

# Knowledge Discovery guided by Domain Knowledge

Amedeo Napoli & Orpailleur Team,  
LORIA

BP 239, F-54506 Vandoeuvre les Nancy,

<http://www.loria.fr/napoli/>

<http://www.loria.fr/LORIA/EXT/equipes/ORPAILLEUR/>

The Orpailleur Team at LORIA Nancy  
Journée nationale en Chemoinformatique,  
Université Louis Pasteur, Strasbourg, 23 mars 2007

# The KDD process

## **Data** (data level: signs and syntax)

↓ preparation (cleaning)

↓ transformation (selection, formatting)

## **Prepared data**

↓ data mining operations (pattern discovery)

↓ numerical and symbolic KDD methods

## **Discovered patterns** (information level: data and meaning)

↓ post-processing

↓ interpretation / evaluation

## **Knowledge units** (knowledge level: information and instruction of use)

# Symbolic and numerical methods for KDD

... used by the Orpailleur...

- ▶ Symbolic methods for KDD include (among others) *frequent itemset search* and *association rule extraction*, and *lattice-based classification* (design of concept lattices).
- ▶ Numerical methods for KDD include methods from data analysis, statistics and probabilities, and among others, *Hidden Markov Models of order 1 and 2*, initially designed for pattern recognition.

# The CORON Platform

CORON is a platform for symbolic KDD including:

- ▶ a collection of algorithms for *itemset search*, such as Apriori, Close, Titanic, Charm, Eclat, etc., together with adapted algorithms, such as Zart, Eclat-Z, and rare itemsets.
- ▶ a collection of algorithms for the *extraction of association rules*: informative rules, generic basis, informative basis, and rare rules.
- ▶ a set of modules for *data preparation* and *result interpretation*, for filtering data and results by object, attribute, support, confidence, etc.

# Classification has a central place in KDDK

- ▶ The process of “turning complex data into knowledge units” is based on the core idea of **classification**.
- ▶ Classification is a polymorphic process involved in all tasks of knowledge processing, namely for:
  - **modeling components**
  - **mining data**
  - **representing knowledge units**
  - **classification-based reasoning**

## KDDK: a work in progress

or the role played by knowledge for guiding the KDD process...

- ▶ extraction of rare itemsets and rare association rules,
- ▶ meta-knowledge discovery or extraction of meta-rules from a knowledge base,
- ▶ knowledge-based discovery of patient pathways in healthcare networks,
- ▶ extraction of knowledge units from DBs of chemical reactions,
- ▶ ...