Međunarodna konferencija
VERY LARGE DATA BASES

U Berlinu je između 9. i 12. rujna održana 29. međunarodna konferencija o velikim bazama podataka (29th International Conference on Very Large Databases). Ta konferencija, uz ACM SIGMOD/PODS (Special Interest Group on Management of Data/Principles of Databases Systems), jedna je od najznačajnijih i najvećih svjetskih konferencija iz područja baza podataka. Organizira se svake godine, a organizator je Zaklada VLDB Inc. (neprofitna organizacija iz SAD-a) radi *"promoviranja i razmjene znanstvenog rada na području baza podataka na svjetskoj razini"*. S obzirom na to da je riječ o znanosti i tehnologiji, krucijalnoj za geoinformacijsku znanost, tehnologije i sustave, jasno je zašto bi ta i slične konferencije i geodetima trebala biti u fokusu i od posebnog interesa.

Organizatori tog impozantnog skupa bili su: Humboldt-Universität zu Berlin, Freie Universität Berlin, Universität Karlsruhe (TH) i Deutsche Gesellschaft für Informatik (Njemačko informatičko društvo).

Konferencija je održana na poznatom Humboldtovu sveučilištu u Berlinu (Humboldt-Universität zu Berlin), koji ima dugu tradiciju i zanimljivu povijest.

Sveučilište je utemeljeno 1810. godine, a prema utemeljiteljskom konceptu teoretičara i državnika Wilhelma von Humboldta nazvano je "Majka svih modernih sveučilišta". Koncept je predvidio *"Universitas litterarum"*, koji bi s jedne strane trebao ujediniti izobrazbu i istraživanje, a s druge strane osigurati studentima opću humanističku izobrazbu. On je poslije prihvaćen u cijelom svijetu, i na njemu se temelje gotovo sva nova sveučilišta. Na taj su koncept utjecale, među ostalim, reformističke ideje filozofa Johanna Gottlieba Fichtea, prvoga prorektora Sveučilišta, odnosno teologa i filozofa Friedricha Schleiermachera. Sveučilište je na početku imalo četiri klasična fakulteta: pravni, medicinski, filozofski i teološki. Pod utjecajem Wilhelmovoga brata, znanstvenika Alexandra von Humboldta, to sveučilište prvo uvodi mnoge druge discipline. Kemičar August Wilhelm von Hofmann, fizičar Hermann von Helmholtz, matematičari Ernst Kummer, Leopold Kronecker, Karl Theodor Weierstrass ("zvjezdana trojka matematike"), kao i Johannes Müller i Rudolf Virchow u oblasti medicinskih znanosti, postaju poznati i izvan Njemačke. Sveučilište je dalo čak 29 nobelovaca: Albert Einstein, Max Planck, Gustav Hertz, Robert Koch i Max Born samo su neki od njih. I mnogi drugi, poslije poznati ljudi, studirali su ondje: Heinrich Heine, Adelbert von Chamisso, Ludwig Feuerbach, Otto von Bismarck, Karl Liebknecht, Franz Mehring, Alice Salomon, Karl Marx i dr.

Humboldtovo sveučilište u Berlinu – glavna zgrada

Prvu zgradu, koja je u to doba bila palača princa Heinricha, Sveučilištu je poklonio pruski kralj Friedrich Wilhelm III. Zgrada je izgrađena na prekrasnom berlinskom bulevaru Unter den Linden (Ispod lipa). Kada kraljevska knjižnica više nije zadovoljavala nastavne potrebe, 1831. je utemeljena Sveučilišna knjižnica. Usporedo sa svojim razvojem, Sveučilište je integriralo i neke druge institucije koje su postojale u Berlinu. Jedna od njih, Charité, poznata je i danas.

Na konferenciji je sudjelovalo 594 sudionika iz 44 zemlje svijeta (Zapadna/Južna Europa: 300, Istočna Europa: 52, Azija/Australija: 67, Sjeverna/Južna Amerika: 175). Iako takva podjela Europe može biti upitna, ona je od organizatora specificirana kao geografska. Me-

đutim, neke su zemlje (među kojima i Hrvatska), tradicionalno svrstane u zemlje Istočne Europe. Možda bi preciznija podjela bila na zemlje članice EU i one koje to nisu. Ta podjela i ne bi bila toliko značajna da ove godine organizator nije učinio značajan iskorak, kako bi što više znanstvenih institucija i pojedinaca iz zemalja bivše, političke Istočne Europe, uzelo učešća na toj respektabilnoj konferenciji. Naime, registracija (530 €) za sudionike iz Poljske, baltičkih država, Bjelorusije, Ukrajine, Rusije, Češke, Slovačke, Mađarske, Rumunjske, Bugarske i zemalja bivše Jugoslavije nije bila obvezna. Ako se ima u vidu da je iz Zapadne Europe (ne računajući Njemačku sa 158) bilo 78 registriranih, može se reći da je organizator uspio privući pozornost znanstvenika i stručnjaka iz tog dijela Europe. Autor ovog izvješća, premda registriran kao djelatnik austrijske tvrtke *GISquadrat AG*, bio je jedini sudionik iz Hrvatske.

Program konferencije tematski je podijeljen na tri komplementarna dijela:

- Temeljna tehnologija baza podataka (Core Database Technology)
- Infrastruktura za informacijske sustave (Infrastructure for Information Systems)
- Industrijske aplikacije i iskustva (Industrial Applications and Experiences).

Program je uključivao tri pozvana predavanja, prezentaciju najboljeg rada u posljednjih 10 godina, 84 istraživačko-znanstvena rada prezentirana u 25 sesija, 9 prezentacija u 3 industrijske sesije, 3 panel sesije, 5 tutorskih satova u 15 sesija, 17 demonstracija u 6 sesija i 5 doktorskih poster-prezentacija. Programsko povjerenstvo prihvatilo je 84 (18%) od ukupno 457 prijavljenih radova, što jasno govori o kriterijima i kvaliteti ove konferencije.

Zbornik radova (ISBN 0-12-722442-4), koji je izdao Morgan Kaufmann Publishers, u tiskanom obliku sadrži 1151 stranicu, a urednici su mu: Johann-Christoph Freytag (Humboldt-Universität zu Berlin, Njemačka), Peter C. Lockemann (Universität Karlsruhe, Njemačka), Serge Abiteboul (INRIA, Francuska), Michael Carey (BEA Systems, SAD), Pat Selinger (IBM Silicon Valey Lab, SAD) i Andreas Heuer (Universität Rostock, Njemačka).

Prethodne konferencije održane su ovim redom: Framingham (1975), Bruxelles (1976), Tokyo (1977), Berlin (1978), Rio de Janeiro (1979), Montreal (1980), Cannes (1981), Mexico City (1982), Florence (1983), Singapore (1984), Stockholm (1985), Kyoto (1986), Brighton (1987), Los Angeles (1988), Amsterdam (1989), Brisbane (1990), Barcelona (1991), Vancouver (1992), Dublin (1993), Santiago de Chile (1994), Zürich (1995), Mumbai (1996), Atena (1997), New York (1998), Edinburgh (1999), Cairo (2000), Rome (2001), Hong Kong (2002).

Iduća, 30. po redu konferencija održat će se sljedeće godine u Torontu (http://www.vldb04.org/).

U nastavku je dan pregled programa konferencije bez autora radova. Detaljan program konferencije može se vidjeti na internetskoj stranici http://www.vldb.informatik.hu-berlin.de/progr_program.html.

Keynotes

- A Nanotechnology-based Approach to Data Storage
- Integrating Information for On Demand Computing
- The Data-Centric Revolution in Networking

Ten-Year Best Paper Award

- The History of Histograms

Research Sessions

Internet/WWW

- Complex Queries over Web Repositories
- XSEarch: A Semantic Search Engine for XML
- An Efficient and Resilient Approach to Filtering and Disseminating Streaming Data

Data Mining/Streams

- Efficient Mining of XML Query Patterns for Caching
- A Framework for Clustering Evolving Data Streams
- A Regression-Based Temporal Pattern Mining Scheme for Data Streams

Query Processing in the Web

- On the Costs of Multilingualism in Database Systems
- Distributed Top-N Query Processing with Possibly Uncooperative Local Systems
- Optimized Query Execution in Large Search Engines with Global Page Ordering

XML Queries Processing I

- Path Queries on Compressed XML
- On the minimization of Xpath queries
- Covering Indexes for XML Queries: Bisimulation – Simulation = Negation

XML Matching & Storage

- Phrase Matching in XML
- RRXS: Redundancy reducing XML storage in relations
- MARS: A System for Publishing XML from Mixed and Redundant Storage

XML Query Processing II

- Projecting XML Documents
- Mixed Mode XML Query Processing
- From Tree Patterns to Generalized Tree Patterns: On Efficient Evaluation of XQuery

XML Query Processing III

- Efficient Processing of Expressive Node-Selecting Queries on XML Data in Secondary Storage: A Tree Automata-based Approach
- Query Processing for High-Volume XML Message Brokering
- Holistic Twig Joins on Indexed XML Documents

Streaming

- Maximizing the Output Rate of Multi-Way Join Queries over Streaming Information Sources
- Scheduling for shared window joins over data streams
- Load Shedding in a Data Stream Manager

Distributed Data & Streams

- Querying the Internet with PIER
- Tuple Routing Strategies for Distributed Eddies
- AQuery: Query Language for Ordered Data, Optimization Techniques, and Experiments

Web & Statistics

- WISE-Integrator: An Automatic Integrator of Web Search Interfaces for E-Commerce
- SASH: A Self-Adaptive Histogram Set for Dynamically Changing Workloads
- VIPAS: Virtual Link Powered Authority Search in the Web

Caching

- Balancing Performance and Data Freshness in Web Database Servers
- Buffering Accesses to Memory-Resident Index Structures
- Data Morphing: An Adaptive, Cache-Conscious Storage Technique

Data Mining

- COMBI-Operator: Database Support for Data Mining Applications
- A Shrinking-Based Approach for Multi-Dimensional Data Analysis
- Data Bubbles for Non-Vector Data: Speeding-up Hierarchical Clustering in Arbitrary Metric Spaces

OLAP & Data Mining

- Finding Hierarchical Heavy Hitters in Data Streams
- Star-Cubing: Computing Iceberg Cubes by Top-Down and Bottom-Up Integration
- Coarse-Grained Optimization: Techniques for Rewriting SQL Statement Sequences

Advanced Query Processing

- Processing Sliding Window Multi-Joins in Continuous Queries over Data Streams
- Continuous K-Nearest Neighbor Queries for Continuously Moving Points with Updates
- Staircase Join: Teach a Relational DBMS to Watch its (Axis) Steps

Data Quality, Data Mining

- Checks and Balances: Monitoring Data Quality Problems in Network Traffic Databases
- Systematic Development of Data Mining-Based Data Quality Tools
- Adaptive, Hands-Off Stream Mining

Managing derived Data

- Composing Mappings Among Data Sources
- Mapping Adaptation under Evolving Schemas
- Locking Protocols for Materialized Aggregate Join Views

Access Methods & Temporal Data

- Supporting Frequent Updates in R-Trees: A Bottom-Up Approach
- The ND-Tree: A Dynamic Indexing Technique for Multidimensional Non-ordered Discrete Data Spaces
- Temporal Slicing in the Evaluation of XML Queries

Aggregation, Prediction & Constraints

- The Generalized Pre-Grouping Transformation: Aggregate-Query Optimization in the Presence of Dependencies
- Estimating the Output Cardinality of Partial Preaggregation with a Measure of Clusteredness
- BHUNT: Automatic Discovery of Fuzzy Algebraic Constraints in Relational Data

Storage Management

- Tabular Placement of Relational Data on MEMS-based Storage Devices
- Memory Requirements for Query Execution in Highly Constrained Devices
- Lachesis: Robust Database Storage Management Based on Device-specific Performance Characteristics

Performance & Benchmarking

- Cache Tables: Paving the Way for an Adaptive Database Cache
- Primitives for Workload Summarization and Implications for SQL
- A Dependability Benchmark for OLTP Application Environments

Query Optimization

- Supporting Top-k Join Queries in Relational Databases
- AniPQO: Almost Non-intrusive Parametric Query Optimization for Nonlinear Cost Functions
- Efficient Approximation of Optimization Queries Under Parametric Aggregation Constraints

Spatial Support

- The TPR-Tree: An Optimized Spatio-Temporal Access Method for Predictive Queries
- Query Processing in Spatial Network Databases
- Multiscale Histograms: Summarizing Topological Relations in Large Spatial Datasets

Advanced Query Processing

- Avoiding Sorting and Grouping In Processing Queries
- Operator Scheduling in a Data Stream Manager
- Efficient IR-Style Keyword Search over Relational Databases

Metadata & Sampling

- Merging Models Based on Given Correspondences
- Locating Data Sources in Large Distributed Systems
- Robust Estimation With Sampling and Approximate Pre-Aggregation

Potpourri – Access Control, Bioinformatics & Security

- Controlling Access to Published Data Using Cryptography
- OASIS: An Online and Accurate Technique for Local-alignment Searches on Biological Sequences
- Privacy-Preserving Indexing of Documents on the Network

Industrial Sessions

Performance

- Data Compression in Oracle
- Improving Performance with Bulk-Inserts in Oracle R-Trees
- Statistics on Views

Multidimensionality & Bioinformatics

- Efficient Query Processing for Multi-Dimensionally Clustered Tables in DB2
- A Platform Based on the Multi-dimensional Data Model for Analysis of Bio-Molecular Structures
- Capturing Global Transactions from Multiple Recovery Log Files in a Partitioned Database System

XML & Mobility

- The BEA/XQRL Streaming XQuery Processor
- XML Schemas in Oracle XML DB
- Integrated Data Management for Mobile Services in the Real World

Web Services

- Web-Services and unified messaging in an Investment Banking environment
- Foundations of Transactions for the Internet
- Using WS-Coordination and WS-Transaction to Ensure the Consistency of B2B Interactions

Grid and Applications

- Open Grid Services Architecture
- Distributed Data Management in the GRID
- Grid Services for e-Science

Commercial Use of Database Technology

- The Zero-Delay Data Warehouse: Mobilizing Heterogeneous Databases
- Tamino & Co. – Software AG's Database Technology

Panels

- A Database Striptease or How to Manage Your Personal Databases
- Who needs XML Databases?
- Illuminating the Dark Side of Web Services

Demo Sessions

Semistructured Data Management

- Xcerpt and visXcerpt: From Pattern-Based to Visual Querying of XML and Semi-structured Data
- OrientStore: A Schema Based Native XML Storage System
- Managing Distributed Workspaces with Active XML
- XQueC: Pushing Queries to Compressed XML Data
- A System for Keyword Proximity Search on XML Databases
- XISS/R: XML Indexing and Storage System Using RDBMS
- Implementing XQuery 1.0: The Galax Experience
- XQuery Optimization (PhD Poster)

Distributed Information Management

- Web Service Composition with O'GRAPE and OSIRIS
- Chameleon: an Extensible and Customizable Tool for Web Data Translation
- NexusScout: An Advanced Location-Based Application on a Distributed, Open Mediation Platform
- Schema-driven Customization of Web Services
- BibFinder/StatMiner: Effectively Mining and Using Coverage and Overlap Statistics in Data Integration
- S-ToPSS: Semantic Toronto Publish/Subscribe System
- From Focused Crawling to Expert Information: an Application Framework for Web Exploration and Portal Generation
- CachePortal II: Acceleration of Very Large Scale Data Center-Hosted Database-driven Web Applications
- Consistency Based Snapshot Management in Data Grids (PhD Poster)
- Adaptive Distributed Query Processing (PhD Poster)

Advanced Models and Languages/Architectures for Data Analysis

- ATLAS: a Small but Complete SQL Extension for Data Mining and Data Streams
- Business Modeling Using SQL Spreadsheets
- An Interpolated Volume Data Model
- Efficacious Data Cube Exploration by Semantic Summarization and Compression
- QUIET: Continuous Query-driven Index Tuning
- Chip-Secured Data Access: Reconciling Access Rights with Data Encryption
- IrisNet: An Architecture for Internet-scale Sensing Services
- Large-Scale, Standards-Based Earth Observation Imagery and Web Mapping Services
- Matching Database Access Patterns to Storage Characteristics (PhD Poster)
- Inference and Prediction of Uncertain Events in Active Systems: A Language and Execution Model (PhD Poster)

Tutorials

- Privacy-Enhanced Data Management for Next-Generation e-Commerce
- The Semantic Web: Semantics for Data on the Web
- Data Stream Query Processing: A Tutorial
- Grid Data Management Systems & Services
- Constructing and integrating data-centric Web applications: methods, tools, and techniques.

Zdravko Galić