



**Hasso
Plattner
Institut**

IT Systems Engineering | Universität Potsdam

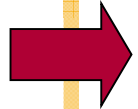
SE: Beauty is our Business

Wissenschaftliche Arbeiten Lesen

26.4.2010

Felix Naumann

2



- Organisatorisches
- Konferenzen und Journale
 - Gutachten
- Gliederung eines Artikels
- Experimente
- Literaturrecherche



Die Themen

3

- Mariposa
 - Tobias Schubotz
- Trio
 - Markus Dreseler
- Fagins Algorithmus
 - Marika Marszalkowski
- Top-K in SQL
 - Sebastian Oergel
- Sorted Neighborhood
 - Matthias Kohnen
- Data Mining
 - Claudia Lehmann

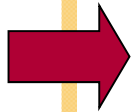
■ Besprechungstermine
mit mir ausmachen!

- E-Mail / Telefon /
Vorbeikommen
- Katrin Heinrich

Termin	Thema	Vortrag
19.4.2010	Einführung	Felix Naumann
26.4.2010	Wissenschaftliche Texte Lesen	Felix Naumann
10.5.2010	Elevatorpitches / Diskussion	Alle
31.5.2010	Vortragstechniken	Felix Naumann
7.6.2010	Vortrag 1: Mariposa (pdf)	Tobias Schubotz
	Vortrag 2: Trio (pdf)	Markus Dreseler
14.6.2010	Einführung in LaTeX	Felix Naumann
21.6.2010	Vortrag 3: Fagins Algorithmus (pdf)	Marika Marszalkowski
	Vortrag 5: Enough Already in SQL (pdf)	Sebastian Oergel
5.7.2010	Vortrag 4: Sorted Neighborhood (pdf)	Matthias Kohnen
	Vortrag 6: Data Mining (pdf)	Claudia Lehmann
12.7.2010	Vorstellung der Gliederungen & Tipps zur Ausarbeitung	Alle
17.9.2010	Abgabe der Ausarbeitungen	Alle

5

- Organisatorisches
- Konferenzen und Journale
 - Gutachten
- Gliederung eines Artikels
- Experimente
- Literaturrecherche



Veröffentlichung auf Konferenzen

6

- Konferenzen wichtig in der Datenbank Forschung
- Journale eher zweitrangig
 - im Gegensatz zu fast allen anderen Forschungsrichtungen
- i.D.R. 12 Seiten, Englisch
- 2 Monate Begutachtung
 - Quote 10%-15% bei guten Konferenzen
- Bei Annahme Reise und Vortrag auf Konferenz
 - Bezahlung: Selbst!
- Workshops ähnlich; nur kürzere Begutachtung
- Journale: Iterative Gutachten; lange Laufzeit bis zur Veröffentlichung (bis zu 3 Jahre)

Wichtige Konferenzen

7

- VLDB
 - Very Large Databases
 - 2003 Berlin
- SIGMOD
 - ACM Special Interest Group – Management of Data
 - GI – Gesellschaft für Informatik
- ICDE
 - IEEE
 - International Conference on Database Engineering
- EDBT
- BTW (2009 in Münster, 2011 in Kaiserslautern)
- Viele kleinere
- Viele Spezial-Workshops

Wichtige Journals

8

- VLDB Journal
- TODS: Transactions on Database Systems
- TOIS: Transactions on Information Systems
- IS: Information Systems
- ACM Computing Surveys
- Journal of the ACM
- CACM
- Auf Einladung / weniger strenge Gutachten
 - IEEE Data Engineering Bulletin
 - SIGMOD Record
 - Datenbankspektrum



VLDB 2004

30th International Conference on Very Large Data Bases



Royal York Hotel
29 August - 3 September 2004
Toronto, Canada

*This site is best viewed at 1024 x 786 or 1280 x 1024

VLDB is the premier international conference on database technology, organized every year by the [VLDB Endowment](#). VLDB 2004 will be held in Toronto, the biggest canadian city and the financial center of the country. Toronto has been called the most multicultural city in the world, hosting more than 50,000 new immigrants per year, and boasting more than 100 languages spoken by some of its citizens. Proceedings from [earlier VLDB Conferences](#) are distributed by [Morgan Kaufmann](#). The VLDB Endowment also publishes the [VLDB Journal](#) through [Springer](#).

Important Dates

Workshops:
August 29 - 30, 2004

Main Conference:
August 31 - September 3, 2004

VLDB 04 Newsletter :

[Issue 1](#)
[Issue 2](#)
[Issue 3](#)

Organization

[Officers](#)
[Conference Organization](#)
[Technical Program](#)
[Local Organization](#)
[Publicity and Publications](#)
[Program Committee](#)
[Core Database Technology](#)
[Infrastructure for Information Systems](#)
[Industrial and Applications](#)
[Workshop Selection](#)

Sponsors

Conference Program Information

[A Word from the PC Chairs](#)
[Program at a Glance](#)
[Full Program](#)
[Keynote Program](#)
[Tutorial Program](#)
[Panel Program](#)
[Demonstrations](#)
[Abstracts](#)
[Workshops](#)

Information for Delegates

[Welcome to VLDB04](#)
[Conference Registration](#)
[Conference Hotel Room Reservation](#)
[Other Hotels](#)
[Conference Venue](#)
[How to Get There](#)
[Social Events](#)
[Arrival Instructions](#)
[VLDB 04 Niagara Falls Tours](#)
[Disclaimer](#)

Information about Toronto

[Toronto for DB Researchers](#)
[Toronto General Information](#)
[Airport Information](#)
[Things to Do in Toronto](#)
[Sightseeing in Toronto](#)
[What's doing in Toronto \(from the NY Times\)](#)

Information for Authors

[Call for Papers](#)
[Call for Workshops](#)
[Call for Tutorials](#)
[Call for Panels](#)
[Paper Submission Site and Guidelines](#)

Organisation einer Konferenz

10

Important Dates

Workshops:

August 29 - 30, 2004

Main Conference:

**August 31 - September 3,
2004**

VLDB 04 Newsletter :

[Issue 1](#)

[Issue 2](#)

[Issue 3](#)

Organization

[Officers](#)

[Conference Organization](#)
[Technical Program](#)
[Local Organization](#)
[Publicity and Publications](#)

[Program Committee](#)

[Core Database Technology](#)
[Infrastructure for Information
Systems](#)
[Industrial and Applications](#)
[Workshop Selection](#)

[Sponsors](#)

Conference Program Information

[A Word from the PC Chairs](#)
[Program at a Glance](#)
[Full Program](#)
[Keynote Program](#)
[Tutorial Program](#)
[Panel Program](#)
[Demonstrations](#)
[Abstracts](#)
[Workshops](#)

Information for Delegates

[Welcome to VLDB04](#)
[Conference Registration](#)
[Conference Hotel Room Reservation](#)
[Other Hotels](#)
[Conference Venue](#)
[How to Get There](#)
[Social Events](#)
[Arrival Instructions](#)
[VLDB 04 Niagara Falls Tours](#)
[Disclaimer](#)

Information about Toronto

[Toronto for DB Researchers](#)
[Toronto General Information](#)
[Airport Information](#)
[Things to Do in Toronto](#)
[Sightseeing in Toronto](#)
[What's doing in Toronto \(from the NY Times\)](#)

Information for Authors

[Call for Papers](#)
[Call for Workshops](#)
[Call for Tutorials](#)
[Call for Panels](#)
[Paper Submission Site and Guidelines](#)

Officers

11

CONFERENCE ORGANIZATION

General Chair

John Mylopoulos, University of Toronto, Canada

VLDB Endowment Liaison

Kyu-Young Whang, KAIST, Korea

Area Coordinator, North America

Philip Bernstein, Microsoft Research, USA

Area Coordinator, South America

Alberto Laender, Federal University of Minas Gerais, Brazil

Area Coordinator, Europe, Mideast & Africa

Avigdor Gal, Technion, Israel

Area Coordinator, Far East & Australia

Hongjun Lu, Hong Kong University of Science and Technology, China

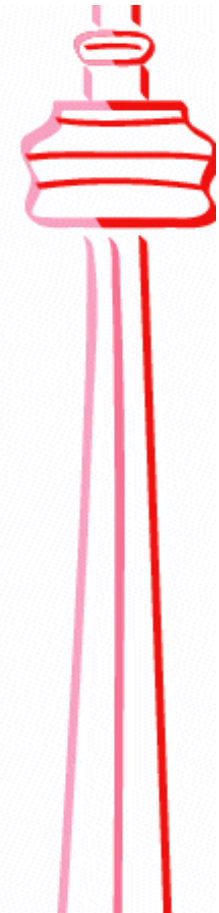
Publicity and Publications Chair

Mariano Consens, University of Toronto, Canada

Web Coordinator

Manuel Kolp, Catholic University of Louvain, Belgium

[Back to Top](#) ↑



TECHNICAL PROGRAM

Technical Program Chair

M. Tamer Özsu, University of Waterloo, Canada

Core Database Technology Program Chair

Donald Kossmann, University of Heidelberg, Germany

Infrastructure for Information Systems Program Chair

Renée Miller, University of Toronto, Canada

Industrial and Applications Program Chairs

Jose Blakeley, Microsoft Corporation, USA

Berni Schiefer, IBM Laboratories, Canada

Tutorial Program Co-Chairs

Raymond Ng, University of British Columbia, Canada

Matthias Jarke, RWTH Aachen, Germany

Panel Program Co-Chairs

Jarek Gryz, York University, Canada

Fred Lochovsky, HKUST, Hong Kong

Demonstrations Chairs

Bettina Kemme, McGill University, Canada

David Toman, University of Waterloo, Canada

Proceedings Editor

Mario Nascimento, University of Alberta, Canada

Workshops Co-Chairs

Alberto Mendelzon, University of Toronto, Canada

S. Sudarshan, Indian Institute of Technology, Bombay, India

[Back to Top](#) ↑

LOCAL ORGANIZATION

Local Arrangement Chair

Nick Koudas, AT&T Labs Research, USA

Exhibits Chair

H.-Arno Jacobsen, University of Toronto, Canada

Treasurer

Kenneth Barker, University of Calgary, Canada

Fund Raising Chair

Victor DiCiccio, University of Waterloo, Canada

Registration Chair

Grant Weddell, University of Waterloo, Canada

[Back to Top](#) ↑

[Webmaster](#)

Last updated: July 26, 2004

Fertig

Typisches Programm einer DB Konferenz

13

PROGRAM AT A GLANCE

Monday at a Glance

Upper Canada

18:00-20:00

Welcome Reception
(Upper Canada)

Tuesday at a Glance

Ballroom

Ontario

Territories

Confederation 5 & 6

Tudor 7 & 8

07:30-08:30

Breakfast (Concert Hall)

08:30-09:00

Opening Ceremony (Ballroom)

09:00-10:30

Keynote Address 1 (Ballroom): Databases in a Wireless World
David Yach, VP Software, Research in Motion

10:30-11:00

Coffee Break (Ballroom Foyer)

11:00-12:30

Research Session 1
Compressing &
Indexing

Research Session 2
XML Views and
Schemas

Research Session 3
Controlling Access

Tutorial 1
Database
Architectures for New
Hardware

NO
SESSION

12:30-14:00

Lunch Break (Concert Hall)

14:00-15:30

Research Session 4
XML (I)

Research Session 5
Stream Mining

Industrial Session 1
Novel SQL
Extensions

Tutorial 1
Database
Architectures for New
Hardware

Demo
Groups 2
and 3

15:30-16:00

Coffee Break (Ballroom Foyer)

16:00-17:30

Research Session 6
XML (II)

Industrial Session 2
New DBMS
architectures and
Performance

Panel 1
Biological Data
Management:
Research, Practice
and Opportunities

Tutorial 2
Security of Shared
Data in Large
Systems: State of the
Art and
Research Directions

Demos
Groups 2
and 3

Workflow zur Veröffentlichung

14

1. Idee(n)
2. Implementierung und Experimente (bzw. Beweis)
3. Artikel schreiben
4. Rechtzeitig einreichen
 - E-Mail oder Web-CMT
5. Begutachtung durch 2-4 peers
 - Double-blind?
6. Entscheidung durch Program Chair
 - E-Mail
7. Bei Annahme: Vorbereitung der Camera-Ready-Copy
8. Reise buchen, Anmelden
9. Vortrag vor Ort

Typisches Gutachtenformular (ICDE 2006)

1. Is the paper relevant to ICDE 2006? Yes, Definitely; Yes, Probably; No
2. Is the paper technically correct? ☐ Yes ☐ No
3. Originality
 - ☐ Strong accept (Very innovative) Accept (Innovative) Weak accept (Marginally innovative) Weak reject (Not very innovative) Reject (No innovation at all)
4. Impact
 - ☐ Strong accept (Very high) Accept (High) Weak accept (Good) Weak reject (Fair) Reject (No impact at all)
5. Technical Depth
 - ☐ Strong accept (Very high) Accept (High) Weak accept (Good) Weak reject (Fair) Reject (No depth at all)
6. Presentation
 - ☐ Strong accept (Excellent) Accept (Good) Weak accept (Average) Weak reject (Fair) Reject (Poor)
7. Overall Rating
 - ☐ Strong accept (Definitely accept) Accept (Probably accept) Weak accept (Could go either way) Weak reject (Probably reject) Reject (Definitely reject)
8. Reviewer Confidence
 - ☐ High (I know this area well) Medium (Moderately confident, I know as much as most) Low (Rather unconfident, I know a bit)
9. How many ICDE attendees are likely to be interested in this paper?
10. Should this paper be considered for a Best Paper Award? ☐ Yes ☐ Probably ☐ No
11. Summary of main contribution and rationale for your recommendation (1-2 paragraphs)
12. Detailed comments to authors
13. Should this paper be considered for a short presentation if accepted as a full paper? ☐ Yes ☐ No
14. Enter comments for the Program Committee (will not be seen by author):

Workflow bei Journalen

16

Wie bei Konferenzen, aber

- Laufende Einreichungen
- Längere Paper
- Längere Gutachtendauer
 - Informelleres (aber ausführlicheres) Gutachten
- Zweite (und dritte) Runde
 - Verbesserungen durch Autoren
 - Neuerliche Gutachten
- Umlaufzeit typischerweise 1-3 Jahre

[Double-Blind doesn't go nearly far enough](#) (Widom)

17

- Organisatorisches
- Konferenzen und Journale
 - Gutachten
- ➔ ■ Gliederung eines Artikels
- Experimente
- Literaturrecherche



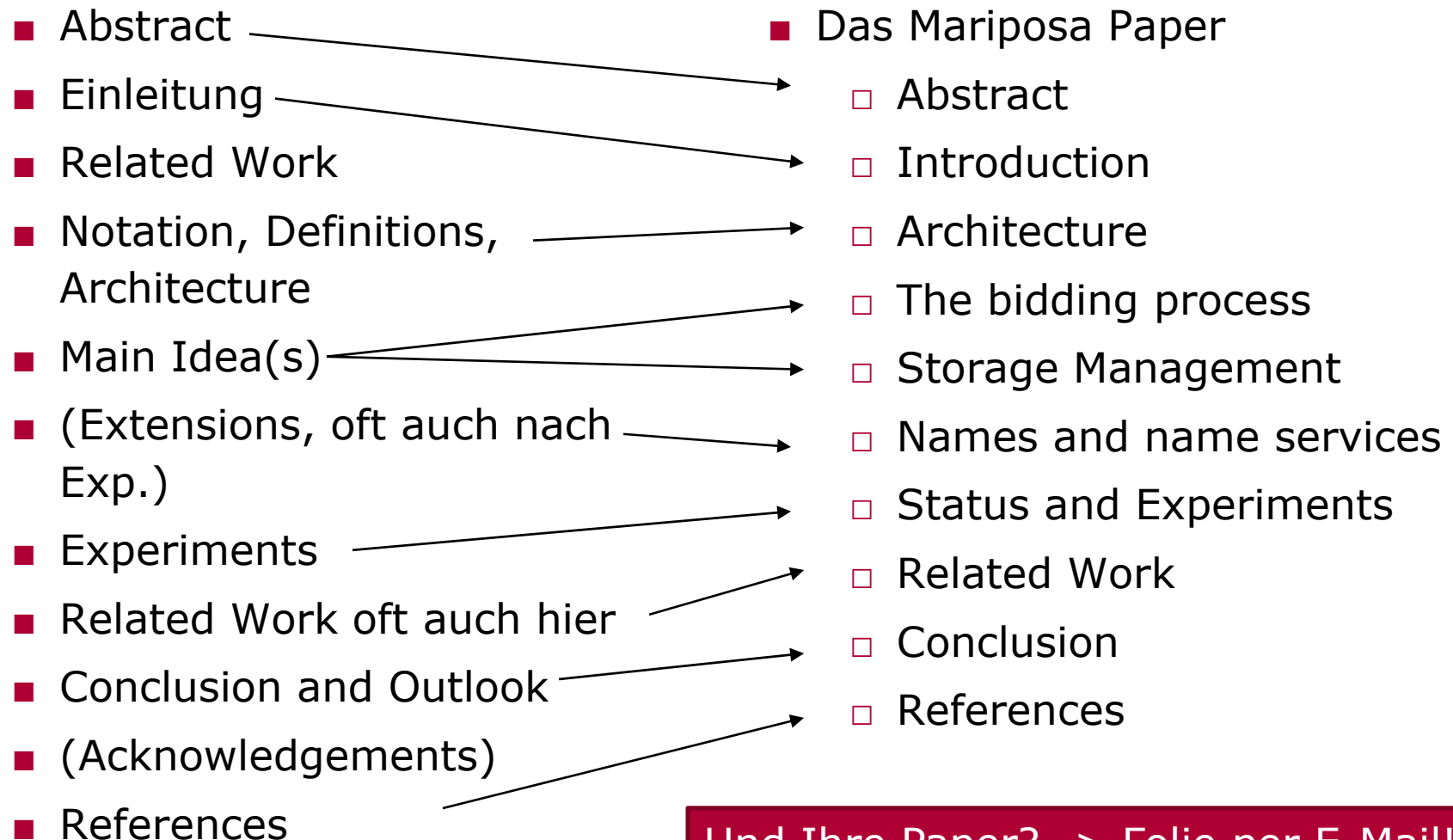
Artikelarten

18

- Journalartikel (10 – 50 Seiten)
 - Oft als Abschluss eines Projektes / einer Dissertation
- Zeitschriftenbeitrag (2 – 10 Seiten)
 - Oft mit künftigen Forschungsvorhaben oder Projektüberblicken
- Konferenzbeitrag (6 – 12 Seiten)
 - Konkrete Forschungsergebnisse
- Demo auf Konferenz (2-4 Seiten)
 - Beschreibung einer Prototyp Demo
- Poster auf Konferenz (3-5 Seiten)
 - „kleines paper“
- Workshopbeitrag (6-12 Seiten)
 - Forschungsergebnisse im Zwischenstadium
- Technischer Bericht (10-30 Seiten)
 - Lange Version eines Konferenzbeitrags
 - Herausgegeben von einer Institution
 - http://www.hpi.uni-potsdam.de/forschung/publikationen/technische_berichte.html

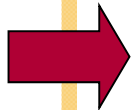
Typische Gliederung

19



Und Ihre Paper? -> Folie per E-Mail!

- Organisatorisches
- Konferenzen und Journale
 - Gutachten
- Gliederung eines Artikels
- Experimente
- Literaturrecherche



Experimente kritisch begutachten

21

- Welche (vereinfachenden) Annahmen wurden getroffen
- Welche Daten wurden verwendet?
 - Real-World-Daten (Szenario?)
 - Künstliche Daten
 - Datenmenge
- Skalen der Grafiken
- Lesbarkeit der Graphiken
- Interpretation
 - Wurden Auffälligkeiten begründet?
- Vollständigkeit der Experimente
 - Wurden alle Aspekte der vorigen Abschnitte getestet?
 - Wurden alle Fragen beantwortet?
 - Funktionalität und Laufzeit

Repeatability

22

- SIGMOD seit 2008: Repeatability
- VLDB seit 2008: Experiments and Evaluation
 - Consolidation and Validation
 - ◇ „Motivated by these surprisingly excellent results, we take a look into the rearview mirror. We have re-implemented the Dwarf index from scratch and make three contributions. First, we successfully repeat several of the experiments of the original paper. Second, we **substantially correct some of the experimental results** reported by the inventors. Some of our results **differ by orders of magnitude.**“
From: Jens Dittrich, Lukas Blunschi, Marcos Antonio Vaz Salles. Dwarfs in the rearview mirror: how big are they really? VLDB 2008
 - ◇ Entlarvung der Adaptive Sorted Neighborhood Method:
“Die Verfahren der ASNM konnten evaluiert und die guten Ergebnisse dieser Verfahren reproduziert werden. Allerdings konnte ebenfalls gezeigt werden, dass die Autoren bei dem Vergleich ihrer Verfahren mit der SNM offensichtlich nicht die transitive Hülle berücksichtigten, denn nur so konnten die großen Unterschiede in den Vergleichen nachvollzogen werden. Unter Berücksichtigung der transitiven Hülle **schneidet die SNM dagegen im Vergleich zu den vorgestellten Verfahren sehr gut oder sogar besser ab.**“
 - By Oliver Wonneberg (HPI, BTW 2009 Studierendenprogramm)

- *It is believed*
 - Ich glaube
- *It is generally believed*
 - Ein paar andere glauben das auch
- *It has long been known*
 - Ich hab mir das Originalzitat nicht herausgesucht
- *In my experience*
 - Einmal
- *In case after case*
 - Zweimal
- *In a series of cases*
 - Dreimal
- *Preliminary experiments showed that...*
 - Wir hoffen, dass...
- *Several lines of evidence demonstrate that...*
 - Es würde uns sehr gut in den Kram passen
- *A definite trend is evident*
 - Diese Daten sind praktisch bedeutungslos
- *While it has not been possible to provide definite answers to the questions*
 - Ein nicht erfolgreiches Experiment, aber ich hoffe immer noch, dass es veröffentlicht wird
- *Three of the samples were chosen for detailed study*
 - Die anderen Ergebnisse machten überhaupt keinen Sinn
- *Typical results are shown in Fig. 1*
 - Das ist die schönste Grafik, die ich habe

Forschersprache 2/2

24

- *Correct within an order of magnitude*
 - Falsch
- *A statistically-oriented projection of the significance of these findings*
 - Eine wilde Spekulation
- *A careful analysis of obtainable data*
 - Drei Seiten voller Notizen wurden vernichtet, als ich versehentlich ein Glas Bier drüber kippte
- *It is clear that much additional work will be required before a complete understanding of this phenomenon occurs*
 - Ich verstehe es nicht
- *After additional study by my colleagues*
 - Sie verstehen es auch nicht
- *Thanks are due to Joe Blotz for assistance with the experiment and to Cindy Adams for valuable discussions*
 - Herr Blotz hat die Arbeit gemacht, und Frau Adams erklärte mir, was das alles bedeutet
- *The purpose of this study was...*
 - Es hat sich hinterher herausgestellt, dass ...
- *Our results confirm and extend previous conclusions that...*
 - Wir fanden nichts neues
- *It is hoped that this study will stimulate further investigation in this field*
 - Ich geb's auf!



Hasso
Plattner
Institut

IT Systems Engineering | Universität Potsdam

*Thoughts about the Experimental
Culture*

in Our Community

An Experiment:

How to Plan it, Run it,
and Get it Published

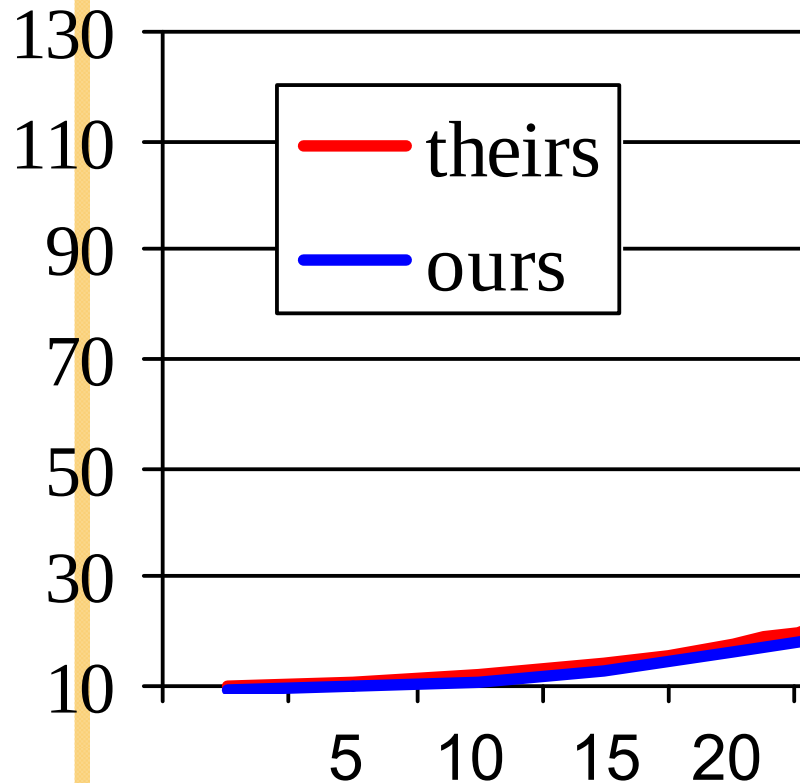
Gerhard Weikum

Performance Experiments (1)

throughput, response time, #IOs, CPU, wallclock,
„DB time“, hit rates, space-time integrals, etc.

26

speed (RT, CPU, etc.)



There are
lies, damn lies, and
workload assumptions

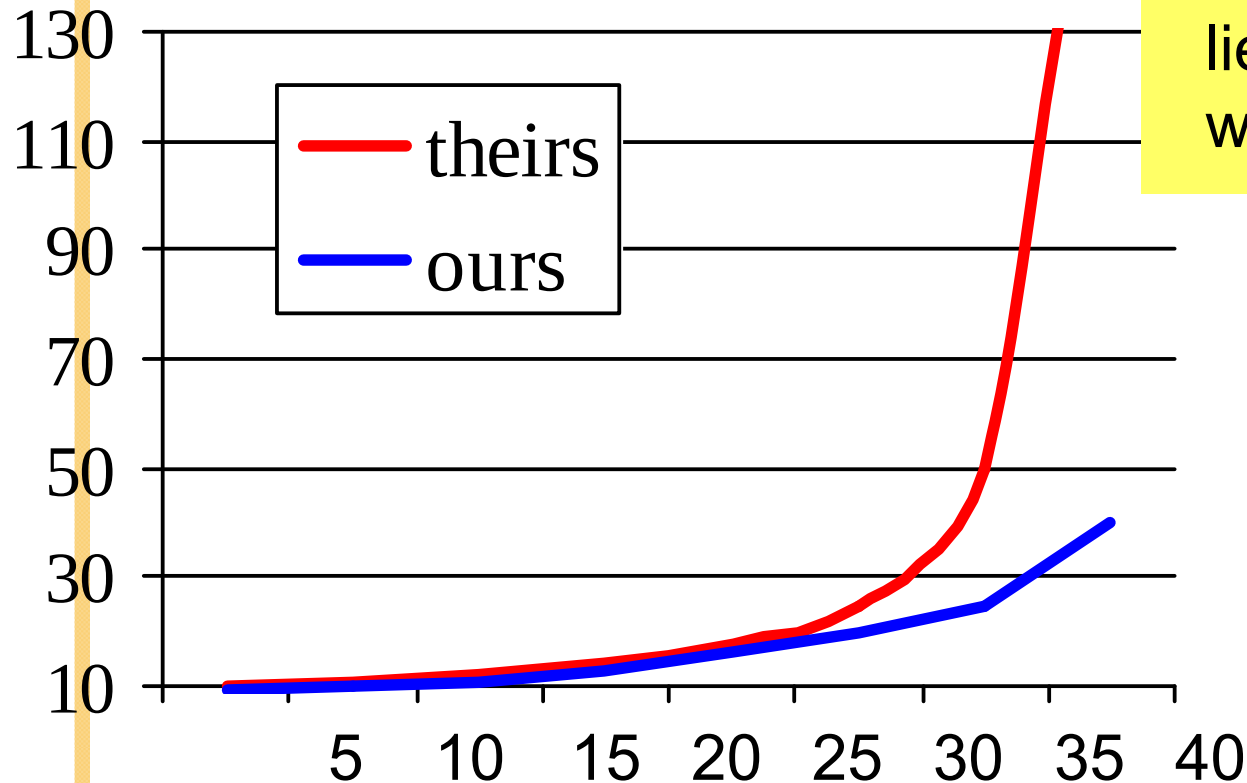
load (MPL,
arrival rate, etc.)

Performance Experiments (1)

throughput, response time, #IOs, CPU, wallclock,
„DB time“, hit rates, space-time integrals, etc.

27

speed (RT, CPU, etc.)

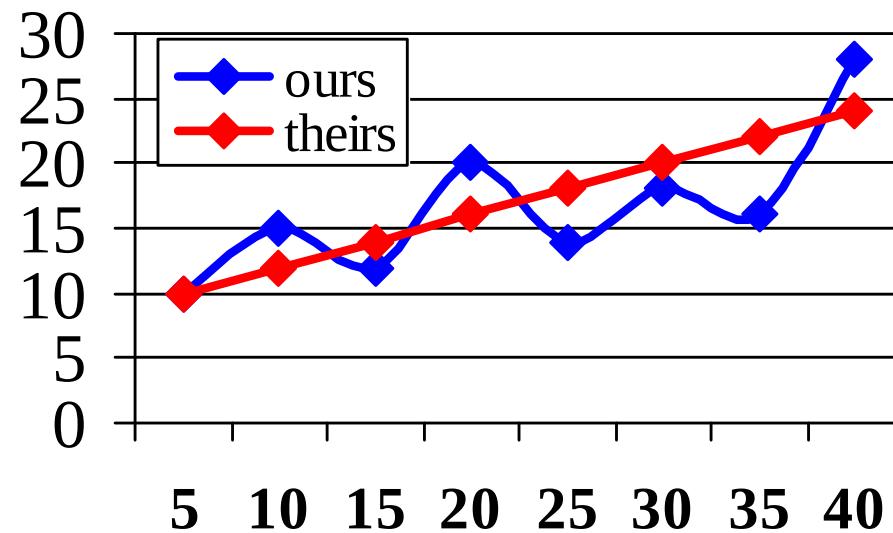


There are
lies, damn lies, and
workload assumptions

load (MPL,
arrival rate, etc.)

Performance Experiments (2)

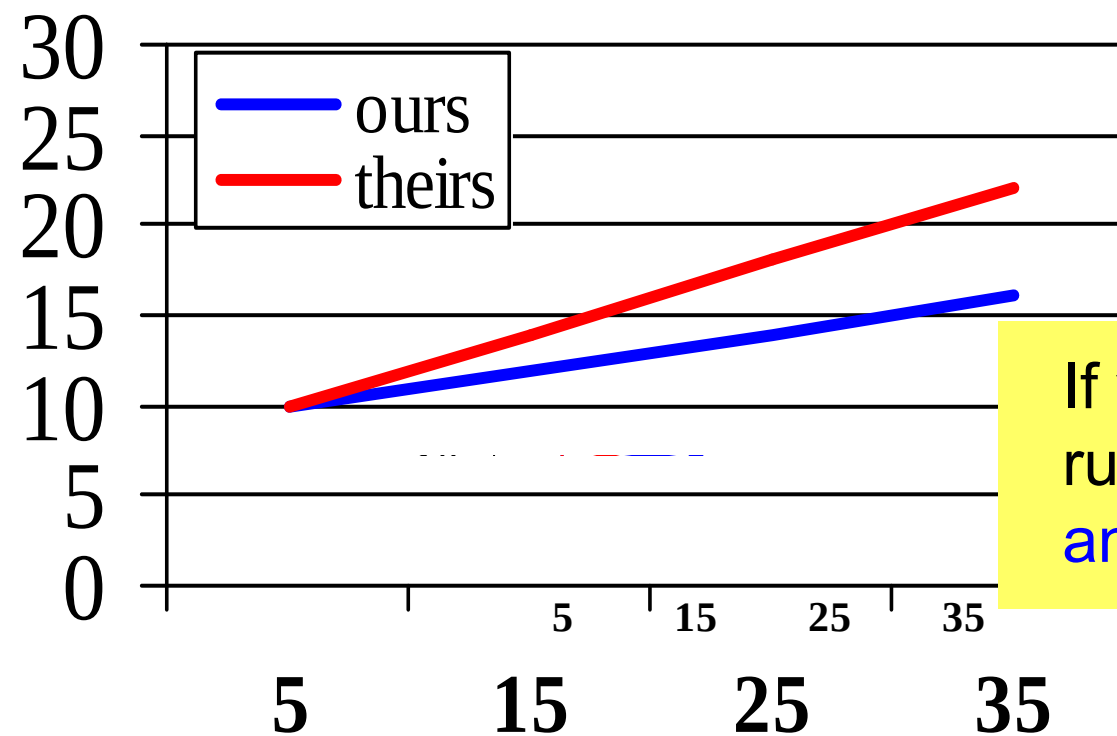
28



If you can't reproduce it,
run it only once

Performance Experiments (2)

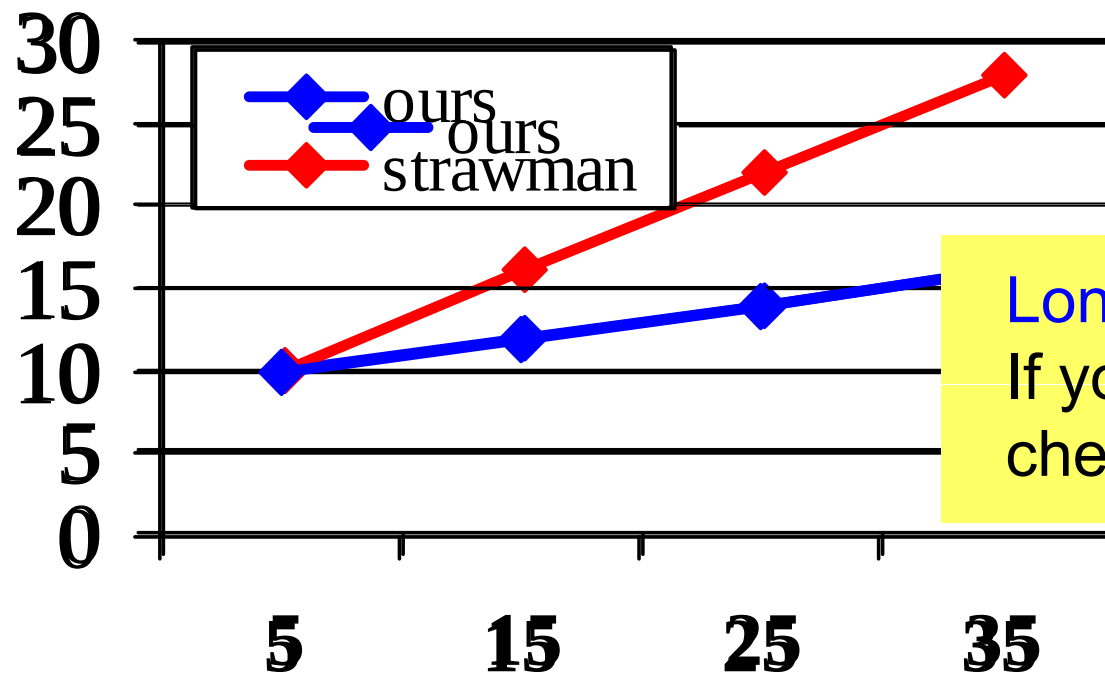
29



If you can't reproduce it,
run it only once
and smoothe it

Performance Experiments (3)

30

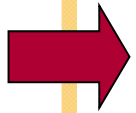


Lonesome winner:
If you can't beat them,
cheat them

90% of all algorithms
are among the best 10%

93.274% of all statistics
are made up

- Organisatorisches
- Konferenzen und Journale
 - Gutachten
- Gliederung eines Artikels
- Experimente
- Literaturrecherche



Suche

32

Rückwärtssuche

- Suche nach referenzierten Artikeln
- Suche nach längeren Versionen
- Suche nach früheren Versionen

Vorwärtssuche

- Suche nach Artikeln, die den vorliegenden referenzieren
 - Vom gleichen Autor
 - ◇ Journal-Version
 - Von anderen Autoren
 - In einem survey (Überblicksartikel)

Suche am Beispiel

33

dblp .uni-trier.de

Recherche auf DBLP

- <http://www.informatik.uni-trier.de/~ley/db/index.html>
- [Recherche\vldb94-487.html](#)

Recherche auf Citeseer

- <http://citeseer.ist.psu.edu/>
- [Recherche\392628.html](#)

CiteSeer.IST
Scientific Literature Digital Library

Recherche auf Google Scholar

- <http://scholar.google.de/>
- [Recherche\scholar.htm](#)

GoogleTM
Scholar BETA

Weitere Quellen

34

- Vorträge
 - Folien + Video manchmal verfügbar
- ACM (Association for Computing Machinery)
 - Digital Lib
 - <http://portal.acm.org/portal.cfm>
- SpringerLink
 - <http://www.springerlink.de/>
- IEEE (Institute of Electrical and Electronics Engineers)
 - <http://www.computer.org/>
- Homepages der Autoren!
- E-Mail Adressen der Autoren
- Und: Bücher