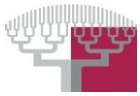
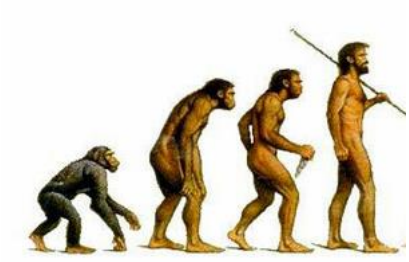
Data Analytics

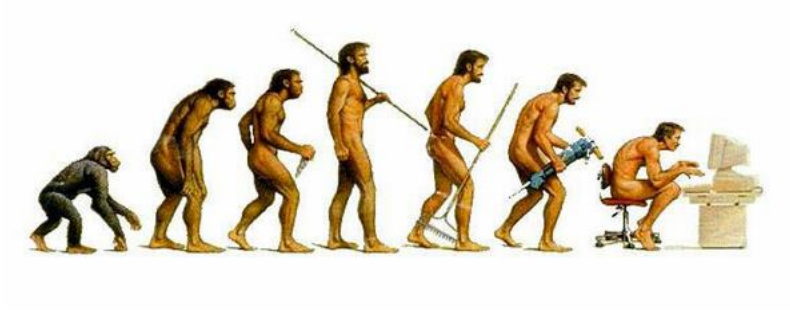


Themis Palpanas
Paris Descartes University

Evolution

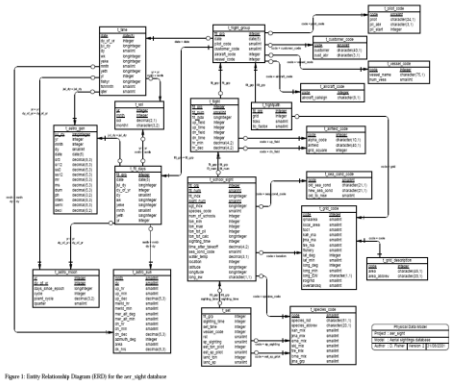


Evolution



Data Evolution

- structured
- precise
- static (almost)

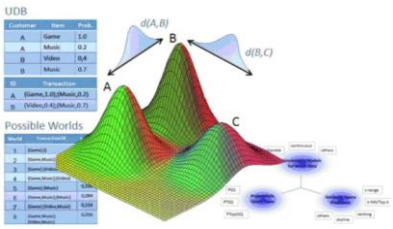


Data Evolution

- non-structured

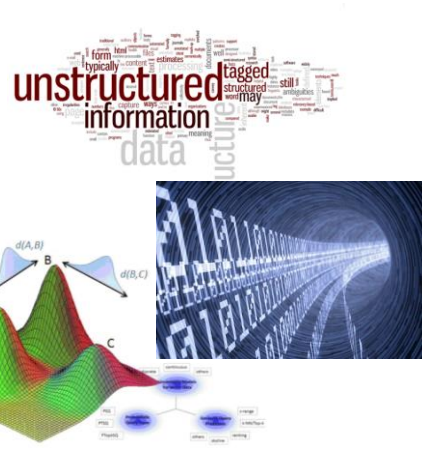
Data Evolution

- non-structured
- uncertain



Data Evolution

- non-structured
- uncertain
- streaming



Things to Come

- large scale data
  - very large collections (i.e., terabytes to exabytes)
  - scientific, business, user generated

Things to Come

- large scale data
- streaming data
  - sensors, feeds, continuous analytics
  - response times in seconds to nanoseconds

Things to Come

- large scale data
- streaming data
- heterogeneous data
  - structured, non-structured, text, multimedia
  - variety of sources, schemas, representations, models
  - computer-generated, human-generated

Things to Come

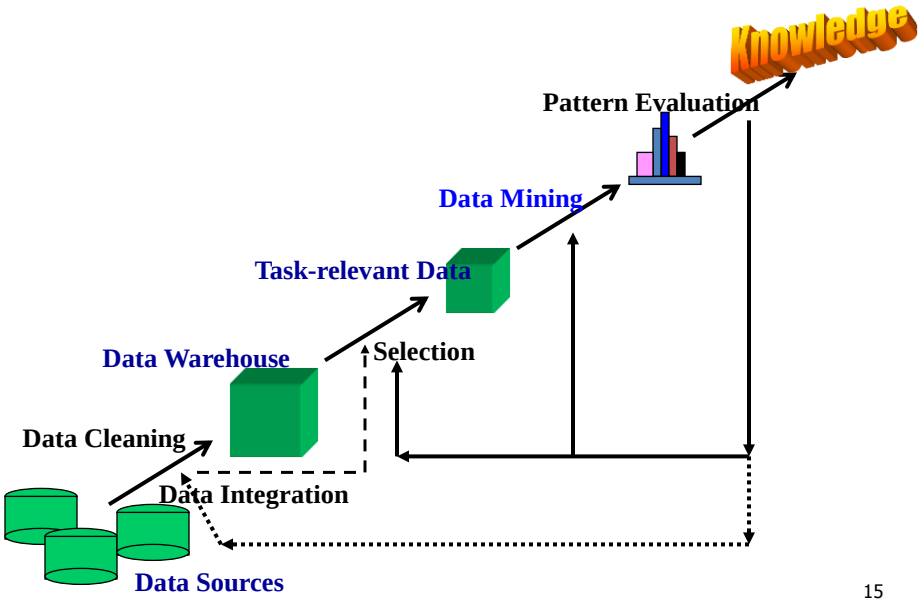
- large scale data
- streaming data
- heterogeneous data
- private data
  - correlating data from multiple sources poses risks
    - credit card history, mobile phone usage, GPS tracking
  - privacy and accountability
    - access granted only to specific person, at specific time, for specific purpose, only for necessary data

Things to Come

- large scale data
- streaming data
- heterogeneous data
- private data
- uncertain data
  - imprecision, inconsistencies, incompleteness, ambiguities, latency, deception, approximations, privacy preserving transformations
  - process/data/model uncertainty

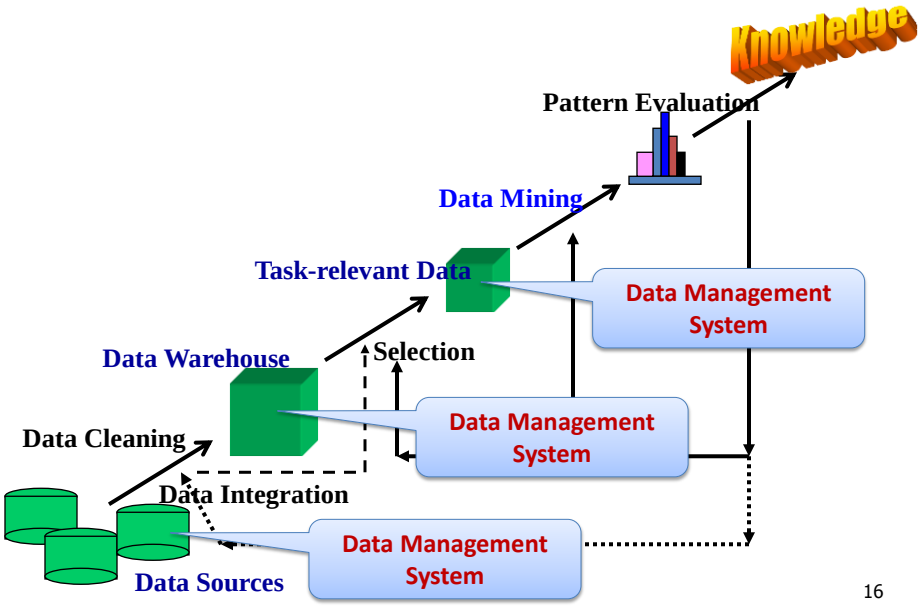
data vs knowledge

Knowledge Discovery (KDD) Process



15

Knowledge Discovery (KDD) Process



16