

Σημασιολογική Ολοκλήρωση Δεδομένων με τη χρήση Οντολογιών



Λίνα Μπουντούρη - Μανόλης Γεργατσούλης
Ιόνιο Πανεπιστήμιο
15ο Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών



Διαδίκτυο και Επίπεδα ετερογένειας δεδομένων

- Διαδίκτυο: Βασικός φορέας πληροφόρησης, Όγκος Δεδομένων, Ποικίλες ανάγκες πληροφόρησης
- Ετερογένεια Συστημάτων
- Ετερογένεια στη Σύνταξη
- Ετερογένεια Σχημάτων
- Σημασιολογική Ετερογένεια



Προβλήματα στην πρόσβαση



Συστήματα Ολοκλήρωσης Δεδομένων

- Ενιαία πρόσβαση σε συλλογές αυτόνομων πηγών ως ενιαίο σύνολο
 - Σχήμα διαμεσολάβησης ή καθολικό σχήμα (mediated schema or global schema)
- Σχέσεις (Views) ανάμεσα στις τοπικές πηγές και το καθολικό σχήμα
 - Global – As – View (GAV)
 - Local – As – View (LAV)



Συστήματα Ολοκλήρωσης Δεδομένων

- Κανόνες Συσχέτισης (Mapping Rules): σχέση ανάμεσα σε δύο σχήματα (πηγαίο και τελικό σχήμα)
- Θέματα προς εξέταση
 - **Ορισμός σχέσεων ανάμεσα στα τοπικά σχήματα και το καθολικό σχήμα**
 - Μετασχηματισμός των ερωτημάτων από το καθολικό σχήμα προς τα τοπικά σχήματα



Ερώτημα χρήστη

Σχήμα Διαμεσολάβησης

Μετασχηματισμός
ερωτήματος προς
την τοπική πηγή



Τοπική Πηγή

Τοπική Πηγή

Τοπική Πηγή

Ενοποίηση
αποτελεσμάτων



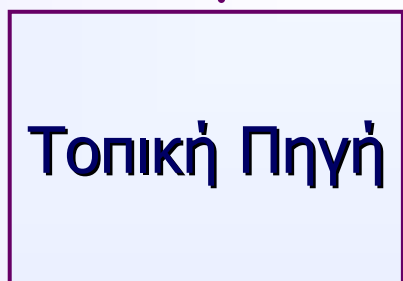
Απάντηση τοπικής
πηγής



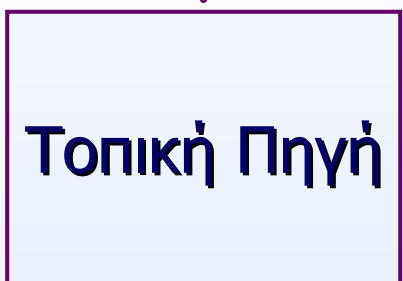
Απάντηση στο χρήστη



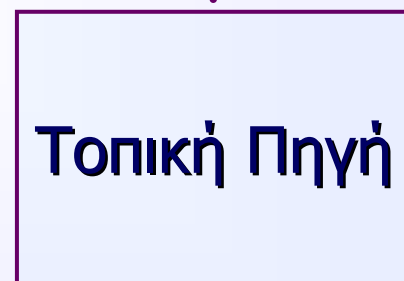
Σχήμα Διαμεσολάβησης



Τοπική Πηγή



Τοπική Πηγή



Τοπική Πηγή





Σημασιολογική Ολοκλήρωση Δεδομένων

- Παγκόσμιος Ιστός → Σημασιολογικός Ιστός
- Ορισμός: *Διαδικασία της χρήσης εννοιολογικών αναπαραστάσεων των δεδομένων και των σχέσεών τους με στόχο της εξάλειψης της σημασιολογικής ετερογένειας*
- Προβλήματα σημασιολογικής ετερογένειας σε επίπεδο σχημάτων (schema level) και σε επίπεδο δεδομένων (data level)



Οντολογίες

- Ορισμός: *Μία τυπική (formal), κατηγορηματική (explicit) προδιαγραφή μίας διαμοιρασμένης (shared) εννοιολογικής αναπαράστασης (conceptualization)*
- Μηχανισμός διαλειτουργικότητας
- Εννοιολογική αναπαράσταση ενός τομέα



Οι Οντολογίες στη Σημασιολογική Ολοκλήρωση Δεδομένων

- Οντολογία ως σχήμα διαμεσολάβησης
 - Μπορεί να ορίσει πολύπλοκες σχέσεις ενός θεματικού πεδίου
 - Αυστηρό φορμαλισμό → εξαγωγή συμπερασμάτων
 - Ανάπτυξη οντολογιών για ένα θεματικό χώρο ("*Domain Ontologies*"), π.χ. CIDOC και ABC για πολιτιστικό περιεχόμενο



Οι Οντολογίες στη Σημασιολογική Ολοκλήρωση Δεδομένων

■ Προσέγγιση Μονής Οντολογίας

- Καθολική οντολογία ως σχήμα διαμεσολάβησης, η οποία συνήθως περιγράφει ένα θεματικό τομέα

■ Προσέγγιση Πολλαπλών Οντολογιών

- Κάθε τοπική πηγή περιγράφεται από μία ξεχωριστή τοπική οντολογία

■ Προσέγγιση Υβριδικής Οντολογίας

- Κάθε μία πηγή περιγράφεται από μία ξεχωριστή τοπική οντολογία της οποίας η δημιουργία έχει βασιστεί στις βασικές έννοιες μίας καθολικής οντολογίας



Κανόνες συσχέτισης

- Τα δεδομένα συνήθως δεν είναι δομημένα σύμφωνα με τις γλώσσες οντολογιών
- Μεγάλος όγκος δεδομένων σε γλώσσα XML
- Ανάγκη για κανόνες συσχέτισης ανάμεσα σε XML και γλώσσες οντολογιών
 - «Διαδικαστική» προσέγγιση: Μετατροπή XML δεδομένων σε κλάσεις και ιδιότητες μίας οντολογίας
 - «Δηλωτική» προσέγγιση: Ορισμός σχέσεων ανάμεσα σε XML δεδομένα και οντολογίες, δίχως να υπάρχει κατά ανάγκη μετασχηματισμός



Σημασιολογική Διαλειτουργικότητα Μεταδεδομένων

- Αρχεία, Βιβλιοθήκες και Μουσεία: διαθέτουν ετερογενές υλικό
- Ανάγκη για περιγραφή και τεκμηρίωση ποικίλων τύπων τεκμηρίων: χρήση συγκεκριμένων σχημάτων περιγραφής ανάλογα με τον τύπο του υλικού
- Ανάγκη αναζήτησης – ανάκτησης σε ετερογενή μεταδεδομένα
 - Μετατροπή από ένα σχήμα μεταδεδομένων σε ένα άλλο
 - Συσχέτιση σχημάτων μεταδεδομένων σε ένα καθολικό σχήμα



Προβλήματα στους κανόνες συσχέτισης

- Ορισμός ποικίλων σημασιολογικών σχέσεων μέσα από έναν κανόνα
 - Συνωνυμία
 - 1: Πολλά
 - 1: Κανένα
 - Δομή – Ιεραρχία
 - Δομή – Σειρά
- Διαφορετική λογική μοντέλων XML και Γλωσσών Οντολογιών (π.χ. RDF)



Προβλήματα στους κανόνες συσχέτισης

- Περιγραφική λογική των Μεταδεδομένων
 - Για ένα ή πολλά αντικείμενα
 - Επιμέρους περιορισμούς (π.χ. υποχρεωτικά / προαιρετικά πεδία, επαναλαμβανόμενα πεδία κτλ)
 - Δύσκολη η αποτύπωση όλων των πληροφοριών σε ένα εννοιολογικό μοντέλο



Σχετικές εργασίες

- Laks V.S. Lakshmanan και Fereidoon Sadri
 - Μοντέλο συσχέτισης που βασίζεται στην XPath
- Isabel F. Cruz
et al. (The University of Illinois at Chicago)
 - Μετατροπή XML δεδομένων σε τοπικές RDF οντολογίες και συσχέτισή τους με μία καθολική RDF οντολογία
- B. Amann et al.
 - Κανόνες συσχέτισης ανάμεσα σε κλάσεις και ιδιότητες της οντολογίας και σε XML διαδρομές
 - Βασίζεται στην XPath



Μελλοντικές κατευθύνσεις

- Σημασιολογικές σχέσεις που πρέπει να οριστούν μέσα από τους κανόνες συσχέτισης
- Επιλογή ή σχεδιασμός γλώσσας για την απόδοση των κανόνων