PANEL: Knowledge Hour!

VENICE
April 2023

DigitalWorld 2023 & NexComm 2023

Theme

Challenges for Deriving Useful Information from Huge Data

PANEL: Knowledge Hour!

VENICE
April 2023

Dynamics on deriving information from Huge Data

Context, Situation and Data Patterns in Various Domains

Evolution: Data → Knowledge in Different Domains

Volatile Data and Stable Knowledge

Super-Computing under Awareness of Incomplete (useless) Data

CONTRIBUTORS

VENICE
April 2023

Moderator

Prof. Dr. Herwig Mannaert, University of Antwerp, Belgium

Panelists

Prof. Dr. Paulo E. Cruvinel, Ph.D., Embrapa Instrumentation, Brazil
Prof. Dr. Martijn Zoet, Zuyd University of Applied Sciences, Sittard, the Netherlands
Lect. Dr. rer. nat. Uwe Riss, OST - Eastern Switzerland University of Applied Sciences, Switzerland
Prof. Dr. Guadalupe Ortiz Bellot, Universidad de Cádiz, Spain
guadalupe.ortiz@uca.es

Chair Position

VENICE
April 2023

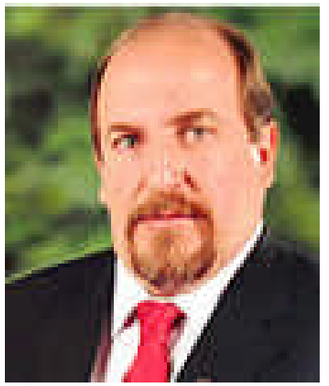
Challenges for deriving useful information from huge data

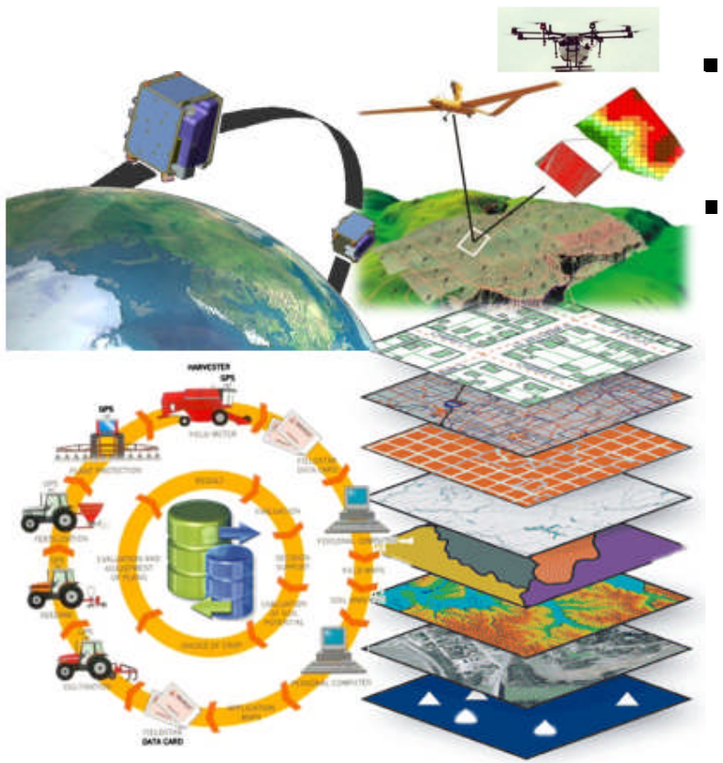
- Information systems create a *mirror of the world* that enables scientific analysis
  - **Digital twins are very valuable to monitor and analyze** various kinds of business processes, and physical infrastructures like power grids
- The tools to mine or recognize patterns in these digital twins or data lakes are becoming increasingly powerful and impressive, but should not make us forget that:
  - **correlations are not always causal**, e.g., gray hair and falling down the stairs
  - **past patterns will not necessarily repeat themselves**, or we would all be rich on the stock market
  - pattern mining and unexplained AI can lead to **phantom logic**, e.g., huskies and wolves
- The data patterns are not always reliable knowledge, and should be used with care, certainly for predictions and generative artificial intelligence:
  - **physical facts** are worth more than statistical predictions, e.g., a parking space prediction vs. sensor
  - **laws of physics** are worth more than statistical predictions, e.g., airplane statistics vs. Bernoulli's Law
  - **facts and semantics** are worth more than statistically generated text, e.g., ChatGPT
- As huge data is used in an ever increasing way to deal with humans beings, the ethical consequences of these fundamental flaws cannot be overestimated

Panelist Position

VENICE
April 2023

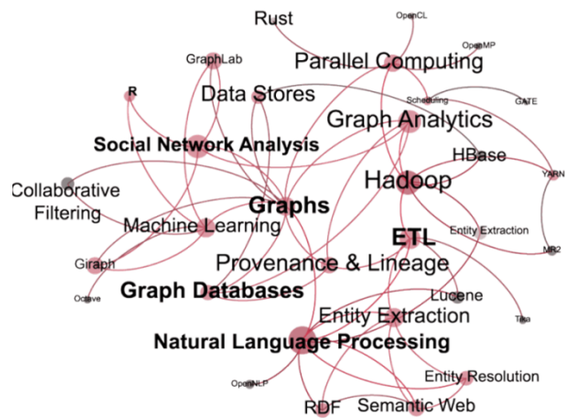
Agricultural decision making, Context – Risk management and sustainable development





- **Phenomenology related to the soil-plant-environmental systems → undressing and modelling;**
- **Sensors and data quality for decision making → diversification of solutions due diversity in agroecosystems and climate conditions;**
- **Huge amount of data and available computation infrastructure → use of BigData and advanced computer architectures and software (Cloud, parallel computing, Machine learning, Deep learning, Artificial intelligence).**

- Challenges
  - New sensors and networking;
  - Interoperability;
  - Field truth and customization;
  - Proprietary and public data;
  - Standards and public policy.

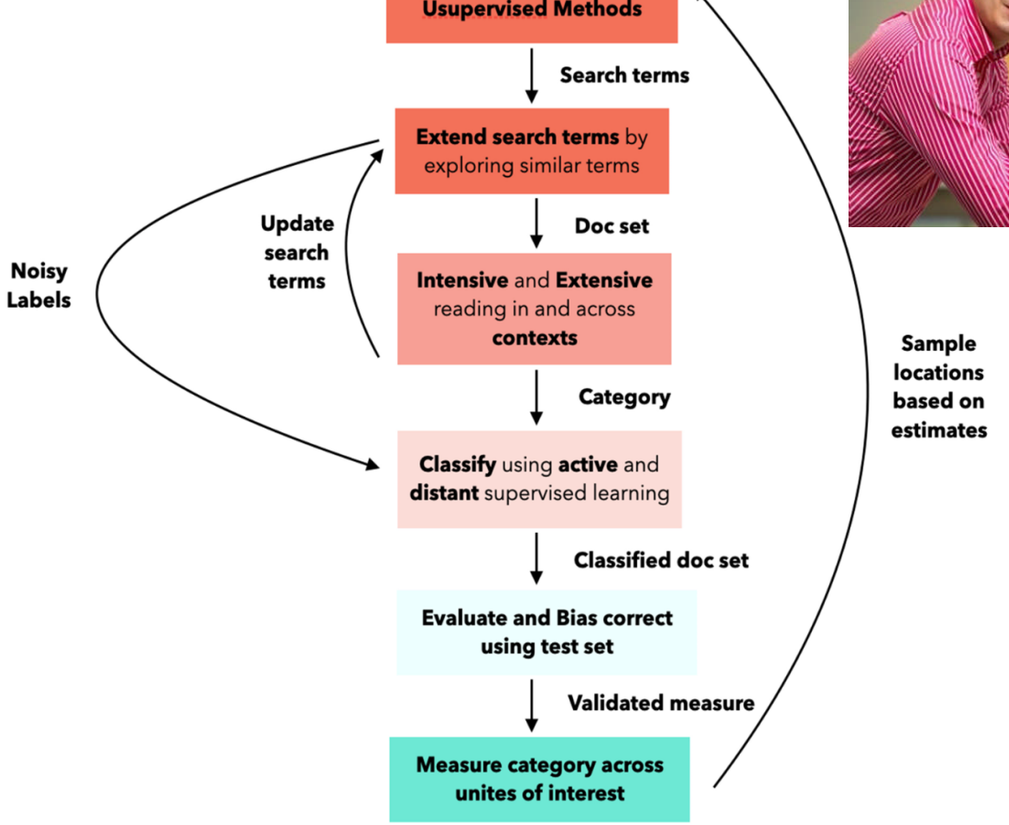


Panelist Position

VENICE
April 2023

Large sets of natural language

- Context and situation awareness;
- Pattern detection and interaction;
- Automatic Corpus Expansion.



Panelist Position

VENICE
April 2023

Uwe Riss

Dealing with the Challenges of New Intelligent Data-based Application

- The problem we face is that with huge amounts of data and powerful analytical tools we get results that **look completely convincing** but do nevertheless **not always represent reliable knowledge**
- Typical application of this kind is ChatGPT
- Can approaches based on these ***analytical tools be used to accomplish knowledge tasks*** in organizations and beyond or ***can't we trust them***?
  - We have to monitor the results with **new IS tools** (e.g., **Data & Process-oriented Digital Twins**) that help us to establish reliability by additional semantic structure
- Can we use such intelligent **tools without hesitation**?
  - On the one hand, such intelligent applications can be used in more areas, as they more flexibly adapt to the users (even those without specific IT knowledge) and are able to learn
  - On the other hand, users become even more transparent than they already are today
    - New infrastructures to protect users' privacy are required

Panelist Position

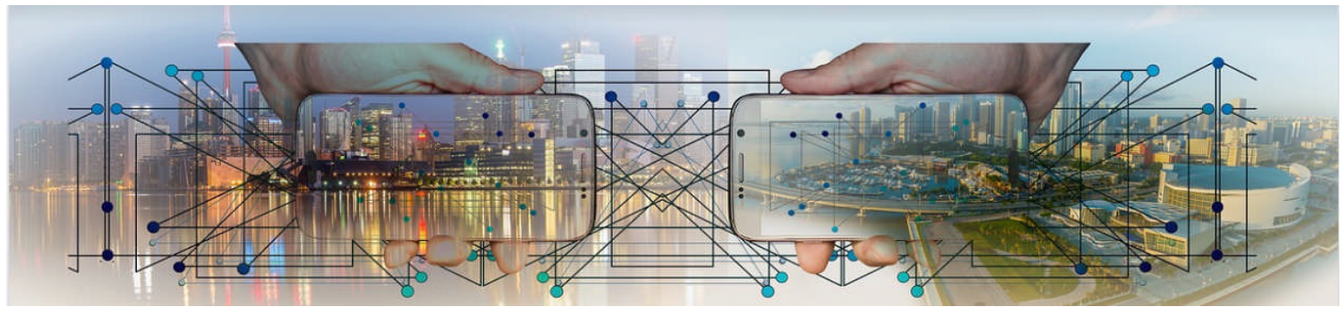
VENICE
April 2023

Collaborative IoT, Context –Awareness and Internet of Behavior

- Correlation between data from different domains→ Knowledge enrichment
- Context and situation awareness →Early warning and personalization
- Pattern detection and interaction →Internet of Behaviour

Guadalupe Ortiz,
University of Cádiz

- Challenges
  - Collaborative-IoT
  - Interoperability
  - Correlation
  - Privacy



Output highlights

- Today's systems must be able to process and correlate multi-domain data in real time, and add artificial intelligence techniques to support the sustainability of the world we live in, while compromising on data privacy and security.
- Mining and recognizing patterns should be cautiously considered
- Quality of data is relevant and often overlooked (collection frequencies – freshness-, noisy data, obsolete data, raw vs processed data, statistics data)
- Humans should exercise caution in evaluating the reliability of knowledge/information and the output of AI-based tools.
- Context/situation awareness leads to automatic corpus expansion (large sets of natural language patterns)