Preface

As we enter the true digital information era, one of the greatest challenges facing organizations and individuals is how to turn their rapidly expanding data stores into accessible, and actionable knowledge. Digital data sources are ubiquitous, created by a variety of means spanning a spectrum of activities: from a supermarket's electronic scanner, to a bank's automated teller machine, from a credit card reader, to a world wide web server, and the most intricate of technical instruments. While advances on data storage and retrieval continue at a breakneck pace, (several organizations have databases that today contain several hundreds of gigabytes, and in some instances terabytes of online data with millions of rows and hundreds of columns; within two years the multi-terabyte database will be common-place) the same cannot be asserted about the advances in information and knowledge extraction from large data sets. Only a very small percentage of the captured data is ever converted to actionable knowledge. The traditional approach of a human analyst, intimately familiar with a data set, serving as a conduit between raw data and synthesized knowledge by producing useful analyses and reports, is breaking down.

Responding to this need, researchers from fields such as pattern recognition, statistics, artificial intelligence, very large databases, and visualization in the mid-1980s started developing tools and techniques to discover knowledge from large, complex data stores. These researchers share a set of core issues: representation of discovered knowledge, search complexity, the use of prior knowledge, statistical inference, algorithms that scale to analysis of massive amounts of data both in size and dimensionality, managing uncertainty, and interactive (human-oriented) presentation. What started in 1989 as a workshop aimed at bringing together these research teams, culminated last year into the First International Conference on Knowledge Discovery and Data Mining (KDD-95), which was held in Montreal, Canada, August 20-21, 1995 in conjunction with IJCAI-95 and was attended by over 350 individuals from academia and industry.

The Second International Conference on Knowledge Discovery and Data Mining (KDD-96), held on August 2-4, 1996 in Portland, Oregon, USA, in conjunction with AAAI-96, provides a forum for KDD researchers and practitioners to present their latest work in this field. KDD-96 is a truly international conference. Of the 215 papers received for review 60% came from outside the United States with the following distribution: Australia (12), Austria (1), Belgium (2), Canada (14), Egypt (1), Finland (4), France (10), Germany (15), Hong Kong (2), Ireland (2), Israel (3), Italy (3), Japan (10), Mexico (1), Northern Ireland (3), Poland (5), Singapore (4), Taiwan (1), The Netherlands (6), and UK (10). Only 42 of the submitted papers were accepted for presentation to the conference; an acceptance rate of just under 20%. In addition, 30 papers were accepted for poster presentation. Posters were designated as short papers; each was allocated four pages in the proceedings, and a two-minute presentation as a "technology highlight." The KDD-96 program was further supplemented with five invited speakers and two special sessions each with an associated panel: "KDD and High Performance Computing" and "Systems for Mining Large Databases." The conference closes with a panel entitled "What Have We Discovered?"

Underlying the need to convert their data into actionable knowledge, organizations have started an aggressive effort to deploy KDD applications. Many such applications are now in production in industries such as finance, insurance, retail, telecommunications, health care, astronomy, planetary sciences, biology, etc. As is reported in the press, the early adopters of KDD are reaping significant financial benefits. With the number of knowledge discovery and data warehousing pilot projects doubling every six months, undoubtedly many more KDD applications will be in production by the end of 1996 and beyond. Knowledge discovery and data mining have captured the peoples' imagination. However, oftentimes we find that the excitement and promise offered by KDD is accompanied by high expectations for success that are unreasonable, unjustifiable, and premature. Obviously there is significant risk and exposure to the long term viability of the field by failing to manage expectations. KDD-96 aims at addressing these issues by providing a clear understanding of what represents the state of the art, and the state of practice in each of the various disciplines comprising KDD.

A conference such as this can only succeed as a team effort. We would first like to acknowledge the contributions of the program committee for their reviewing efforts as well as their invaluable input and advice. Our veteran chairs for Publicity (Padhraic Smyth), Corporate Sponsorship (Gregory Piatetsky-Shapiro), and Demonstration Sessions (Tej Anand) have continued the excellent track record they set in KDD-95. We are grateful to our sponsors for their generous support of KDD-96.

Finally, we would like to extend a special appreciation to the American Association for Artificial Intelligence (AAAI) for sponsoring KDD-96, the only conference AAAI has solely sponsored apart from the National Conference on Artificial Intelligence and the Innovative Applications of Artificial Intelligence Conference. We thank the AAAI staff for all their help and especially Annette Eldredge for handling submissions and local arrangements, and Carol Hamilton for her help, advice, and for managing the operations of this conference.

Evangelos Simoudis (IBM) *&*
Jiawei Han (Simon Fraser University)
Program Committee Cochairs

Usama Fayyad (Microsoft Research)
General Conference Chair

Portland, Oregon, USA, August 2, 1996