



**Hasso
Plattner
Institut**

IT Systems Engineering | Universität Potsdam

SE: Beauty is our Business

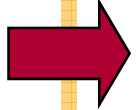
Wissenschaftliche Arbeiten Lesen

23.4.2008

Felix Naumann

Überblick

2



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



Die Themen (und Termine)

3

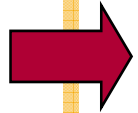
- Mariposa
 - Stefan Richter (+ Alexander Kuscher)
- Trio
 - Alexander Kuscher (+ Stefan Richter)
- Fagins Algorithmus
 - Felix Elliger (+ Fabian Lindenberg)
- Enough Already in SQL
 - Fabian Lindenberg (+ Felix Elliger)
- Sorted Neighborhood
 - Stefanie Reinicke (+ Jan Oberst)
- Data Mining
 - Jan Oberst (+ Stefanie Reinicke)

Termine

4

- Besprechungstermine mit mir ausmachen!
 - E-Mail / Telefon / Vorbeikommen - Pat Hobro
- 14.5.2008 Literaturkritik / Diskussion
- 21.5.2008 Vortragstechniken
- 4.6.2008
 - Vortrag 1: Mariposa
 - Vortrag 2: Trio
- 11.6.2008 Einführung in LaTeX
- 18.6.2008
 - Vortrag 3: Fagins Algorithmus
 - Vortrag 4: Enough Already in SQL
- 25.6.2008
 - Vortrag 5: Sorted Neighborhood
 - Vortrag 6: Data Mining
- 2.7.2008 Vorstellung der Gliederungen & Tipps zur Ausarbeitung
- 15.8.2008 Abgabe der Ausarbeitungen

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

- SIGMOD
 - ACM Special Interest Group – Management of Data
 - GI – Gesellschaft für Informatik
- ICDE
 - IEEE
 - International Conference on Database Engineering
- VLDB
 - Very Large Databases
 - 2003 Berlin
- EDBT
- BTW (2007 in Aachen)
- 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

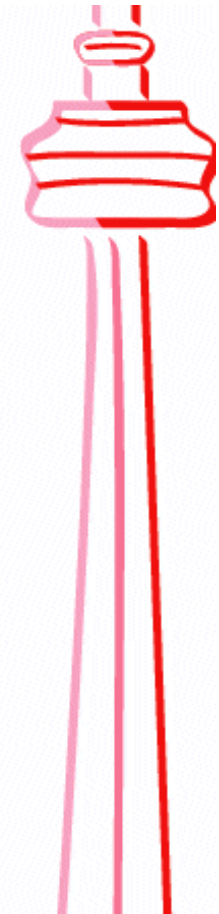
Publicity and Publications Chair

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					
<u>Upper Canada</u>					
18:00-20:00	<u>Welcomes Reception (Upper Canada)</u>				
Tuesday at a Glance					
	<u>Ballroom</u>	<u>Ontario</u>	<u>Territories</u>	<u>Confederation 5 & 6</u>	<u>Tudor 7 & 8</u>
07:30-08:30	Breakfast (<u>Concert Hall</u>)				
08:30-09:00	<u>Opening Ceremony (Ballroom)</u>				
09:00-10:30	<u>Keynote Address 1 (Ballroom): Databases in a Wireless World</u> David Yach, VP Software, Research in Motion				
10:30-11:00	Coffee Break (<u>Ballroom Foyer</u>)				
11:00-12:30	<u>Research Session 1</u> Compressing & Indexing	<u>Research Session 2</u> XML Views and Schemas	<u>Research Session 3</u> Controlling Access	<u>Tutorial 1</u> Database Architectures for New Hardware	NO SESSION
12:30-14:00	Lunch Break (<u>Concert Hall</u>)				
14:00-15:30	<u>Research Session 4</u> XML (I)	<u>Research Session 5</u> Stream Mining	<u>Industrial Session 1</u> Novel SQL Extensions	<u>Tutorial 1</u> Database Architectures for New Hardware	<u>Demo</u> Groups 2 and 3
15:30-16:00	Coffee Break (<u>Ballroom Foyer</u>)				
16:00-17:30	<u>Research Session 6</u> XML (II)	<u>Industrial Session 2</u> New DBMS architectures and Performance	<u>Panel 1</u> Biological Data Management: Research, Practice and Opportunities	<u>Tutorial 2</u> Security of Shared Data in Large Systems: State of the Art and Research Directions	<u>Demos</u> Groups 2 and 3

Typisches Programm einer DB Konferenz

14

Wednesday at a Glance					
	Ballroom	Ontario	Territories	Confederation 5 & 6	Tudor 7 & 8
07:45-08:45	Breakfast (Concert Hall)				
09:00-10:30	Keynote Address 2 (Ballroom) : Structures, Semantics and Statistics Alon Halevy (U. of Washington)				Demo Groups 1 and 2
10:30-11:00	Coffee Break (Ballroom Foyer)				
11:00-12:30	Research Session 7 XML and Relations	Research Session 8 Stream Mining (II)	Industrial Session 3 Semantic Query Approaches	Tutorial 2 Security of Shared Data in Large Systems: State of the Art and Research Directions	Demo Groups 1 and 2
12:30-14:00	Lunch Break (Concert Hall)				
14:00-16:00	Research Session 9 Stream Query Processing	Research Session 10 Managing Web Information Sources	Research Session 11 Distributed Search and Query Processing	Tutorial 3 Self-Managing Technology in Database Management Systems	Demo Groups 4 and 5
16:00-16:30	Coffee Break (Ballroom Foyer)				
16:30-17:30	Research Session 12 Stream Data Management Systems	Research Session 13 Auditing	Research Session 14 Data Warehousing	Tutorial 3 Self-Managing Technology in Database Management Systems	Demo Groups 4 and 5
18:00-21:00	Conference Banquet (Royal Ontario Museum)				

Thursday at a Glance					
	Ballroom	Ontario	Territories	Confederation 5 & 6	Tudor 7 & 8
07:45-08:45	Breakfast (Concert Hall)				
09:00-10:30	10 Year Best Paper Award (Ballroom)				Demo Groups 3 and 4
10:30-11:00	Coffee Break (Ballroom Foyer)				
11:00-12:30	Research Session 15 Link Analysis	Research Session 16 Sensors, Grid, Publish/Subscribe Systems	Industrial Session 4 Automatic Tuning in Commercial DBMSs	Panel 2 Where is Business Intelligence taking today's database systems?	Demo Groups 3 and 4
12:30-14:00	Lunch Break (Concert Hall)				
14:00-15:30	Research Session 17 Top-K Ranking	Research Session 18 DBMS Architecture and Performance	Industrial Session 5 XML implementations, automatic physical design, and indexing	Tutorial 4 Architectures and Algorithms for Internet-Scale (P2P) Data Management	Demo Groups 1 and 6
15:30-16:00	Coffee Break (Ballroom Foyer)				
16:00-17:30	Research Session 19 Privacy	Research Session 20 Nearest Neighbor Search	Industrial Session 6 Data Management with RFIDs and Ease of Use	Tutorial 4 Architectures and Algorithms for Internet-Scale (P2P) Data Management	Demo Groups 1 and 6

Friday at a Glance					
	Ballroom	Ontario	Territories	Confederation 5 & 6	Tudor 7 & 8
07:45-08:45	Breakfast (Concert Hall)				
09:00-10:30	Research Session 21 Similarity Search and Applications	Research Session 22 Query Processing	Industrial Session 7 Data management challenges in Life Sciences and Email systems	Tutorial 5 The continued saga of DB-IR integration	Demo Groups 5 and 6
10:30-11:00	Coffee Break (Ballroom Foyer)				
11:00-12:30	Research Session 23 Novel Models	Research Session 24 Query Processing and Optimization	Industrial Session 8 Issues in data warehousing	Tutorial 5 The continued saga of DB-IR integration	Demo Groups 5 and 6

Workflow zur Veröffentlichung

15

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

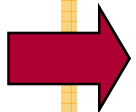
17

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)

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



Artikelarten

19

- 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

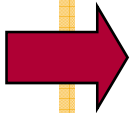
20

- | | | |
|---------------------------------------|---|---------------------------|
| ■ Abstract | → | ■ Das Mariposa Paper |
| ■ Einleitung | → | □ Abstract |
| ■ Related Work | → | □ Introduction |
| ■ Notation, Definitions, Architecture | → | □ Architecture |
| ■ Main Idea(s) | → | □ The bidding process |
| ■ (Extensions, oft auch nach Exp.) | → | □ Storage Management |
| ■ Experiments | → | □ Names and name services |
| ■ Related Work oft auch hier | → | □ Status and Experiments |
| ■ Conclusion and Outlook | → | □ Related Work |
| ■ (Acknowledgements) | → | □ Conclusion |
| ■ References | → | □ References |

Und die anderen Paper?

21

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



Experimente kritisch begutachten

22

- 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

23

- SIGMOD 2008: Repeatability
- VLDB 2008: Experiments and Evaluation
 - Consolidation
 - Validation

- *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

25

- *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

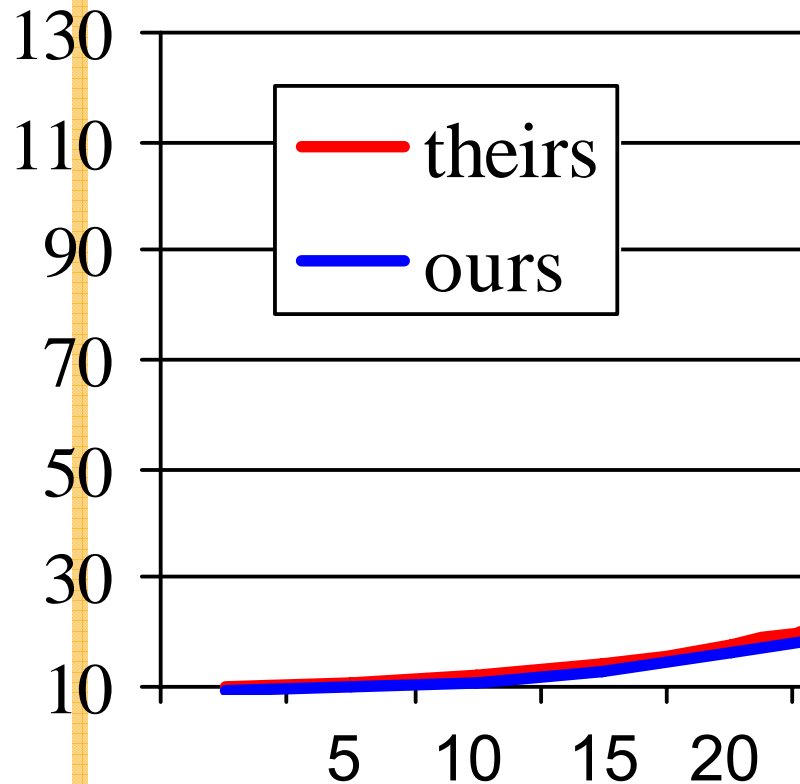
Gerhard Weikum

Performance Experiments (1)

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

27

speed (RT, CPU, etc.)



There are lies, damn lies, and workload assumptions

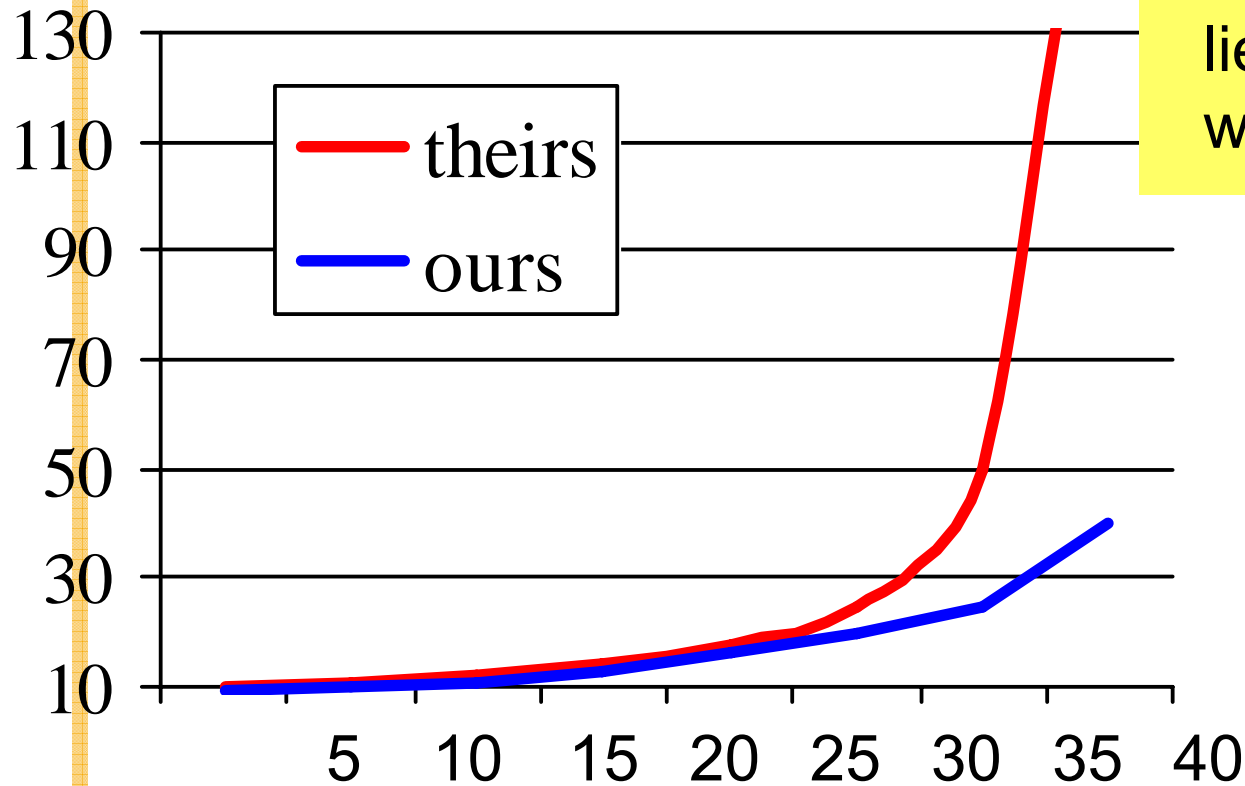
load (MPL, arrival rate, etc.)

Performance Experiments (1)

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

28

speed (RT, CPU, etc.)

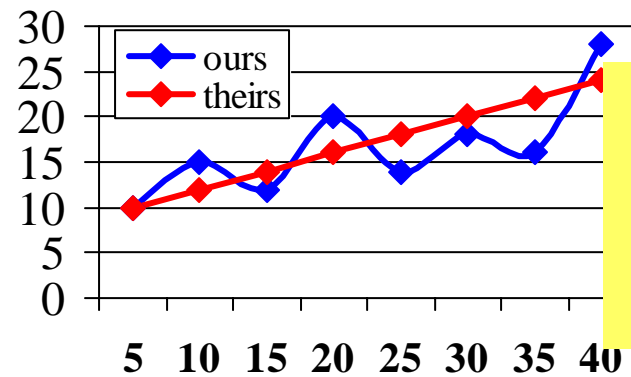


There are lies, damn lies, and workload assumptions

load (MPL, arrival rate, etc.)

Performance Experiments (2)

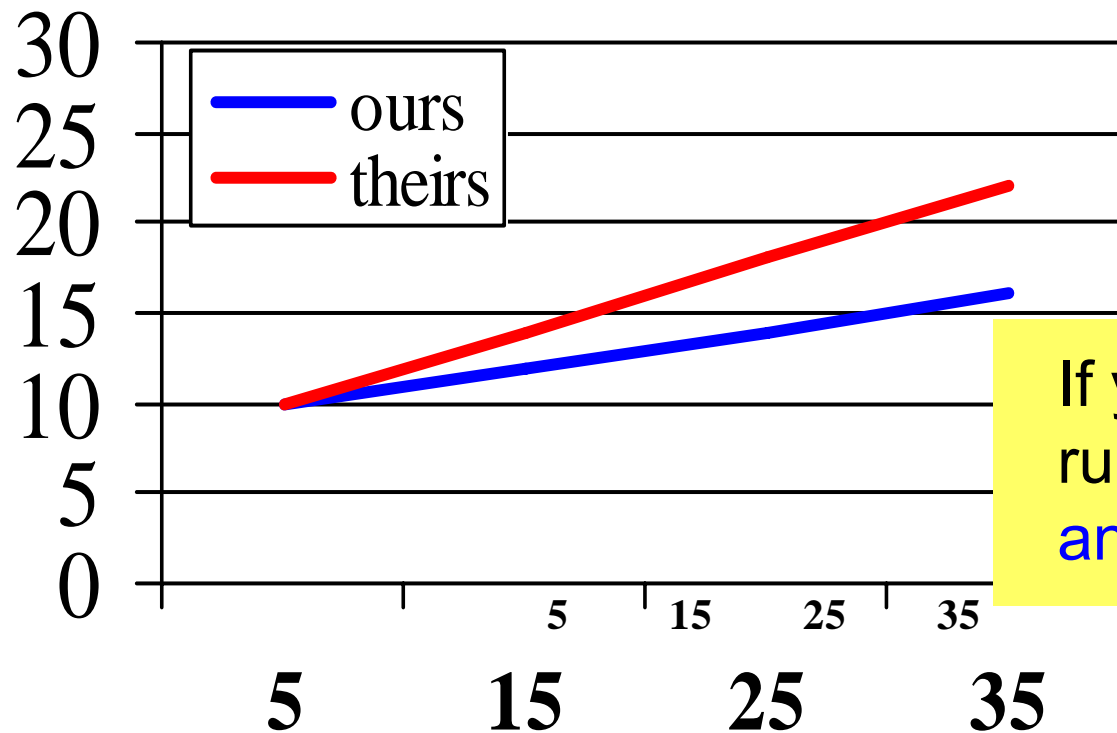
29



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

Performance Experiments (2)

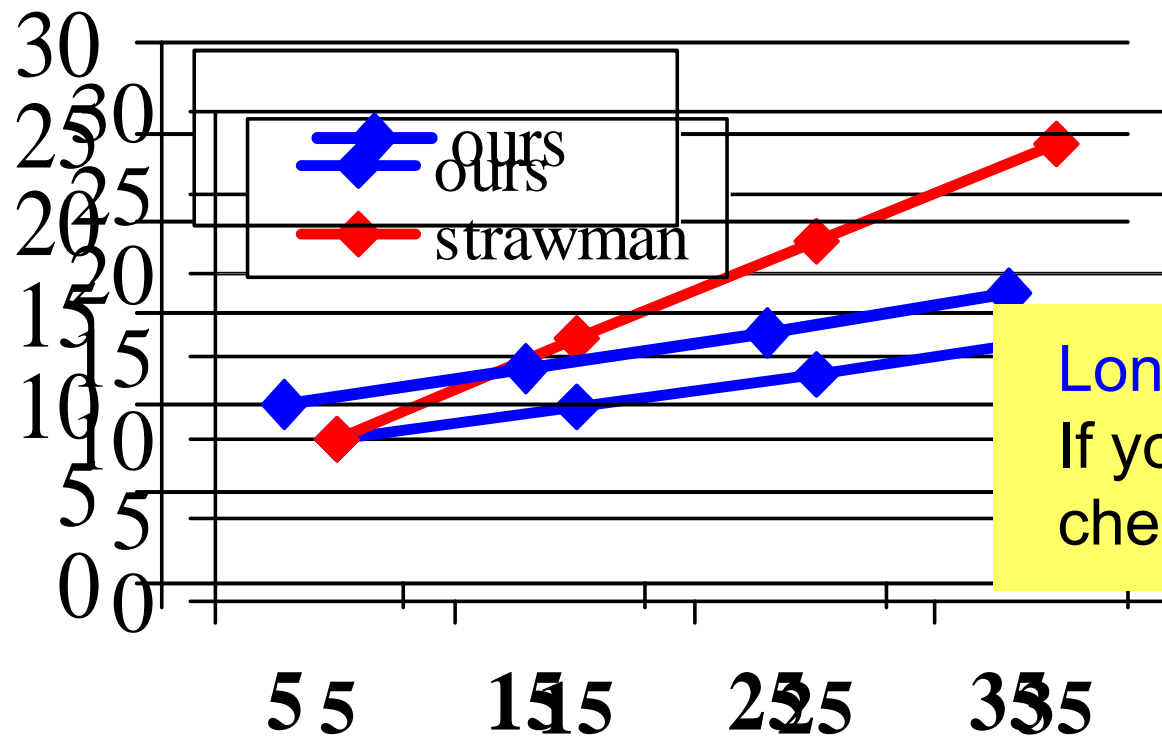
30



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

Performance Experiments (3)

31

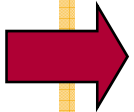


Lonesome winner:
If you can't beat them,
cheat them

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

33

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

34

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](#)

Recherche auf Google Scholar

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



CiteSeer.IST
Scientific Literature Digital Library



GoogleTM
Scholar BETA

Weitere Quellen

35

- ACM (Association for Computing Machinery)
 - Digital Library
 - <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