

Προγραμματισμός Σημασιολογικού Ιστού

Ενότητα 4: Από το μοντέλο EAV στους γράφους

Μ.Στεφανιδάκης

6-3-2018

Από την τελευταία μας άσκηση...

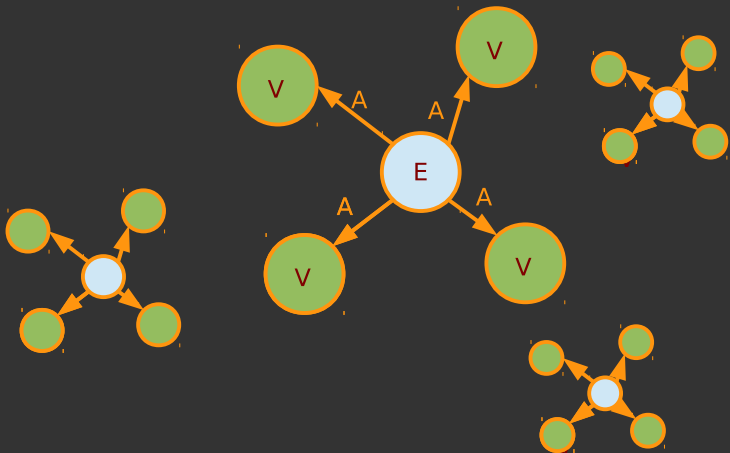
- ▶ Καταλήξαμε σε μια λίστα τριάδων (s, p, o)
 - ▶ Η πληροφορία του ωρολογίου προγράμματος

```
[ [0, 'Ημέρα', 'Δευτέρα'],  
  [0, 'Έναρξη', '18:00'],  
  [0, 'Λήξη', '21:00'],  
  [0, 'Μάθημα', 'Εξόρυξη δεδομένων και διαχείριση γνώσης'],  
  # .. κλπ..  
  [1, 'Ημέρα', 'Τρίτη'],  
  [1, 'Έναρξη', '15:00'],  
  [1, 'Λήξη', '18:00'],  
  [1, 'Μάθημα', 'Προγραμματισμός Σηματολογικού Ιστού'],  
  # .. κλπ..  
]
```

- ▶ Το πρώτο μέρος (id διάλεξης) είναι κοινό ανά ομάδες τριάδων!
 - ▶ Πώς εκφράζεται σχηματικά;

Το μοντέλο EAV σχηματικά

Τα δεδομένα έχουν τη μορφή μεμονωμένων “νιφάδων” (ή “αστέρα”) με το id της διάλεξης στο κέντρο



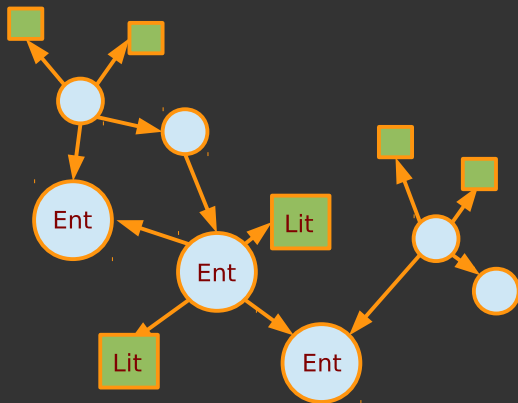
Είναι όμως πράγματι έτσι;

Οι τιμές (values) στο μοντέλο EAV

- ▶ Σημασιολογικά, όλες οι τιμές (V) δεν είναι ίδιες
 - ▶ Στο παράδειγμα του ωρολογίου προγράμματος
- ▶ Υπάρχουν τιμές που θα μπορούσαν να είναι επίσης “οντότητες” (entities);
 - ▶ Με τις δικές τους ιδιότητες και τιμές;
 - ▶ Τα Μαθήματα, οι Αίθουσες, οι Διδάσκοντες...
- ▶ Σε αντίθεση με
 - ▶ Τις Ημέρες, τις Ώρες (και ίσως το Εξάμηνο)...
 - ▶ Τα τελευταία είναι απλές τιμές (literals)

Ο μετασχηματισμός σε γράφο

- ▶ Θεωρώντας έναν κόμβο ανά οντότητα (και literal)
- ▶ Ο γράφος περιγράφει τις σχέσεις μεταξύ οντοτήτων



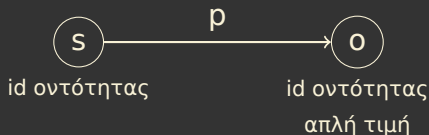
Η ιδέα είναι παλιά: βλ. “semantic networks” της Τεχνητής Νοημοσύνης (δεκαετίες 50-60)

Πώς αναγνωρίζουμε τους κόμβους;

- ▶ Έμμεσα παραδεχόμαστε ότι το ίδιο “όνομα” (αλφαριθμητικό αναγνωριστικό) αναφέρεται στην ίδια οντότητα
 - ▶ Διαφορετικές εμφανίσεις του ίδιου ονόματος συγχωνεύονται σε έναν μοναδικό κόμβο
- ▶ Η μέθοδος εφαρμόζεται όσο θεωρούμε ότι τα δεδομένα μας είναι μοναδικά στον κόσμο
 - ▶ Μη ρεαλιστική υπόθεση, θα ασχοληθούμε αργότερα με το θέμα αυτό...
- ▶ Προσοχή: για τις απλές τιμές (**literals**), το ίδιο αλφαριθμητικό δεν σημαίνει απολύτως τίποτα
 - ▶ Π.χ. δύο εμφανίσεις του literal “Πέμπτη” δεν συγχωνεύονται σε μοναδικό κόμβο στον γράφο!

Οι τριάδες ξανά – με άλλο όνομα

- ▶ Ισοδύναμο με EAV μοντέλο για την περιγραφή της οργάνωσης των δεδομένων
 - ▶ Χρήση **τριάδων** για την περιγραφή **γράφων δεδομένων**
 - ▶ Κάθε τριάδα αποτελεί μια “δήλωση” (statement) πληροφορίας
 - ▶ **Subject – Predicate – Object** ή απλά **(s,p,o)**
 - ▶ πολύ κοντά στην απλή φυσική μορφή “υποκείμενο – ρήμα – αντικείμενο”



Σχετικά με τη φορά των ακμών p

- ▶ Ο γράφος που παράγουν οι τριάδες (s,p,o) είναι **κατευθυνόμενος**
 - ▶ Πώς διαλέγουμε τη φορά;
- ▶ Εξαρτάται από τις ανάγκες της εφαρμογής μας!
 - ▶ Λειτουργικά, η τριάδα **(Διδάσκων X , διδάσκει, Μάθημα Y)** είναι ισοδύναμη με την **(Μάθημα Y , διδάσκεται_από, Διδάσκων X)**
 - ▶ Αφήνοντας κατά μέρος δικαιολογίες όπως “το διδάσκει φαίνεται στην εφαρμογή μου καλύτερα από το διδάσκεται_από”...
- ▶ Προσοχή: για τις απλές τιμές (**literals**), δεν έχουμε επιλογές
 - ▶ Εμφανίζονται πάντα στη θέση **o** (object) (ως στόχος της ακμής p)!

Ανώνυμοι κόμβοι (blank nodes)

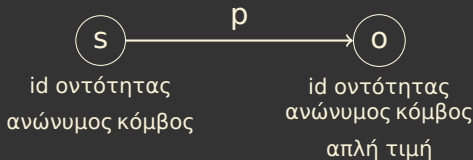
- ▶ Κάθε οντότητα χαρακτηρίζεται από ένα μοναδικό αναγνωριστικό
 - ▶ Θα χρησιμοποιηθεί και στα ερωτήματα
 - ▶ “Τί ξέρω για την οντότητα X;”
- ▶ Υπάρχουν όμως οντότητες “εσωτερικές” σε κάθε μοντέλο οργάνωσης δεδομένων
 - ▶ Βρίσκονται εκεί απλά και μόνο για να διασυνδέουν άλλες οντότητες
 - ▶ Δεν υπάρχει περίπτωση να είναι ο (κύριος) στόχος μιας ερώτησης
 - ▶ Το αναγνωριστικό τους δεν θα μεταδοθεί ποτέ “προς τα έξω”
 - ▶ το αναγνωριστικό αυτό έχει τοπική (local) μόνο σημασία
- ▶ Οι κόμβοι των εσωτερικών αυτών οντοτήτων ονομάζονται **ανώνυμοι** (blank nodes)

Παράδειγμα ανώνυμων κόμβων

- ▶ Η κομβική οντότητα **Διάλεξη** στο παράδειγμα του ωρολογίου προγράμματος
 - ▶ Υπάρχει για να διασυνδέει Μαθήματα, Αίθουσες, Διδάσκοντες, Ημέρες, Ώρες και Εξάμηνο
 - ▶ Δεν θα υπάρξει ερώτηση ειδικά για μία συγκεκριμένη διάλεξη
 - ▶ Παρά μόνο στα πλαίσια μιας ερώτησης σχετικής με τις άλλες οντότητες
 - ▶ Το (τεχνητό) αναγνωριστικό των διαλέξεων (0, 1, 2, 3...) δεν έχει ιδιαίτερο νόημα εκτός της εφαρμογής μας
- ▶ Συνεπώς, οι κόμβοι των διαλέξεων είναι ένα τυπικό παράδειγμα **ανώνυμων κόμβων!**

Η θέση των ανώνυμων κόμβων στις τριάδες

- ▶ Οι **ανώνυμοι κόμβοι** αναγνωρίζουν οντότητες
 - ▶ Έστω και με “αδύναμο” τρόπο
 - ▶ Με τοπική μόνο σημασία
- ▶ Συνεπώς μπορούν να εμφανιστούν σε θέση υποκειμένου (subject) ή αντικειμένου (object)



Το τελευταίο κομμάτι στο παζλ: το κατηγορημα

- ▶ Το **κατηγορημα** (predicate) (δεύτερο μέρος τριάδων)
 - ▶ Είναι κι αυτό ένα αναγνωριστικό όνομα (id)
 - ▶ Το id αντιπροσωπεύει **μονοσήμαντα** το κατηγορημα
 - ▶ Το είδος της **σχέσης** μεταξύ υποκειμένου και αντικειμένου

