05421 Abstracts Collection
Data Always and Everywhere
— Dagstuhl Seminar —

Gustavo Alonso[1], Christian S. Jensen[2] and Bernhard Mitschang[3]

[1] ETH Zürich, CH
alonso@inf.ethz.ch
[2] Aalborg University, DK
csj@cs.auc.dk
[3] Univ. Stuttgart, DE
mitsch@informatik.uni-stuttgart.de

Abstract. From 16.10.05 to 21.10.05, the Dagstuhl Seminar 05421, Data Always and Everywhere - Management of Mobile, Ubiquitous, Pervasive, and Sensor Data, was held in the International Conference and Research Center, Schloss Dagstuhl. During the seminar, all participants were given the opportunity to present their current research, and ongoing activities and open problems were discussed. This document is a collection of the abstracts of the presentations given during the seminar. Some abstracts offer links to extended abstracts, full papers, and other supporting documents. A separate companion document summarizes the seminar.
The authors wish to acknowledge Victor Teixeira de Almeida, who served as collector for the seminar and thus played a key role in collecting materials from the seminar participants.

Keywords. Mobile ubiquitous and pervasive computing, sensor data, data streams, content integration, replication, caching, and consistency, service orientation, process models, peer-to-peer computing, mobile ad-hoc networking, context awareness and preferences, moving objects, location-based mobile services, query and update processing, indexing, tracking

05421 Executive Summary – Data Always and Everywhere – Management of Mobile, Ubiquitous, and Pervasive Data

This report summarizes the important aspects of the workshop on "Management of Mobile, Ubiquitous, and Pervasive Data", which took place from October 16th to October 21st, 2005. Thirty-seven participants from thirteen countries met during that week and discussed a broad range of topics related to the management of data in relation to mobile, ubiquitous, and pervasive applications of information technology. The wealth of the contributions is available at the seminar page at the Dagstuhl server. Here, we provide a short overview.

Dagstuhl Seminar Proceedings 05421
Data Always and Everywhere - Management of Mobile, Ubiquitous, Pervasive, and Sensor Data
http://drops.dagstuhl.de/opus/volltexte/2006/796

Keywords: Mobile Data Management, Ubiquitous Computing, Pervasive Computing, Streaming Data, Middleware, Data Integration, Data Placement, Ad-hoc Networking, Micro DBMSs, Context-Aware Applications

Joint work of: Alonso, Gustavo; Jensen, Christian S.; Mitschang, Bernhard

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2006/794

Data from sensor networks

Gustavo Alonso (ETH Zürich, CH)

Database work exploring sensor networks tends to make the assumption that teh data streams produced by these networks are complete and accurate. In this brief presentation I show several examples to the contrary. In practice, sensor networks tend to be very unreliable and there is a data cleaning and pre-processing step that is unavoidable in almost all applications. The question is then how much real data streaming can be done on such data and where the different processing functionalities should be located. There is also the question of why trying to use techniques designed for close worlds and complete data when sensor networks only provide incomplete and inaccurate information.

Keywords: Sensor networks

AmbientDB: P2P database middleware for ubiquitous computing

Peter A. Boncz (CWI - Amsterdam, NL)

The future generation of consumer electronics devices is envisioned to provide automatic cooperation between devices and run applications that are sensitive to people's likings, personalized to their requirements, anticipatory of their behavior and responsive to their presence. We see this 'Ambient Intelligence' as a key feature of future pervasive computing. We focus here on one of the challenges in realizing this vision: information management. This entails integrating, querying, synchronizing and evolving structured data, on a heterogeneous and ad-hoc collection of (mobile) devices. Rather than hard-coding data management functionality in each individual application, we argue for adding highlevel data management functionalities to the distributed middleware layer. Our AmbientDB P2P database management system addresses this by providing a global database abstraction over an ad-hoc network of heterogeneous peers

Keywords: Ubiquitous computing, pervasive computing, ambient intelligence, P2P databases

Full Paper:
http://www.cwi.nl/∼boncz/ambientdb.html

Building a Spatial Data Integration System: Lessons Learned

Omar Boucelma (LSIS - Marseille, F)

With the proliferation of Geographic Information Systems (GIS) and spatial resources over the Internet, there is an increasing demand for efficient data integration solutions that allow federation/interoperation of massive repositories of heterogeneous spatial data and metadata.

During this talk, we will attempt to discuss some lessons learned when building VirGIS, an integration system that follows a mediation approach and complies with OpenGIS consortium (OGC) recommendation.

Keywords: GIS, Mediation, Heterogeneity, Schema Matching, Query Rewriting, GML, WFS

Data in Crisis Management: An Exercise for Mobile databases, Sensor databases and Peer-to-Peer databases

Panos Kypros Chrysanthis (University of Pittsburgh, USA)

A key problem in disaster and emergency management is the ability to efficiently search, exchange, integrate, analyze, and update large amounts of data needed to support decision-making by first responders, emergency managers in federal, state, and local agencies operating in dynamic environments. In this talk, we present a new architecture for enabling reliable data dissemination, message passing and filtering for disaster response management. Our architecture, which is called Whiteboard P2P (WBp2p), attempts to aggregate the benefits that exist from two predominant architectures, the client/server and the peer-to-peer architecture. The system we are designing acts as an innovative data stream processing system and integrates data from sensors (human and devices). WBp2p supports both continues and ad-hoc sensor queries with in-network processing and in-network storage for energy efficiency. Finally, our system was designed not only to provide highly sclable delivery of messages but also allow for disconnected operation while facilitating the dynamic restructuring of the network as required. Our underlying design goal is to improve disaster response, by providing quick, easy, and accurate information to a variety of mobile and stationary actors in disaster response while enabling coordination following the necessary rules, QoS and QoD requirements.

This is collaborative work with Alexandros Labrinidis and the members of the Advanced Data Management Technologies Lab at the University of Pittsburgh. It is funded in part by the NSF ITR Medium Project on "Secure CITI: A Secure Critical Information Technology Infrastructure for Disaster Management". Additional information can be found at http://db.cs.pitt.edu

Content Management for Mobile and Ubiquitous Computing

Nigel Davies (Lancaster University, GB)

For over ten years researchers at Lancaster have been building and deploying prototype mobile and ubiquitous comptuing applications. Many of these deployments are critically depenedant on content to ensure an appropriate user experience. However, to date there are no content management systems for mobile or ubiquitous computing. In this talk we explore the problems of providing content in these environments and present requirements for a new research effort in the area of content management systems for mobile and pervasive environments.

Keywords: Mobile computing, ubiquitous computing, content management systems

Probabilistic Fusion of Sensor Data for Mobile Object Tracking

Anatole Gershman (Accenture Labs - Chicago, USA)

The Sensor Fusion project at Accenture Technology Labs is addressing the problem of people and object tracking and behavior recognition in an indoor environment. We have a network of 30+ cameras deployed around the lab's floor, infrared badges worn by some of the people some of the time and a fingerprint reader used by some of the people to check in when they enter the lab. In addition, we have access to some of the people's calendars and statistical information about some of the people patterns of behavior. Information gained from these sources is very noisy, unreliable and poorly synchronized. Yet, fairly reliable automatic recognition and tracking is possible through analysis and fusion of multiple streams of such information.

We believe that solving the sensor fusion problem will become increasingly important as the proliferation of a wide variety of sensors (video cameras, microphones, infrared badges, RFID tags, etc.) in public places such as airports, train stations, streets, parking lots, hospitals, governmental buildings, and shopping malls creates many opportunities for homeland security and business applications. Surveillance for threat detection, monitoring sensitive areas and detecting unusual events, tracking customers in retail stores, controlling and monitoring the movement of assets, and monitoring elderly and sick people at home are just some of the applications that require the ability to automatically detect, recognize and track people and other objects by analyzing multiple streams of often unreliable and poorly synchronized sensory data.

Video surveillance has been in use for decades but systems that can automatically detect and track people (or objects) in multiple locations using multiple streams of heterogeneous and noisy sensory data is still a great challenge and an

active research area. The focus of our project is not a better video surveillance system per se, but a scalable and robust framework for logical integration of noisy sensory data from multiple heterogeneous sensory sources that combines probabilistic and knowledge-based approaches. The probabilistic part is used for object identification and tracking and the knowledge-based part is used for maintaining overall coherence of reasoning. Our framework exploits local semantics from the environment of each sensor (e.g., if a camera is pointed at a location where people usually tend to stand, the local semantics enable the system to use the "standing people" statistical models, as opposed to a camera pointing at an office space where people are usually sitting) and takes advantage of data and sensor redundancy to improve accuracy and robustness while avoiding the combinatorial explosion.

In this presentation I will discuss our sensor fusion framework and the reliminary experimental results of applying our approach to creating a people localization system.

Keywords: Sensor fusion, probabilistic reasoning, bayesian networks, video surveillance

B-tree indexes for high update rates

Goetz Graefe (Microsoft Research, USA)

In some applications, data capture dominates query processing. For example, monitoring moving objects often requires more insertions and updates than queries. Data gathering using automated sensors often exhibits this imbalance. More generally, indexing streams apparently is considered an unsolved problem.

For those applications, B-tree indexes are reasonable choices if some trade-off decisions are tilted towards optimization of updates rather than of queries. This paper surveys techniques that let B-trees sustain very high update rates, up to multiple orders of magnitude higher than traditional B-trees, at the expense of query processing performance. Perhaps not surprisingly, some of these techniques are reminiscent of those employed during index creation, index rebuild, etc., while others are derived from other well known technologies such as differential files and log-structured file systems.

Keywords: B-tree, high update rates

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2006/763

Efficiently Managing Context Information for Large-scale Scenarios

Matthias Großmann (Universität Stuttgart, D)

In this paper, we address the data management aspect of large-scale pervasive computing systems. We aim at building an infrastructure that simultaneously supports many kinds of context-aware applications, ranging from room level up to nation level. This all-embracing approach gives rise to synergetic benefits like data reuse and sensor sharing. We identify major classes of context data and detail on their characteristics relevant for efficiently managing large amounts of it. Based on that, we argue that for large-scale systems it is beneficial to have special-purpose servers that are optimized for managing a certain class of context data. In the Nexus project we have implemented five servers for different classes of context data and a very flexible federation middleware integrating all these servers. For each of them, we highlight in which way the requirements of the targeted class of data are tackled and discuss our experiences.

Joint work of: Grossmann, Matthias; Bauer, Martin; Hönle, Nicola; Käppeler, Uwe-Philipp; Nicklas, Daniela; Schwarz, Thomas

See also: In: Proceedings of the 3rd IEEE Conference on Pervasive Computing and Communications: PerCom2005; Kauai Island, Hawaii, March 8-12, 2005

PALADIN: Pattern-based Approach to Large-scale Dynamic Information Integration

Jürgen Göres (TU Kaiserslautern, D)

To utilize the full potential of structured or semi-structured data stored across different information systems, users and applications must not be confronted directly with the individual, heterogeneous data sources, but instead be supplied with a customized integrated view on the data. Traditional information integration is relying on a human-driven process to accomplish this task. While feasible in static, closed-world scenarios, this approach fails in settings like the nascent data grids, which are characterized by a large, permanently changing set of autonomous data sources. The PALADIN project aims at reducing and ultimately eliminating the dependency on human experts in the integration process in order to provide fast and cost-effective integration services for these dynamic environments.

In order to automate the creation of mappings from the data sources to the integrated view, we propose a declarative notation to capture information integration knowledge. Using graph tranformations, we describe integration patterns that consist of an abstract problem description and an approach to a solution.

This problem description can later be discovered in a specific integration scenario, where the solution can then be adapted to the specifics of the scenario. By combining different patterns, an abstract integration plan that transforms the schema and data of the data sources is deducted. This plan can then be mapped to a concrete integration plan for a specific runtime environment.

Keywords: PALADIN dynamic information integration data grid

Web and Database Caching - Accelerating the Entire User-to-Data Path in the Internet

Theo Härder (TU Kaiserslautern, D)

A Web client request traverses four types of Web caches, before the Web server as the origin of the requested document is reached. This client-to-server path is continued to the backend DB server if timely and transaction-consistent data is needed to generate the document. Web caching typically supports identifier-based access to single Web objects kept ready somewhere in caches up to the server, whereas database caching, applied in the remaining path to the DB data, allows declarative query processing in the cache. Database caching uses a full-fledged DBMS as cache manager to adaptively maintain sets of records from a remote database and to evaluate queries on them. Using so-called cache groups, we introduce the new concept of constraint-based database caching. These cache groups are constructed from parameterized cache constraints, and their use is based on the key concepts of value completeness and predicate completeness. We show how cache constraints affect the correctness of query evaluations in the cache and which optimizations they allow. Cache groups supporting practical applications must exhibit controllable load behavior for which we identify necessary conditions. Important open problems are transactional updates to the caches and models to evaluate cache performance. Finally, we comment on future research problems.

Keywords: Database caching, value completeness, predicate completeness, cache update

Data Management for Moving Objects

Christian S. Jensen (Aalborg University, DK)

Much of my current research concerns a variety of aspect of data management in relation to moving objects. The basic setting assumed is one where individuals in a population of location-aware, on-line, mobile individuals use or take part in mobile services that exploit knowledge of the users' locations. A current instantiation of this setting is the one where each vehicle in a fleet of vehicles (e.g., trucks, public or school buses, taxis, emergency vehicles, police cars, rental

cars, service vehicles) is equipped with an on-board computer (e.g., a navigation computer, a PDA, a mobile phone), a GPS receiver, and a GPRS connection to a central server.

This setting poses a range of challenges:

- How to cost effectively maintain a reasonably accurate record at the server side of the inherently inaccurate location of each moving object.
- How to use location-related data, based on data obtained from the moving objects, for rendering the services geo-context aware. For example, past location data from an object may be used for learning the routes taken by the object and the destinations traveled to by the object.
- How to exploit in query processing that objects are constrained to a transportation network, that the routes the objects follow are known, and/or that the likely destinations are known.
- How to index the positions (past, present, and/or near-future) of large populations of moving objects. Here, both efficient updates and queries are important, and the trade-offs among (perhaps most notably) update efficiency, query efficiency, and query correctness (recall) are of interest.
- How to use location data, e.g., sequences of historical GPS positions, for applications in areas such as traffic management, collective transport, and telematics.
- How to manage large volumes of content for the purpose of delivery via mobile services, including context-aware push services.

Depending on the time available, I will cover the solutions we are studying for some or all of these challenges.

Keywords: Moving objects, indexing, query processing, tracking, routes, geo-context

Is the deployment of context- and user-aware technologies necessary to proof their concepts and success?

Matthias Joest (Europ. Media Lab. - Heidelberg, D)

In the past years many researchers from various domains within the field of information technologies have dealt with context- and user-awareness of systems. Many approaches have been shown to model various aspects of contextual information in order to serve the needs of the users. But mostly those systems have not left their laboratory environments where they were born. In my talk I would like to challenge the participants in order the think of way how we can deploy context-aware applications and proof their concepts with real users. I will present our attempt by introducing a commercial mobile information portal that features some context-awareness for a broader audience.

Keywords: Context-awareness

MonetDB/DataCell: database technology for the ambient home

Martin Kersten (CWI - Amsterdam, NL)

MonetDB/DataCell is an innovative solution to provide a database access point for sensor-networks.

In this talk, we outline its architecture and functionality based on the concept of data pumps, which collect, filter, aggregate, log, and transform data to be picked up by actuators. Its realization is illustrated using requirements derived from real-world applications to create an ambient home setting.

We describe how the system can be constructed with modest extension of a modern database kernel, how it challenges the query plan generation, and what to expect from the performance.

Keywords: Streaming databases, publish-subscribe, embedded databases

Adaptive Workload-Aware Overlay Networks in Pervasive Environments

Georgia Koloniari (University of Ioannina, GR)

Pervasive computing refers to an emerging trend towards numerous casually accessible devices connected to an increasingly ubiquitous network infrastructure. An important challenge in this context is discovering the appropriate data and services efficiently.

We assume that services and data are described using hierarchically structured metadata. We present a distributed workload-aware procedure for building clustered overlay networks of nodes that provide similar data and services. Clustering aims at improving query processing performance by reducing the communication cost through placing similar data at neighboring nodes.

To summarize the content of the nodes we use Counting Depth Bloom filters, specialized compact structures suitable for dynamic distributed environments. We present an efficient algorithm for creating a single filter for each cluster, by merging the filters of the nodes that belong to it. This filter is then enhanced with the query workload to derive the cluster description, which includes a set of representative path expressions selected based on their popularity in the nodes content and the query workload. Furthermore, we present an adaptive procedure that incrementally adjusts the cluster overlay network to reflect the changes that occur in the query workload and the topology of the system.

Incentives for Cooperation: Why do we need them? How can they be engineered?

Birgitta König-Ries (Universität Jena, D)

This talk first explains, why incentives for cooperation are needed in peer to peer systems. We then explore, why some obvious approaches (i.e., tamper resistant hardware and trusted third parties) are not applicable in the scenarios we are interested in. We go on to explain distributed reputation systems and show their problems. Finally, we present our approach which overcomes these problems by introducing the notion of type belief and unrepudiable evidences.

Keywords: Peer to Peer, P2P, Incentives, reputation systems

Joint work of: König-Ries, Birgitta; Obreiter, Philipp

Location Privacy in Mobile Location-based Services

Ling Liu (Georgia Institute of Technology, USA)

Continued advances in mobile networks and positioning technologies have created a strong market push for location-based applications. Examples include location-aware emergency response, location-based advertisement, and location-based entertainment. An important challenge in wide deployment of location-based services (LBSs) is the privacy-aware management of location information, providing safeguards for location privacy of mobile users against vulnerabilities for misuse and abuse. In this talk I will give an overview of location privacy problems and describe a scalable architecture for protecting location privacy from various privacy threats. A unique characteristic of our location privacy architecture is the use of a flexible privacy personalization framework to support location k-anonymity for a wide range of users with context-sensitive privacy requirements. This framework enables each mobile node to specify the minimum level of anonymity it desires and the maximum temporal and spatial tolerances it is willing to accept when requesting for k-anonymity preserving location-based services. We devise an efficient message perturbation engine to implement the proposed location privacy framework. Our experiments show that the personalized location k-anonymity model together with our location perturbation engine can achieve high guarantee of location k-anonymity and high resilience to location privacy threats without introducing significant performance penalty.

Data Management Frameworks for Sensor Networks

Pedro Jose Marron (Universität Stuttgart, D)

Data management is a crucial topic of the sensor network research area and has received a lot of attention in the past years. In this talk, we describe relevant examples of frameworks that deal with the problem of managing data in resource-limited environments, such as sensor networks.

Keywords: Data management, sensor networks, system software

Federating Location-based Data Services

Bernhard Mitschang (Universität Stuttgart, D)

With the emerging availability of small and portable devices which are able to determine their position and to communicate wirelessly, mobile and spatially-aware applications become feasible. These applications rely on information that is bound to locations and managed by so-called location-based data services. Based on a classification of location-based data services we introduce a service-oriented architecture that is built on a federation approach to efficiently support location-based applications.

Keywords: Federation Architecture, Data Services

Knowledge Applications on P2P

Jano Moreira de Souza (UFRJ - Rio de Janeiro, BR)

Most of the methodologies for collaborative knowledge building have as the goal to achieve a common understanding, such as a common ontology. Knowledge exchange in that context, take as assumption intentionality and the existence of a centralized, respected and certified source of knowledge. As in the sort of application that we are developing users can share their view of the world and reuse portions of it, without the need to agree on the meaning of the whole.

We will discuss three peer-to-peer applications which follow that principle: COE - A Cooperative Ontology Editor; KCE - A Cooperative Editor for Knowledge Chains, and Cooman2 - A cooperative tool for project management.

We believe that tools such as these will contribute to help people to build, manage and strength their personal knowledge and social networks.

RDBMS Support for Indexing of Historical Spatio-Temporal Data

Mario A. Nascimento (University of Alberta, CA)

Despite pressing need, current RDBMS support for spatiotemporal data is limited and inadequate, and most existing spatiotemporal access methods cannot be readily integrated into an RDBMS. In this short presentation we discuss two solution we have proposed to address this. First we discuss SPIT, an adaptive technique for spatiotemporal storage, indexing and query support that can be fully integrated within any off-the-shelf RDBMS as long as it support a B+-tree. Next we discuss a technique for splitting trajectories into smaller trajectories in order to use existing R-trees. Neither approach proposes a new indexing structure but rather re-use, in an optmized way, the resources that a DBMS would already have. (Work done jointly with: V. Botea, J. Elding, D. Mallett and J. Sander.)

Keywords: Spatiotemporal indexing, spatiotemporal data management

Data, Context and Situation: Interpretation Layers of Context Models

Daniela Nicklas (Universität Stuttgart, D)

Context-aware applications adapt their behavior depending on the state of the physical world along with other information representing context. This requires context management, i.e., the efficient management of context information and feasible context representations in order to allow reasoning.

The different context sources and characteristics of context information, e.g., type and representation, has led to a number of different approaches to supply applications with context information.

Besides specialized approaches, e.g., the context toolkit for sensor integration or the location stack for positioning systems, two major classes of generic context management exist. Context models provide a database-style management of context information and typically offer interfaces for applications to query context information or receive notifications on context changes. Contextual ontologies address the need of applications to access a thorough representation of knowledge to reason about context information and to react accordingly.

These approaches differ in the level of interpretation on the physical world, which is also reflected in the use of the terms "lower context" and "higher context": lower context is information that can be directly observed by sensors, while higher context is an interpretation of lower context to derive situations. For example, a GPS sensor gives the coordinates of a person, which have to be compared with a map to find out whether the person is in a certain location, e.g. a shopping mall (which would be higher context).

We can show how different levels of context information can be integrated and used in a common context model, that still can be managed efficiently, if we exploit the spatial scope of the information. As an example we present the Nexus Augmented World Model that serves both as a common ontology for different types of mobile, context-aware applications and as an integration scheme for a federated management platform.

Keywords: Context model, contextual ontology, interpretion, sensor data

Transferring Database Technology to Mobile Ad-Hoc Networks

Sebastian Obermeier (Universität Paderborn, D)

My current research discusses two problems of the application of database technology in mobile ad-hoc networks (MANETs), i.e. data caching and atomicity. Compared to fixed- wired networks, message costs are high and network failures, like network partitioning or node failures, make global knowledge concerning the operational status of devices difficult or even impossible. Therefore, within MANETs, there are some interesting new challenges, among which my research focuses on the following:

- Which kind of atomic guarantees can be given for distributed transactions?
- Which requirements must be fulfilled by mobile atomic commit protocols regarding compensation, transaction models, and blocking time?
- How to efficiently employ different kinds of mobile devices as caches for mobile databases, if devices act egoistically and contribute only if they profit from their efforts?

In my talk, I will introduce research topics that I am currently investigating and discuss the problems which possible solutions must answer.

Keywords: Mobile ad-hoc networks caching atomic commit transactions

Autonomic Sensor Network for Ecological Waters Supervision

Peter L. Peinl (Fachhochschule Fulda, D)

Sensor networks, especially mobile ones, can play an important role in the monitoring and supervison of the environment, for example in agriculture and ecology, as proposed in the AsNews project. Modules equipped with sensors to measure chemical, physical and biological parameters on the one hand and with mobile communication technologies on the other hand can be deployed in waters

(rivers, lakes, sea). Autonomic networks of those sensors can continuously gather, forward and partially assess ecological data. These can be used for generating alerts, forecasts or day-to-day statistics. Managing those networks, communicating those data, combining them with geographical and other data to calculate ecological models either directly in the sensor network or in land based computing centers poses many new and challenging problems in data modelling, integration and management and control.

Indexing the Past, Present and Anticipated Future Positions of Moving Objects

Simonas Saltenis (Aalborg University, DK)

With the proliferation of wireless communications and geo-positioning, e-services are envisioned that exploit the positions of a set of continuously moving users to provide context-aware functionality to each individual user.

Because advances in disk capacities continue to outperform Moore's Law, it becomes increasingly feasible to store on-line all the position information obtained from the moving e-service users. With the much slower advances in I/O speeds and many concurrent users, indexing techniques are of essence in this scenario.

Existing indexing techniques come in two forms. Some techniques capture the position of an object up until the time of the most recent position sample, while other techniques represent an object's position as a constant or linear function of time and capture the position from the current time and into the (near) future. This paper offers an indexing technique capable of capturing the positions of moving objects at all points in time. The index substantially extends partial persistence techniques, which support transaction time, to support valid time for monitoring applications. The performance of a timeslice query is independent of the number of past position samples stored for an object. No existing indices exist with these characteristics.

Keywords: Continuous variable, indexing, moving object, polyline, querying, trajectory, update

Efficient Domain-Specific Information Integration for Context-Aware Applications

Thomas Schwarz (Universität Stuttgart, D)

In this talk, we present the Nexus approach to efficient domain-specific integration of many loosely coupled data sources. A so called information maximizing mediation middleware (IMMM) has to cope with large data volumes and many queries, and at the same time achieve a tight semantic integration for the data

instances. For efficiency and practicability reasons, we propose to use an extensible global schema and a limited domain-specific query language. This facilitates employing domain-specific semantic knowledge in the middleware: detect duplicates, merge multiple representations, aggregate and generalize information.

We highlight the benefits of using a custom-made integration system adapted to the targeted application domain of context-aware applications compared to using a general purpose off-the-shelf system. We are able to leverage the characteristics of the application domain to provide specific access paths, allow for declarative caching of the data, and integrate semantically rich data transformation services into the system.

Data Streams Always and Everywhere

Bernhard Seeger (Universität Marburg, D)

The huge amount of data received as high-speed data streams from autonomous data providers require adequate methods for an efficient online processing without storing the entire streams persistently in a database system. A large variety of applications like traffic and environmental monitoring has caused a vastly emerging research interest in data streams recently.

At first in this talk, we will outline the differences between processing persistent data and transient streams as well as the general issues arising in the latter. Then, we will proceed in giving a brief overview of our research project PIPES, an infrastructure designed for building a prototype of a data stream management system PIPES adopts from traditional database systems the successful concept of differentiating between a logical algebra and a physical algebra. Unlike traditional systems, the algebra operators offer an integrated publish-subscribe interface that allows building complex query graphs. Their precisely defined semantics serve as a foundation for an effective algebraic query optimization. The physical operators perform in a data-driven manner where associated sliding windows ensure their non-blocking behaviour. A highly dynamic data structures with efficient insertion, deletion, and reorganization capabilities is used for organizing the data of a sliding window. The window size of an operator is adjustable at runtime to control its resource requirements. As a demonstration example for the query capabilities of PIPES, we will introduce our novel approach to maintaining complex stochastic estimators over data streams that are particularly useful for continuous monitoring important system parameters in PIPES. A complete implementation of PIPES is available within our Java library XXL (eXtensible and fleXible Library) for advanced query processing. PIPES extends XXL's scope towards a seamless integration of queries over data streams and persistent databases.

Keywords: Data streams, continuous queries, data integration

Media Distribution in a Pervasive Computing Environment

Alexander Sinitsyn (Philips Research - Eindhoven, NL)

Distribution of media in the fast growing world of digital stored content and multimedia supporting devices with connectivity, calls for a new media distribution architecture. The user should be provided with the experience of having an overview of his full media collection, regardless of the time, the place, and the connectivity. Transparent distributed data management is crucial to Ambient Intelligent applications. The proposed media distribution architecture offers a possible solution. It provides the user with the experience of having all his media collections available at any time, in any place, and managing them regardless of connection availability in the heterogeneous environment. This experience is enabled in our system by the separation of metadata and content handling. Other features are efficient handling of snapshots, usage of various database technologies, and leveraging device and service discovery mechanisms.

Keywords: Data management

Joint work of: Berkvens, Winfried A. H. ; Claassen, Arjan ; van Gassel, Joep P.; Sinitsyn, Alexander

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2006/762

Processing of Ontologies in Mobile Environments

Günther Specht (Universität Ulm, D)

Today information systems profit from using Ontologies.

Also mobile information systems could do that, but todays reasoners are much to huge to work on mobile devices, with there limitations on CPU-power, main memory, and storage capability.

In this talk three architectural variants for mobile processing of ontologies are presented. It turns out that transforming ontologies from OWL Light into a logic program and further into SQL scales the problem of processing down to views and queries on a mobile database. There are two approaches for this mapping available. The first on, we call it "direct mapping" was introduced by Grosof et al. It has several drawbacks in expressiveness and computation time for preprocessing if evaluated in a database. The second one, we call it "meta mapping" was introduced by Weithöner and Specht and overcomes these limitations. It has lower computational complexity and more representational flexibility. The main benefit is, that the rule set is now fix, independent from the concrete ontology. Thus it can be precompiled and preoptimized and that is the reason why it also scales down for the usage in mobile devices.

This can be even shown in some benchmarking results.

Keywords: Mobile Databases, Ontologies, Meta Mapping

Conceptual Modeling of Moving Objects: Why Is It Still A Hard Problem?

Jianwen Su (Univ. California - Santa Barbara, USA)

It is desirable to have conceptual data models so that the process of expressing a query or manipulation does not rely too much on the physical data representation. The current most popular conceptual data model for moving object trajectories in the community is to view them as (vectors of) linear functions of time. So what are the problems? Well, there are plenty, even if we assume that objects do not have spatial dimensions (i.e., they are moving points) and that we are happy with linearity. For examples, how do we represent the likelihood of an object being at a fixed location at a fixed time instant? In a fixed region at a fixed time instant? How do we represent the trajectory of an object whose location at every time instant is uncertain? With or without the uncertain locations of objects, what do we really want to know about moving objects when querying them? Can we have nice query languages for moving objects? Last but not least, what is the computation complexity of evaluating queries in these languages? In this talk, we will discuss some of the technical issues towards developing conceptual models for moving objects.

Modeling and Querying Moving Objects in Networks

Victor Teixeira de Almeida (FernUniversität in Hagen, D)

Moving Objects Databases have become an important research issue in recent years. For modeling and querying moving objects, there exists a comprehensive framework of abstract data types to describe objects moving freely in the 2D plane, providing data types such as moving point or moving region. However, in many applications people or vehicles move along transportation netw̆orks. It makes a lot of sense to model the network explicitly and to describe movements relative to the network rather than unconstrained space, because then it is much easier to formulate in queries relationships between moving objects and the network. Moreover, such models can be better supp̆orted in indexing and query processing. In the talk, I plan to present an ADT approach by modeling networks explicitly and providing data types and operations (an algebra) for static and moving network positions and regions. In a highway network, example entities corresponding to these data types are motels, construction areas, cars, and traffic jams. The network model is not too simplistic; it allows one to distinguish simple roads and divided highways and to describe the possible traversals of junctions precisely. Such an algebra may be embedded into an extensible DBMS data model to obtain a complete data model and query language for moving objects in networks.

Keywords: Spatio-Temporal Databases, Moving Objects, Abstract Data Types

Joint work of: Teixeira de Almeida, Victor; Hartmug Güting, Ralf; Ding, Zhiming

Full Paper: http://www.informatik.fernuni-hagen.de/import/pi4/papers/PaperMon.pdf

See also: VLDB Journal, 15(2):165-190, June 2006

Autonomy versus guarantees in Mobile P2P environment

Jari Veijalainen (University of Jyväskylä, FIN)

There has been a lot of research done in Mobile Peer-to-Peer systems and data management in them. In this paper we study aspects of the node autonomy, its degree and its influence on the data management tasks within a collection of autonomous nodes.

Keywords: Node autonomy, global guarantees, distributed file system, Mobile P2P

Joint work of: Veijalainen, Jari; Chrysantis, Panos

Energy consumption tradeoffs for compressed wireless data at a mobile terminal

Jari Veijalainen (University of Jyväskylä, FIN)

The high-end telecom terminal and PDAs, sometimes called Personal Trusted Devices (PTDs) are programmable, have tens of megabytes memory, and rather fast processors. In this paper we analyze, when it is energy-efficient to transfer application data compressed over the downlink and then decompress it at the terminal or compress it first at the terminal and then send it compressed over up-link. These questions are meaningful in the context of usual application code or data and streams that are stored before presentation and require lossless compression methods to be used. We deduce an analytical model and assess the model parameters based on experiments in 2G (GSM) and 3G (FOMA) network. The results indicate that if the reduction through compression in size of the file to be downloaded is higher than four per cent, energy is saved as compared to receiving the file uncompressed. For the upload case even two percent reduction in size is enough for energy savings at the terminal with the current transmission speeds and observed energy parameters. If time is saved using compressed files during transmission, then energy is certainly saved. From energy savings at the terminal we cannot deduce time savings, however. Energy and time consumed at the server for compression/decompression is considered negligible in this context and ignored. The same holds for the base stations and other fixed telecom infrastructure components.

The deduced formulae should be valid also in Mobile P2P environment. This is for further study.

Keywords: Personal trusted device, energy consumption, compression, wireless data transmission,

Joint work of: Veijalainen, Jari; Ojanen, Eetu; Haq, Mohammad Aminul; Vahteala, Ville-Pekka; Matsumoto, Mitsuji

See also: IEICE TRANSACTIONS. VOL. E87-B, No. 5, May 2004, pp. 1123-1130

MOBI-DIk: MOBIle DIscovery of Knowledge about local resources in peer-to-peer wireless networks

Ouri Wolfson (Univ. of Illinois - Chicago, USA)

In this talk we examine management of databases distributed among moving objects. The objects are interconnected by a Mobile Ad Hoc Network. Several inherent characteristics of this environment, including the dynamic and unpredictable network topology, the limited peer-to-peer communication bandwidth, and the need for incentive for peer-to-peer cooperation, impose challenges to data management. In this talk we discuss these challenges in the context of a database that represents resource information. The information is disseminated and queried by the moving objects in search of resources. We are currently building such a resource discovery engine called MOBI-DIK.

MOBI-DIK will enable quick building of matchmaking or resource discovery services in many application domains, including social networks, transportation, mobile electronic commerce, emergency response, and homeland security. For example, in a large professional, political, or social gathering, the technology is useful to automatically facilitate a face-to-face meeting based on matching profiles. In transportation, MOBI-DIK incorporated in navigational devices can be used to disseminate to other similarly-equipped vehicles information about relevant resources such as free parking slots, traffic jams and slowdowns, available taxicabs, and ride sharing. In mobile electronic commerce, MOBI-DIK is useful to match buyers and sellers in a mall, or to disseminate information about a marketed product. In emergency response, MOBI-DIK can be used by first responders to support rescue efforts (locate victims, and match responder capability with needs) even when the fixed infrastructure is inoperative. In homeland security, sensors mounted on neighboring containers can communicate and transitively relay alerts to remote check-points.

Keywords: P2P, mobile computing

Joint work of: Wolfson, Ouri; Xu, Bo ; Yin, Huabei ; Cao, Hu

Full Paper:
http://sites.computer.org/debull/A05sept/issue1.htm