# What Is Data Mining?

Data mining (knowledge discovery from data)

– Extraction of interesting (<u>non-trivial,</u> <u>implicit</u>, <u>previously unknown</u> and <u>potentially useful)</u> patterns or knowledge from huge amount of data

Alternative names

– Knowledge discovery (mining) in databases (KDD), knowledge extraction, data/pattern analysis, data archeology, data dredging, information harvesting, business intelligence, etc.

Watch out: Is everything "data mining"?

– Simple search and query processing

– (Deductive) expert systems

# What is (not) Data Mining?

- What is not Data Mining?

  — Look up phone number in phone directory

  — Query a Web search engine for information about "Amazon"

- What is Data Mining?

  — Certain names are more prevalent in certain US locations (O'Brien, O'Rurke, O'Reilly... in Boston area)

  — Group together similar documents returned by search engine according to their context (e.g. Amazon rainforest, Amazon.com,)

# Why Data Mining?

The Explosive Growth of Data: from terabytes to petabytes

- Data collection and data availability
  - Automated data collection tools, database systems, Web, computerized society
- Major sources of abundant data
  - Business: Web, e-commerce, transactions, stocks, …
  - Science: Remote sensing, bioinformatics, scientific simulation, …
  - Society and everyone: news, digital cameras, YouTube
  
  ***We are drowning in data, but starving for knowledge!***

"Necessity is the mother of invention"—Data mining—Automated analysis of massive data sets