

Dependencies identification in large data sets

Zhivko Angelov

ОПЕРАТИВНА ПРОГРАМА

„РАЗВИТИЕ НА КОНКУРЕНТОСПОСОБНОСТТА НА БЪЛГАРСКАТА ИКОНОМИКА” 2007-2013 г.

**Анализиране и идентификация на
зависимости в големи масиви от данни
– приложение за икономически и
технологични анализи**

ADISS Lab Ltd.

&

IICT – BAS Hierarchical Systems Department

What I'm going taking about

- What is done
- Why we are doing it
- What we are going to do

Dependency identification

- **Example:** What we mean
- **Philosophy:** In most cases a phenomena can be recognised observing summrized data and can be explained by raw data
- **Implicit dependency:** We will understand judgment, which can be expressed after performing an operation on a set of data

Dependency identification

- **Hypothesis:** Only a human being can identify a unexpected dependency (till now)
- **Proposal:** We need a tool which summarizes data and easy displays the raw data from which an summarized item come from

Summrize data

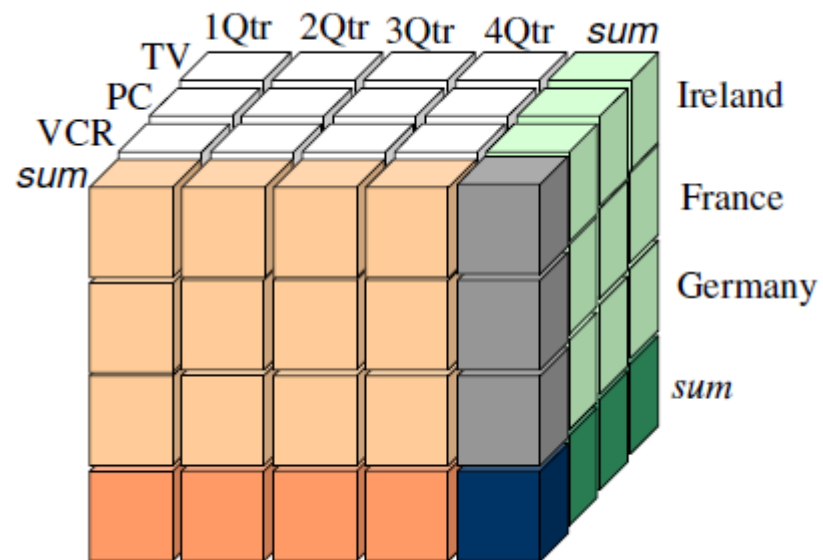
- **Aggregation:** Produce a new fact connected to a subset of records based on existing facts (by operations like sum, max, avarige & etc.)
- **Grouping:** Produce a new dimension – a new splitting of the whole set of records into several subsets (by operations like classification, clustering & etc.)

Aggregation

- OLAP (OnLine Analytical Processing) cube
- Domains' lattice

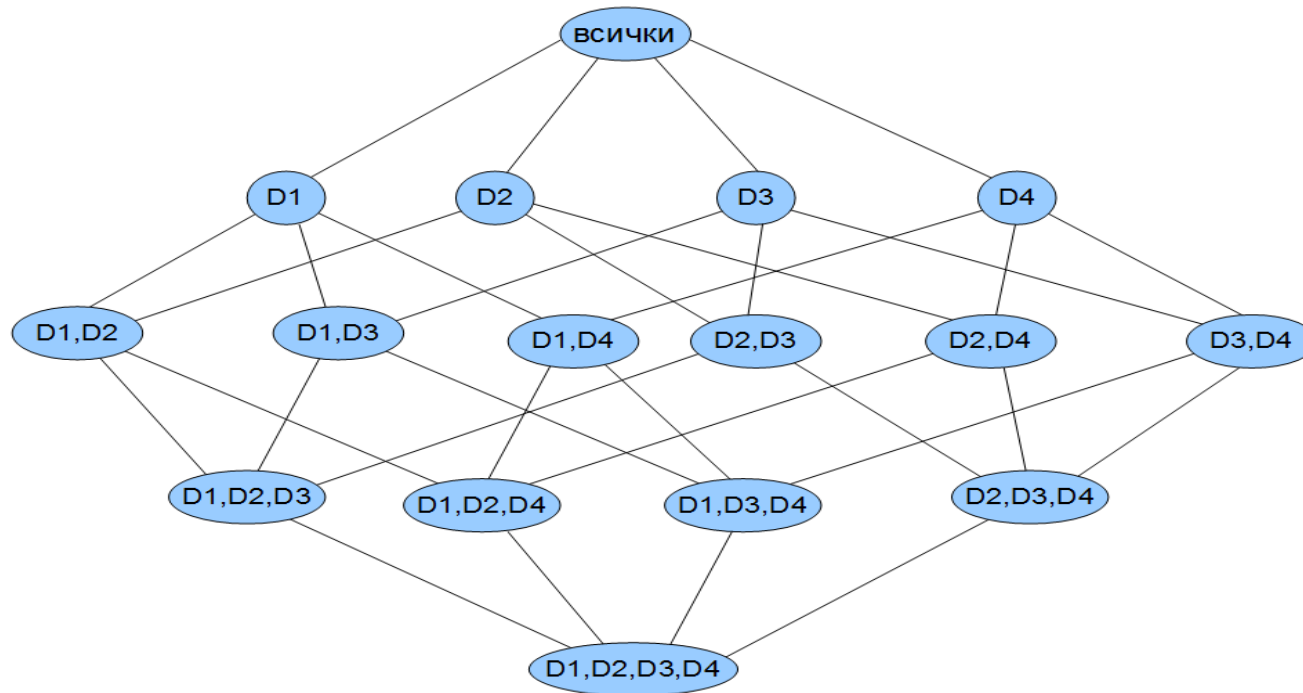
Aggregation

OLAP (OnLine Analytical Processing) cube



Aggregation

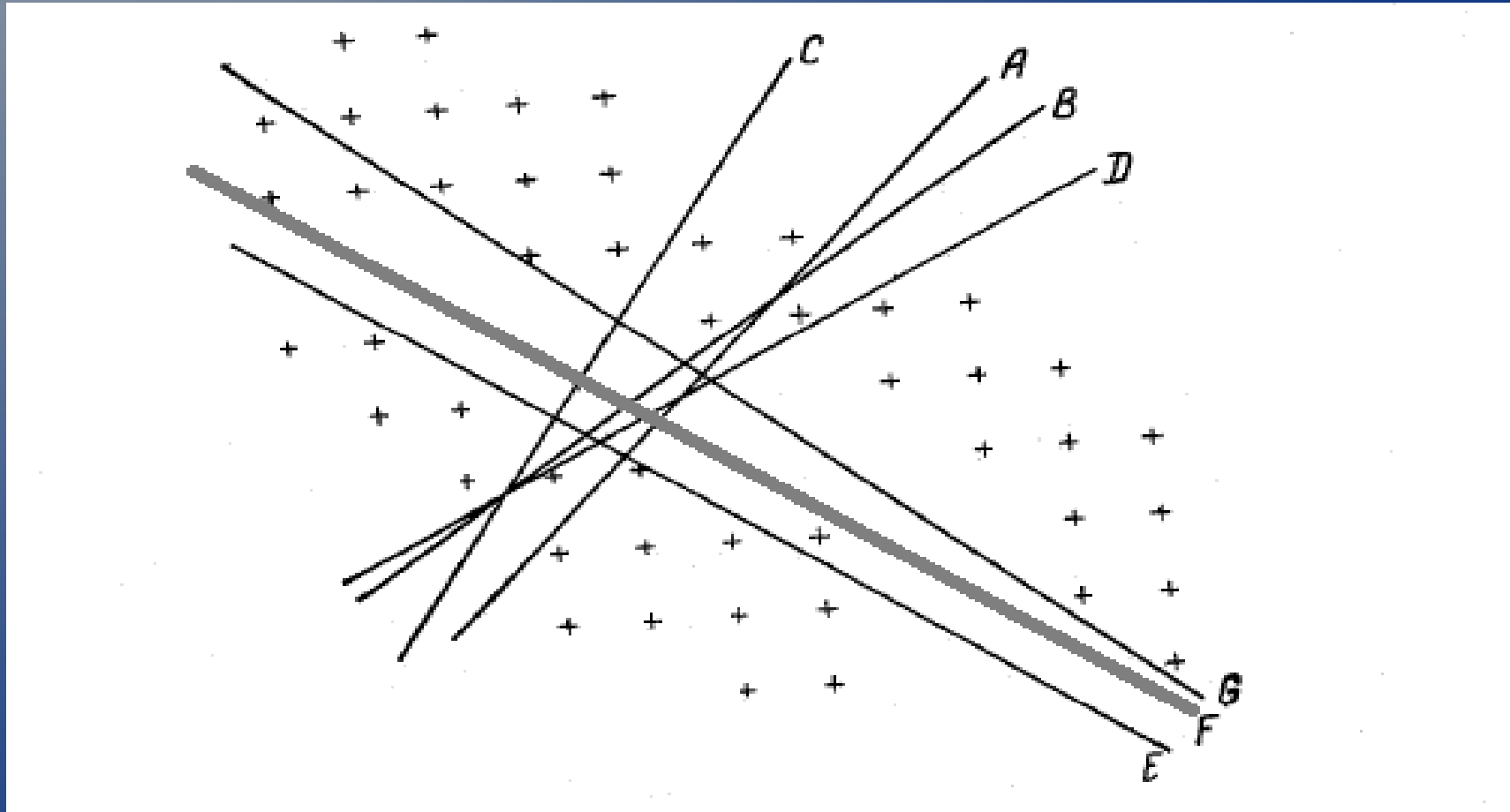
Domains' lattice



Grouping

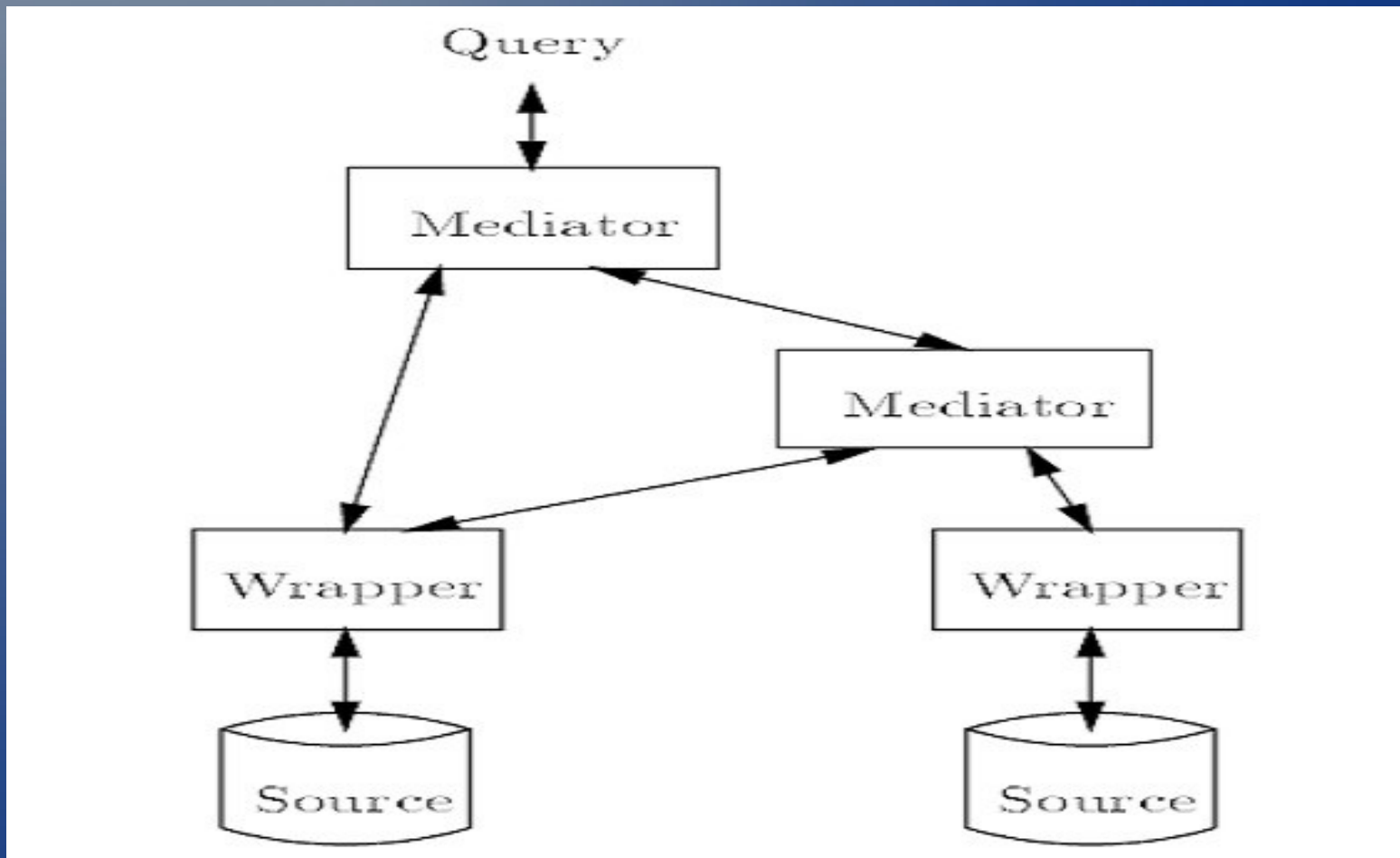
- Classification rules
- Clustering algorithms

Clustering

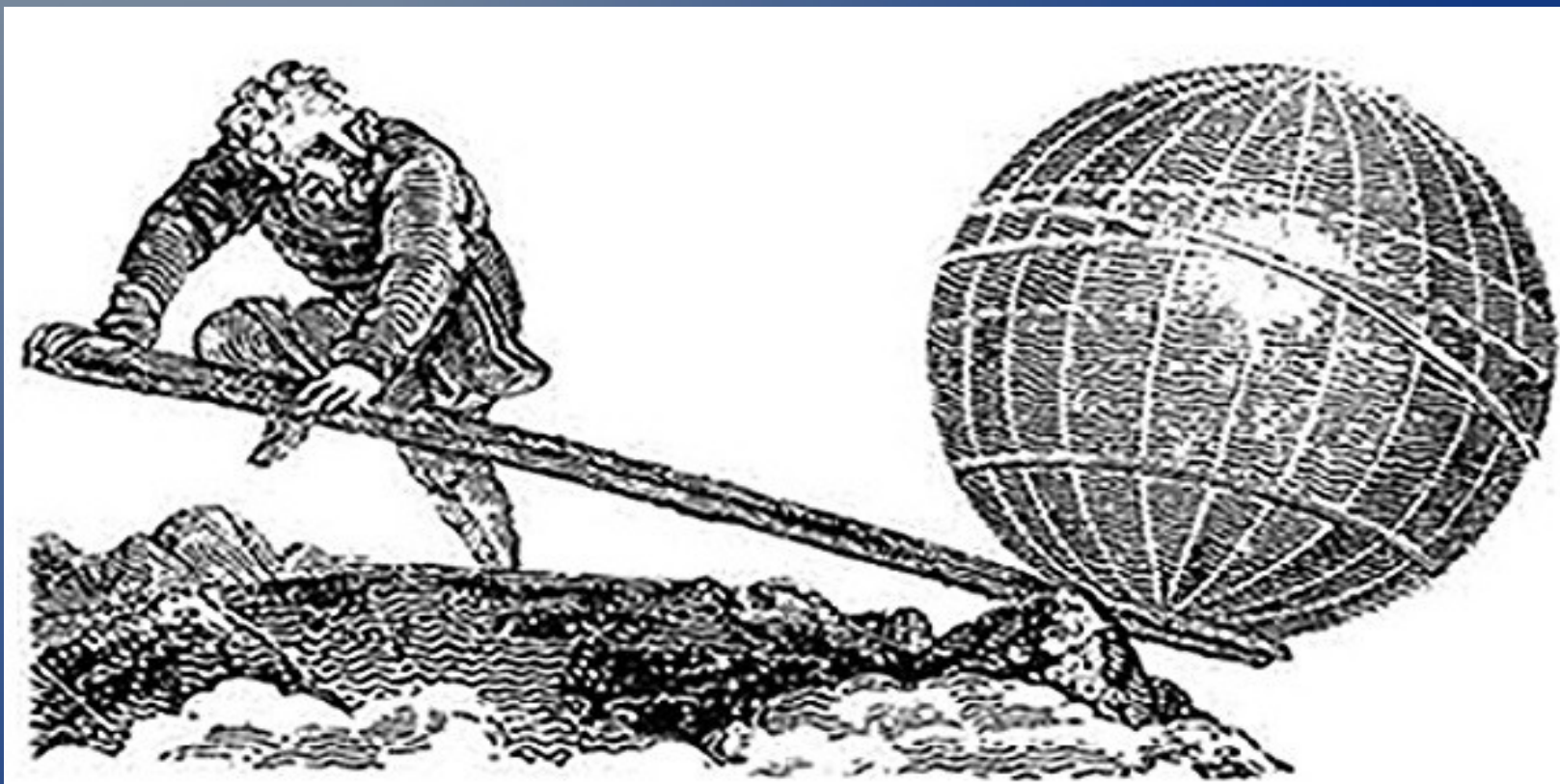


The last but not at least

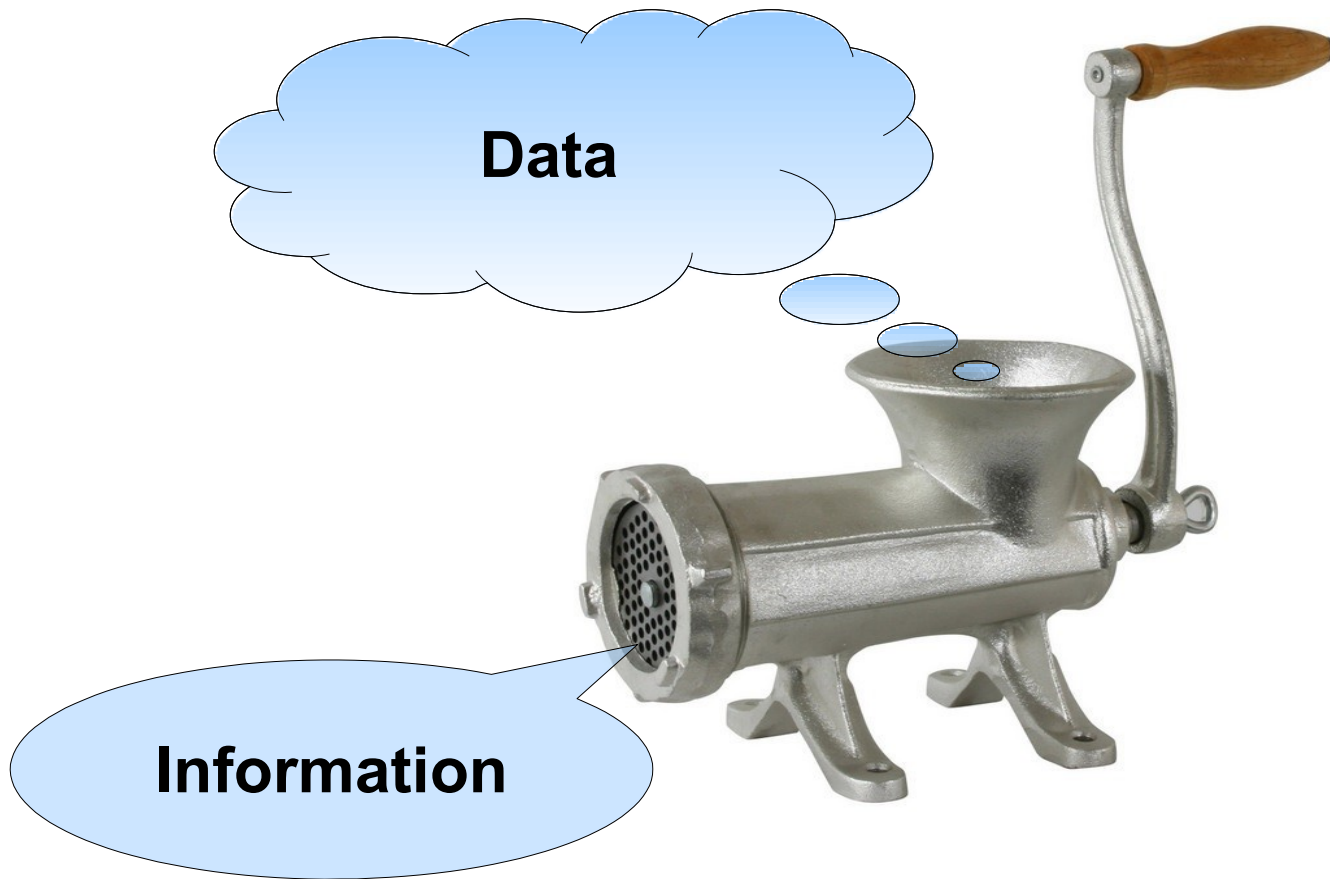
Who to collect the raw data?



Archimedes' times



Nowadays ... maybe this



Thanks a lot for your attention!