

# The Second International Conference on Knowledge Discovery and Data Mining

Sponsored by the American Association for Artificial Intelligence  
Portland, Oregon, USA, August 2-4, 1996

## Conference Brochure

### 1996 KDD Program Committee

#### General Conference Chair

Usama M. Fayyad, *Microsoft*

#### Program Cochairs

Jiawei Han, *Simon Fraser University*

Evangelos Simoudis, *IBM Almaden Research Center*

#### Publicity Chair

Padhraic Smyth, *University of California, Irvine*

#### Sponsorship Chair

Gregory Piatetsky-Shapiro, *GTE Laboratories*

#### Program Committee

Rakesh Agrawal, *IBM Almaden Research Center*

Tej Anand, *AT&T Global Information Solutions*

Ron Brachman, *AT&T Bell Laboratories*

Wray Buntine, *Heuristicrats Research*

Nick Cercone, *University of Regina*

Peter Cheeseman, *NASA AMES Research Center*

Bruce Croft, *University of Massachusetts at Amherst*

Stephen G. Eick, *AT&T Bell Laboratories*

Usama Fayyad, *Microsoft*

Dan Geiger, *Technion*

Clark Glymour, *Carnegie-Mellon University*

George Grinstein, *University of Lowell*

David Hand, *Open University*  
David Heckerman, *Microsoft Corporation*  
Se June Hong, *IBM T. J. Watson Research Center*  
Tomasz Imielinski, *Rutgers University*  
Larry Jackel, *AT&T Bell Laboratories*  
Larry Kerschberg, *George Mason University*  
Willi Kloesgen, *GMD*  
David Madigan, *University of Washington*  
Heikki Mannila, *University of Helsinki*  
Chris Matheus, *GTE Laboratories*  
Sham Navathe, *Georgia Institute of Technology*  
Raymond Ng, *University of British Columbia*  
Gregory Piatetsky-Shapiro, *GTE Laboratories*  
Daryl Pregibon, *AT&T Bell Laboratories*  
Pat Riddle, *Boeing Computer Services*  
Ted Senator, *US Department of the Treasury*  
Wei-Min Shen, *University of Southern California*  
Arno Siebes, *CWI*  
Avi Silberschatz, *AT&T Bell Laboratories*  
Andrzej Skowron, *University of Warsaw*  
Steve Smith, *Dun and Bradstreet*  
Padhraic Smyth, *University of California, Irvine*  
Ramakrishnan Srikant, *IBM Almaden Research Center*  
Sal Stolfo, *Columbia University*  
Paul Stolorz, *Jet Propulsion Laboratory*  
Alex Tuzhilin, *NYU Stern School*  
Ramasamy Uthurusamy, *GM Research Laboratories*  
Xindong Wu, *Monash University*  
Wojciech Ziarko, *University of Regina*  
Jan Zytkow, *Wichita State University*

## KDD-96 Opening Reception

The KDD-96 opening reception will be held August 2 from 6:00 - 8:00 PM in the Oregon Convention Center in conjunction with a poster session. The reception is open to KDD-96 registrants only.

## KDD-96 Conference Banquet

A banquet will be held for interested KDD-96 attendees on Saturday evening, August 3, 1996. Information will be sent to KDD-96 registrants as soon as it becomes available. The banquet will require payment of an additional fee.

# Registration Information

Your KDD-96 program registration includes admission to the KDD-96 sessions, the KDD-96 invited talks, the AAI-96 Exhibition, the KDD-96 opening reception, and the KDD-96 conference *Proceedings*.

Onsite registration will be located in  
Concourse A, lower level  
Oregon Convention Center  
777 Northeast Martin Luther King Jr. Boulevard  
Portland, Oregon 97212

Registration hours will be Friday, August 2 through Sunday, August 4, from 7:30 AM - 6:00 PM. All attendees must pick up their registration packets for admittance to programs.

## KDD-96 Registration Fees

### **Early Registration (Postmarked by June 7)**

#### *AAAI Members*

Regular \$185

Students \$85

#### *Nonmembers*

Regular \$245

Students \$115

### **Late Registration (Postmarked by July 5)**

#### *AAAI Members*

Regular \$240

Students \$110

#### *Nonmembers*

Regular \$300

Students \$140

### **On-Site Registration (Postmarked after July 5 or onsite.)**

#### *AAAI Members*

Regular \$295

Students \$135

#### *Nonmembers*

Regular \$355

Students \$165

## Payment and Registration Information

Prepayment of registration fees is required. Checks, international money orders, bank transfers, and traveler's checks must be in US dollars. American Express, MasterCard, VISA, and government purchase orders are also accepted. Registration applications postmarked after the early registration deadline will be subject to the late registration fees. Registration applications postmarked after the late registration deadline will be subject to on-site registration fees.

## Refund Requests

All refund requests must be made in writing. A \$75.00 processing fee will be assessed for all refunds. Student registrations must be accompanied by proof of full-time student status. The deadline for refund requests is July 12, 1996.

## Registration forms and inquiries should be directed to :

KDD-96  
445 Burgess Drive  
Menlo Park, CA 94025 USA  
Phone: (650) 328-3123, Fax: (650) 321-4457  
email: [kdd@aaai.org](mailto:kdd@aaai.org)  
URL: <http://www.aaai.org>

## Invited Talks

### Harnessing the Human in Knowledge Discovery

*Georges Grinstein, University of Massachusetts at Lowell, MA and The MITRE Corporation*

Knowledge, the primary goal of data analysis and exploration, is most often discovered by generating information (structure) from data, and then abstracting non-trivial patterns (rules or associations for example) from the information. The discovery process can be done using visualization, data mining, statistics, neural networks, or mathematical modeling and simulation. Visualization is different from the rest, however, in that it is also the actual mechanism by which the analyses and their results can be presented to the user. In this talk, I will present a brief history of alternative visualizations and how they have been applied to various data visualization problems. The emphasis will be on how exploratory visualization can support the knowledge discovery process, including concept development for database management, database visualizations, and minimally structured dataset visualizations.

**Georges Grinstein** is professor of computer science at the University of Massachusetts Lowell, Director of the Institute for Visualization and Perception Research and principal engineer with the MITRE Corporation's Center for Air Force C3I Systems. He received his BS from the City College of New York in 1967, his MS from the Courant Institute of Mathematical Sciences of New York University in 1969 and his PhD in mathematics from

the University of Rochester in 1978. He has chaired several conferences on visualization, and is a member of IEEE, and ACM.

## Materializing Views in Data Warehouses

*Jeffrey D. Ullman, Stanford University*

Data warehouses are collections of materialized views of source data. The optimal set of views to materialize depends on the assumed distribution of queries that will be posed about the data. Given a query distribution, a "greedy" approach to selecting materialized views picks a sequence of views, each of which provides the maximum "benefit" (reduction in average query cost), given the set of views previously chosen for materialization. Under a variety of assumptions about the way possible views relate to one another, greedy approaches are guaranteed to come within 63 percent of the optimum benefit. In fact, in some of these cases, such as the important case of a "data cube" storing multidimensional data, it can be shown that no polynomial algorithm can be guaranteed to come closer than 63 percent.

**Jeffrey D. Ullman** (PhD, Princeton University, 1966), is a professor at Stanford University, and a leading researcher in database systems, knowledge-base systems, theoretical computer science, analysis of algorithms, and programming languages and compilers. He served as the chairman of the Department of Computer Science, Stanford University, from 1990 to 1994 and served or is serving on the editorial boards of many scientific journals, including *SIAM Journal Computing*, *Journal ACM*, *Journal Computer and System Sciences*, and *Journal Logic Programming*. He has authored or coauthored 14 books and 160 technical publications, including *Foundations of Computer Science* (1992, 1994), *Principles of Database and Knowledge-Base Systems* (two volumes, 1988, 1989), *Introduction to Automata Theory, Languages, and Computation* (1979), and *The Design and Analysis of Computer Algorithms* (1974). Recently, he is also working on the efficient implementation of data cubes for data warehousing and data mining.

This talk presents work of Chandra Chekuri, Himanshu Gupta, Venky Harinarayan, and Anad Rajaraman, as well as the author.

## Small Sample Size Paradigm in Statistical Inference

*Vladimir Vapnik, AT&T Research Laboratories*

Vladimir Vapnik will describe (from both the theoretical and the applied point of view) a new approach to statistical inference that is based on the minimization of the guaranteed risk for a fixed sample size, which provides a high level of generalization ability and in many cases contradicts the existing classical paradigms.

**Vladimir Vapnik**, currently a member of the technical staff at AT&T Research Laboratories, is one of the creators of generalization theory in statistical inference, the so-called VC-theory (Vapnik-Chervonenkis theory). In particular, he developed the statistical theory for inference from small sample size (in contrast to classical asymptotic statistics). Vapnik is the author of seven books, including *Estimation of Dependencies Based on Empirical Data* (Springer, 1982), *The Nature of Statistical Learning Theory* (Springer, 1995), and *Statistical Learning Theory* (Wiley, 1996).

## Data Integration and Analysis in a Client Server Environment: The Sara Lee Meat Experience

*Perry K. Youngs, Sara Lee Corporation*

The role of marketing research is currently going through dramatic changes in the United States as census based syndicated scanner data is becoming available to retailers and manufacturers. This change is being lead by ECR and Category Management initiatives that are removing costs from distribution channels. In an attempt to manage the ever increasing amounts of information needed for this endeavor, client server based information systems are being developed with new data warehousing technology.

Sara Lee Meats has just successfully implemented the conversion of a main frame based system to a client server based system using a three tier object technology from Information Advantage, Incorporated and data warehousing technology from Red Brick Systems, Incorporated. Youngs will discuss Sara Lee Meat's experiences relating to data integration and analysis in a client server environment.

**Perry K. Youngs** began his career in the early 1980s working for an office furniture manufacturer in Grand Rapids, Michigan as a marketing analyst. Working on a TRS-80 computer with 64K of RAM and eight inch floppy disks, Youngs soon became the company PC expert. This is a time when IS people laughed at marketing people needing anything more than a PC with ten megabytes of disk space. He then was hired by Bil Mar Foods, Inc. where he was responsible for developing internal marketing/sales information systems. In 1988, with the introduction of external syndicated services from Information Resources, Inc. and A. C. Nielsen, as well as the purchase of Bil Mar Foods by the Sara Lee Corporation, Youngs was given the challenge of building a decision support system for all of the 10 domestic Sara Lee Meat based food companies. Youngs is currently located at the Sara Lee Meat Group corporate headquarters in Cordova, Tennessee.

Youngs has a BA in marketing from Grand Valley University as well as postgraduate studies. He has been an adjunct instructor at a small private business college and currently sits on the board of directors of the Memphis chapter of the American Marketing Association.