



Ειδικά Θέματα Βάσεων Δεδομένων

Ενότητα 13: Σύγχρονα θέματα βάσεων δεδομένων

Δρ. Τιμπίρης Αλκιβιάδης

Τμήμα Μηχανικών Πληροφορικής ΤΕ



Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο ΤΕΙ Κεντρικής Μακεδονίας» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ενότητα 13

Σύγχρονα θέματα βάσεων δεδομένων

Δρ. Τσιμπίρης Αλκιβιάδης

Περιεχόμενα ενότητας

- Εισαγωγή στις έννοιες ΒΔ πολυμέσων
- Χαρακτηριστικά των τύπων πηγής πολυμέσων
- Πηγή πολυμέσου-Εικόνα
- Πηγή πολυμέσου-Βίντεο
- Πηγή πολυμέσου-Κειμένου
- Πηγή πολυμέσου-Ήχου
- Χωρικές βάσεις δεδομένων
- Κατηγορίες τυπικών χωρικών επερωτήσεων
- Γεωγραφικά συστήματα πληροφοριών
- Απαιτήσεις διαχείρισης δεδομένων για τα ΓΣΠ
- ARC-INFO

Σκοποί ενότητας

Στην ενότητα αυτή γίνεται μία γενική εισαγωγή στις έννοιες των χωρικών και πολυμεσικών βάσεων δεδομένων που αποτελούν έναν τομέα σύγχρονο με μεγάλο ενδιαφέρον και προοπτική.

Εισαγωγή στις έννοιες ΒΔ πολυμέσων

Οι βάσεις δεδομένων πολυμέσων παρέχουν χαρακτηριστικά που επιτρέπουν στους χρήστες να αποθηκεύουν και να ζητούν διευκρινήσεις για διαφορετικούς τύπους πληροφοριών πολυμέσων που περιλαμβάνουν **εικόνες** (όπως φωτογραφίες ή σχέδια), **ακολουθίες βίντεο** (ταινίες ειδήσεις κλπ), **ακολουθίες ήχου** (όπως τραγούδια, τηλεφωνικά μηνύματα, ή αγορεύσεις) και **έγγραφα** (όπως βιβλία και άρθρα).

Εισαγωγή στις έννοιες ΒΔ πολυμέσων

Οι βασικοί τύποι ικανοποιητικών ερωτήσεων (**εδώ έγραφε επερωτήσεις**) βάσης δεδομένων που απαιτούνται περιλαμβάνουν τον εντοπισμό των πηγών πολυμέσων οι οποίες περιέχουν κάποια αντικείμενα που σας ενδιαφέρουν. Οι τύποι αυτοί αναφέρονται σαν **ανάκτηση με βάση το περιεχόμενο**.

Επομένως μια ΒΔ πολυμέσων **πρέπει** να χρησιμοποιεί κάποιο μοντέλο για την οργάνωση και τα ευρετήρια των πηγών πολυμέσων που θα βασίζεται στα περιεχόμενα τους. Ο προσδιορισμός των περιεχομένων των πηγών πολυμέσων είναι μια δύσκολη και χρονοβόρα προσπάθεια.

Εισαγωγή στις έννοιες ΒΔ πολυμέσων

Υπάρχουν **δύο βασικές προσεγγίσεις**:

- Βασίζεται στην **αυτόματη ανάλυση των πηγών πολυμέσων** για τον προσδιορισμό κάποιων μαθηματικών χαρακτηριστικών του περιεχομένου τους. Η προσέγγιση αυτή χρησιμοποιεί διαφορετικές τεχνικές ανάλογα με τον τύπο της πηγής του πολυμέσου (εικόνα, κείμενο, βίντεο ή ήχος).
- Η δεύτερη προσέγγιση εξαρτάται από τον **χειρονακτικό προσδιορισμό των αντικειμένων** και των δραστηριοτήτων που ενδιαφέρουν σε κάθε πηγή πολυμέσων και στην χρησιμοποίηση αυτών των πληροφοριών για τη δημιουργία ευρετηρίων. Η προσέγγιση αυτή μπορεί να εφαρμοσθεί σε όλες τις διαφορετικές πηγές πολυμέσων, αλλά απαιτεί μια φάση χειρονακτικής προεπεξεργασίας όπου κάποιο άτομο πρέπει να σαρώσει κάθε πηγή πολυμέσου για να προσδιορίσει και να καταγράψει τα αντικείμενα και τις δραστηριότητες που περιέχει ώστε να μπορούν να χρησιμοποιηθούν για εύρεση αυτών των πηγών

Χαρακτηριστικά των τύπων πηγής πολυμέσων

Χαρακτηριστικά κάθε τύπου πηγής πολυμέσου (εικόνες, βίντεο, ήχος και πηγές κειμένων)

Εικόνα αποθηκεύεται σε μη επεξεργασμένη μορφή σαν ένα σύνολο από τιμές πλέγματος η κελιών, ή σε συμπιεσμένη μορφή για εξοικονόμηση χώρου. Η περιγραφή σχήματος της είναι ένα ορθογώνιο από κελιά συγκεκριμένου πλάτους και ύψους. Κάθε εικόνα μπορεί να αναπαρασταθεί από ένα m επί n πλέγμα από κελιά. Κάθε κελί περιέχει την τιμή ενός εικονοστοιχείου που περιγράφει το περιεχόμενο του κελιού.

Μαυρόασπρες εικόνες \rightarrow 1bit

Γκρίζα κλίμακα ή στις έγχρωμες \rightarrow πολλά bits

Σχόλιο! Οι εικόνες απαιτούν μεγάλο χώρο και αποθηκεύονται σε συμπιεσμένη μορφή.

Πηγή πολυμέσου-Εικόνα

Δύο βασικές τεχνικές τύπων αναζήτησης εικόνας:

- **Συνάρτηση απόστασης** συγκρίνει τη δοθείσα εικόνα με τις αποθηκευμένες εικόνες και τα τμήματά τους. Αν η τιμή της απόστασης που επιστρέφεται είναι μικρή, τότε η πιθανότητα εύρεσης του προτύπου είναι μεγάλη. Μπορούν να δημιουργηθούν ευρετήρια που να ομαδοποιούν τις αποθηκευμένες εικόνες με περίπου ίδια μετρική απόσταση έτσι που να περιορίζεται ο χώρος της αναζήτησης.
- **Προσέγγιση μετασχηματισμού** μετράει την ομοιότητα των εικόνων εκτελώντας ένα μικρό αριθμό μετασχηματισμών οι οποίοι μπορούν να μετασχηματίσουν τα κελιά μιας εικόνας ώστε να καταφέρουν να τα ταιριάξουν με την άλλη εικόνα. Οι μετασχηματισμοί περιλαμβάνουν στροφές, μεταφράσεις και κλιμάκωση.

Πηγή πολυμέσου-Βίντεο

Ένα βίντεο αναπαρίσταται από μια ακολουθία από καρτέ, όπου κάθε καρτέ είναι μια παγωμένη εικόνα. **Το βίντεο:**

- Τοποθετεί τα αντικείμενα και τις δραστηριότητες σε κάθε ξεχωριστό καρτέ
- Χωρίζεται σε τμήματα, όπου κάθε τμήμα αποτελείται από μια ακολουθία διαδοχικών καρτέ που περιέχουν τα ίδια αντικείμενα και τις ίδιες δραστηριότητες
- Κάθε τμήμα προσδιορίζεται από το αρχικό και το τελικό καρτέ.
- Τα αντικείμενα και οι δραστηριότητες που προσδιορίζονται σε κάθε τμήμα βίντεο μπορούν να χρησιμοποιηθούν για την ευρετηριοποίηση των τμημάτων.

Πηγή πολυμέσου-Βίντεο

Για την ευρετηριοποίηση των βίντεο έχει προταθεί μια τεχνική ευρετηριοποίησης που ονομάζεται **δένδρα τμημάτων καρέ**. Το ευρετήριο περιλαμβάνει τόσο αντικείμενα, όπως άτομα, σπίτια, αυτοκίνητα, όσο και δραστηριότητες, όπως ένα άτομο που κάνει μια ομιλία ή δύο άτομα που συζητούν.

Πηγή πολυμέσου-Πηγή κειμένου

Πηγή κειμένου: είναι το πλήρες κείμενο κάποιου άρθρου, βιβλίου ή περιοδικού. Για την ευρετηριοποίηση των πηγών χρησιμοποιείται ο προσδιορισμός των λέξεων-κλειδιών που εμφανίζονται στο κείμενο καθώς και των σχετικών συχνοτήτων τους. Για τον σκοπό αυτό χρησιμοποιείται μια τεχνική που βασίζεται σε μετασχηματισμούς πινάκων και ονομάζεται **διασπάσεις μοναδικών τιμών (singular value decompositions (SVD))**. Στη συνέχεια μπορεί να χρησιμοποιηθεί μια τεχνική ευρετηριοποίησης που ονομάζεται **δένδρα τηλεσκοπικών διανυσμάτων ή TV-δένδρα** για την ομαδοποίηση κειμένων.

Πηγή πολυμέσου-Ήχου

Πηγή ήχου:

- Περιέχουν ηχογραφημένα μηνύματα όπως ομιλίες, παρουσιάσεις διαλέξεων ή ακόμη καταγραφές παρακολούθησης τηλεφωνικών μηνυμάτων ή συνδιαλέξεων που έχουν επιβληθεί από τον νόμο

Σχόλιο! Για τον εντοπισμό των κύριων χαρακτηριστικών της φωνής κάποιου ατόμου μπορούν να χρησιμοποιηθούν διακριτοί μετασχηματισμοί , ώστε να έχουμε ευρετηριοποίηση και ανάκτηση βασισμένες στην ομοιότητα. Τα χαρακτηριστικά στοιχεία του ήχου περιλαμβάνουν ένταση, τόνο και καθαρότητα.

Χωρικές βάσεις δεδομένων

Οι χωρικές βάσεις δεδομένων:

- Υποστηρίζουν έννοιες για βάσεις δεδομένων που καταγράφουν αντικείμενα σ' έναν πολυδιάστατο χώρο
- Αποθηκεύουν αντικείμενα που περιγράφονται με χωρικά χαρακτηριστικά
- Χρειάζονται χωρικές πράξεις για να εφαρμοσθούν στα χωρικά χαρακτηριστικά των αντικειμένων
- Χρειάζονται μια περιγραφή των χωρικών θέσεων πολλών τύπων αντικειμένων

Κατηγορίες τυπικών χωρικών επερωτήσεων

Επερώτηση διαστήματος: Βρίσκει τα αντικείμενα ενός συγκεκριμένου τύπου που βρίσκονται μέσα σε δεδομένη χωρική περιοχή ή σε συγκεκριμένη απόσταση από δοθείσα θέση.

Επερώτηση για τον κοντινότερο γείτονα: Βρίσκει ένα αντικείμενο συγκεκριμένου τύπου που βρίσκεται πιο κοντά σε δεδομένη θέση.

Χωρικές συνενώσεις ή επικαλύψεις: Συνενώνει αντικείμενα δύο τύπων με βάση κάποια χωρική συνθήκη, όπως αντικείμενα που τέμνονται ή επικαλύπτονται χωρικά ή που βρίσκονται σε κάποια απόσταση το ένα από το άλλο.

Γεωγραφικά συστήματα πληροφοριών

Τα γεωγραφικά συστήματα πληροφοριών (ΓΣΠ-GIS) χρησιμοποιούνται για συλλογή, μοντελοποίηση, αποθήκευση και ανάλυση πληροφοριών που περιγράφουν φυσικές ιδιότητες του γεωγραφικού κόσμου.

Τύποι δεδομένων:

- Χωρικά δεδομένα, προερχόμενα από χάρτες, ψηφιακές εικόνες, διοικητικά και πολιτικά όρια, δρόμους, δίκτυα μεταφορών, φυσικά δεδομένα, ποτάμια, εδαφολογικά χαρακτηριστικά, κλιματολογικές περιοχές, λόφους.
- Μη χωρικά δεδομένα, όπως μετρήσεις απογραφής οικονομικά δεδομένα και πληροφορίες πωλήσεων και προώθησης προϊόντων.

Το ΓΣΠ αποτελούν ένα γρήγορα αναπτυσσόμενο πεδίο που προσφέρει πολύ πρωτοποριακές προσεγγίσεις για την αντιμετώπιση μερικών τεχνικών απαιτήσεων.

Απαιτήσεις διαχείρισης δεδομένων για τα ΓΣΠ

- Μοντελοποίηση Δεδομένων και Αναπαράσταση
- Ανάλυση Δεδομένων
- Ενοποίηση Δεδομένων
- Συλλογή Δεδομένων

ARC-INFO

Είναι απαραίτητο να σχεδιασθούν συστήματα που να υποστηρίζουν τη διανυσματική και την ψηφιοποιημένη αναπαράσταση, τη χωρική λειτουργικότητα καθώς και τα απαραίτητα χαρακτηριστικά ΣΔΒΔ γιατί τα ΣΣΔΒΔ ή τα ΑΣΔΒΔ **δεν καλύπτουν** τις ειδικές ανάγκες των γεωγραφικό σύστημα πληροφοριών. Ένα δημοφιλές γεωγραφικό σύστημα πληροφοριών ονομάζεται **ARC-INFO** και το οποίο δεν είναι ΣΔΒΔ αλλά ενοποιεί τη λειτουργικότητα των ΣΣΔΒΔ στο INFO τμήμα του συστήματος.