



DATABASES

<http://db.cs.utwente.nl>

Peter M.G. Apers

DATABASES ORGANIZATION



■ Staff

- Peter Apers (full prof)
- Willem Jonker (full prof)
- Djoerd Hiemstra (associate prof; information retrieval)
- Maurice van Keulen (associate prof; data interoperability)
- Maarten Fokkinga (assistant prof; formal semantics)
- Andreas Wombacher (assistant prof; sensor data management)
- Jan Flokstra (scientific programmer)
- Ida den Hamer (secretary)

■ 16 temporary (14 PhDs & 2 postdocs)

■ Cooperation

- HMI (Nijholt & de Jong); DIES (Hartel), SE (Ferreira-Pires), IS (Van Sinderen), ITC (De By)

DATABASES TEACHING



Inf & Create Bachelor

- Databases
- Data-driven Applications

CS, HMI, & BIT Master

- Advanced Database Systems
- Database transactions and processes
- Datawarehousing & Data Mining
- XML & Databases 1 & 2
- Sensor Data Management 1 & 2
- Distributed Data Processing (MapReduce)
- Secure Data Management
- Information Retrieval
- Knowledge Representation

General Bachelor

- Design project
- Bachelor paper

General Master

- Research topics
- Traineeship
- Final project



DATABASES RESEARCH FOCUS

Mission: to provide data management to create added value on top of autonomous data sources.

1. Search on semi-structured data
2. Enriching uncertain data
3. Data security and privacy

THEME 1: SEARCH SEMI-STRUCTURED DATA

- Web search today:
 - Huge data centers
 - Invades our privacy
 - Not objective
- P2P search challenges:
 - Economic models, Learning from search result pages, Distributed caching
 - Deep web search, free-text to structured queries



THEME 2:

ENRICHING UNCERTAIN DATA



- Principle: “Data is uncertain”
 - Sensor data
 - Data interoperability (coupling / integration / annotation / cleaning / heterogeneity / ...)
- App areas: e-Science / web&social netw. / ...
- Approaches
 - Manage data in *Probabilistic* DBs (rel/xml/rdf)
 - Data provenance & annotations for sensor data
 - “Good-is-good-enough” & “Pay-as-you-go”

THEME 3:

DATA SECURITY & PRIVACY



■ Private data?

- Stored indefinitely / limited retention / encrypted
- ... hackers / abuse of weak privacy policies / negligence / malicious db admins, ...

■ Progressive data degradation by DBMS

second → minute → hour → week → month → year

address → street → district → city → country

- Less privacy-sensitive – still usable
- Challenges: storage structures, indexing, transaction management, logging

FUNDING & COOPERATION



■ EU & International

- FP7 Strep PuppyIR: (1 PhD, with HMI: Nijholt & De Jong)
- FP7 IP AXES: Audiovisual Archives (1 postdoc; with HMI: De Jong)
- Tsinghua, prof. Chi Hung-Chi (2 PhDs with IS: v. Sinderen, SE: Pires)

■ National

- NWO Vidi: Distributed Search (2 PhD)
- NWO Effort: Focused Retrieval (1 postdoc)
- COMMIT Spatio-temporal DB (2 PhD vacancies, with ITC: De By)
- COMMIT Entity information retrieval (1 PhD vacancy)

■ Industry

- Philips: Secure patient data (1 PhD; with DIES: Hartel)

■ Local UT

- Neogeography (2 PhDs, with ITC: De By)
- Sensor data management: (1 PhD)
- NICE: PayDB (1 PhD, with HMI: v.d. Vet)
- Protection & security of content (1 PhD; with DIES: Hartel)

PHD THESES SINCE 2000



- | | |
|--|--|
| 2010 Abdel Kader: Optimization of XQueries | 2006 Smiljanic: XML schema matching |
| 2010 Kelkboom: Performance of privacy techniques | 2005 Derks: Concurrency and Recovery in DB Systems |
| 2010 Aly: Multimedia Retrieval | 2005 Meratnia: DB Support for Moving Object data |
| 2010 van Heerde: Privacy-aware data management | 2004 Xinming: Spatial object modeling |
| 2009 Evers: Sensor Data Management | 2003 Campos dos Santos: Biological INF systems |
| 2009 Serdyukov: IR - Expert search | 2003 Petkovic: DB Supported Video Retrieval |
| 2008 Rode: IR - Entity Retrieval | 2002 Van Rein: Process Validation |
| 2008 Van Bunningen: Context-aware Querying | 2002 Van Zwol: web-based IR |
| 2008 De Keijzer: Management of Uncertain Data | 2002 Blok: Database Optimization for IR |
| 2007 Brinkman: Searching in encrypted data | 2001 Hiddink: Educational Multimedia Databases |
| 2006 Mihajlovic: IR - exploiting structure | 2001 Raza: GIS for urban applications |
| 2006 Van Ruth: Query Optimization | 2000 Saadi Mesgari: three dimensional spatial data |

On the average 2.5 per year

LEADERSHIP



- Appointed full professor:
 - Siebes (UU), Grefen (TUE), De Vries (TUD), Feng (Tsinghua), Petkovic (TUE)
- Two NWO Vidi
 - Feng (2004), Hiemstra (2008)
- International Industry
 - Petkovic (Philips), Van Zwol (Yahoo USA), Serdyukov (Yandex Russia), Abdel Kader (ASML), Mihajlovic (Philips)
- NIRICT, D-COMMIT, STW, ICT Labs
 - Apers, Jonker