** létezését kimutatjuk.

üres klóz = ellentmondás

$$\frac{\begin{array}{c}\neg P \\ P\end{array}}{\text{üres}}$$

Az **ellentmondás formálisan** (interpretációtól függetlenül) **ismerhető fel** (más – kielégíthető - eset sajnos nem!)

Problémák az ítéletlogikai ágenssel

Ítéletlogika ábrázoló ereje olyan gyenge, nem képes olyan „egyszerű” feladat, mint a Wumpus világ kezelésére sem.

Fő probléma: **túl sok ítéletszimbólumot kell kezelni**.
Egy szabály: „ne menj egyenesen előre, ha a Wumpus előtted van”

ítéletlogika = 64 szabály (16 négyzet x 4 irány).

Szabályok ezreire van szükség: - szabályok leírása?
- **lassú a következtetési eljárás**.
(igazságtábla mérete)

Másik probléma a **változások kezelése**.

Egy adott időpontban: a TB összes mondata igaz abban a pillanatban.
Amikor az ágens lép, az $\mathbf{A_{1,1}}$ állítás hamissá válik és az $\mathbf{A_{2,1}}$ igazzá.
Lehet, hogy fontos emlékezni, volt korábban: nem felejtheti el az $\mathbf{A_{1,1}}$-et
(pl. $\mathbf{A_{1,1,0}}$-ként).

Különböző ítéletszimbólumokra lesz szükségünk minden egyes időlépésben:
- nem tudjuk, hogy a játék **meddig fog tartani**.
- újra kell írni minden szabályt **időfüggő változatban**.