

Call for Papers

First International Workshop on

Knowledge Discovery in Inductive Databases (KDID'02)

In conjunction with [ECML/PKDD-2002](http://www.cinq-project.org/ecmlpkdd2002/) (see also <http://www.cinq-project.org/ecmlpkdd2002/>)

13th European Conference on Machine Learning (ECML'02) and 6th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'02)

Helsinki, Finland, August 19/20, 2002

Workshop supported by the "cInQ" European Project
[cInQ](http://www.cinq-project.org/), Consortium on Discovering Knowledge with Inductive Queries

See also

Tutorial by Jean-François Boulicaut and Luc De Raedt on

Inductive Databases and Constraint-Based Mining

In conjunction with [ECML/PKDD-2002](http://www.cinq-project.org/), Helsinki, Finland, August 19/20, 2002

General Scope

During the past decade, the field of data mining has emerged as a novel field of research, investigating interesting research issues and developing challenging real-life applications. Researchers and practitioners in both academia and industry have realized that data mining is the key to success when analyzing databases in order to discover patterns and regularities from data.

Ever since the start of the field of data mining, it has been realized that the data mining process should be supported by database technology. In recent years, this idea has been formalized in the concept of *inductive databases* introduced by Imielinski and Mannila in a seminal paper that appeared in CACM 1996 (Vol. 39, Issue 11, pages 58-64). Inductive databases are databases that, in addition to data, also contain generalizations, i.e., patterns, extracted from the data. Within the inductive database framework KDD is modelled as an interactive process in which users can query both data and patterns to gain insight about the data. To this aim a so-called *inductive query language* is used. Very often inductive queries impose constraints on the patterns of interest (e.g. w.r.t. frequency, syntax, generality, accuracy, significance). The constraint-based approach to data mining is also closely related to the issue of *declarative bias* in machine learning, e.g., to syntactic bias, which imposes certain (syntactic) constraints on the patterns learned.

Today, a number of specialized inductive query languages have already been proposed and implemented (e.g., MINE RULE by Meo et al., MSQL by Imielinski et al., DMQL by Han et al.). Most of these inductive query languages extend SQL. This is motivated by the industrial perspective of relational database mining, which has focused on efficient and portable implementations of SQL. Recently the database community has also become interested in the use of XML (eXtensible Markup Language), which has rapidly become an important standard for representing and exchanging information through its applications. Thus, developing XML aware data mining techniques and query languages is also a significant part of the research.

In addition to specific proposals for inductive query languages, there has also been a large body of research work concerned with identifying key concepts to be used in inductive databases. On the one hand, this includes the many types of constraints that have been employed, and on the other hand, the many efficient algorithms and representations for solving constraint-based queries that have been developed in both machine learning and data mining. This includes, e.g., the level-wise algorithm, Apriori, the version space algorithm, boundary set representations, condensed representations, etc.

Thirdly, there exists a large body of research on database techniques that enable and optimize the data mining process. This includes various new database primitives that have been suggested, e.g., to efficiently compute statistics needed by data mining algorithms.

Despite these many contributions, we are still far away from a deep understanding of the issues in inductive databases. In this respect, one can see an analogy with the state-of-the-art in databases in the early seventies. Today, there exist many different approaches and implementations of data mining systems but our theoretical understanding is still imperfect. In a sense, we are looking for the equivalent of Codd's relational database theory for data mining.

Topics

We are looking for all possible contributions related to inductive databases and inductive querying. More specifically, the workshop will focus on the following topics:

- Query languages for data mining
- Constraint-based data mining
- Pattern languages and primitives for data mining
- Efficient search algorithms for constraint-based mining
- Declarative bias formalisms in machine learning
- Version spaces and boundary set representations
- Database support, coupling and primitives for data mining
- Coupling of database and data mining systems
- Integration of database languages such as SQL and XML with data mining

Research works presenting theoretical results, basic research, perspective solutions and practical developments are welcome, provided that they address the topic of the workshop. Position papers are also welcome and encouraged.

Organization and Intended Audience

The full-day KDID' 02 workshop consists of invited talks by experts of the field, paper sessions with technical and position papers, and a panel discussion. In addition, short 5-minute presentations of new ideas and visions will be included in the program.

In conjunction with [ECML/PKDD-2002](#), also a tutorial on Inductive Databases and Constraint-Based Mining will be organized by Jean-François Boulicaut and Luc De Raedt. In the tutorial, the concept and architecture of inductive databases is introduced with state-of-the-art discussions. Also, constraint-based mining as a mechanism for knowledge discovery is covered. Finally, some future research and application issues in inductive databases are discussed.

The workshop as well as the tutorial hope to gather researchers and practitioners interested in the key questions of data mining, and hopes to attract interest from a wide range of possible fields, including: data mining, machine learning, databases, constraint programming, etc.

Submission of Papers

Two types of basic submissions will be considered :

1. **Technical contributions** (an extended abstract with up to 10 pages in the Lecture Notes in Computer Science (LNCS) format).
2. **Position papers** (up to 4-page abstracts in the Lecture Notes in Computer Science (LNCS) format).

Papers should be original and not previously published elsewhere. Basic, applied research and position papers are solicited.

In addition to the basic submissions, there is a separate late call for "New ideas and visions statements":

3. **New ideas and visions statement papers** (up to 2-page abstracts in the Lecture Notes in Computer Science (LNCS) format).

The intention behind the "New ideas and visions statements" call is to complement the work presented in the technical contributions and position papers by fresh, new ideas that represent starting points for future research. Please note that the deadline for the statements is later than for the basic paper submissions 1. and 2.

Submit the papers and vision statements electronically with a submission form of the workshop web site <http://www.cinq-project.org/ecmlpkdd2002/>. Electronic submission in PDF format is preferred, but Postscript will also be accepted. Please note that hard copy and fax submissions will not be accepted. Precise submission instructions can be found on the workshop web site.

The final camera-ready copy versions of the papers follow the same length and style guidelines as the original submissions. Papers will be published in the workshop informal proceedings and electronically in the workshop web page <http://www.cinq-project.org/ecmlpkdd2002/>. Tentatively, only submissions of type 1. and 2. will be considered for inclusion in the informal workshop proceedings.

Successively, a book on Database Technologies for Data Mining will be published with a selection of best workshop papers; the book will be published by an international publisher.

Important Dates

Technical and position papers:

Paper submission:	May 24, 2002
Notification of acceptance:	June 14, 2002
Camera-ready copy due:	July 1, 2002

New ideas and visions statements:

Paper submission:	July 19, 2002
Notification of acceptance:	August 2, 2002

Workshop:	August 19/20, 2002
------------------	---------------------------

Organizing Committee

Mika Klemettinen (chair)	Nokia Research Center, Finland e-mail: mika.klemettinen@nokia.com
Rosa Meo (co-chair)	Università degli Studi di Torino, Italy e-mail: meo@di.unito.it
Fosca Giannotti	CNUCE-CNR, Italy
Luc De Raedt	University of Freiburg, Germany

The workshop is supported by the European Union project clnQ IST-2000-26469 (May 2001-May 2004), funded by the Future and Emerging Technologies arm of the IST Programme. The partners of clnQ (consortium on discovering knowledge with Inductive Queries) are Nokia Research Center (Helsinki, Finland), Politecnico di Milano (Italy), University of Torino (Italy), University of Freiburg (Germany) and INSA Lyon (France).

Program Committee

Chair:	Mika Klemettinen , Nokia Research Center, Finland e-mail: mika.klemettinen@nokia.com
Co-chair:	Rosa Meo , Università degli Studi di Torino, Italy e-mail: meo@di.unito.it
Rakesh Agrawal(?)	IBM Almaden, USA
Roberto Bayardo	IBM Almaden, USA
Marco Botta	Università di Torino, Italy
Jean-François Boulicaut	INSA-Lyon, France
Stefano Ceri	Politecnico di Milano, Italy
Saso Dzeroski	Jozef Stefan Institute, Slovenia
Fosca Giannotti	CNUCE-CNR, Italy
Daniel A. Keim	AT&T Shannon Research Labs and University of Constance, Germany
Stefan Kramer	University of Freiburg, Germany
Pier Luca Lanzi	Politecnico di Milano, Italy
Giuseppe Manco	ISI-CNR, Italy
Heikki Mannila	Nokia Research Center, Finland and Helsinki University of Technology, Finland
Giuseppe Psaila	University of Bergamo, Italy
Luc De Raedt	University of Freiburg, Germany
Christophe Rigotti	INSA-Lyon, France
Hannu TT Toivonen	University of Helsinki, Finland and Nokia Research Center, Finland
Jeff D. Ullman	Stanford University, USA
Mohammed J. Zaki	Rensselaer Polytechnic Institute, USA