11ο Ελληνικό Συμπόσιο Διαχείρισης Δεδομένων

ΕΣΔΔ

Χανιά

28-29 Ιουνίου 2012

http://hdms2012.softnet.tuc.gr/

Περιεχόμενα

Χαιρετισμός Προέδρου & Οργανωτικής Επιτροπής

Αγαπητά Μέλη της Ελληνικής Κοινότητας Διαχείρισης Δεδομένων,

Με ιδιαίτερη χαρά σας καλωσορίζουμε στα Χανιά και στο 11ο Ελληνικό Συμπόσιο Διαχείρισης Δεδομένων, ΕΣΔΔ (Hellenic Data Management Symposium, HDMS), τον ετήσιο θεσμό της διεθνούς ελληνικής κοινότητας διαχείρισης δεδομένων για την παρουσίαση και συζήτηση των τελευταίων εξελίξεων του χώρου. Μετά από δέκα επιτυχημένα συμπόσια σε Αθήνα, Θεσσαλονίκη, Κρήτη, και Κύπρο, το 11ο ΕΣΔΔ διεξάγεται στα Χανιά, μεταξύ 28 και 29 Ιουνίου 2012.

Το πρόγραμμα του 11ου ΕΣΔΔ περιλαμβάνει 27 παρουσιάσεις ερευνητικών εργασιών (τρεις σύντομης διάρκειας), τρεις επιδείξεις λογισμικού, μια προσκεκλημένη ομιλία και μία συζήτηση στρογγυλής τραπέζης. Τα πρακτικά του 11ου ΕΣΔΔ είναι ανεπίσημα και ως εκ τούτου το βασικό κριτήριο για την επιλογή εργασιών προς παρουσίαση είναι το τεχνικό περιεχόμενό τους.

Η εναρκτήρια ομιλία από τον προσκεκλημένο ομιλητή Καθηγητή Νεοκλή Πολυζώτη (UCSC) έχει θέμα «*Κλιμακώνοντας τη Μηχανική Μάθηση σε Μεγάλα Δεδομένα»* -- θέμα εξαιρετικά ενδιαφέρον και επίκαιρο.

Οι 27 ερευνητικές εργασίες και οι τρεις επιδείξεις έχουν οργανωθεί σε 8 συνεδρίες:
(1) Κινητικότητα, Χρόνος και Χώρος (Mobility, Time and Space)
(2) Επεξεργασία και Βελτιστοποίηση Επερωτήσεων (Query Processing & Optimization)
(3) Φυσική και Λογική Οργάνωση Δεδομένων (Physical & Logical Data Organization)
(4) Κατανεμημένη Επεξεργασία Δεδομένων (Distributed Data Processing)
(5) Ροές Δεδομένων και Συνεχείς Επερωτήσεις (Streaming Data & Continuous Queries)
(6) Παρουσιάσεις Επιδείξεων
(7) Ανάκτηση Πληροφορίας, Εξόρυξη Δεδομένων και Ιδιωτικότητα (Information Retrieval, Data Mining & Privacy) και
(8) Ακολουθίες και Συστάδες (Sequences & Clusters).

Τέλος, η συζήτηση στρογγυλής τραπέζης έχει θέμα «Big Data: ερευνητικές προκλήσεις, εφαρμογές, εργαλεία.» Οι συμμετέχοντες θα φωτίσουν τις διαφορετικές ερευνητικές πτυχές της συναρπαστικής και επιδραστικής αυτής περιοχής.

Θα θέλαμε να ευχαριστήσουμε όλους όσους συνέβαλαν για να πραγματοποίηση του φετινού ΕΣΔΔ: όλους τους συγγραφείς των υποβληθεισών εργασιών, τα μέλη της Επιτροπής Προγράμματος για τη λεπτομερή και έγκαιρη αξιολόγηση καθώς και τους συντονιστές των συνεδριών. Θα θέλαμε να εκφράσουμε ιδιαίτερες ευχαριστίες στην Αικατερίνη Ιωάννου και τον Οδυσσέα Παπαπέτρου για τη σημαντική συνεισφορά στη συνολική οργάνωση του συνεδρίου και την επιμέλεια του Ιστοχώρου του, στην Ελευθερία Νυφλή για την επιμέλεια των πρακτικών του συνεδρίου, και στην υποστήριξη του Microsoft CMT Conference Management Tool. Ευχαριστούμε τα πανεπιστήμια και τμήματα μας, Τμήμα Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών του Πολυτεχνείου Κρήτης και Τμήμα Πληροφορικής του Οικονομικού Πανεπιστημίου Αθηνών για την τεχνική υποστήριξη. Ευχαριστούμε ιδιαίτερα την εταιρία Virtual Trip (Χρυσός Χορηγός) και την εταιρία Microsoft Hellas (Χορηγός) που με την οικονομική τους υποστήριξη ενίσχυσαν σημαντικά το συνέδριο. Τέλος, θα θέλαμε να ευχαριστήσουμε όλους εσάς που με τη συμμετοχή σας κάνατε πραγματικότητα το 11ο Ελληνικό Συμπόσιο Διαχείρισης Δεδομένων.

Μίνως Γαροφαλάκης, Πρόεδρος Συμποσίου

Βασίλειος Βασσάλος, Πρόεδρος Επιτροπής Προγράμματος

Ιούνιος 2012

Οργάνωση

Πρόεδρος:

Μίνως Γαροφαλάκης, Πολυτεχνείο Κρήτης

minos@softnet.tuc.gr

Πρόεδρος Επιτροπής Προγράμματος:

Βασίλης Βασσάλος, Οικονομικό Πανεπιστήμιο Αθηνών

vassalos@aueb.gr

Επιτροπή Προγράμματος:

Αναστασία Αϊλαμάκη	Ecole Polytechnique Federale de Lausanne
Μιχαέλα Βορνέα	IBM TJ Watson Research Center
Βασίλης Χριστοφίδης	Πανεπιστήμιο Κρήτης & ITE
Αντώνιος Δεληγιαννάκης	Πολυτεχνείο Κρήτης
Αλέξης Δελής	Πανεπιστήμιο Αθηνών
Χρήστος Δουλκερίδης	Πανεπιστήμιο Πειραιώς και NTNU
Γιώργος Φάκας	Manchester Metropolitan University
Ειρήνη Φουντουλάκη	ITE-ΙΠ, Ελλάδα
Δημήτριος Γουνόπουλος	Πανεπιστήμιο Αθηνών
Μάριος Χατζηελευθερίου	AT&T Shannon Labs
Βαγγέλης Χρηστίδης	University of California, Riverside
Στράτος Υδραίος	CWI
Αικατερίνη Ιωάννου	Πολυτεχνείο Κρήτης
Πάνος Καλνής	King Abdullah University of Science and Technology
Βάνα Καλογεράκη	Οικονομικό Πανεπιστήμιο Αθηνών
Γιώργος Κόλλιος	Boston University
Γιάννης Κωτίδης	Οικονομικό Πανεπιστήμιο Αθηνών
Μανόλης Κουμπαράκης	Πανεπιστήμιο Αθηνών

Νίκος Κούδας	University of Toronto
Γεωργία Κούτρικα	IBM Almaden Research Center
Αλέξανδρος Λαμπρινίδης	University of Pittsburgh
Νίκος Μαμουλής	University of Hong Kong
Κυριάκος Μουρατίδης	Singapore Management University
Αλέξανδρος Ντούλας	Zynga
Θέμης Παλπάνας	University of Trento
Δημήτρης Παπαδιάς	Hong Kong University of Science and Technology
Όλγα Παπαεμανουήλ	Brandeis University
Γιάννης Παπακωνσταντίνου	University of California, San Diego
Οδυσσέας Παπαπέτρου	Πολυτεχνείο Κρήτης
Μιχάλης Πετρόπουλος	SUNY Buffalo
Ευαγγελία Πιτουρά	Πανεπιστήμιο Ιωαννίνων
Νεοκλής Πολυζώτης	UC Santa Cruz
Νίκος Ρουσόπουλος	UMIACS
Τίμος Σελλής	Ερευνητικό Κέντρο ΑΘΗΝΑ
Γιάννης Θεοδωρίδης	Πανεπιστήμιο Πειραιώς
Παναγιώτης Τριανταφύλλου	Πανεπιστήμιο Πατρών
Πάνος Βασιλειάδης	Πανεπιστήμιο Ιωαννίνων
Στρατής Βίγλας	University of Edinburgh
Ακριβή Βλάχου	NTNU
Δημήτρης Ζεϊναλιπούρ	Πανεπιστήμιο Κύπρου

Τοπική Οργάνωση:

Αντώνιος Δεληγιαννάκης	Πολυτεχνείο Κρήτης
Αικατερίνη Ιωάννου	Πολυτεχνείο Κρήτης
Οδυσσέας Παπαπέτρου	Πολυτεχνείο Κρήτης

Webmaster:

Αικατερίνη Ιωάννου	Πολυτεχνείο Κρήτης

Πρόγραμμα Συμποσίου

11ο Ελληνικό Συμπόσιο Διαχείρισης Δεδομένων

Χανιά

28-29 Ιουνίου 2012

Πέμπτη, 28 Ιουνίου 2012

8:30 – 8:55 Εγγραφή

8:55 – 9:00 Καλωσόρισμα

9:00 – 10:15 1η Συνεδρία – Κεντρική ομιλία
Κλιμακώνοντας τη Μηχανική Μάθηση σε Μεγάλα Δεδομένα
Νεοκλής (Άλκης) Πολυζώτης (UCSC)

10:15 – 10:30 Διάλειμμα – Καφές

10:30 – 12:00 2η Συνεδρία – Mobility, Time and Space
Session Chair: Stratos Idreos

- *Κλιμακώσιμη Αναζήτηση κ Πλησιεστέρων Γειτόνων μεταξύ Καθέτως Αποθηκευμένων Χρονοσειρών*
  Shrikant Kashyap (National University of Singapore), Panagiotis Karras (Rutgers University)

- *Προγραμματισμός Κινητών Συστημάτων Πραγματικού Χρόνου χρησιμοποιώντας MapReduce*
  Adam Dou (Google), Vana Kalogeraki (Athens University of Economics and Business), Dimitris Gunopulos (University of Athens), Taneli Mielikainen (Nokia), Ville Tuulos (Nokia)

- *Ενιαίο Πλαίσιο για την Προστασία Ιδιωτικότητας από Τεχνικές Εξόρυξης Προτύπων και Επερωτήσεων σε Βάσεις Τροχιών Κινούμενων Αντικειμένων*
  Despina Kopanaki (University of Piraeus), Nikos Pelekis (University of Piraeus), Aris Gkoulalas-Divanis (IBM Research– Ireland), Marios Vodas (University of Piraeus), Yannis Theodoridis (University of Piraeus)

- *Αποδοτικοί Αλγόριθμοι για Επερωτήσεις Κορυφαίων-Κ Χωρικών Προτιμήσεων*
  Joao B. Rocha-Junior (NTNU), Akrivi Vlachou (NTNU), Christos Doulkeridis (University of Piraeus), Kjetil Noervaag (NTNU)

12:00 – 14:00 Γεύμα

14:00 – 15:30 3η Συνεδρία – Query Processing & Optimization
Session Chair: Nikos Mamoulis

- *NoDB: Αποτελεσματική εκτέλεση επερωτήσεων σε ακατέργαστα αρχεία δεδομένων*

Ioannis Alagiannis (EPFL), Renata Borovica (EPFL), Miguel Branco (EPFL), Stratos Idreos (CWI), Anastasia Ailamaki (Ecole Polytechnique Federale de Lausanne)

- *Πέρα από τα 100 εκατομμύρια οντότητες: μεγάλης κλίμακας σύζευξη ετερογενών δεδομένων με βάση τεχνικές blocking*
  George Papadakis (L3S Research Center), Ekaterini Ioannou (Technical University of Crete), Claudia Niederee (L3S Research Center), Themis Palpanas (University of Trento), Wolfgang Nejdl (L3S Research Center)

- *Επιλογή υλοποιημένων όψεων για φόρτους εργασίας XQuery*
  Asterios Katsifodimos (INRIA Saclay & Univ. Paris-Sud), Ioana Manolescu (INRIA Saclay & Univ. Paris-Sud), Vasilis Vassalos (Athens University of Economics and Business)

- *Επιτάχυνση επερωτήσεων εύρους για προσομοιώσεις του εγκεφάλου*
  Farhan Tauheed (EPFL), Laurynas Biveinis (University of Aalborg), Thomas Heinis (EPFL), Anastasia Ailamaki (Ecole Polytechnique Federale de Lausanne), Felix Schurmann (EPFL), Henry Markram (EPFL)

15:30 – 16:00 Διάλειμμα για καφέ και συζήτηση

16:00 – 17:15 4η Συνεδρία – Physical & Logical Data Organization

Session Chair: Panagiotis Karras

- *Διάδοση Αλλαγών Σε Πληροφοριακά Οικοσυστήματα (Σύντομη παρουσίαση)*
  George Papastefanatos (Institute for the Management of Information Systems), Panos Vassiliadis* (University of Ioannina), Alkis Simitsis (HP labs)

- *Ετερογενή Σχέδια για την Ρύθμιση Συστημάτων με Αντίγραφα*
  Mariano Consens (U of Toronto), Kleoni Ioannidou (UC Santa Cruz), Jeff LeFevre (UC Santa Cruz), Alkis Polyzotis (UC Santa Cruz)

- *Έλεγχος Ταυτοχρονισμού σε Προσαρμοστική Ευρετηρίαση*
  Goetz Graefe (HP Labs), Felix Halim (National University of Singapore), Stratos Idreos* (CWI), Harumi Kuno (Hp Labs), Stefan Manegold (CWI)

- *Στοχαστικός Κατακερματισμός: Εύρωστη Προσαρμοστική Ευρετηρίαση σε Συστήματα Βάσεων Δεδομένων Οργανωμένα ανα Στήλες στην Κύρια Μνήμη*
  Felix Halim (National University of Singapore), Stratos

Idreos (CWI), Panagiotis Karras (Rutgers University), Roland Yap (National University of Singapore)

17:15 – 17:30 Διάλειμμα για καφέ και συζήτηση

17:30 – 18:35 5η Συνεδρία – Distributed data processing
Session Chair: Christos Doulkeridis

- *Αποδοτική Εξισορρόπηση Φόρτου σε Διαμερισμένα Ερωτήματα κάτω από Τυχαίες Μεταβολές*
  Anastasios Gounaris (Aristotle University), Christos Yfoulis (ATEI of Thessaloniki), Norman Paton (University of Manchester)

- *Κατανεμημένη επεξεργασία SPARQL ερωτημάτων πάνω από υπολογιστικά νέφη.*
  Nikolaos Papailiou (Cslab (ntua), Ioannis Konstantinou (ntua), Dimitrios Tsoumakos (ntua), Panagiotis Karras (Rutgers University), Nectarios Koziris (ntua)

- *Γεωμετρική Παρακολούθηση Κατανεμημένων Ρευμάτων Δεδομένων Βασισμένη στην Πρόβλεψη*
  Nikos Giatrakos (University of Piraeus), Antonios Deligiannakis (Technical University of Crete), Minos Garofalakis (Technical University of Crete), Izchak Sharfman (Faculty of Computer Science Technion), Assaf Schuster (Faculty of Computer Science Technion)

20:00 Δείπνο Συνεδρίου

Παρασκευή, 29 Ιουνίου 2012

8:30 – 10:00 6η Συνεδρία – Streaming Data & Continuous Queries
Session Chair: Akrivi Vlachou

- *Συνεχής Αναζήτηση των Κ Πλησιέστερων Έξυπνων Κινητών Συσκευών για Κάθε Χρήστη*
  Georgios Chatzimilioudis (University of Cyprus), Demetris Zeinalipour (University of Cyprus), Wang-Chien Lee (Pennsylvania State University), Marios Dikaiakos (University of Cyprus)

- *Δυναμική Υποστήριξη Ποικιλομορφίας σε Ροές Δεδομένων (Dynamic Diversification of Continuous Data)*
  Marina Drosou (University of Ioannina), Evaggelia Pitoura (University of Ioannina)

- *Αναγνώριση bursts σε χωροχρονικές ροές εγγράφων*
  Theodoros Lappas (UC Riverside), Marcos Vieira (UC Riverside), Dimitris Gunopulos (University of Athens), Vassilis Tsotras (UC Riverside)

- *Τριεπίπεδη Εκτέλεση Πολλαπλών Συγκεντρωτικών Ερωτημάτων Διάρκειας*
  Shenoda Guirguis (University of Pittsburgh), Mohamed Sharaf (University of Queensland), Panos K. Chrysanthis (University of Pittsburgh), Alexandros Labrinidis (University of Pittsburgh)

10:00 – 11:15 7η Συνεδρία – Demo Presentations *(με καφέ)*
Session Chair: Theodore Dalamagas

- *Αλληλεπιδράσεις Εγγύτητας με το Crowdcast*
  Marios Constantinides (University of Cyprus), G Constantinou (University of Cyprus), Andreas Panteli (University of Cyprus), Theofilos Phokas (University of Cyprus), Georgios Chatzimilioudis (University of Cyprus), Demetris Zeinalipour (University of Cyprus)

- *Κλιμακώσιμη και Ευέλικτη Υποστήριξη Στιγμιαίας Επισκοπικής Αναζήτησης*
  Pavlos Fafalios (Institute of Computer Science, FORTH-ICS), Ioannis Kitsos (Institute of Computer Science, FORTH-ICS), Yannis Tzitzikas (Institute of Computer Science, FORTH-ICS)

- *EVISUGE: Οπτικοποίηση Γεγονότων στο Google Earth*
  Nikolaos Tarantilis (Technical University of Crete), Chrisa Tsinaraki (Technical University of Crete), Fotis Kazasis (Technical University of Crete), Nektarios Gioldasis (Technical University of Crete), Stavros Christodoulakis (Technical University of Crete)

11:15 – 12:30 8η Συνεδρία – Information Retrieval, Data Mining & Privacy
Session Chair: Demetris Zeinalipour

- *Μεγέθους-λ Περιλήψεις Αντικειμένων για Σχεσιακή Αναζήτηση με Λέξεις-κλειδιά*
  Georgios Fakas (Manchester Metropolitan Univer), Nikos Mamoulis (University of Hong Kong), Zhi Cai (Manchester Metropolitan University)

- *Εκπαίδευση Αναζήτησης Αποτελεσμάτων με Βάση το Σκοπό Αναζήτησης (Σύντομη παρουσίαση)*
  Giorgos Giannopoulos (NTU Athens - Athena R.C.)

- *Διατήρηση της κ-ανωνυμίας σε κατανεμημένα δεδομένα*
  Katerina Doka (NTUA), Dimitrios Tsoumakos (NTUA), Nectarios Koziris (NTUA)

- *Υποχώροι υψηλής αντίθεσης για τη βαθμολόγηση εκτόπων με βάση την πυκνότητα*
  Fabian Keller (Karlsruhe Institute of Technology (KIT)), Emmanuel Müller (Karlsruhe Institute of Technology

(KIT)), Klemens Böhm (Karlsruhe Institute of Technology (KIT)

12:30 – 13:15 9η Συνεδρία – Ομιλίες Χορηγών
Session Chair: Yannis Kotidis

- *SQL Azure – Η Βάση Δεδομένων ως Cloud υπηρεσία*
  Δημήτρης Κουτσαναστάσης, Microsoft Hellas

- Ομιλία Χρυσού Χορηγού Virtual Trip (Θα καθοριστεί)

13:15 – 14:30 Γεύμα

14:30 – 16:00 10η Συνεδρία – Sequences & Clusters
Session Chair: Ioannis Konstantinou

- *Μια νέα τεχνική υπολογισμού ομοιότητας ακολουθιών με εφαρμογή σε Query-By-Humming*
  Alexios Kotsifakos (University of Athens), Panagiotis Papapetrou (Aalto University), Jaakko Hollmen (Aalto University), Dimitris Gunopulos (University of Athens)

- *Συσταδοποίηση προβολής βάσει πυκνότητας σε ροές δεδομένων υψηλών διαστάσεων*
  Eirini Ntoutsi (LMU), Arthur Zimek (LMU), Themis Palpanas (University of Trento), Peer Kroeger (LMU), Hans-peter Kriegel (LMU)

- *Προσεγγιστικό τμηματικό ταίριασμα ακολουθιών*
  Thanasis Vergoulis (NTUA & Athena RC), Theodore Dalamagas (NTUA & Athena RC), Dimitris Sacharidis (NTUA & Athena RC), Timos Sellis (NTUA & Athena RC)

16:00 – 16:15 Διάλειμμα για καφέ

16:15 – 17:30 11η Συνεδρία - Συζήτηση στρογγυλής τραπέζης

- **Big Data: ερευνητικές προκλήσεις, εφαρμογές, εργαλεία.**
- **Συντονιστής: Βασίλης Βασσάλος, ΟΠΑ**

17:30 – 17:45 Κλείσιμο Συνεδρίου

Οργανωτική Υποστήριξη:

Πολυτεχνείο Κρήτης

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Πρόεδρος: Μίνως Γαροφαλάκης, Πολυτεχνείο Κρήτης

Πρόεδρος Επιτροπής Προγράμματος: Βασίλης Βασσάλος, Οικονομικό Πανεπιστήμιο Αθηνών

Οικονομική Υποστήριξη

ΧΡΥΣΟΣ ΧΟΡΗΓΟΣ

virtualtrip
materializing innovation

Virtual Trip is a leading entrepreneurial ecosystem of eleven dynamic companies in the ICT sector providing a broad spectrum of products and services to the international market. We have grown organically since our establishment in year 2000 at Heraklion-Crete, Greece, based on the strong entrepreneurial drive of our team.

ΧΟΡΗΓΟΣ

Microsoft

1η Συνεδρία – Κεντρική ομιλία

Κλιμακώνοντας τη Μηχανική Μάθηση σε Μεγάλα Δεδομένα Άλκης Πολυζώτης (UCSC)

Παρόλο που η μηχανική μάθηση είναι πολύ σημαντική για συστήματα μεγάλων δεδομένων, η εφαρμογή αλγορίθμων μηχανικής μάθησης σε τέτοια συστήματα εμπεριέχει σημαντικές προκλήσεις, κυρίως λόγω της χαμηλής υποστήριξης επαναληπτικών και αναδρομικών διαδικασιών. Στην ομιλία αυτή θα κάνουμε μια ανασκόπηση των ερευνητικών προσπαθειών για αντιμετώπιση των προκλήσεων αυτών, συμπεριλαμβανομένης και της δικής μας προσέγγισης που βασίζεται σε αρχές και τεχνικές από επεξεργασία ερωτημάτων σε βάσεις δεδομένων. Η αποδοτικότητα της προσέγγισης μας είναι συγκρίσιμη ή και ανώτερη της αποδοτικότητας εξειδικευμένων εργαλείων μηχανικής μάθησης, διατηρώντας παράλληλα ένα γενικό και επεκτάσιμο πλαίσιο για εφαρμογή μηχανικής μάθησης σε μεγάλα δεδομένα.

2η Συνεδρία – Κινητικότητα, Χρόνος και Χώρος

Κλιμακώσιμη Αναζήτηση κ Πλησιεστέρων Γειτόνων μεταξύ Καθέτως Αποθηκευμένων Χρονοσειρών

Shrikant Kashyap (National University of Singapore), Panagiotis Karras (Rutgers University)

Η αναζήτηση πλησιεστέρων γειτόνων μεταξύ χρονοσειρών έχει λάβει μεγάλη προσοχή ως βασική λειτουργία εξόρυξης δεδομένων. Εντούτοις, οι μέχρι τούδε προταθείσες μέθοδοι δεν κλιμακώνουν με την αύξηση του μήκους των χρονοσειρών. Αυτό οφείλεται στο γεγονός ότι παρέχουν μόνο μια εφάπαξ δυνατότητα για κλάδεμα. Ειδικότερα, οι παραδοσιακές μέθοδοι χρησιμοποιούν ένα ευρετήριο για να ψάξουν σε έναν χώρο γνωρισμάτων με μειωμένες διαστάσεις. Ωστόσο, για μεγάλο μήκος χρονοσειρών, η αναζήτηση με ένα τέτοιο ευρετήριο δίνει πολλά εσφαλμένα θετικά αποτελέσματα που πρέπει να εξαλειφθούν προσπελάζοντας τα πλήρη δεδομένα. Μια προσπάθεια για μείωση των σφαλμάτων με την ευρετηρίαση περισσότερων γνωρισμάτων επιδεινώνει την κατάρα της διαστατικότητας, και τούμπαλιν. Μια πρόσφατη εναλλακτική λύση, το iSAX, χρησιμοποιεί συμβολικές προσεγγιστικές αναπαραστάσεις με έναν απλό κατάλογο του συστήματος αρχείων ως ευρετήριο. Εντούτοις, και το iSAX παράγει εσφαλμένες συμπεριλήψεις, οι οποίες και πάλι εξαλείφονται με την πρόσβαση των δεδομένων στο ακέραιο: άπαξ και παραχθεί μία εσφαλμένη συμπερίληψη, δεν υπάρχει δεύτερη ευκαιρία για να κλαδευθεί. Έτσι, η δυνατότητα κλαδέματος του iSAX είναι επίσης εφάπαξ. Αυτό το άρθρο προτείνει μια εναλλακτική λύση για την αναζήτηση πλησιεστέρων γειτόνων μεταξύ χρονοσειρών, ακολουθώντας έναν μη παραδοσιακό τρόπο κλαδέματος. Αντί της πλοήγησης μεταξύ των υποψηφίων μέσω ενός ευρετηρίου, προσπελάζουμε τα γνωρίσματά τους, αποκεκτημένα από έναν μετασχηματισμό πολλαπλής

ανάλυσης, με μια σταδιακή ακολουθιακή σάρωση, ένα επίπεδο ανάλυσης σε κάθε βήμα, πάνω σε μια κάθετη αναπαράσταση. Οι περισσότεροι υποψήφιοι αφαιρούνται προοδευτικά αφού προσπελαστούν μερικά από τα γνωρίσματά τους, χρησιμοποιώντας εκ των προτέρων υπολογισθείσες στατιστικές και μια άνευ προηγουμένου στενή διπλή οριοθέτηση αποστάσεων. Η πειραματική μας μελέτη με ευμεγέθη δεδομένα χρονοσειρών μεγάλου μήκους επιβεβαιώνει το πλεονέκτημα της προσέγγισής μας σε σχέση τόσο με το iSAX όσο και με κλασικές μεθόδους.

Προγραμματισμός Κινητών Συστημάτων Πραγματικού Χρόνου χρησιμοποιώντας MapReduce

Adam Dou (Google), Vana Kalogeraki (Athens University of Economics and Business), Dimitris Gunopulos (University of Athens), Taneli Mielikainen (Nokia), Ville Tuulos (Nokia)

Η δημοτικότητα των φορητών ηλεκτρονικών συσκευών όπως τα smartphones και PDAs και η αύξηση των δυνατοτήτων επεξεργασίας τους, έχει επιτρέψει την ανάπτυξη πολλών εφαρμογών οι οποίες απαιτούν χαμηλές καθυστερήσεις χρόνου, υψηλή απόδοση και επεκτασιμότητα. Η υποστήριξη εφαρμογών πραγματικού χρόνου στις κινητές συσκευές είναι ιδιαίτερη πρόκληση λόγω των περιορισμένων πόρων, αποτυχίες των κινητών συσκευών καθώς και σημαντικές διακυμάνσεις της ποιότητας του ασύρματου μέσου. Στο άρθρο αυτό αντιμετωπίζουμε το πρόβλημα της υποστήριξης κατανεμημένων εφαρμογών πραγματικού χρόνου χρησιμοποιώντας MapReduce υπό την παρουσία αποτυχιών. Παρουσιάζουμε το σύστημά μας Mobile MapReduce (MiscoRT), με στόχο την υποστήριξη της εκτέλεσης κατανεμημένων εφαρμογών με απαιτήσεις πραγματικού χρόνου απόκρισης. Προτείνουμε ένα σύστημα προγραμματισμού δύο επιπέδων, σχεδιασμένο για το MapReduce μοντέλο προγραμματισμού, το οποίο ουσιαστικά προβλέπει τον χρόνο εκτέλεσης των εφαρμογών και δυναμικά χρονοδρομολογεί την εκτέλεση των προγραμμάτων. Έχουμε κάνει εκτενείς δοκιμές σε ένα κινητό σύστημα αποτελούμενο από κινητά της Nokia N95 8GB smartphones. Έχουμε αποδείξει ότι το σύστημα μας είναι αποδοτικό, έχει χαμηλή επιβάρυνση και εκτελείται έως και 32% γρηγορότερα από ό, τι οι ανταγωνιστές του.

Ενιαίο Πλαίσιο για την Προστασία Ιδιωτικότητας από Τεχνικές Εξόρυξης Προτύπων και Επερωτήσεων σε Βάσεις Τροχιών Κινούμενων Αντικειμένων

Despina Kopanaki (University of Piraeus), Nikos Pelekis (University of Piraeus), Aris Gkoulalas-Divanis (IBM Research– Ireland), Marios Vodas (University of Piraeus), Yannis Theodoridis (University of Piraeus)

Οι υπάρχουσες τεχνικές για την προστασία της ιδιωτικότητας κατά τον διαμοιρασμό δεδομένων κίνησης, στοχεύουν στη δημοσίευση μιας ανώνυμης έκδοσης του συνόλου των δεδομένων

κίνησης, υποθέτοντας ότι το μεγαλύτερο μέρος της πληροφορίας στο αρχικό σύνολο δεδομένων μπορεί να αποκαλυφθεί χωρίς να προκληθούν παραβιάσεις της ιδιωτικότητας. Στην εργασία αυτή, υποθέτουμε ότι η πληροφορία που υπάρχει σε ένα σύνολο δεδομένων κίνησης πρέπει να παραμείνει ιδιωτική και τα δεδομένα πρέπει να παραμείνουν στον οργανισμό που τα φιλοξενεί. Για να διευκολυνθεί η προστασία της ιδιωτικότητας κατά την ανταλλαγή δεδομένων κίνησης αναπτύξαμε μια μηχανή επερωτήσεων για τροχιές κινούμενων αντικειμένων που επιτρέπει στους εγγεγραμμένους χρήστες να αποκτήσουν περιορισμένη πρόσβαση στην βάση δεδομένων για να πραγματοποιήσουν διάφορες εργασίες ανάλυσης. Η προτεινόμενη μηχανή (i) ελέγχει επερωτήσεις για δεδομένα τροχιών κινούμενων αντικειμένων για να εμποδίσει πιθανές επιθέσεις κατά της ιδιωτικότητας των χρηστών, (ii) υποστηρίζει χωρικές και χωροχρονικές επερωτήσεις εύρους, απόστασης και k-κοντινότερων γειτόνων, και (iii) διατηρεί την ανωνυμία των χρηστών κατά τις απαντήσεις των επερωτήσεων (α) επιστρέφοντας ένα σύνολο προσεκτικά δημιουργημένων ρεαλιστικών ψεύτικων τροχιών, και (β) διασφαλίζοντας ότι καμία ευαίσθητη περιοχή που να αφορά σε οποιονδήποτε χρήστη δεν θα απεικονίζεται ως μέρος των επιστρεφόμενων τροχιών. Επιπρόσθετα, παρουσιάζουμε την πλατφόρμα Private-HERMES, που παρέχει ένα δισδιάστατο πλαίσιο αναφοράς που περιλαμβάνει: (α) την προαναφερθείσα μηχανή επερωτήσεων που παρέχει λειτουργικότητα για την προστασία της ιδιωτικότητας και (β) ένα προοδευτικό πλαίσιο ανάλυσης, το οποίο εκτός από μεθόδους ανωνυμοποίησης για την δημοσίευση δεδομένων, περιλαμβάνει διάφορες τεχνικές εξόρυξης γνώσης για δεδομένα κίνησης για την αξιολόγηση της επίδρασης της ανωνυμοποίησης στα αποτελέσματα αναζήτησης και εξόρυξης.

Αποδοτικοί Αλγόριθμοι για Επερωτήσεις Κορυφαίων-Κ Χωρικών Προτιμήσεων

Joao B. Rocha-Junior (NTNU), Akrivi Vlachou (NTNU), Christos Doulkeridis (University of Piraeus), Kjetil Noervaag (NTNU)

Οι επερωτήσεις κορυφαίων-Κ χωρικών προτιμήσεων επιστρέφουν ένα σύνολο από τα Κ καλύτερα αντικείμενα με κατάταξη βάσει των σκορ άλλων αντικειμένων ενδιαφέροντος που βρίσκονται στη γειτονιά τους. Παρά το μεγάλο εύρος εφαρμογών με βάση τη θέση που στηρίζονται σε επερωτήσεις χωρικών προτιμήσεων, οι υπάρχοντες αλγόριθμοι έχουν σημαντικό κόστος επεξεργασίας καταλήγοντας σε υψηλό χρόνο απόκρισης. Αυτό οφείλεται στο γεγονός ότι για τον υπολογισμό του σκορ οποιουδήποτε αντικειμένου απαιτείται εξέταση της γειτονιάς του για να βρεθούν τα αντικείμενα ενδιαφέροντος με υψηλότερο σκορ. Σε αυτό το άρθρο, προτείνουμε μια καινοτόμα τεχνική που επιταχύνει την επεξεργασία των επερωτήσεων κορυφαίων-Κ χωρικών προτιμήσεων. Για το σκοπό αυτό, προτείνουμε μια αντιστοίχιση ζευγαριών αντικειμένων και αντικειμένων ενδιαφέροντος σε ένα χώρο απόστασης-σκορ, που μάς επιτρέπει να εντοπίσουμε και να αποθηκεύσουμε το ελάχιστο υποσύνολο ζευγαριών που είναι ικανό να απαντήσει οποιαδήποτε

επερώτηση χωρικών προτιμήσεων. Επιπλέον, παρουσιάζουμε ένα καινοτόμο αλγόριθμο που βελτιώνει την απόδοση της επεξεργασίας επερωτήσεων αποφεύγοντας να εξετάσει τις γειτονιές ορισμένων αντικειμένων κατά την εκτέλεση των επερωτήσεων. Επιπρόσθετα, προτείνουμε ένα αποδοτικό αλγόριθμο για την αποθήκευση και περιγράφουμε χρήσιμες ιδιότητες που μειώνουν το κόστος συντήρησης. Με εκτενή πειράματα δείχνουμε ότι η προσέγγισή μας μειώνει σημαντικά το πλήθος των μεταφερόμενων μπλοκ καθώς και το χρόνο εκτέλεσης συγκριτικά με τους καλύτερους γνωστούς αλγόριθμους για ποικίλα σενάρια.

3η Συνεδρία – Επεξεργασία και Βελτιστοποίηση Επερωτήσεων

NoDB: Αποτελεσματική εκτέλεση επερωτήσεων σε ακατέργαστα αρχεία δεδομένων
Ioannis Alagiannis (EPFL), Renata Borovica (EPFL), Miguel Branco (EPFL), Stratos Idreos (CWI), Anastasia Ailamaki (Ecole Polytechnique Federale de Lausanne)

Καθώς οι συλλογές δεδομένων μεγαλώνουν, η φόρτωση δεδομένων έχει εξελιχθεί σε σημαντικό παράγοντα καθυστέρησης. Πολλές εφαρμογές όπως ανάλυση επιστημονικών δεδομένων και κοινωνικών δικτύων, αποφεύγουν τη χρήση βάσεων δεδομένων εξαιτίας της πολυπλοκότητας και του αυξημένου χρόνου ανάκτησης δεδομένων. Σε τέτοιες εφαρμογές τα δεδομένα αυξάνονται, ακόμα και σε ημερήσια βάση. Επιπλέον, βρισκόμαστε ήδη σε μια εποχή κατακλυσμού από δεδομένα όπου έχουμε περισσότερα δεδομένα από ότι μπορούμε να μετακινήσουμε, να αποθηκεύσουμε αλλά και να αναλύσουμε. Σε αυτή την εργασία παρουσιάζουμε ένα νέο παράδειγμα σχεδίασης συστημάτων βάσεων δεδομένων, που ονομάζεται NoDB, το οποίο δεν προϋποθέτει φόρτωση δεδομένων ενώ διατηρεί όλα τα χαρακτηριστικά ενός σύγχρονου ΣΔΒΔ. Συγκεκριμένα, δείχνουμε πως τα ακατέργαστα αρχεία δεδομένων μπορούν να μετατραπούν σε επιλογή πρωταρχικής σημασίας, πλήρως ενοποιημένα με τη μηχανή επερωτήσεων. Μέσω του σχεδιασμού και των διδαγμάτων από την εφαρμογή της NoDB φιλοσοφίας σε ένα σύγχρονο ΣΔΒΔ, παρουσιάζουμε θεμελιώδεις περιορισμούς καθώς και δυνατότητες που προκύπτουν για έρευνα προς αυτή την κατεύθυνση. Αναγνωρίζουμε συγκεκριμένους παράγοντες καθυστέρησης για επεξεργασία δεδομένων στην αρχική τους θέση, όπως την επαναλαμβανόμενη ανάλυση λέξεων και συμβολική επεξεργασία, και το κόστος μετατροπής τύπων δεδομένων. Για να αντιμετωπίσουμε αυτά τα προβλήματα, εισάγουμε έναν προσαρμοστικό μηχανισμό δεικτοδότησης που διατηρεί πληροφορίες θέσης ώστε να παρέχει αποτελεσματική πρόσβαση σε ακατέργαστα αρχεία δεδομένων, καθώς και μια ευέλικτη δομή μνήμης. Η υλοποίηση μας πάνω από την PostgreSQL, ονομάζεται PostgresRaw και αποφεύγει το κόστος φόρτωσης δεδομένων ενώ συναγωνίζεται και

πολλές φορές ξεπερνά σε απόδοση την κανονική PostgreSQL. Συμπεραίνουμε πως είναι εφικτό να σχεδιαστούν και να υλοποιηθούν NoDB συστήματα πάνω από σύγχρονες αρχιτεκτονικές βάσεων δεδομένων, φέρνοντας μια άνευ προηγουμένου θετική επίδραση όσον αφορά τη λειτουργικότητα και την απόδοση.

Πέρα από τα 100 εκατομμύρια οντότητες: μεγάλης κλίμακας σύζευξη ετερογενών δεδομένων με βάση τεχνικές blocking

George Papadakis (L3S Research Center), Ekaterini Ioannou (Technical University of Crete), Claudia Niederee (L3S Research Center), Themis Palpanas (University of Trento), Wolfgang Nejdl (L3S Research Center)

Προϋπόθεση για την αξιοποίηση των τεράστιων ποσοτήτων δεδομένων που είναι διαθέσιμα στο Δίκτυο είναι η Σύζευξη Οντοτήτων, δηλαδή η διαδικάσια εντοπισμού και σύνδεσης των δεδομένων που περιγράφουν τα ίδια αντικείμενα του πραγματικού κόσμου. Για να εφαρμοστεί αυτή η υψηλής πολυπλοκότητας διαδικασία σε μεγάλα σύνολα δεδομένων συνήθως επιστρατεύονται τεχνικές blocking: οι οντότητες ομαδοποιούνται με βάση την ομοιότητα τους σε υποσύνολα, τα blocks, ώστε να αρκούν οι συγκρίσεις μόνο μεταξύ οντοτήτων που συμπίπτουν σε ένα τουλάχιστον block. Για να διαχειριστούμε, όμως, τα θορυβώδη, ετερογενή, ημιδομημένα σύνολα δεδομένων του Δικτύου απαιτούνται καινούριες τεχνικές blocking, αφού οι υπάρχουσες μέθοδοι είναι ανεφάρμοστες λόγω της εξάρτησης τους από πληροφορίες αναπαράστασης. Η χρήση πλεονασματικότητας μπορεί να βελτιώσει την ευρωστία των τεχνικών blocking, αλλά επιφέρει επιπλέον υπολογιστικό κόστος. Σε αυτή τη δημοσίευση παρουσιάζουμε μεθόδους που ενισχύουν την αποδοτικότητα των τεχνικών blocking που βασίζονται σε πλεονασματικότητα. Εισηγούμαστε τη δημιουργία blocks με βάση μια πληθώρα ενδείξεων, όπως τα αναγνωριστικά των οντοτήτων και οι συνδέσεις μεταξύ τους. Οι τεχνικές αυτές περιορίζουν σημαντικά τον απαιτούμενο αριθμό συγκρίσεων, διατηρώντας παράλληλα την αποτελεσματικότητα σε υψηλά επίπεδα. Επιπλέον, εισηγούμαστε δυο θεωρητικές μετρικές που παρέχουν αξιόπιστες εκτιμήσεις της επίδοσης μιας μεθόδου blocking, χωρίς να απαιτούν την αναλυτική επεξεργασία των blocks της. Με βάση αυτές τις μετρικές, αναπτύσσουμε δυο τεχνικές για την βελτίωση τις αποδοτικότητας του blocking: τον συνδυασμό ατομικών, αλλά συμπληρωματικών τεχνικών blocking και την διαγραφή blocks μέχρι την ικανοποίηση συγκεκριμένων κριτηρίων. Δοκιμάσαμε τις μεθόδους μας μέσω μιας εκτενούς πειραματικής αξιολόγησης χρησιμοποιώντας ένα ογκώδες σύνολο δεδομένων με 182 εκατομμύρια ετερογενείς οντότητες. Τα αποτελέσματα της μελέτης μας μαρτυρούν την εφαρμοσιμότητα και την υψηλή επίδοση των προσεγγίσεών μας.

Επιλογή υλοποιημένων όψεων για φόρτους εργασίας XQuery

Asterios Katsifodimos (INRIA Saclay & Univ. Paris-Sud), Ioana Manolescu (INRIA Saclay & Univ. Paris-Sud), Vasilis Vassalos (Athens University of Economics and Business)

Η αποτελεσματική επεξεργασία ερωτημάτων XQuery εξακολουθεί να θέτει σημαντικές προκλήσεις. Μια ιδιαίτερα αποτελεσματική τεχνική βελτίωσης της επίδοσης επεξεργασίας XQuery ερωτημάτων, αποτελείται από τη χρήση υλοποιημένων όψεων. Στην παρούσα εργασία, ασχολούμαστε με το πρόβλημα της επιλογής των καλύτερων όψεων που πρέπει να υλοποιηθούν σε μία βάση δεδομένων XML δεδομένου ενός ορίου στο χώρο που καταλαμβάνουν. Ο στόχος της βελτιστοποίησης είναι να βελτιωθούν οι επιδόσεις κατά την επεξεργασία ενός δεδομένου φόρτου εργασίας ερωτημάτων. Η εργασία αυτή είναι η πρώτη που αντιμετωπίζει το συγκεκριμένο πρόβλημα επιλογής όψεων για ερωτήματα και όψεις δεντρικών προτύπων που υποστηρίζουν ένωση ως προς την τιμή και πολλαπλούς κόμβους επιστροφής. Οι προκλήσεις που αντιμετωπίζουμε προέρχονται τόσο από την εκφραστική δύναμη και τα χαρακτηριστικά της γλώσσας έκφρασης ερωτημάτων και όψεων, όσο και από το μέγεθος του χώρου αναζήτησης των υποψηφίων όψεων προς υλοποίηση. Ενώ το πρόβλημα στη γενική του μορφή χαρακτηρίζεται από απαγορευτική πολυπλοκότητα, προτείνουμε και μελετάμε έναν αλγόριθμο ο οποίος επιδεικνύει την καλύτερη επίδοση σε σύγκριση με τις μέχρι στιγμής εργασίες στη βιβλιογραφία.

Επιτάχυνση επερωτήσεων εύρους για προσομοιώσεις του εγκεφάλου

Farhan Tauheed (EPFL), Laurynas Biveinis (University of Aalborg), Thomas Heinis (EPFL), Anastasia Ailamaki (Ecole Polytechnique Federale de Lausanne), Felix Schurmann (EPFL), Henry Markram (EPFL)

Οι νευροεπιστήμονες χρησιμοποιούν όλο και περισσότερο υπολογιστικά εργαλεία για τη δημιουργία και προσομοίωση μοντέλων του εγκεφάλου. Ο όγκος των δεδομένων που εμπλέκονται στις προσομοιώσεις αυτές είναι τεράστιος και η σημασία της αποτελεσματικής διαχείρισης τους καθίσταται πάρα πολύ σημαντική. Ένα ιδιαίτερο πρόβλημα όσον αφορά την ανάλυση αυτών των δεδομένων είναι η κλιμάκωση της εκτέλεσης των επερωτήσεων σε χωρικά μοντέλα του εγκεφάλου. Γνωστές προσεγγίσεις ευρετηρίασης δεν αποδίδουν καλά, ακόμα και στα σημερινά χωρικά μοντέλα μικρής κλίμακας που περιέχουν μόνο μερικά εκατομμύρια πυκνά χωρικά στοιχεία. Το πρόβλημα των σημερινών προσεγγίσεων είναι ότι με την αύξηση του επιπέδου των λεπτομερειών στα μοντέλα, η επικάλυψη στη δενδρική δομή αυξάνει, επιβραδύνοντας το ρυθμό εκτέλεσης των επερωτήσεων. Η ανάγκη των νευροεπιστημόνων να εργαστούν με ολοένα και πιο λεπτομερή

(πυκνότερα) μοντέλα, ενθαρρύνει την ανάπτυξη μιας νέας προσέγγισης ευρετηρίασης. Για το σκοπό αυτό αναπτύξαμε το FLAT, μια κλιμακούμενη μέθοδο ευρετηρίασης για πυκνά σύνολα δεδομένων. Η ανάπτυξη του FLAT βασίστηκε στην καίρια παρατήρηση ότι οι τρέχουσες προσεγγίσεις υποφέρουν από επικάλυψη σε περίπτωση πυκνών συνόλων δεδομένων. Σχεδιάσαμε, συνεπώς, το FLAT ως μια προσέγγιση με δύο φάσεις, καθεμία ανεξάρτητη από την πυκνότητα. Πειραματικά αποτελέσματά επιβεβαιώνουν ότι το FLAT επιτυγχάνει απόδοση ανεξάρτητη από τον όγκο των δεδομένων καθώς και από την πυκνότητα. Εξάλλου, ξεπερνά τις παραλλαγές του R-tree όσον αφορά το κόστος εισόδου/εξόδου από δύο έως οκτώ φορές.

4η Συνεδρία – Φυσική και Λογική Οργάνωση Δεδομένων

Διάδοση Αλλαγών Σε Πληροφοριακά Οικοσυστήματα (Σύντομη παρουσίαση)
George Papastefanatos (Institute for the Management of Information Systems), Panos Vassiliadis* (University of Ioannina), Alkis Simitsis (HP labs)

Στο άρθρο αυτό ασχολούμαστε με τη διαχείριση γεγονότων εξέλιξης σε συστήματα διαχείρισης δεδομένων. Συγκεκριμένα, μας απασχολεί ένα προσανατολισμένο στα δεδομένα οικοσύστημα (data centric ecosystem) που στη μοντελοποίησή του περιλαμβάνει σχέσεις, όψεις και ερωτήσεις. Οι ερωτήσεις είναι μια καλή αρχική αφαίρεση για να μοντελοποιήσουμε πολύ πιο σύνθετες ενότητες (modules), οι οποίες είναι είτε εσωτερικές στη βάση δεδομένων (όπως για παράδειγμα αποθηκευμένες διαδικασίες – stored procedures) ή εξωτερικές (όπως για παράδειγμα, εφαρμογές που προσπελάζουν τη βάση δεδομένων, αναφορές, κλπ). Με άλλα λόγια, το οικοσύστημα επεκτείνεται και εκτός της βάσεως και περιλαμβάνει και τις εφαρμογές που προσπελάζουν τη βάση στη μοντελοποίησή του. Επιπλέον, επισημειώνουμε το πληροφοριακό οικοσύστημα με πολιτικές διαχείρισης εξελικτικών γεγονότων – δηλαδή, για κάθε πιθανή αλλαγή που μπορεί να φτάσει σε μια ενότητα λογισμικού (πίνακας, όψη, ή ερώτηση), η εν λόγω ενότητα είναι επισημειωμένη με πολιτικές που καθορίζουν την αντίδρασή της στο γεγονός της αλλαγής. Θεωρούμε ένα γράφημα που αναπαριστά το όλο οικοσύστημα (το οποίο ονομάζουμε Γράφημα Αρχιτεκτονικής του οικοσυστήματος) και διερευνούμε το πώς είναι εφικτή η διαχείριση της διάδοσης των αλλαγών στο γράφημα της αρχιτεκτονικής. Πρακτικά, κάθε αλλαγή που καταφτάνει σε μια ενότητα ενεργοποιεί και μια πολιτική, η οποία με τη σειρά της είτε τερματίζει τη διάδοση του γεγονότος, είτε καθορίζει την περαιτέρω μετάδοση σε επόμενες ενότητες. Δείχνουμε ότι ο μηχανισμός που προτείνουμε τερματίζει πάντα και στο τέλος κάθε διάδοσης, κάθε ενότητα είναι επισημειωμένη με μία μοναδική κατάσταση (status), ασχέτως της σειράς με την οποία καταφτάνουν σε αυτή οι αλλαγές που διαδίδονται στο γράφημα.

Ετερογενή Σχέδια για την Ρύθμιση Συστημάτων με Αντίγραφα

Mariano Consens (U of Toronto), Kleoni Ioannidu (UC Santa Cruz), Jeff LeFevre (UC Santa Cruz), Alkis Polyzotis (UC Santa Cruz)

Εισάγουμε τα ετερογενή σχέδια ως ένα καινούριο παράδειγμα για την ρύθμιση συστημάτων βάσεων δεδομένων που αναπαράγουν την βάση σε διαφορετικούς υπολογιστικούς κόμβους. Ένα ετερογενές σχέδιο χτίζει διαφορετικές φυσικές δομές (π.χ., ευρετήρια και υλοποιημένες όψεις) σε κάθε αντίγραφο της βάσης, εξειδικεύοντας έτσι το αντίγραφο για την αποτίμηση συγκεκριμένων ομάδων ερωτημάτων. Ορίζουμε το πρόβλημα εύρεσης του βέλτιστου σχεδίου, αναλύουμε την πολυπλοκότητά του προβλήματος και χαρακτηρίζουμε τις ιδιότητες των λύσεων. Στην συνέχεια προτείνουμε έναν ευριστικό αλγόριθμο για τον υπολογισμό καλών ετερογενών σχεδίων, ο οποίος εκμεταλλεύεται υπάρχοντα εργαλεία όπως το Index Tuning Wizard του MS SQL Server ή το db2advis του IBM DB2. Παρουσιάζουμε μια εκτενή πειραματική μελέτη που επιβεβαιώνει την αποτελεσματικότητα του αλγόριθμου καθώς και τα πλεονεκτήματα των ετερογενών σχεδίων.

Έλεγχος Ταυτοχρονισμού σε Προσαρμοστική Ευρετηρίαση

Goetz Graefe (HP Labs), Felix Halim (National University of Singapore), Stratos Idreos* (CWI), Harumi Kuno (Hp Labs), Stefan Manegold (CWI)

Η προσαρμοστικής ευρετηρίαση κατασκευάζει ευρετήρια δυναμικά και σταδιακά. Ο στόχος είναι να πετύχουμε τα οφέλη της παραδοσιακής ευρετηρίασης χωρίς το κόστος της κατασκευής των ευρετηρίων. Παρά ταύτα μια παρενέργεια είναι ότι οι επερωτήσεις ανάγνωσης μετατρέπονται σε ενημερώσεις. Αυτή η εργασία μελετάει τεχνικές ελέγχου ταυτοχρονισμού σε προσαρμοστική ευρετηρίαση. Δείχνουμε ότι ο σχεδιασμός της προσαρμοστικής ευρετηρίασης χωρίζει τα περιεχόμενα από τη δομή των ευρετηρίων. Αυτό μειώνει τους περιορισμούς και τα προβλήματα κατά τον έλεγχο ταυτοχρονισμού για την προσαρμοστική ευρετηρίαση σε σχέση με τις παραδοσιακές ενημερώσεις. Η σχεδίαση μας εκμεταλλεύεται το γεγονός ότι η προσαρμοστική ευρετηρίαση χτίζει ευρετήρια σταδιακά. Παρουσιάζουμε μια λεπτομερή πειραματική ανάλυση που αποδεικνύει ότι η προσαρμοστική ευρετηρίαση διατηρεί της ιδιότητες της ακόμα και κατά τη διάρκεια ταυτόχρονων επερωτήσεων.

Στοχαστικός Κατακερματισμός: Εύρωστη Προσαρμοστική Ευρετηρίαση σε Συστήματα Βάσεων Δεδομένων Οργανωμένα ανά Στήλες στην Κύρια Μνήμη

Felix Halim (National University of Singapore), Stratos Idreos (CWI), Panagiotis Karras (Rutgers University), Roland Yap (National University of Singapore)

Σήμερα απαιτούνται εγγενώς δυναμικά περιβάλλοντα αποθήκευσης δεδομένων. Τέτοια περιβάλλοντα χαρακτηρίζονται από δύο απαιτητικά γνωρίσματα: (α) λίγο αδρανή χρόνο συστήματος, και (β) περιορισμένη εκ των προτέρων γνώση του φόρτου εργασίας, ενώ αυτός ο φόρτος αλλάζει διαρκώς. Σε τέτοια περιβάλλοντα δεν μπορούν να εφαρμοστούν παραδοσιακές λύσεις για την οικοδόμηση και συντήρηση ευρετηρίων. Ο κατακερματισμός βάσεων δεδομένων είναι μια πρόταση που επιτρέπει την φυσική αναδιοργάνωση των δεδομένων στη στιγμή, ως παράπλευρη συνέπεια της επεξεργασίας επερωτήσεων. Ο κατακερματισός επιδιώκει να προσαρμόζει το ευρετήριο στον φόρτο εργασίας συνεχώς και αυτόματα, χωρίς ανθρώπινη παρέμβαση. Το ευρετήριο χτίζεται σταδιακά, προσαρμοστικά, και κατά ζήτηση. Ωστόσο, οι υφιστάμενες μέθοδοι προσαρμοστικής ευρετηρίασης δεν παρέχουν ευρωστία ως προς τον φόρτο εργασίας: αποδίδουν καλά με τυχαίους φόρτους, αλλά όχι με άλλους. Αυτή η αδυναμία απορρέει από την ανελαστικότητα με την οποία ερμηνεύουν κάθε ερώτημα ως έναν υπαινιγμό για το πώς να αποθηκεύονται τα δεδομένα. Τα σημερινά συστήματα κατακερματισμού αναδιοργανώνουν τα δεδομένα εντός της εμβέλειας κάθε ερωτήματος στα τυφλά, ακόμη και αν έτσι προκύπτουν διαδοχικοί δαπανηροί χειρισμοί με ελάχιστο όφελος για την ευρετηρίαση. Στην παρόν άρθρο εισάγουμε τον στοχαστικό κατακερματισμό, μια σημαντικά πιο ανθεκτική λύση για προσαρμοστική ευρετηρίαση. Ο στοχαστικός κατακερματισμός επίσης χρησιμοποιεί κάθε ερώτημα ως έναν υπαινιγμό για το πώς να αναδιοργανώσει τα δεδομένα, αλλά όχι στα τυφλά. Κερδίζει αντοχή και αποφεύγει σημεία συμφόρησης εφαρμόζοντας εσκεμμένα ορισμένες αυθαίρετες επιλογές κατά την λήψη των αποφάσεών του. Έτσι, φέρνουμε την προσαρμοστική ευρετηρίαση σε μια πιο ώριμη μορφοποίηση που παρέχει τη ευρωστία που έλειπε από προηγούμενες προσεγγίσεις. Η πειραματική μας μελέτη επιβεβαιώνει ότι ο στοχαστικός κατακερματισμός διατηρεί τις επιθυμητές ιδιότητες του αρχικού, και επιπλέον αποδίδει καλά με ποικίλους ρεαλιστικούς φόρτους εργασίας.

5η Συνεδρία – Κατανεμημένη Επεξεργασία Επερωτήσεων

Αποδοτική Εξισορρόπηση Φόρτου σε Διαμερισμένα Ερωτήματα κάτω από Τυχαίες Μεταβολές

Anastasios Gounaris (Aristotle University), Christos Yfoulis (ATEI of Thessaloniki), Norman Paton (University of Manchester)

Η συγκεκριμένη εργασία εξετάζει ένα στιγμιότυπο του προβλήματος της σχεδίασης αποδοτικών προσαρμόσιμων συστημάτων, υπό τη συνθήκη ότι κάθε προσαρμογή επιφέρει μη αμελητέο κόστος όταν πραγματοποιείται. Πιο συγκεκριμένα, ασχολούμαστε με το πρόβλημα της

εξισορρόπησης φόρτου σε διαμερισμένα παράλληλα ερωτήματα βάσεων δεδομένων που εκτελούνται σε ετερογενείς κόμβους. Στην εργασία μας, λαμβάνουμε υπόψιν το κόστος των προσαρμογών της κατανομής του φόρτου. Με αυτόν τον τρόπο αποφεύγουμε προβλήματα υπερ-προσαρμογής και λήψης αποφάσεων με εντελώς ευριστικές μεθόδους. Αυτά τα προβλήματα είναι δυνατόν να οδηγήσουν σε σημαντική υποβάθμιση της απόδοσης σε διάφορα σενάρια, όπως τυχαίες μεταβολές του φόρτου. Η προσέγγισή μας βασίζεται στην θεωρία αυτομάτου ελέγχου: ειδικότερα, προτείνουμε έναν γραμμικό τετραγωνικό ελεγκτή, ο οποίος αποτυπώνει τον συμβιβασμό μεταξύ της επίτευξής εξισορρόπησης και του κόστους γι αυτήν την επίτευξη. Η προσέγγισή μας, εκτός του ότι χαρακτηρίζεται και βασίζεται σε αυστηρές θεωρητικές βάσεις, παρουσιάζει καλύτερη απόδοση συγκρινόμενη με τις πλέον σύγχρονες ευριστικές μεθόδους σε ρεαλιστικές καταστάσεις, όπως επιβεβαιώνει η αξιολόγησή μας.

Κατανεμημένη επεξεργασία SPARQL ερωτημάτων πάνω από υπολογιστικά νέφη.
Nikolaos Papailiou (Cslab (ntua), Ioannis Konstantinou (ntua), Dimitrios Tsoumakos (ntua), Panagiotis Karras (Rutgers University), Nectarios Koziris (ntua)

Η μεγάλη αύξηση των διαθέσιμων RDF δεδομένων επιβάλει την εύρεση αποδοτικών και κλιμακώσιμων λύσεων για την διαχείρισή τους. Πρόσφατες προσεγγίσεις έχουν προτείνει κατανεμημένες μεθόδους διαχείρισης των RDF δεδομένων, οι οποίες μπορούν να κλιμακώσουν σε απεριόριστα μεγάλο αριθμό δεδομένων χρησιμοποιώντας πλατφόρμες υπολογιστικών νεφών. Παρόλα αυτά, τα υπάρχοντα συστήματα αποτυγχάνουν να αξιοποιήσουν πλήρως τις δυνατότητες κατανεμημένης επεξεργασίας και αποθήκευσης που παρέχονται από τα υπολογιστικά νέφη. Σε αυτή την εργασία παρουσιάζουμε το H2RDF, μια πλήρως κατανεμημένη βάση αποθήκευσης RDF δεδομένων, η οποία συνδυάζει το πλαίσιο επεξεργασίας του MapReduce με μια κατανεμημένη NoSQL βάση. Υλοποιώντας 3 ευρετήρια RDF τριπλετών σε HBASE πίνακες, το H2RDF μπορεί να απαντήσει SPARQL ερωτήματα σε θεωρητικά απεριόριστο αριθμό RDF δεδομένων. Επιπλέον, σε αντίθεση με τις υπάρχουσες προσεγγίσεις, το H2RDF μπορεί να επεξεργαστεί σύνθετα ερωτήματα με κλιμακώσιμο τρόπο κάνοντας προσαρμοστικές αποφάσεις για την σειρά και τον τρόπο εκτέλεσης των join. Τα join εκτελούνται κατανεμημένα ή κεντρικά, σε έναν υπολογιστή, ανάλογα με το κόστος τους. Η εκτεταμένη πειραματική αξιολόγησή μας δείχνει ότι το H2RDF απαντάει αποδοτικά σε απλά και σύνθετα ερωτήματα, καταφέρνοντας να διαχειριστεί μέχρι 14 δισεκατομμύρια triples χρησιμοποιώντας μια συστοιχία 15 υπολογιστών. Το H2RDF έχει καλύτερη απόδοση από ανταγωνιστικές κατανεμημένες λύσεις σε σύνθετα, μεγάλα ερωτήματα, έχει συγκρίσιμη απόδοση με κεντρικά συστήματα σε μικρά, επιλεκτικά ερωτήματα και κερδίζει σε ρυθμό απάντησης ερωτημάτων εκτελώντας ταυτόχρονα ερωτήματα.

Γεωμετρική Παρακολούθηση Κατανεμημένων Ρευμάτων Δεδομένων Βασισμένη στην Πρόβλεψη

Nikos Giatrakos (University of Piraeus), Antonios Deligiannakis (Technical University of Crete), Minos Garofalakis (Technical University of Crete), Izchak Sharfman (Faculty of Computer Science Technion), Assaf Schuster (Faculty of Computer Science Technion)

Πολλές σύγχρονες εφαρμογές ρευμάτων δεδομένων, όπως η online ανάλυση οικονομικών, δικτυακών, δεδομένων αισθητήρων και άλλων μορφών δεδομένων είναι κατανεμημένης φύσεως. Ένας σημαντικός τύπος επερώτησης στο οποίο εστιάζουν τέτοια σενάρια εφαρμογών αφορά ερωτήματα ενεργοποίησης, όπου λαμβάνεται αντίστοιχη δράση με βάση μια συνθήκη επί της τιμής που λαμβάνει κάποια υπό παρακολούθηση συνάρτηση. Πρόσφατες εργασίες μελετούν το πρόβλημα της παρακολούθησης (μη γραμμικών) πολύπλοκων συναρτήσεων με κατανεμημένο τρόπο. Η βασική ιδέα πίσω από τη γεωμετρική παρακολούθηση που προτείνεται εκεί, είναι κάθε απομακρυσμένος υπολογιστής να παρακολουθεί τη συνάρτηση μέσω ενός κατάλληλου υποσυνόλου του πεδίου ορισμού της. Στην παρούσα εργασία, εξετάζουμε την αύξηση της αποδοτικότητας του μηχανισμού της γεωμετρικής παρακολούθησης, σε όρους του αριθμού των ανταλλασσόμενων μηνυμάτων, επεκτείνοντας το πλαίσιο της γεωμετρικής παρακολούθησης έτσι ώστε να χρησιμοποιεί μοντέλα πρόβλεψης. Αρχικά, περιγράφουμε κάποιους τοπικούς προβλεπτές (predictors) οι οποίοι είναι χρήσιμοι για τις εφαρμογές που εξετάζουμε και έχουν αποδειχθεί χρήσιμοι σε προηγούμενες μελέτες. Έπειτα επιδεικνύουμε τη δυνατότητα ενσωμάτωσης των προβλεπτών στη γεωμετρική μέθοδο και δείχνουμε ότι η παρακολούθηση βασισμένη στην πρόβλεψη στην πραγματικότητα γενικεύει το αρχικό γεωμετρικό πλαίσιο. Προτείνουμε μεγάλη γκάμα διαφορετικών μοντέλων παρακολούθησης βασισμένων στην πρόβλεψη για την κατανεμημένη ιχνηλάτηση πολύπλοκων συναρτήσεων με βάση κάποιο δοθέν κατώφλι. Η εξαντλητική μας πειραματική μελέτη με πληθώρα συνόλων δεδομένων, συναρτήσεων και ρυθμίσεων παραμέτρων δείχνει ότι η προσεγγίσεις μας μπορούν να παρέχουν σημαντικά επικοινωνιακά οφέλη που κυμαίνονται από 2 φορές έως 3 τάξεις μεγέθους σε σύγκριση με το κόστος μετάδοσης του αρχικού γεωμετρικού πλαισίου.

6η Συνεδρία – Ροές Δεδομένων και Συνεχείς Επερωτήσεις

Συνεχής Αναζήτηση των Κ Πλησιέστερων Έξυπνων Κινητών Συσκευών για Κάθε Χρήστη

Georgios Chatzimilioudis (University of Cyprus), Demetris Zeinalipour (University of Cyprus), Wang-Chien Lee (Pennsylvania State University), Marios Dikaiakos (University of Cyprus)

Μία άκρως επιθυμητή λειτουργία των κινητών συσκευών είναι η συνεχής παροχή των K γεωγραφικά πλησιέστερων γειτόνων. Αυτή η λειτουργία ονομάζεται συνεχής επερώτηση K-πλησιέστερων γειτόνων για όλους (CAkNN). Μια τέτοια λειτουργία θα ενίσχυε τις δημόσιες υπηρεσίες έκτακτης ανάγκης, επιτρέποντας στον χρήστη να στείλει σήματα SOS στους πλησιέστερους χρήστες, ή εφαρμογές παιγνίων, επιτρέποντας στο χρήστη να αλληλεπιδράσει με τους γύρω του παίκτες. Στην παρούσα εργασία, μελετάμε το πρόβλημα της αποτελεσματικής επεξεργασίας μιας CAkNN επερώτησης σε ασύρματα δίκτυα έξυπνων κινητών συσκευών. Παρουσιάζουμε τον αλγόριθμος μας, Proximity, ο οποίος απαντά μια επερώτηση CAkNN σε χρόνο O(n(K+λ)), όπου το n δηλώνει τον αριθμό των χρηστών και το λ την χωρητικότητα ενός δικτυακού σημείου σύνδεσης (λ << n). Ο Proximity δεν κάνει χρήση πρόσθετων υποδομών ή εξειδικευμένου υλικού και η αποτελεσματικότητα του επιτυγχάνεται κυρίως λόγω της έξυπνης κοινής χρήσης του πεδίου αναζήτησης. Η υλοποίηση βασίζεται σε μια νέα δομή δεδομένων με όνομα K+-σωρός, η οποία επιτυγχάνει σταθερό χρόνο O(1) πρόσβασης και λογαριθμικό χρόνο O(log(K*λ)) εισαγωγής και ενημέρωσης. Ο Proximity, δεν χρειάζεται παραμέτρους και είναι αποτελεσματικός σε περιβάλλοντα υψηλής κινητικότητας και ασύμμετρης κατανομής χρηστών (π.χ., η υπηρεσία λειτουργεί εξίσου καλά στο κέντρο μιας πόλης, στα προάστια, και σε αγροτικές περιοχές). Έχουμε αξιολογήσει τον Proximity χρησιμοποιώντας τροχιές από δύο διαφορετικές πηγές και συμπεραίνουμε ότι η προσέγγισή μας είναι τουλάχιστον μία τάξη μεγέθους γρηγορότερη σε σύγκριση με προσαρμοσμένες υφιστάμενες τεχνικές.

Δυναμική Υποστήριξη Ποικιλομορφίας σε Ροές Δεδομένων (Dynamic Diversification of Continuous Data)

Marina Drosou (University of Ioannina), Evaggelia Pitoura (University of Ioannina)

Η ποικιλομορφία (diversity) των αποτελεσμάτων έχει αποδειχθεί ότι αυξάνει την ικανοποίηση των χρηστών σε εφαρμογές όπως τα συστήματα συστάσεων αλλά και στην αναζήτηση στον παγκόσμιο ιστό και σε βάσεις δεδομένων. Σε αυτήν την εργασία, εστιαζόμαστε στο πρόβλημα της επιλογής των k πιο διαφορετικών από ένα σύνολο αποτελεσμάτων. Σε αντίθεση με την προηγούμενη βιβλιογραφία η οποία εξετάζει κυρίως την στατική έκδοση του προβλήματος, σε αυτήν την εργασία, εξετάζουμε την δυναμική περίπτωση όπου τα δεδομένα αλλάζουν με την πάροδο του χρόνου, όπως στην περίπτωση των υπηρεσιών αποστολής ειδοποιήσεων. Ορίζουμε το συνεχές πρόβλημα της k-ποικιλομορφίας σε συνδυασμό με κατάλληλους περιορισμούς οι οποίοι παρέχουν ιδιότητες συνέχειας στα αποτελέσματα που παρουσιάζονται στους χρήστες. Η μέθοδος που προτείνουμε βασίζεται στα δέντρα επικάλυψης (cover trees) και υποστηρίζει δυναμικές εισαγωγές και διαγραφές δεδομένων. Το πρόβλημα της k-ποικιλομορφίας είναι γενικά NP-πλήρες. Δίνουμε θεωρητικά φράγματα για την ποιότητα της λύσης μας σε σχέση με τη βέλτιστη λύση. Τέλος, παρουσιάζουμε

πειραματικά αποτελέσματα για την απόδοση της μεθόδου μας χρησιμοποιώντας πραγματικά και συνθετικά δεδομένα.

Αναγνώριση bursts σε χωροχρονικές ροές εγγράφων

Theodoros Lappas (UC Riverside), Marcos Vieira (UC Riverside), Dimitris Gunopulos (University of Athens), Vassilis Tsotras (UC Riverside)

Χιλιάδες έγγραφα γίνονται διαθέσιμα στους χρήστες μέσω του διαδικτύου σε καθημερινή βάση. Ένα από τα πιο εκτενώς μελετημένα προβλήματα είναι η αναγνώριση bursts σε ροές εγγράφων. Για ένα όρο t, ένα burst θεωρείται γενικά ότι συμβαίνει όταν παρατηρείται μια ασυνήθιστα υψηλή συχνότητα για το t. Ενώ η αναγνώριση χωρικών ή χρονικών bursts έχει μελετηθεί στο παρελθόν, αυτή η δουλειά είναι η πρώτη που γίνεται αναγνώριση και μέτρηση χωροχρονικών bursts. Επιπλέον, χρησιμοποιούμε την πληροφορία που εξορύσσεται από τα bursts σε μια μηχανή αναζήτησης εγγράφων: για κάθε ερώτημα ενός χρήστη, η μηχανή επιστρέφει μια λίστα κατάταξης των εγγράφων με ισχυρό αντίκτυπο χωροχρονικό αντίκτυπο. Δείχνουμε την αποτελεσματικότητα των μεθόδων μας, με μια εκτεταμένη πειραματική αξιολόγηση σε πραγματικά και συνθετικά σύνολα δεδομένων.

Τριεπίπεδη Εκτέλεση Πολλαπλών Συγκεντρωτικών Ερωτημάτων Διάρκειας

Shenoda Guirguis (University of Pittsburgh), Mohamed Sharaf (University of Queensland), Panos K. Chrysanthis (University of Pittsburgh), Alexandros Labrinidis (University of Pittsburgh)

Τα Συγκεντρωτικά Ερωτήματα Διάρκειας (ΣΕΔ) είναι μια σημαντική κατηγορία Ερωτημάτων Διάρκειας η οποία εκτελείται πολύ συχνά με πιθανό μεγάλο κόστος εκτέλεσης. Με το δεδομένο αυτό, η βελτιστοποίηση των ΣΕΔ είναι απαραίτητη ώστε τα συστήματα διαχείρισης ροών δεδομένων (ΣΔΡΔ) να μπορέσουν να υποστηρίξουν στο μέγιστο κρίσιμες εφαρμογές παρακολούθησης. Τα υφιστάμενα συστήματα βελτιστοποίησης πολλαπλών ΣΕΔ με διαφορετικές προδιαγραφές στα παράθυρα και φίλτρα προ---συγκέντρωσης υποθέτουν ένα μοντέλο επεξεργασίας όπου κάθε ΣΕΔ υπολογίζεται ως τελικό άθροισμα υπο--αθροισμάτων. Στην εργασία αυτή προτείνουμε ένα νέο μοντέλο για την επεξεργασία κάθε ΣΕΔ, που ονομάζεται TriOps, με στόχο την ελαχιστοποίηση της επανάληψης των υπολογισμών στο επίπεδο των υπο-αθροισμάτων. Προτείνουμε επίσης τον βελτιστοποιητή TriWeave για πολλαπλά ΣΕΔ που χρησιμοποιούν το TriOps και την έννοια της Πλεξιμότητας ("Weaveability"). Έχουμε δείξει αναλυτικά και πειραματικά τα οφέλη από τις επιδόσεις των προτεινόμενων μοντέλων μας. Η ανάλυσή μας δείχνει την υπεροχή τους έναντι των εναλλακτικών μοντέλων. Τέλος, γενικεύουμε το TriWeave ενσωματώνοντας τις τεχνικές υπαγωγής της κλασικής βελτιστοποίησης πολλαπλών

συγκεντρωτικών ερωτημάτων για να υποστηρίζει επικαλυπτόμενα πεδία ομαδοποίησης. Αυτή η εργασία εμφανίζεται στα πρακτικά του 2012 IEEE International Conference on Data Engineering (ICDE).

7η Συνεδρία – Επιδείξεις

Αλληλεπιδράσεις εγγύτητας με το Crowdcast

Marios Constantinides (University of Cyprus), G Constantinou (University of Cyprus), Andreas Panteli (University of Cyprus), Theofilos Phokas (University of Cyprus), Georgios Chatzimilioudis (University of Cyprus), Demetris Zeinalipour (University of Cyprus)

Παρουσιάζουμε το Crowdcast, μια σουίτα εφαρμογών που βασίζονται στην εγγύτητα και έχουν ως πυρήνα το πλαίσιο αλγορίθμων του Proximity. Το Proximity συνδέει αποτελεσματικά έναν χρήστη με τους πλησιέστερους γείτονές του ανά πάσα στιγμή, ανεξάρτητα από το πού βρίσκεται ο χρήστης και πόσο μακριά είναι οι γείτονές του. Με μια τέτοια υπηρεσία, υλοποιεί τον ειδικό τελεστή μας που λύνει *Συνεχείς Επερωτήσεις των Κ-κοντινότερων Γειτόνων (CAkNN)* αποτελεσματικά. Το Proximity δεν απαιτεί οιαδήποτε πρόσθετη υποδομή ή εξειδικευμένο υλικό και η αποδοτικότητά του επιτυγχάνεται κυρίως λόγω της έξυπνης κοινής χρήσης του πεδίου αναζήτησης που αναπτύσσουμε.

Η σουίτα εφαρμογών Crowdcast δείχνει πως οι αλγόριθμοι διαχείρισης δεδομένων του Proximity μπορούνε να δώσουν ώθηση σε νέες υπηρεσίες με βάση την εγγύτητα. Κατά τη διάρκεια του συνεδρίου, οι συμμετέχοντες θα έχουν την δυνατότητα να χρησιμοποιήσουν τις εφαρμογές Crowdcast σε όλο το χώρο του συνεδρίου. Θα είναι σε θέση να (i) ανακοινώσουν κείμενο ή φωνητικό μηνύματα στον πίνακα ανακοινώσεων της γειτονίας τους, ο οποίος θα είναι ορατός στους Κ πλησιέστερους γείτονες. Για παράδειγμα, ένας συμμετέχων μπορεί να ξεκινήσει μια συζήτηση με άλλους συμμετέχοντες στην ίδια συνεδρίαση για να διευκρινίσει τα ζητήματα σχετικά με την παρουσίαση χωρίς να ενοχλήσει, (ii) επεκτείνουν την όραση ή την ακοή τους σχετικά με τις δραστηριότητες των παρουσιάσεων με τις κάμερες και τα μικρόφωνα των γειτόνων τους, (iii) ανακοινώνουν τοπικές εργασίες στη γειτονία τους ως μέρος της οργάνωσης μιας δραστηριότητας, κλπ.

Κλιμακώσιμη και Ευέλικτη Υποστήριξη Στιγμιαίας Επισκοπικής Αναζήτησης

Pavlos Fafalios (Institute of Computer Science, FORTH-ICS), Ioannis Kitsos (Institute of Computer Science, FORTH-ICS), Yannis Tzitzikas (Institute of Computer Science, FORTH-ICS)

Τα τελευταία χρόνια υπάρχει αυξανόμενο ενδιαφέρον για υπηρεσίες που εμφανίζουν στιγμιαία τα κορυφαία αποτελέσματα αναζήτησης καθώς ο χρήστης πληκτρολογεί το ερώτημά του χαρακτήρα-χαρακτήρα. Στην εργασία αυτή παρουσιάζουμε (και επιδεικνύουμε) μια οικογένεια εφαρμογών στιγμιαίας αναζήτησης (instant search) οι οποίες εκτός των κορυφαίων αποτελεσμάτων της πληκτρολογούμενης ερώτησης, εμφανίζουν στιγμιαία στο χρήστη πρόσθετη προϋπολογισμένη πληροφορία συγκεντρωτικής φύσεως (εξ' ου ο όρος «στιγμιαία επισκοπική αναζήτηση»). Η λειτουργικότητα αυτή είναι πιο βοηθητική από τις υπάρχουσες αφού μπορεί να συνδυάσει την αυτόματη συμπλήρωση ερωτήσεων (query autocompletion), την ομαδοποίηση αποτελεσμάτων (results clustering), τη πολυδιάστατη αναζήτηση (faceted search), την εξόρυξη οντοτήτων (entity mining), κ.α. Συνάμα η λειτουργικότητα αυτή είναι υπολογιστικά επωφελής για το διακομιστή. Εντούτοις, η στιγμιαία παροχή τέτοιων υπηρεσιών για μεγάλο αριθμό επερωτήσεων, μεγάλη ποσότητα προϋπολογισμένης πληροφορίας και πολλούς ταυτόχρονα εξυπηρετούμενους χρήστες, αποτελεί πρόκληση. Θα επιδείξουμε πώς αυτό μπορεί επιτευχθεί με συμβατικό υλικό (hardware) χρησιμοποιώντας (α) διαμερισμένα ευρετήρια προθεμάτων (Tries) τα οποία αξιοποιούν τη διαθέσιμη κύρια και δευτερεύουσα μνήμη, και (β) ειδικές τεχνικές προσωρινής αποθήκευσης (caching). Αναφέρουμε τις επιδόσεις της μεθόδου επί ενός συμβατικού υπολογιστή (με 3 GigaBytes κύρια μνήμη) ο οποίος προσφέρει υπηρεσίες στιγμιαίας επισκοπικής αναζήτησης για εκατομμύρια διαφορετικές επερωτήσεις και αξιοποιεί προϋπολογισμένη πληροφορία κλίμακας terabyte. Επιπροσθέτως, οι παρεχόμενες υπηρεσίες είναι ανεκτικές σε ορθογραφικά λάθη και στη σειρά των πληκτρολογούμενων λέξεων.

EVISUGE: Οπτικοποίηση Γεγονότων στο Google Earth

Nikolaos Tarantilis (Technical University of Crete), Chrisa Tsinaraki (Technical University of Crete), Fotis Kazasis (Technical University of Crete), Nektarios Gioldasis (Technical University of Crete), Stavros Christodoulakis (Technical University of Crete)

Σε αυτό το άρθρο παρουσιάζεται το EVISUGE, ένα σύστημα που επιτρέπει την οπτικοποίηση και διαχείριση σεναρίων του πραγματικού κόσμου στο Google Earth. Ένα EVISUGE σενάριο απαρτίζεται από γεγονότα, τα οποία αναπαριστώνται σύμφωνα με το μοντέλο αναπαράστασης γεγονότων MOME (MObile Multimedia Event Capturing and Visualization) που έχουμε αναπτύξει. Τα γεγονότα των σεναρίων οπτικοποιούνται με βάση τα χωρικά και χρονικά χαρακτηριστικά τους πάνω από τους τρισδιάστατους αλληλεπιδραστικούς χάρτες του Google Earth. Οι δυνατότητες του συστήματος EVISUGE επιδεικνύονται μέσω πραγματικών σεναρίων: (α) Τον καθορισμό, τη διάσχιση και την οπτικοποίηση μιας φυσιολατρικής διαδρομής; και (β) Τη χωροχρονική αναπαράσταση και οπτικοποίηση μαχών.

8η Συνεδρία – Ανάκτηση Πληροφορίας, Εξόρυξη Δεδομένων, Ιδιωτικότητα

Μεγέθους-λ Περιλήψεις Αντικειμένων για Σχεσιακή Αναζήτηση με Λέξεις-κλειδιά
Georgios Fakas (Manchester Metropolitan Univer), Nikos Mamoulis (University of Hong Kong), Zhi Cai (Manchester Metropolitan University)

Μια μεθοδολογία αναζήτησης με λέξεις--κλειδιά, που προτάθηκε στο παρελθόν, παράγει ως αποτέλεσμα μιας επερώτησης, ένα κατάλογο περιλήψεων αντικειμένων (ΠΑ). Η ΠΑ είναι μια δομή δέντρου η οποία συνενώνει πλειάδες, συνοψίζοντας όλα την πληροφορία σε μια σχεσιακή βάση δεδομένων που σχετίζονται με ένα συγκεκριμένο υποκείμενο δεδομένων (ΥΔ). Ωστόσο, μερικές ΠΑ είναι πολύ μεγάλες σε μέγεθος και ως εκ τούτου μη φιλικές προς τους χρήστες που προτιμούν αρχικά συνοπτικότερες πληροφορίες. Στην παρούσα εργασία, διερευνάται η αποτελεσματική και αποδοτική ανάκτηση συνοπτικών ΠΑ (μεγέθους-λ ΠΑ). Θεωρούμε ότι μια καλή μεγέθους-λ ΠΑ θα πρέπει να είναι μια αυτόνομη και ουσιαστική σύνοψη των πιο σημαντικών πληροφοριών σχετικά με το συγκεκριμένο ΥΔ. Πιο συγκεκριμένα, ορίζουμε την μεγέθους-λ ΠΑ σαν μια μερική ΠΑ που περιέχει τις λ πιο σημαντικές πλειάδες. Προτείνουμε τρείς αλγορίθμους για τον αποδοτικό υπολογισμό μεγέθους--- λ ΠΑ (επιπλέον του βέλτιστου αλγορίθμου που απαιτεί εκθετικό χρόνο). Η πειραματική μας μελέτη χρησιμοποιώντας τις βάσεις DBLP και TPC-H επαληθεύει την αποδοτικότητα της μεθόδου μας.

Εκπαίδευση Αναζήτησης Αποτελεσμάτων με Βάση το Σκοπό Αναζήτησης (Σύντομη παρουσίαση)
Giorgos Giannopoulos (NTU Athens - Athena R.C.)

Τα εξατομικευμένα μοντέλα ανάκτησης πληροφορίας στοχεύουν στον εντοπισμό των ενδιαφερόντων αναζήτησης των χρηστών με σκοπό να παρέχουν εξατομικευμένα αποτελέσματα, προσαρμοσμένα στις αντίστοιχες ανάγκες αναζήτησης. Τα ενδιαφέροντα των χρηστών, όμως, είναι ετερογενή, ευμετάβλητα και δύσκολα αναγνωρίσιμα, καθιστώντας δύσκολο το πρόβλημα της εφαρμογής εξατομικευμένων μοντέλων. Σε αυτή τη δουλειά ακολουθούμε μία διαφορετική προσέγγιση, μελετώντας μοντέλα με βάση το σκοπό αναζήτησης. Εκμεταλλευόμαστε την ανάδραση του χρήστη, σε μορφή αποτελεσμάτων που πάτησε, ώστε να ομαδοποιήσουμε μοντέλα ταξινόμησης αποτελεσμάτων με βάση τη συμπεριφορά και το σκοπό αναζήτησης. Κάθε ομάδα

(συστάδα) που προκύπτει αντιστοιχίζεται σε ένα διαφορετικό μοντέλο ταξινόμησης, το οποίο αντιπροσωπεύει συγκεκριμένα ενδιαφέροντα-συμπεριφορές αναζήτησης χρηστών. Κάθε νέο ερώτημα που τίθεται, αντιστοιχίζεται στις ομάδες αυτές, έτσι ώστε η ταξινόμηση των αποτελεσμάτων του να εκμεταλλεύεται διαφορετικά μοντέλα ταξινόμησης, συνδυάζοντας διαφορετικές συμπεριφορές αναζήτησης. Πειραματική αποτίμηση της μεθόδου μας σε σετ δεδομένων μεγάλης διαδικτυατής εταιρίας για αναζήτησης αποδεικνύει ότι η μέθοδος μας είναι πιο αποτελεσματική από υπάρχουσες προσεγγίσεις.

Διατήρηση της κ-ανωνυμίας σε κατανεμημένα δεδομένα
Katerina Doka (NTUA), Dimitrios Tsoumakos (NTUA), Nectarios Koziris (NTUA)

Σε αυτή την εργασία περιγράφουμε ένα κατανεμημένο σύστημα, σχεδιασμένο να διατηρεί την ανωνυμία πολυδιάστατων, ιεραρχικών δεδομένων. Ενώ επιτρέπει την αποδοτική αποθήκευση, δεικτοδότηση και αναζήτηση των δεδομένων, το σύστημα εφαρμόζει μία προσαρμοστική μέθοδο που αλλάζει με αυτόματο τρόπο το επίπεδο δεικτοδότησης ανάλογα με τους περιορισμούς της ιδιωτικότητας: Αποδοτικές λειτουργίες συσσώρευσης (roll-up) και εμβάθυνσης (drill-down) λαμβάνουν χώρα ώστε να διασφαλίζεται κ-ανωνυμία, ελαχιστοποιώντας ταυτόχρονα την αλλοίωση και την ασυνέπεια των δεδομένων. Η αρχική πειραματική αποτίμηση αποδεικνύει ότι το σύστημα επιτυγχάνει τη διατήρηση της ανωνυμίας των δεδομένων με κατανεμημένο τρόπο, ακόμα και κατά την εισαγωγή νέων δεδομένων, χωρίς να χάνει την ικανότητα να απαντά επερωτήσεις. Συγκρινόμενο με μια δημοφιλή κεντρική μέθοδο ολικής ανωνυμίας, το σύστημά μας βελτιώνει την ποιότητα των δεδομένων μέχρι και 22%, επιτυγχάνοντας σχεδόν βέλτιστη χρησιμότητα ανεξάρτητα από το μέγεθος του δικτύου ή των δεδομένων, με λογικό κόστος επικοινωνίας που διαμοιράζεται ανάμεσα στους συμμετέχοντες κόμβους.

Υποχώροι υψηλής αντίθεσης για τη βαθμολόγηση εκτόπων με βάση την πυκνότητα
Fabian Keller (Karlsruhe Institute of Technology (KIT)), Emmanuel Müller (Karlsruhe Institute of Technology (KIT)), Klemens Böhm (Karlsruhe Institute of Technology (KIT)

Η εύρεση εκτόπων είναι ένα σημαντικό μέρος της ανάλυσης δεδομένων. Έκτοπα είναι εξαιρετικά δεδομένα που διακρίνονται από τα συνήθη δεδομένα. Η βαθμολόγηση εκτόπων προτείνει τη μέτρηση του βαθμού εκτοπικότητας με βάση την πυκνότητα σε τοπικές γειτονιές. Σε πολλές εφαρμογές με πολυδιάστατες βάσεις δεδομένων, η βαθμολόγηση αυτή δεν μπορεί να ανακαλύψει τα έκτοπα. Λόγω της χαμηλής αντίθεσης μεταξύ εκτόπων και υπόλοιπων δεδομένων, παραδοσιακές μέθοδοι δίνουν μια τυχαία σειρά δεδομένων ως αποτέλεσμα. Έκτοπα δεν εμφανίζονται στο διασκορπισμένο ολικό χώρο των δεδομένων, αλλά είναι κρυμμένα σε

υποχώρους υψηλής αντίθεσης. Η μέτρηση της αντίθεσης για την εύρεση εκτόπων είναι ένα άλυτο ερευνητικό πρόβλημα. Προτείνουμε μια νέα μέθοδο για τον εντοπισμό υποχωρών με υψηλή αντίθεση. Η επιλογή σχεδιάστηκε ως προ-επεξεργασία για τη βαθμολόγηση εκτόπων με βάση την πυκνότητα και εντοπίζει υψηλή αντίθεση σε υποχώρους με σημαντικό ποσό εξάρτησης μεταξύ των επιλεγμένων διαστάσεων. Με τη μέθοδό μας προτείνουμε το πρώτο μέτρο αντίθεσης για υποχώρους και βελτιώνουμε την ποιότητα της μέτρησης εκτοπικότητας με την επιλογή υψηλής αντίθεσης. Τα εμπειρικά αποτελέσματα αποδεικνύουν ότι η μέθοδός μας ξεπερνά παραδοσιακές τεχνικές που χρησιμοποιούν μείωση των διαστάσεων, τυχαίες προβολές και άλλες τεχνικές για τον εντοπισμό υποχώρων. Η μέθοδός μας δείχνει υψηλή ποιότητα στην εύρεση εκτόπων σε συνθετικά και πραγματικά δεδομένα.

9η Συνεδρία – Ομιλίες Χορηγών

10η Συνεδρία – Ακολουθίες και Συστάδες

Μια νέα τεχνική υπολογισμού ομοιότητας ακολουθιών με εφαρμογή σε Query-By-Humming
Alexios Kotsifakos (University of Athens), Panagiotis Papapetrou (Aalto University), Jaakko Hollmen (Aalto University), Dimitris Gunopulos (University of Athens)

Προτείνουμε μια νέα τεχνική που υπολογίζει την ομοιότητα ανάμεσα σε δύο ακολουθίες. Η τεχνική επιτρέπει κενά στο ταίριασμα της ακολουθίας-ερώτημα και της ακολουθίας-στόχο, και μεταβλητά επίπεδα ανοχής στο ταίριασμα τιμών. Χρησιμοποιώντας αυτή την τεχνική η μέθοδός μας αναγνωρίζει την ακολουθία σε μια βάση δεδομένων που ταιριάζει καλύτερα με το ερώτημα. Δείχνουμε ότι η προτεινόμενη μέθοδος είναι ιδιαίτερα αποτελεσματική για την αναζήτηση μουσικής. Τα κομμάτια μουσικής αναπαριστούνται από 2-διαστάσεων χρονοσειρές, όπου οι διαστάσεις έχουν πληροφορία για το ύψος και τη διάρκεια της κάθε νότας, αντίστοιχα. Στο χρόνο εκτέλεσης, το ερώτημα τραγούδι μεταμορφώνεται στην ίδια 2-διάστατη αναπαράσταση. Παρουσιάζουμε μια εκτενή πειραματική αξιολόγηση με τη χρήση συνθετικών και πρακτικών ερωτημάτων σε μια μεγάλη βάση δεδομένων μουσικής.

Συσταδοποίηση προβολής βάσει πυκνότητας σε ροές δεδομένων υψηλών διαστάσεων
Eirini Ntoutsi (LMU), Arthur Zimek (LMU), Themis Palpanas (University of Trento), Peer Kroeger (LMU), Hans-Peter Kriegel (LMU)

Η συσταδοποίηση ροών δεδομένων υψηλών διαστάσεων αποτελεί σημαντικό πρόβλημα σε πολλές εφαρμογές. Διάφορες προσεγγίσεις έχουν προταθεί οι οποίες αντιμετωπίζουν ανεξάρτητα τις επιμέρους προκλήσεις του προβλήματος. Συγκεκριμένα, υπάρχουν μέθοδοι για τον εντοπισμό συστάδων σε ρεύματα δεδομένων στον πολυδιάστατο χώρο γνωρισμάτων και μέθοδοι για τον εντοπισμό συστάδων σε στατικά δεδομένα σε υποσύνολα διαστάσεων του χώρου γνωρισμάτων. Ωστόσο ελάχιστες προσεγγίσεις υπάρχουν που να αντιμετωπίζουν ταυτόχρονα και τις δύο προκλήσεις του προβλήματος, δηλαδή το μεγάλο πλήθος διαστάσεων του χώρου γνωρισμάτων και την δυναμική φύση των ρευμάτων δεδομένων. Στην παρούσα εργασία, προτείνουμε έναν νέο αλγόριθμο συσταδοποίησης προβολής βάσει πυκνότητας, τον HDDStream, για ρεύματα δεδομένων υψηλών διαστάσεων. Ο προτεινόμενος αλγόριθμος συνοψίζει τόσο τα σημεία όσο και τις διαστάσεις όπου αυτά τα σημεία συνθέτουν ομάδες και διατηρεί αυτές τις συνόψεις online, καθώς νέα σημεία καταφθάνουν και παλιά σημεία ακυρώνονται λόγω ηλικίας. Τα πειραματικά μας αποτελέσματα αποδεικνύουν την αποτελεσματικότητα και την αποδοτικότητα του προτεινόμενου αλγορίθμου. Επίσης αποδεικνύουν ότι ο HDDStream θα μπορούσε να χρησιμεύσει για τον εντοπισμό δραστικών αλλαγών στον υπό-μελέτη πληθυσμό, π.χ., εντοπισμό έξαρσης επιθέσεων σε ένα δίκτυο.

Προσεγγιστικό τμηματικό ταίριασμα ακολουθιών

Thanasis Vergoulis (NTUA & Athena RC), Theodore Dalamagas (NTUA & Athena RC), Dimitris Sacharidis (NTUA & Athena RC), Timos Sellis (NTUA & Athena RC)

Για την εξυπηρέτηση της έρευνας στις βιοεπιστήμες, προκύπτουν συνεχώς νέες ανάγκες για ανάλυση ακολουθιών και, συνεπώς, νέα προβλήματα. Σε αυτή την εργασία, εισάγουμε ένα νέο πρόβλημα ταιριάσματος ακολουθιών, το πρόβλημα του προσεγγιστικού τμηματικού ταιριάσματος ακολουθιών. Δεδομένης μιας ακολουθίας δεδομένων και μιας ακολουθίας-προτύπου, μια απάντηση σε αυτό το πρόβλημα είναι μια προσεγγιστική εμφάνιση ενός τμήματος του προτύπου μέσα στην ακολουθία δεδομένων, με την προϋπόθεση ότι ισχύουν δύο συνθήκες. Πρώτον, το τμήμα πρέπει να περιέχει μια προκαθορισμένη περιοχή του προτύπου, την οποία ονομάζουμε πυρήνα. Δεύτερον, η επιτρεπόμενη διαφοροποίηση μεταξύ του τμήματος του προτύπου και της εμφάνισής του στα δεδομένα εξαρτάται από το μήκος του τμήματος. Προτείνουμε τη μέθοδο PS-ARSM, η οποία επεξεργάζεται μαζικά όλα τα τμήματα ενός προτύπου για να παράξει τα προσεγγιστικά τμηματικά ταιριάσματα. Η αποδοτικότητά της αποτιμάται σε σχέση με τις υπάρχουσες τεχνολογίες που μπορούν να μετασχηματιστούν για να απαντήσουν στο συγκεκριμένο πρόβλημα.