

Τεχνητή Νοημοσύνη

9η διάλεξη (2024-25)

Ίων Ανδρουτσόπουλος

<http://www.aueb.gr/users/ion/>

Οι διαφάνειες αυτής της διάλεξης βασίζονται εν μέρει στο βιβλίο *Artificial Intelligence – A Modern Approach* των S. Russel και P. Norvig, 2^η και 4^η έκδοση, Prentice Hall, 2003 και 2020.

Τι θα ακούσετε σήμερα

- Εξαγωγή συμπερασμάτων με προτάσεις Horn προτασιακής λογικής.
- Πρωτοβάθμια κατηγορηματική λογική.
 - Συντακτικό ΠΚΛ.
 - Παράσταση γνώσεων με ΠΚΛ.

Προτάσεις Horn προτασιακής λογικής

- **Πρόταση Horn ΠΛ:** $(l_1 \vee \dots \vee l_k)$
 - Κάθε l_i είναι σύμβολο (π.χ. P) ή άρνηση συμβόλου (π.χ. $\neg P$).
 - Το πολύ ένα από τα l_i δεν είναι άρνηση συμβόλου («θετικό»).
 - Π.χ. $(\neg L_{1,1} \vee \neg Breeze \vee B_{1,1})$, $(\neg W_{1,1} \vee \neg W_{1,2})$, $B_{1,1}$
- Μετατρέπεται ισοδύναμα σε «κανόνα» $((\dots \wedge \dots \wedge \dots) \Rightarrow l_i)$
 - $((L_{1,1} \wedge Breeze) \Rightarrow B_{1,1})$, $((W_{1,1} \wedge W_{1,2}) \Rightarrow False)$, $(True \Rightarrow B_{1,1})$
 - Αν δεν έχει **κανένα αρνητικό** σύμβολο, «γεγονός» (fact).
 - Αν **δεν έχει θετικό** σύμβολο, «περιορισμός ακεραιότητας».
 - Δε θα ασχοληθούμε με περιορισμούς ακεραιότητας.
 - Αν **έχει (ένα) θετικό** σύμβολο, «**οριστική πρόταση**» (definite clause). (Τα γεγονότα είναι και αυτά οριστικές προτάσεις.)
- **ΒΓ προτάσεων Horn ΠΛ:**
 - Περιέχει μόνο προτάσεις Horn ΠΛ.
 - Άρα χρησιμοποιούμε **υποσύνολο** της προτασιακής λογικής.

Εξαγωγή συμπερασμάτων με προτάσεις Horn ΠΛ

- Η εξαγωγή συμπερασμάτων μπορεί να γίνει σε **χρόνο γραμμικό** προς το μέγεθος της ΒΓ.
 - Πιο εύκολη από ό,τι αν χρησιμοποιούμε την πλήρη προτασιακή λογική και τον κανόνα της ανάλυσης.
 - Όμως χρησιμοποιούμε υποσύνολο της προτασιακής λογικής.
- Εξαγωγή συμπερασμάτων **προς τα εμπρός** (forward chaining).
 - Διαδοχική «πυροδότηση» (κατά την κατεύθυνση των βελών) κανόνων των οποίων οι υποθέσεις αληθεύουν (με Modus Ponens), μέχρι να καταλήξουμε στο επιθυμητό συμπέρασμα.
- Εξαγωγή συμπερασμάτων **προς τα πίσω** (backward chaining).
 - Χρήση κανόνων κατά την αντίστροφη κατεύθυνση των βελών, ξεκινώντας από το συμπέρασμα μέχρι να καταλήξουμε σε υποθέσεις που γνωρίζουμε (υπάρχουν στη ΒΓ).
- Η βάση του **λογικού προγραμματισμού**.
 - Π.χ. τα προγράμματα **Prolog** είναι συλλογές προτάσεων Horn (αλλά πρωτοβάθμιας κατηγορηματικής λογικής).

Γράφοι σύζευξης-διάζευξης (AND-OR)

ΒΓ σε μορφή προτάσεων
Horn ΠΛ:

$$P \Rightarrow Q$$

$$L \wedge M \Rightarrow P$$

$$B \wedge L \Rightarrow M$$

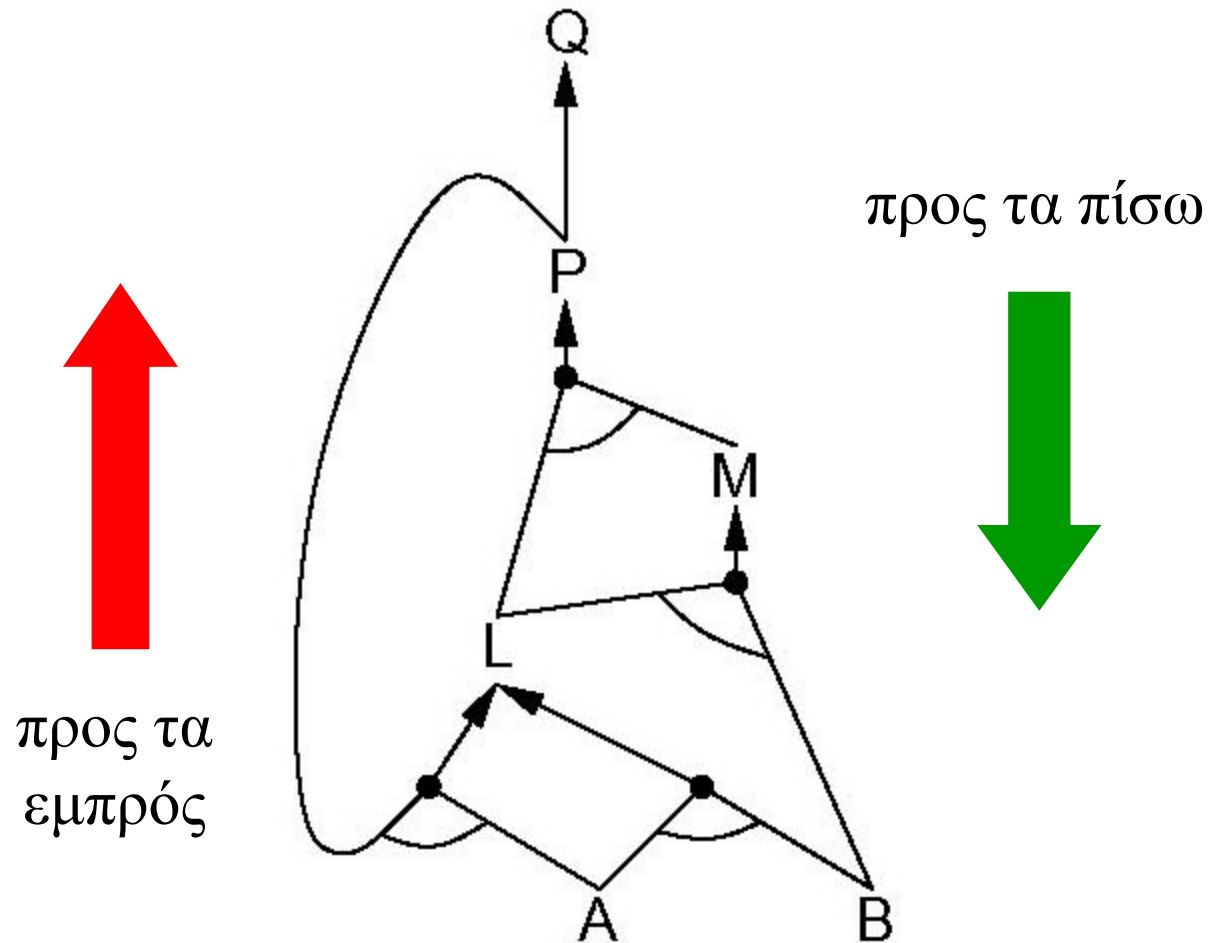
$$A \wedge P \Rightarrow L$$

$$A \wedge B \Rightarrow L$$

A

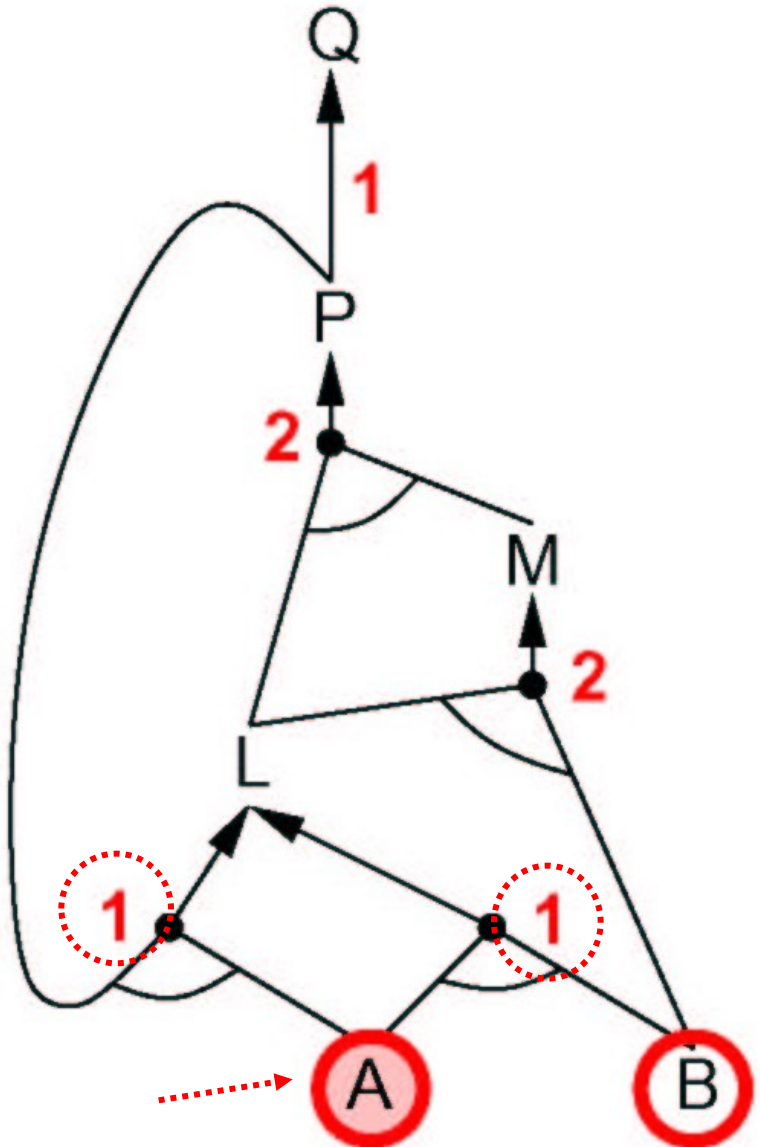
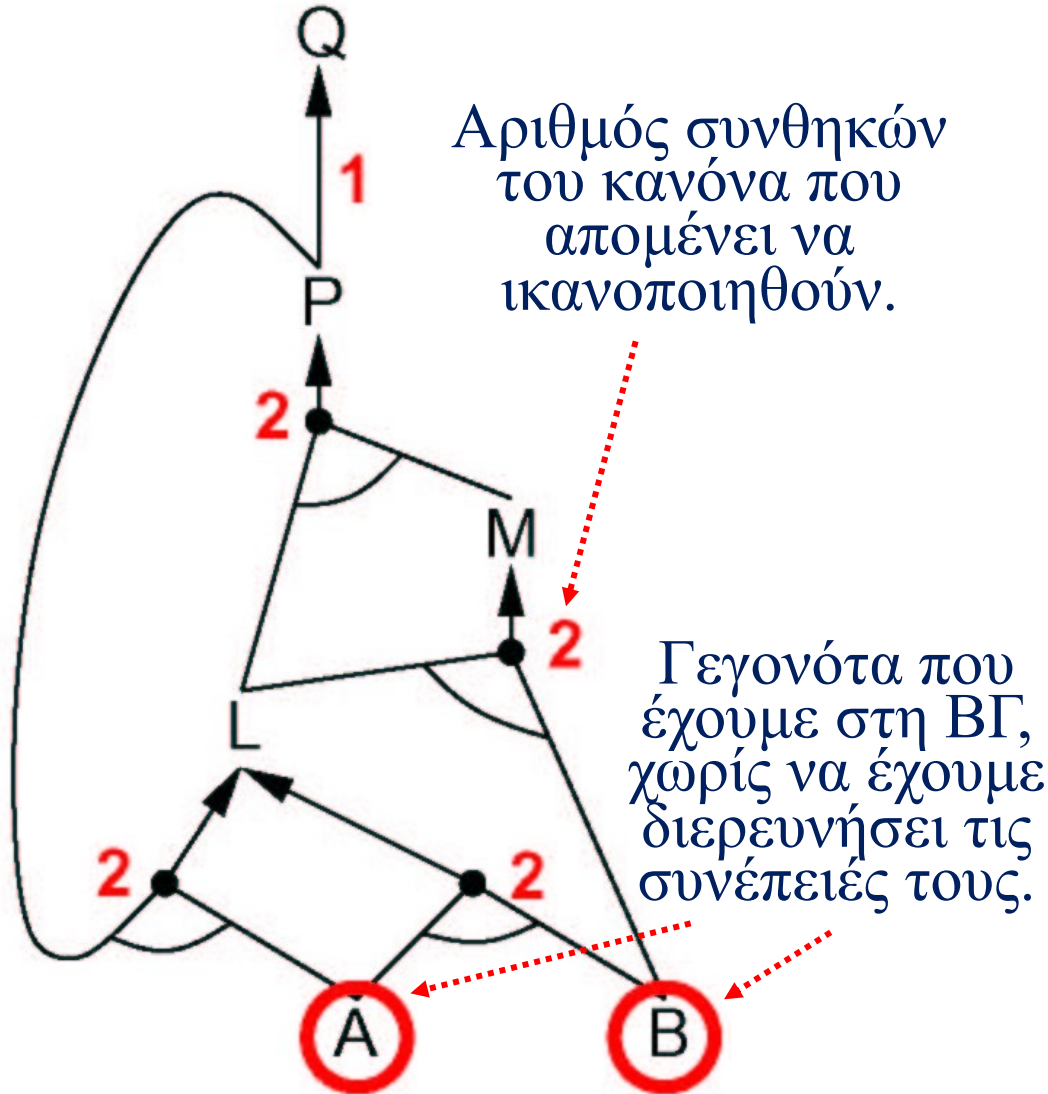
B

ισοδύναμος γράφος AND-OR:

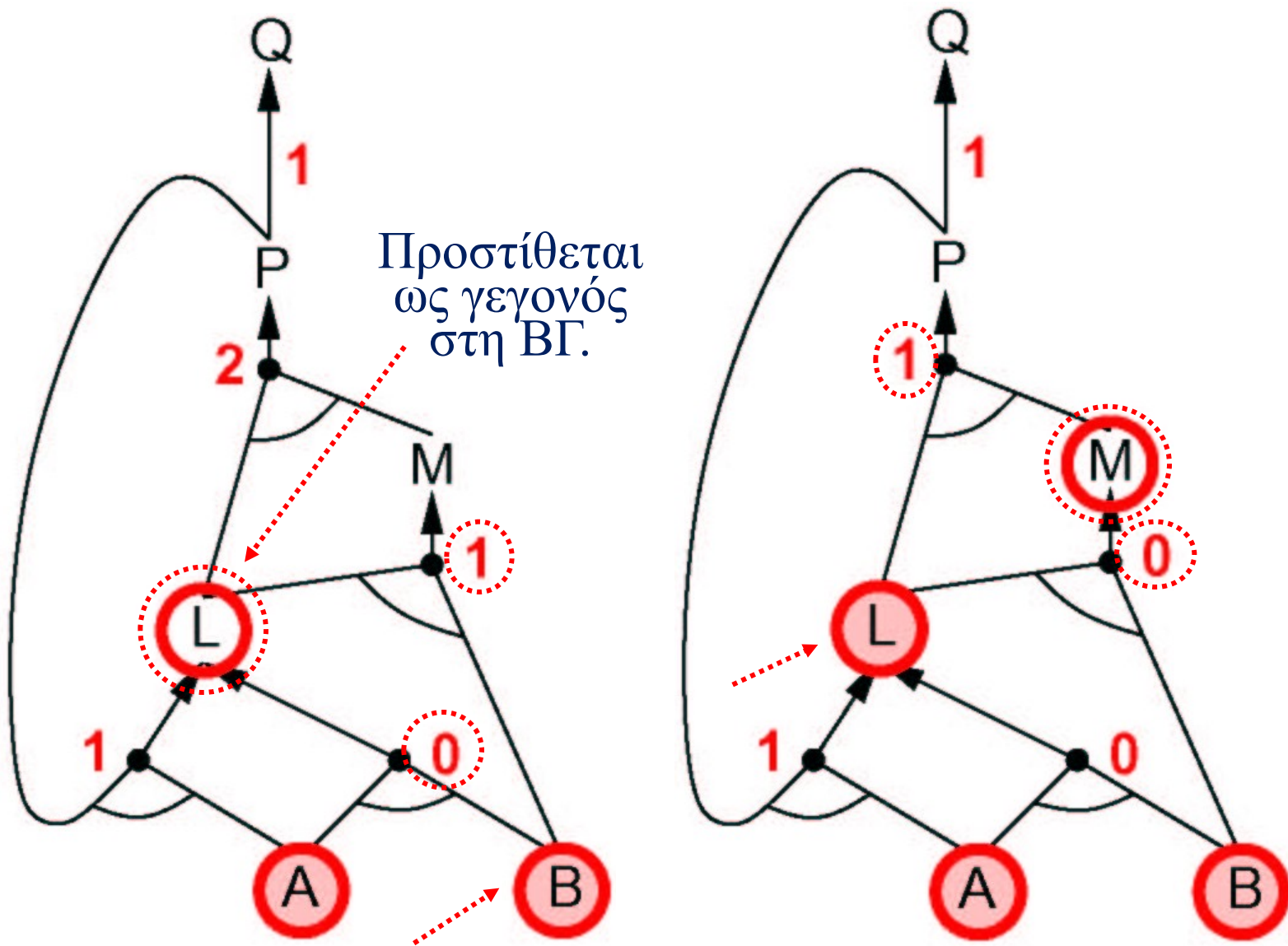


Εξαγωγή συμπεράσματος προς τα εμπρός

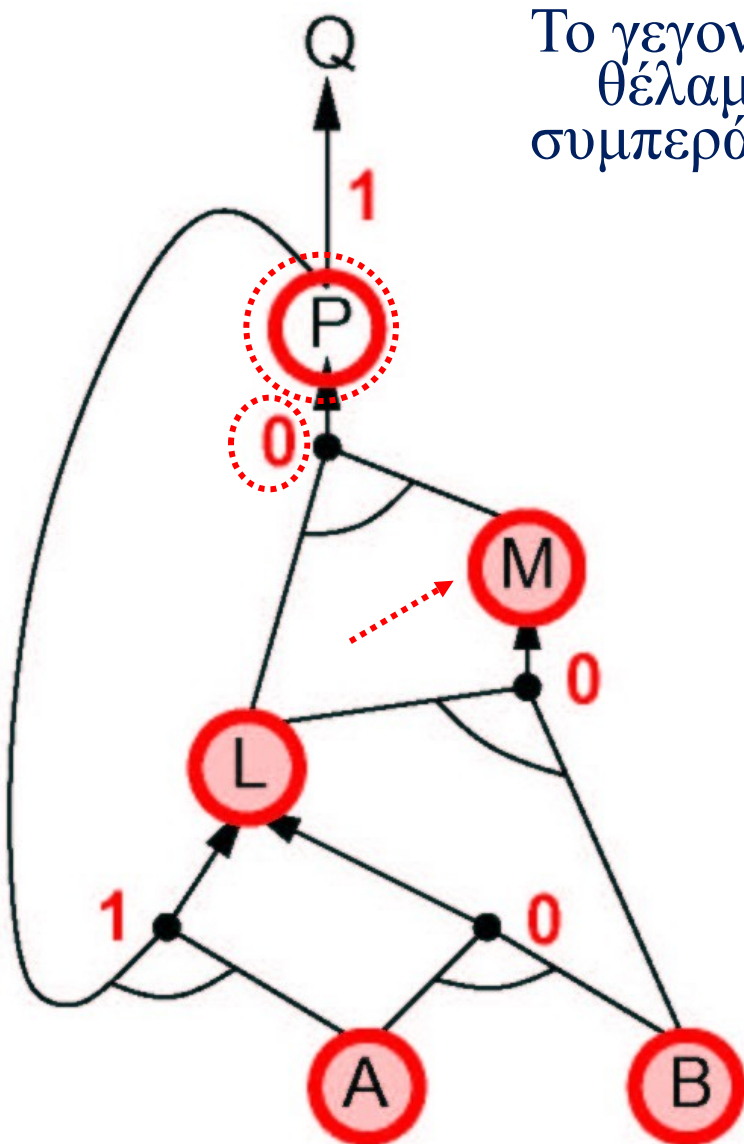
Θέλουμε να δούμε αν μπορούμε να συμπεράνουμε Q από τη ΒΓ.



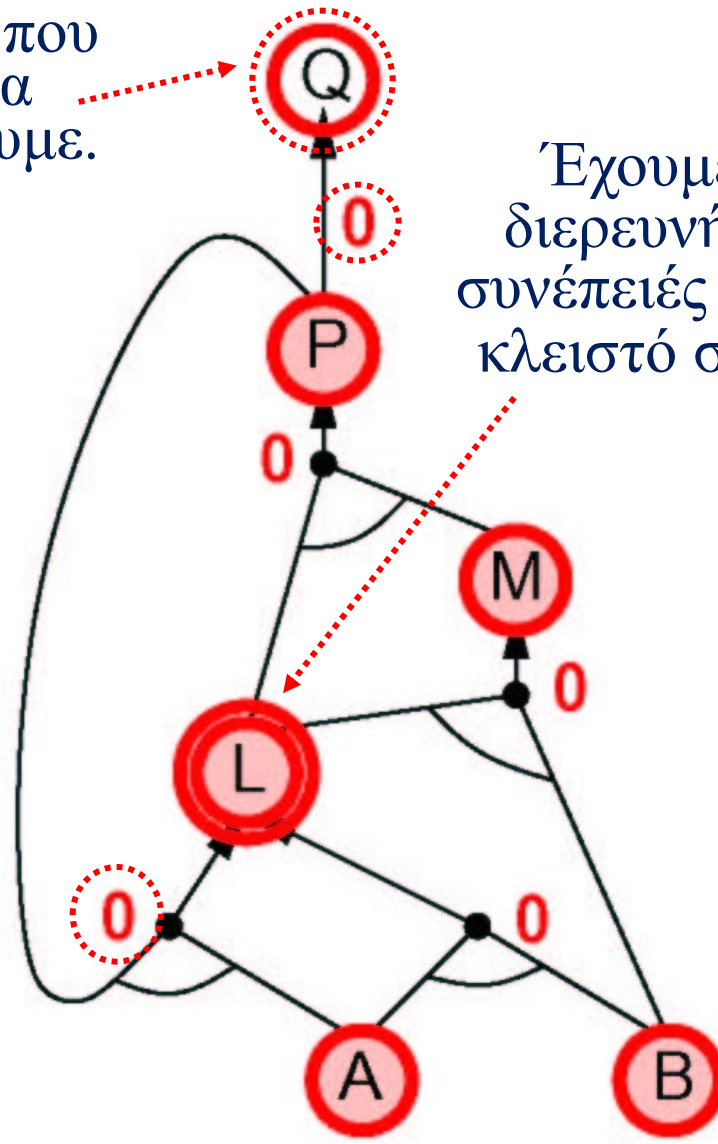
Συνέχεια εξαγωγής συμπεράσματος προς τα εμπρός



Συνέχεια εξαγωγής συμπεράσματος προς τα εμπρός



Το γεγονός που θέλαμε να συμπεράνουμε.



Έχουμε ήδη διερευνήσει τις συνέπειές του (π.χ. κλειστό σύνολο).

Εξαγωγή συμπερασμάτων (με ΒΓ προτάσεων Horn προτασιακής λογικής) προς τα εμπρός

function PL-FC-ENTAILS?(*KB, q*) **returns** *true* or *false*

local variables: *count*, a table, indexed by clause, initially the number of premises

inferred, a table, indexed by symbol, each entry initially *false*

agenda, a list of symbols, initially the symbols known to be true

while *agenda* is not empty **do**

Γεγονότα που έχουμε στη ΒΓ χωρίς να έχουμε διερευνήσει τις συνέπειές τους (μέτωπο).

$p \leftarrow \text{POP}(\textit{agenda})$

unless *inferred*[*p*] **do**

inferred[*p*]: Δείχνει αν έχουμε συμπεράνει το γεγονός *p* ή όχι.

inferred[*p*] \leftarrow *true*

for each Horn clause *c* in whose premise *p* appears **do**

decrement *count*[*c*]

if *count*[*c*] = 0 **then do**

Πυροδότηση κανόνα.

if HEAD[*c*] = *q* **then return** *true*

PUSH(HEAD[*c*], *agenda*)

return *false*

Χαρακτηριστικά του PL-FC-Entails?

- **Τερματίζει.**
 - Η ατζέντα μικραίνει κατά ένα σε κάθε επανάληψη του while.
 - Μεγαλώνει όταν πυροδοτείται ένας κανόνας.
 - Κάθε κανόνας πυροδοτείται το πολύ μία φορά (όταν μηδενιστεί ο μετρητής του) και έχουμε πεπερασμένο αριθμό κανόνων.
 - Άρα τελικά η ατζέντα αδειάζει και ο αλγόριθμος τερματίζει.
- **Χρονική πολυπλοκότητα: γραμμική ως προς το μέγεθος της ΒΓ.**
 - Η απόδειξη παραλείπεται.
- **Ορθότητα:** Χρησιμοποιούμε μόνο Modus Ponens, που είναι ορθός κανόνας εξαγωγής συμπερασμάτων.
- **Πληρότητα** (αν $B\Gamma \models q$, τότε $B\Gamma \vdash_i q$, για q γεγονός).
 - Η απόδειξη παραλείπεται.

Μειονεκτήματα προτασιακής λογικής

- Δύσκολη παράσταση γενικεύσεων.
 - Π.χ. «Υπάρχει ρεύμα αέρος σε ένα τετράγωνο αν υπάρχει όρυγμα σε διπλανό τετράγωνο».
 - Χρειαζόμαστε μια διαφορετική πρόταση για κάθε τετράγωνο:
 $(B_{2,2} \Leftrightarrow (P_{2,1} \vee P_{2,3} \vee P_{1,2} \vee P_{3,2}))$
 $(B_{3,2} \Leftrightarrow (P_{3,1} \vee P_{3,3} \vee P_{2,2} \vee P_{4,2}))$ κ.ο.κ.
- Ενώ εύκολη παράσταση στην **πρωτοβάθμια κατηγορηματική λογική (ΠΚΛ)**.
 - Ένας μόνο τύπος για όλα τα μη ακραία τετράγωνα:
 $\forall x \forall y (B(x, y) \Leftrightarrow$
 $(P(x, y-1) \vee P(x, y+1) \vee P(x-1, y) \vee P(x+1, y)))$
 - Και ειδικοί τύποι για τα ακραία τετράγωνα.

Παραδείγματα τύπων ΠΚΛ

- Σε όλες τις γάτες αρέσει το γάλα.

$$\forall x (\text{IsCat}(x) \Rightarrow \text{Likes}(x, \text{Milk}))$$

- Υπάρχει μια γάτα που της αρέσει το γάλα.

$$\exists x (\text{IsCat}(x) \wedge \text{Likes}(x, \text{Milk}))$$

- Προσοχή: ο τύπος $\exists x (\text{IsCat}(x) \Rightarrow \text{Likes}(x, \text{Milk}))$ λέει «Υπάρχει ένα x που: (i) δεν είναι γάτα ή (ii) του αρέσει το γάλα».

- Η Ψίτα συμπαθεί όλους τους σκύλους.

$$\forall x (\text{IsDog}(x) \Rightarrow \text{Likes}(\text{Psita}, x))$$

- Προσοχή: ο τύπος $\forall x (\text{IsDog}(x) \wedge \text{Likes}(\text{Psita}, x))$ λέει «Τα πάντα είναι σκύλοι και αρέσουν στην Ψίτα».

- Υπάρχει μια γάτα που συμπαθεί όλους τους σκύλους.

$$\exists x (\text{IsCat}(x) \wedge \forall y (\text{IsDog}(y) \Rightarrow \text{Likes}(x, y)))$$

Συμβουλή: μην
παραλείπετε ποτέ
παρενθέσεις!

Παραδείγματα τύπων ΠΚΛ – συνέχεια

- *Ο Μίλος αντιπαθεί όλες τις γάτες.*

$$\forall x (\text{IsCat}(x) \Rightarrow \neg \text{Likes}(\text{Milos}, x))$$

- *Όλοι οι σκύλοι αντιπαθούν όλες τις γάτες.*

$$\forall x (\text{IsDog}(x) \Rightarrow \forall y (\text{IsCat}(y) \Rightarrow \neg \text{Likes}(x, y)))$$

ή ισοδύναμα:

$$\forall x \forall y ((\text{IsDog}(x) \wedge \text{IsCat}(y)) \Rightarrow \neg \text{Likes}(x, y))$$

- *Κάθε άνθρωπος συμπαθεί τον πατέρα του.*

$$\forall x \forall y ((\text{IsHuman}(x) \wedge \text{IsFatherOf}(y, x)) \Rightarrow \text{Likes}(x, y))$$

$$\text{ή: } \forall x (\text{IsHuman}(x) \Rightarrow \text{Likes}(x, \mathbf{\text{FatherOf}(x)}))$$

Συναρτησιακός
όρος

Παραδείγματα τύπων ΠΚΛ – συνέχεια

- Κάθε σκύλος που γαβγίζει φοβάται μια (πιθανώς διαφορετική ή την ίδια) γάτα.

$$\forall x ((\text{IsDog}(x) \wedge \text{Barks}(x)) \Rightarrow$$

$$\exists y (\text{IsCat}(y) \wedge \text{IsAfraidOf}(x, y)))$$

- Κάθε γάτα συμπαθεί ακριβώς έναν (πιθανώς διαφορετικό ή τον ίδιο όλες) σκύλο.

$$\forall y (\text{IsCat}(y) \Rightarrow$$

$$\exists x (\text{IsDog}(x) \wedge \text{Likes}(y, x) \wedge$$

$$\forall z ((\text{IsDog}(z) \wedge \text{Likes}(y, z)) \Rightarrow z = x)))$$

Συντακτικό ΠΚΛ

τύπος \rightarrow ατομικός_τύπος

| (τύπος σύνδεσμος τύπος)

| ποσοδείκτης μεταβλητή τύπος

| \neg τύπος

ατομικός_τύπος \rightarrow σύμβολο_σχέσης(όρος, ...) | όρος = όρος

όρος \rightarrow σταθερά | μεταβλητή |

σύμβολο_συνάρτησης(όρος, ...)

σύνδεσμος \rightarrow \wedge | \vee | \Rightarrow | \Leftrightarrow

ποσοδείκτης \rightarrow \forall | \exists

σταθερά \rightarrow A | X₁ | John | Mary | ...

μεταβλητή \rightarrow a | x | s | ...

σύμβολο_σχέσης \rightarrow IsFatherOf | HasColor | IsKing | ...

σύμβολο_συνάρτησης \rightarrow FatherOf | LeftLeg | ...

Τα σύνολα των σταθερών, μεταβλητών, συμβόλων σχέσεων, συμβόλων συναρτήσεων θεωρούμε ότι είναι ανά δύο ξένα.

Βιβλιογραφία

- Russel & Norvig (4^η έκδοση): ενότητες 7.5.3, 7.5.4, εισαγωγή ενότητας 8.1.
 - Όσοι ενδιαφέρονται μπορούν να διαβάσουν προαιρετικά και τις υπόλοιπες ενότητες του κεφαλαίου 7.
- Βλαχάβας κ.ά: εισαγωγή ενότητας 9.2 (χωρίς την υπο-ενότητα 9.2.1).