AP: Data, Content and Media

Project Area: Data, Content and Media

cnr DIITET
Consiglio Nazionale delle Ricerche
Engineering, ICT and Technologies for Energy and Trasportation Department

Objective: The PA objective is both to advance methods and technologies and to discover piece of knowledge for a better understanding of individual and collective phenomena and behaviors and for the construction of intelligent and autonomous systems in challenging domains.

The main research and development challenges concern the

- **modelling, analysis, and visualization of data,** which cannot be processed with traditional methods;
- **extraction of knowledge and learning predictive models from multi-dimensional, multi-sources, networked, and dynamic data** based on artificial intelligence, data mining and network science methods;
- **intelligent processing of image, audio, and audio-visual content** for the development of applications based on content recognition;
- **analysis and comparison of digital content for 3D models**, and more generally, multi-dimensional representations;
- **development of applied ontologies of socio-technical systems and semantic technologies** for their treatment based on the languages of the semantic web and for semantic interoperability;
- **natural interaction with computer systems** based on multimodal paradigms that make it accessible and usable.

The AP research and development activities involve

- **15 CNR Institutes**: ISTI, IMATI, ITC, IIT, ISTC, IREA, IEIIT, ICAR, IASI, IAC, STIIMA, IFAC, GI, IRC, ISTEC
- **a total of around 600 person-months per year**.

Scientific Impact & Results will cover

- database and semantic web technologies
- knowledge representation and management
- data visualization, data mining and pattern recognition
- machine learning, and artificial intelligence, complex system theory and network science
- information retrieval and text mining, statistics and applied mathematics
- natural language processing, computer vision and computer graphics
- user modelling and cognitive computing

Approach: The **AP approach and research activities** will be developed according to the following research lines

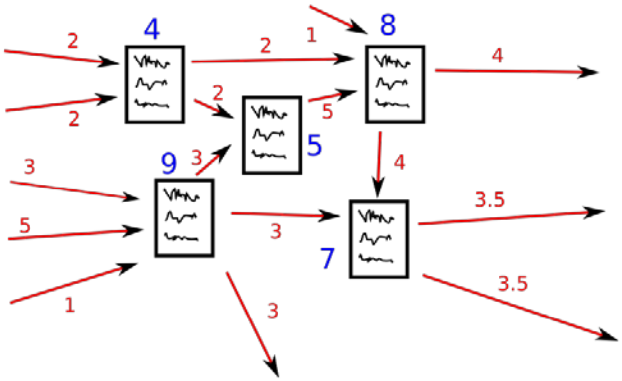
- **DATA**
  - Line1: Big data Sensing and Management
  - Line2: Knowledge Representation, Reasoning and Engineering
  - Line3: Knowledge Extraction and Semantic enrichment
- **CONTENT**
  - Line 4: Data Mining and Machine Learning
  - Line 5: Network Analysis
  - Line 6: Behavior Analysis
- **MEDIA**
  - Line 7: Acquisition, modelling, and analysis of images, videos, 3D and multidimensional data
  - Line 8: Multimodal Interaction and Accessibility

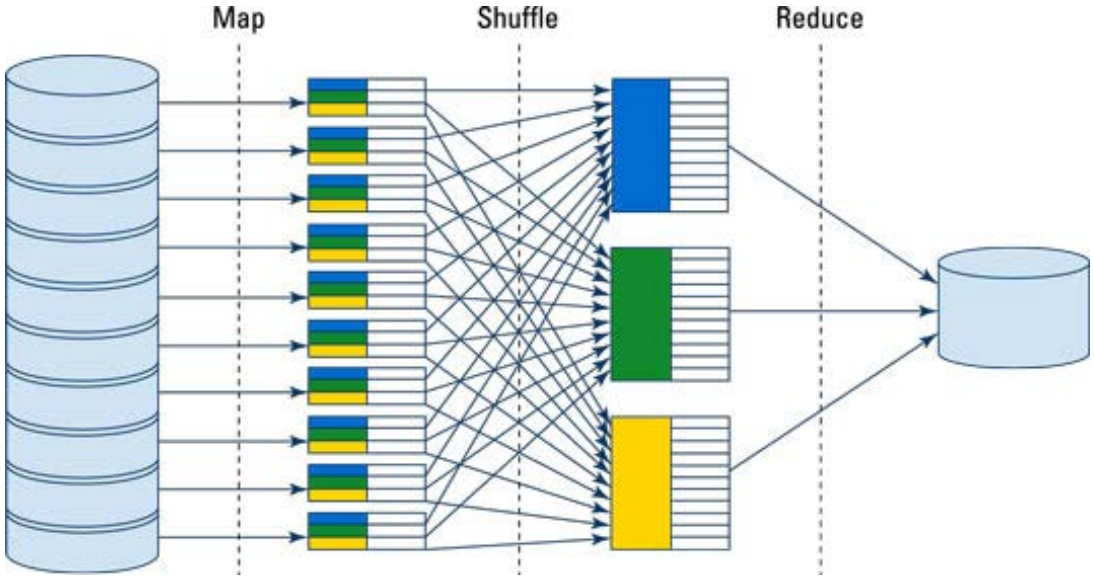
CNR-ISTI

Line 1 – Big Data Sensing and Management

Big data management [HPC Lab]

- Data cleaning at a large scale: deduplication, normalization and feature extraction of text corpora with more than 5 Billion documents
- Management and analysis of very Large graphs: PageRank and connected components computations on 4B-node Web graphs with Map-Reduce/Giraph/Spark





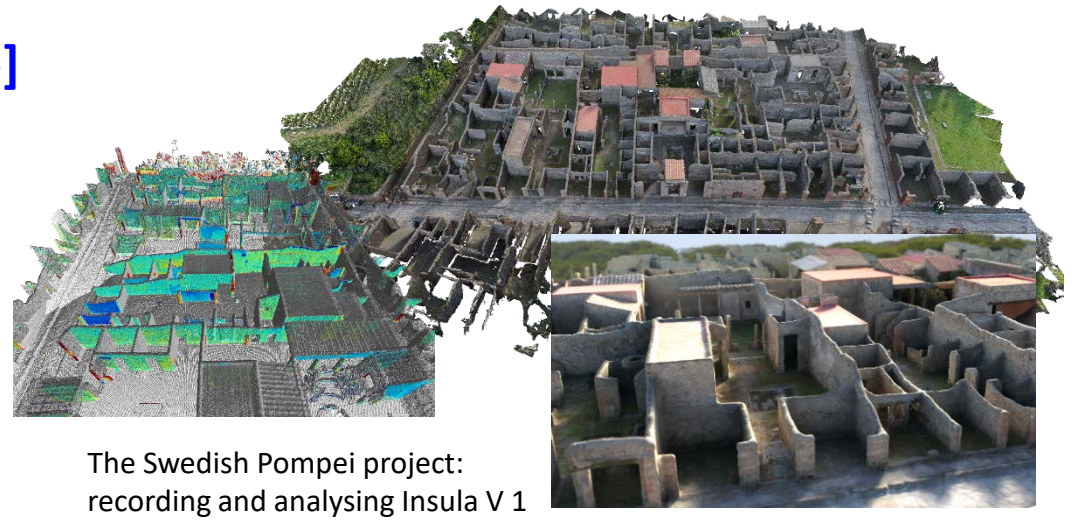
IMPACT and CHALLENGES

The above solutions for Web analysis and Big Data cleaning are deployed and used by the Italian search engine istella.it.

Line 1 – Big Data Sensing and Management

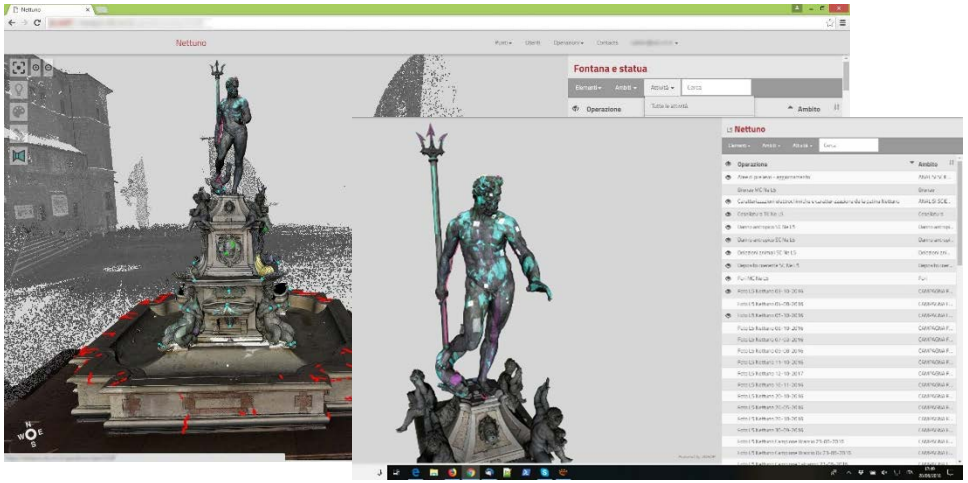
Managing Massive 3D data [VCLab]

- Massive 3D data acquisition, structuring, visualization, compression and distribution
- Impact inside and outside academia, e.g. for cultural heritage documentation, archaeological monitoring, urban planning, industrial surveying



The Swedish Pompei project: recording and analysing Insula V 1
http://vcg.isti.cnr.it/activities/swedishpompeii/

G. Palma et al.: *Enhanced visualization of detected 3D geometric differences*, Computer Graphics Forum 37, 2018



F. Apollonio et al.: *A 3D-centered information system for the documentation of a complex restoration intervention.* Journal of Cultural Heritage 29, 2017

VISUAL COMPUTING LABORATORY

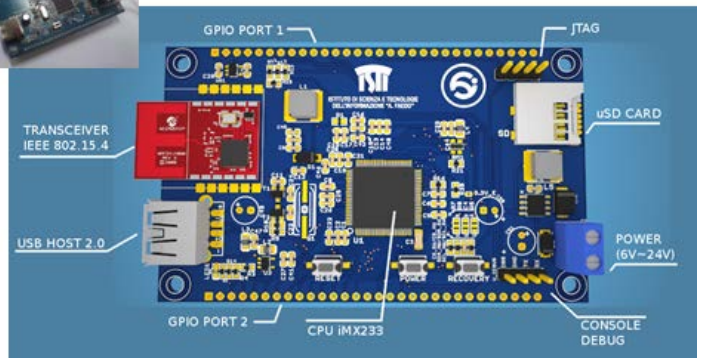
ISTI

Line 1 – Big Data Sensing and Management

Smart camera networks [SILab]

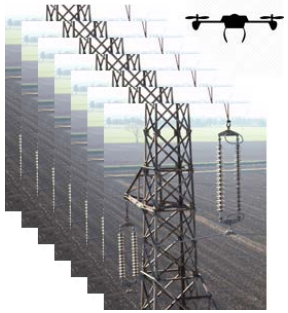
- Computer vision methods for object detection and recognition suitable for deployment on smart camera networks
- Several prototypes of smart camera nodes ranging from fixed surveillance cameras to cameras mounted on drones to provide on board and real-time understanding of the scene
- Large range of applications including the assessment of traffic and mobility patterns in the city and for autonomous inspection of structures such as powerline and historical buildings

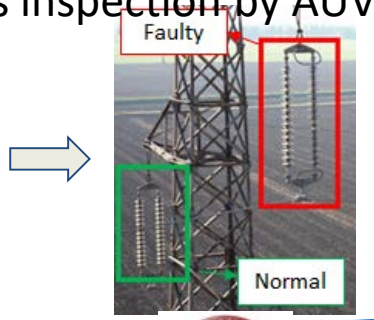
Custom low-energy prototype



Smart cameras for smart parking

Autonomous inspection by AUV





IMPACT

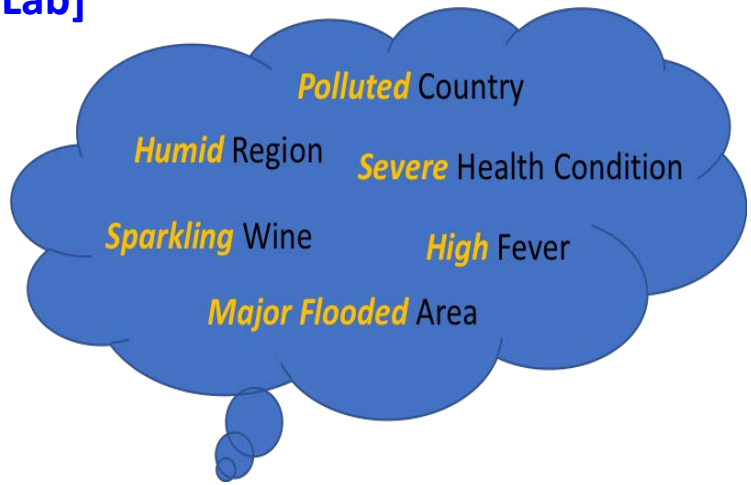
Technology validated for the study of mobility patterns in urban scenarios

CHALLENGES

Devise ad hoc robust algorithms based on deep learning for deployment on resource-constrained HW

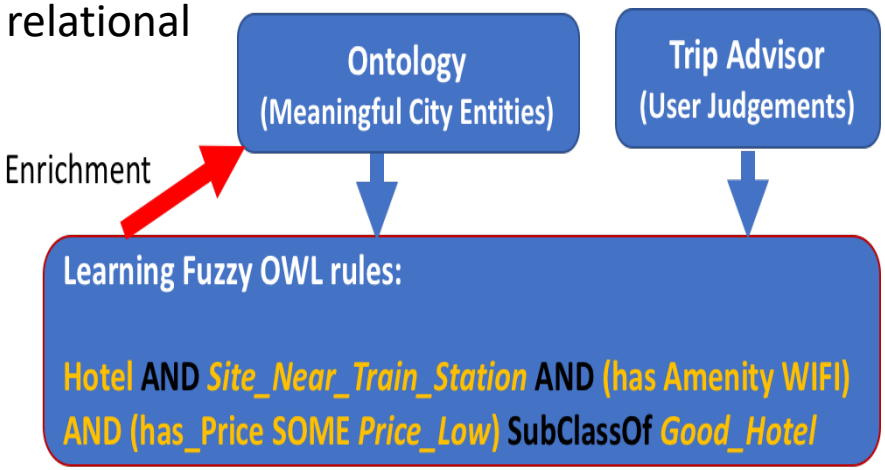
Knowledge representation and Reasoning [NeMIS Lab]

- Fuzzy Logic extensions of both OWL 2 and RDFS family of languages
- Implementation of reasoning tools, among which Fuzzy OWL 2 and fuzzyDL



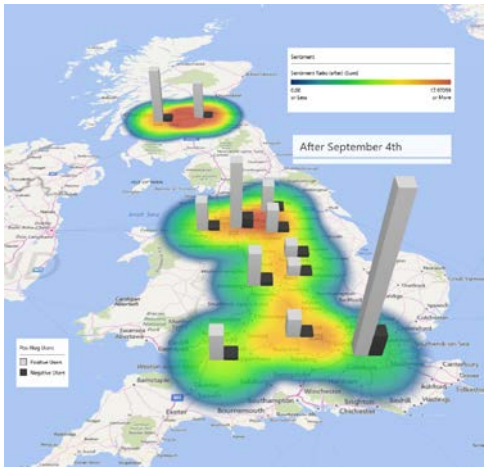
Statistical Relational Learning

- New learning methods stemming from probability theory and statistics to address uncertainty while incorporating tools from logic, databases, and programming languages to learn complex, relational structures from samples

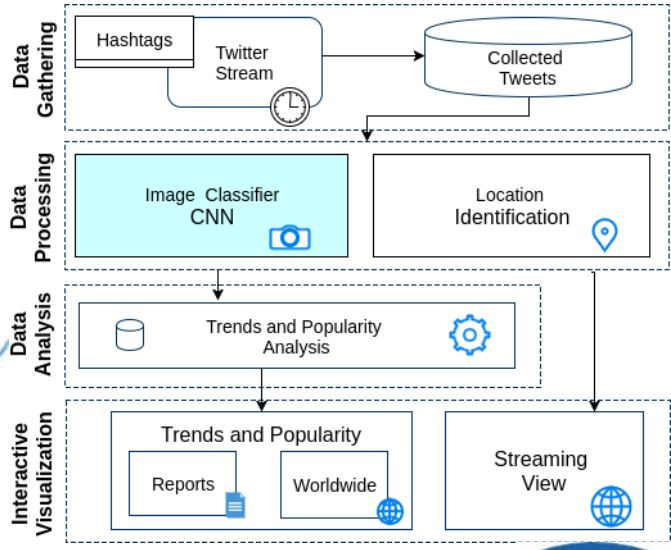


Line 3 – Knowledge extraction and semantic enrichment

Semantic enrichment of texts, social media, mobility data
[HPC Lab, ISTI]



- Open source framework for Entity Linking (dxtr.it) and ML classification of salient entities mentioned in texts, e.g., for text summarization
- Multidimensional analysis of Social Media
  - Large scale analysis of polarized discussions on on Twitter, e.g., the EU perception of the refugees crisis, the Brexit in UK
  - Worldwide food trend analysis with automatic recognition of #foodporn pictures with deep CNNs [in collaboration with NeMis, ISTI]
- Semantic enrichment of mobility data to build holistic trajectories, e.g., for personalized recommendation of budgeted sightseeing tours



IMPACT and CHALLENGES

Best paper award at ACM DocEng 2016

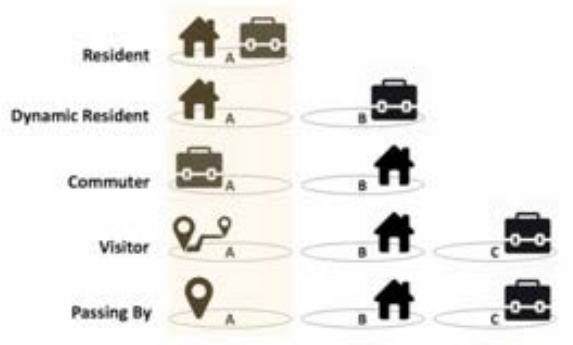
Challenging Marie Curie H2020 project on multi-aspects holistic trajectories for sea monitoring, tourism and public transportation

Brazilian start-up on sightseeing tours *planning*

Line 3 – Knowledge extraction and semantic enrichment

Semantic enrichment of texts, social media, mobility data
[KDD and HPC]

- [KDD] User behaviors classification: using a condensed representation of the his locations (extracted from Call Data Records of his phone).
- [KDD] Labeling the user mobility with the purpose of the movements (e.g. leisure, work, brings&get, etc.) inferred by the spatio-temporal characteristics of his trips.



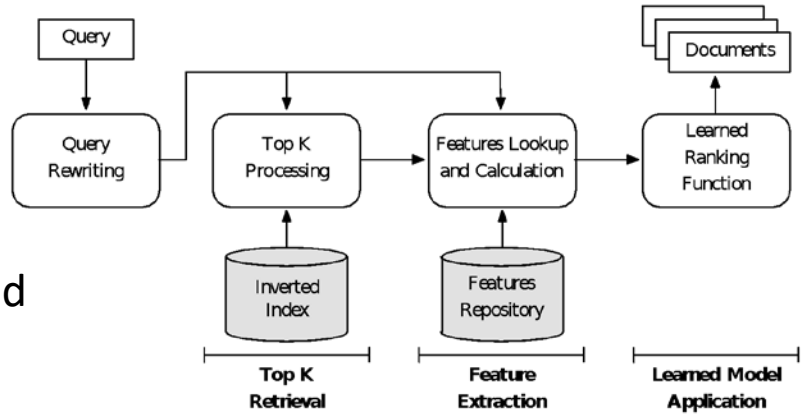
IMPACT and CHALLENGES

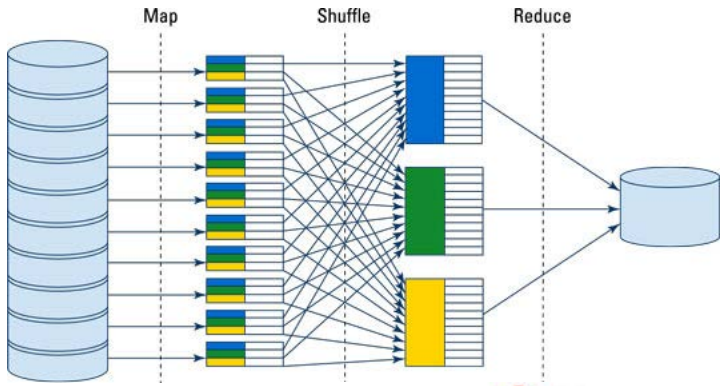
- Enriching the big data with the semantic information of user mobility gives additional dimensions to navigate and interpret the model extracted and analyze the results

Line 4 – Data Mining and Machine Learning

Efficient and Scalable Data Mining and Machine Learning algorithms [HPC Lab, ISTI]

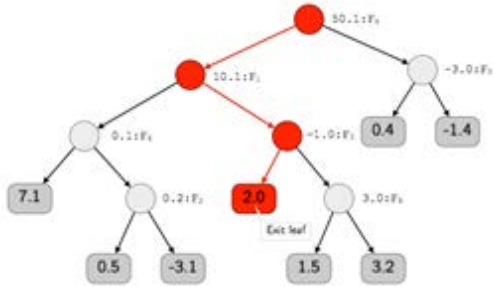
- Learning-to-rank algorithms for web search, OSNs and large-scale classification tasks
  - Focus on building compact and robust models to enhance efficiency and reduce overfitting
- Accurate prediction of query execution times to feed adaptive scheduling strategies for energy saving
- Deep Learning solutions for anomaly detection in the automotive and paper production industries
- Fast traversal of ML models based on ensembles of decision trees for complex classification. regression, and ranking tasks





IMPACT and CHALLENGES

- *The above ML solutions for ranking are currently deployed and used by the Italian search engine istella.it.*
- *Best paper award at ACM SIGIR 2015*
- *The Deep learning solutions for anomaly detection are currently used by FCA at the Centro Ricerche FCA, Turin for testing new car models*



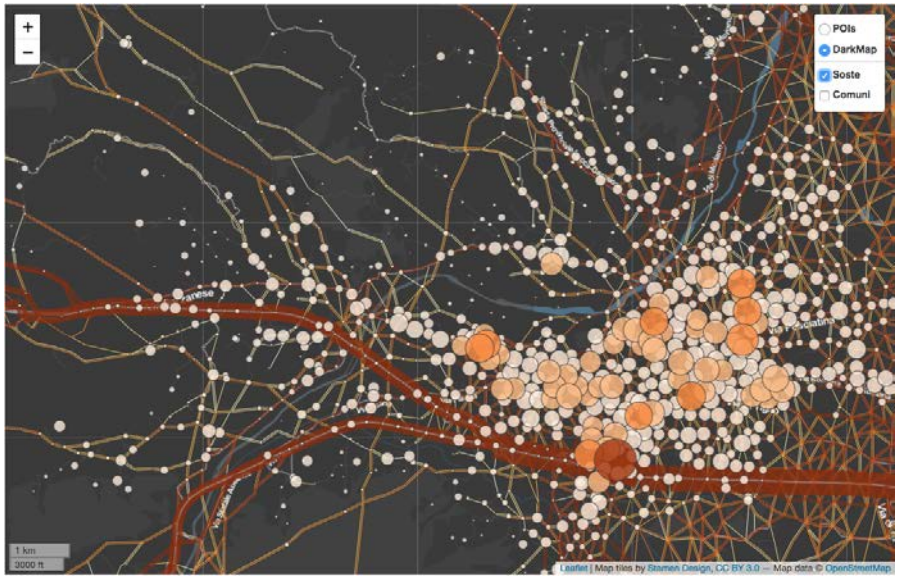
Line 4 – Data Mining and Machine Learning

Mobility Data Mining [KDD]

a stack of spatial, and spatio-temporal algorithms from basic one as trajectory pattern mining, flocks and clustering to richer one as future location prediction, mobility profile discovery, activity recognition, transfer learning among different territories, functional area detection

IMPACT

- Reasoning for decision support for mobility planning, understanding and prediction



CHALLENGES

- Transferable learning of models and patterns in different geographical contexts
- Bridge the knowledge relationship between individual behaviors and global phenomena

Line 4 – Data Mining and Machine Learning

Text and Sentiment Mining [NeMIS Lab]

- Facing Text Analytics problems:
  - sentiment/ topic classification, ordinal regression, quantification, information extraction, named entity recognition...
- With a multidisciplinary approach:
  - natural language processing
  - information retrieval
  - machine learning

IMPACT

- Enabling data mining process on text by giving structure to the unstructured information expressed in natural language.

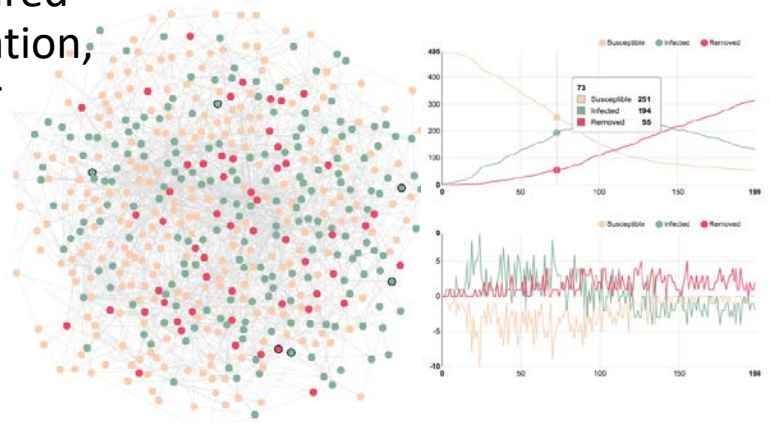
CHALLENGES

- Multilinguality: cross-language learning, cross-language retrieval
- Data scarcity: unsupervised learning, transfer learning.

Line 5 – Network Analysis

Social Network Analysis [KDD]

- Large library of methods to model and analyse structured and unstructured data in terms of network representation, providing support for novel concepts and methods for community discovery in network, models of network diffusion and the definition of novel measures and indicators based on the topology of the network and predictive network models



https://github.com/KDDComplexNetworkAnalysis/

IMPACT

- Modelling and prediction of complex diffusive phenomena, like opinions, diseases, innovation
- Reasoning layer to evaluate what-if scenarios, with programmable strategies for agent based simulations

CHALLENGES

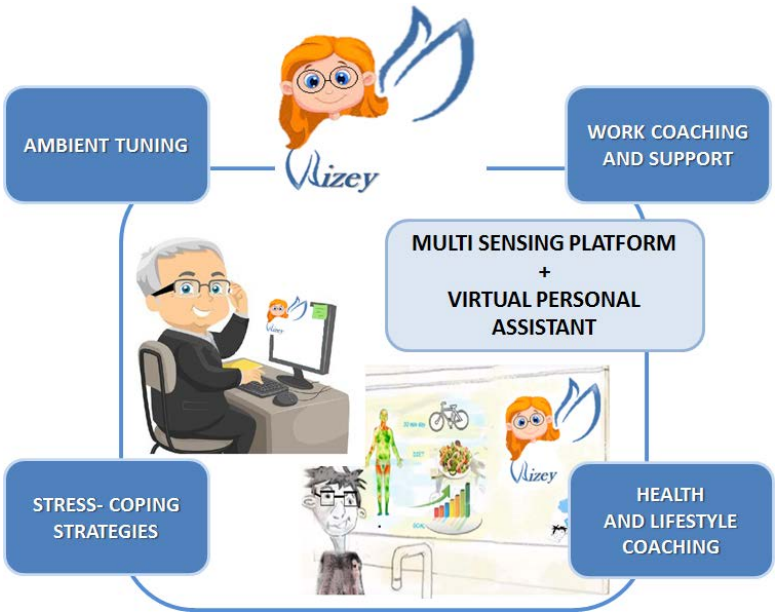
- Management of evolving social and network structures, creation and removal of links, evolution of community structures

Line 6 – Behaviour Analysis

Cognitive Computing [SILab]

Delivery of personalised guidance systems (e.g., Virtual Personal Assistants) underpinning visual-based emotion recognition, reinforcement learning, and natural language processing
and relying on

- user profiling to model users' behavioural and psychological data
- adaptive coaching content and message delivery strategies
- user engagement through a rich and unobtrusive user experience



IMPACT

- Innovative solutions offering unobtrusive support to individuals in normal-life settings (e.g., coaching solutions to foster healthy lifestyles)

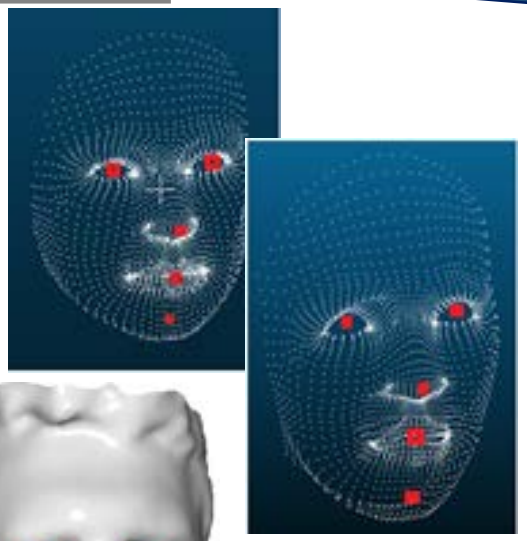
CHALLENGES

- Integration of multidisciplinary approaches (multimedia data processing, multimodal & natural interfaces, behavioural science, computational intelligence)
- High degrees of personalization, trustworthiness, proactiveness and pleasant interaction

Line 7 – Acquisition, modelling, and analysis of images, videos, 3D and multi-dimensional data

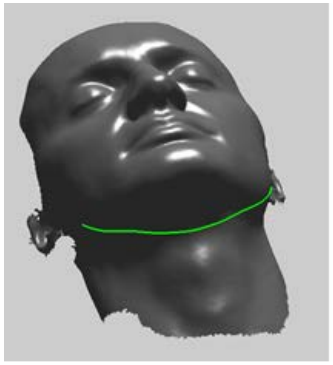
Visual and sensory data analysis and understanding [SILab]

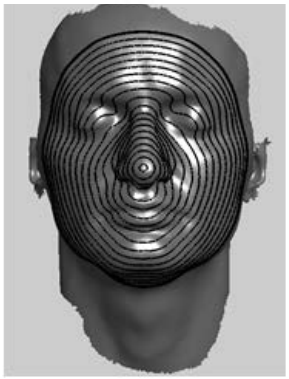
- Analysis and understanding of visual data (2D images,3D and 4D models) based on concepts from computational topology and geometry
- Application to
  - human face characterization to identify facial correlates of health statuses
  - time-series of multimodal signals from environmental sensor monitoring

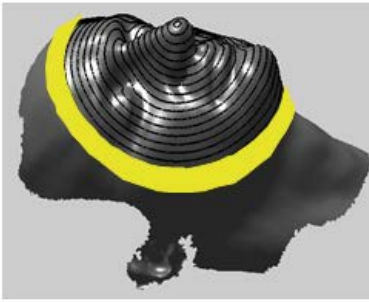


IMPACT

- Solutions to
  - foster citizens' self-empowerment towards self-monitoring of their health and wellbeing
  - improve citizens' comfort at work by a continuous monitoring of their activities





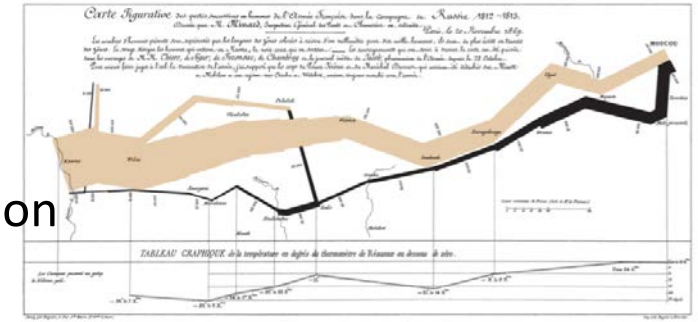


CHALLENGES

- Robustness to noise
- Adaptation to case-specific peculiarities
- Large-population studies

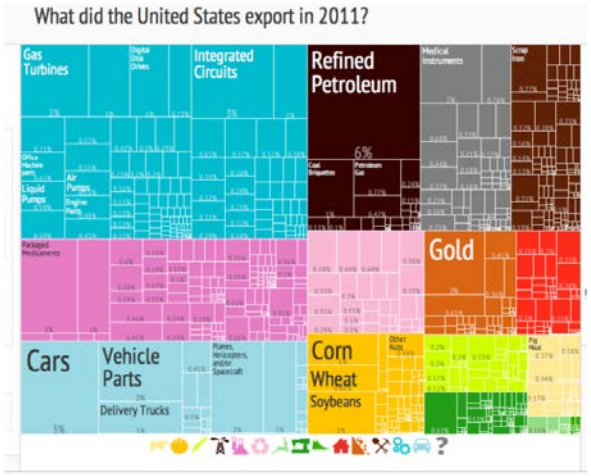
Visual Analytics [KDD]

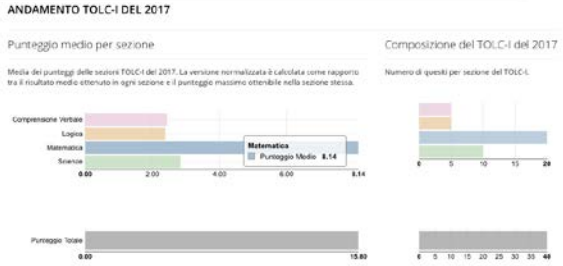
- Visual Interfaces to present and explain complex patterns and model extracted from the data
- Driving the learning process through visual interaction



IMPACT and CHALLENGES (optional)

- Advanced strategies for multi-dimensional data exploration: linked displays, small multiples, browsing





CHALLENGES

- Manage large volumes of data through dimensional reduction and selective exploration

Artificial Intelligence and Multimedia Information Retrieval

[NeMIS Lab]

Deep Learning and Similarity Search for

- Multimedia & Cross-Media Retrieval
- Web-scale and social multimedia mining
- Image/Video Understanding
- Face Recognition and Verification

APPLICATIONS

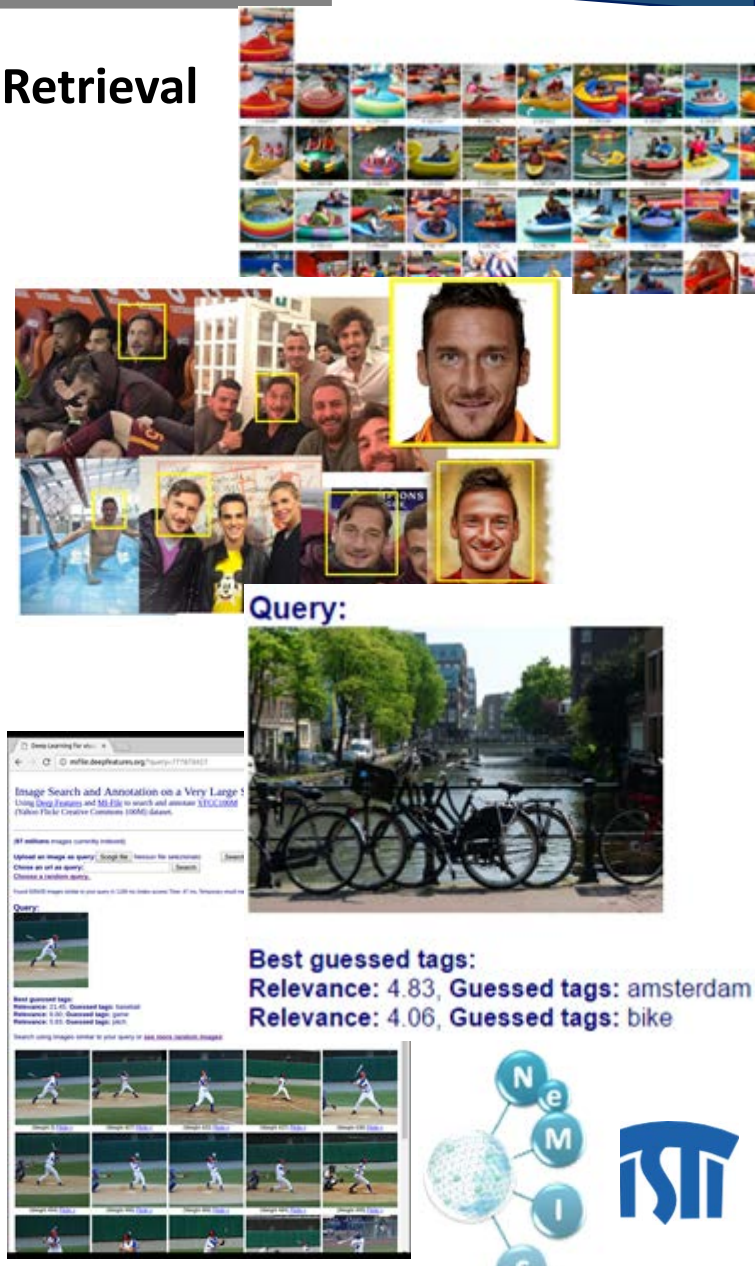
Security, Creative industry, Industry 4.0,
Cultural Heritage, Journalism, Edutainment, etc.

IMPACT

Generating value and extracting knowledge
from big multimedia data

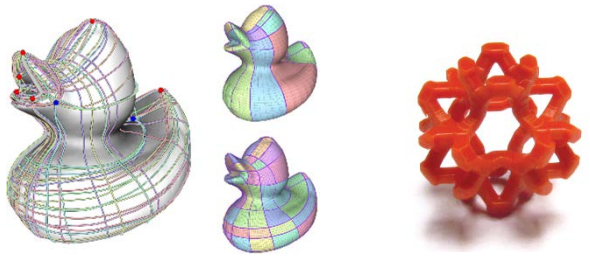
CHALLENGES

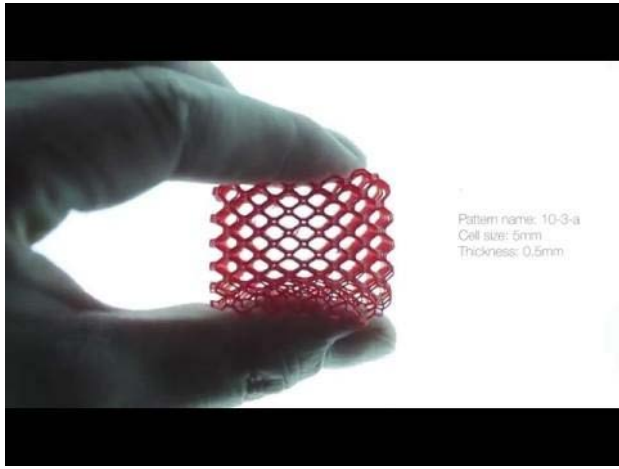
AI on a very large scale
Cross-media and Multimodal Learning and Retrieval
Relational Reasoning
Secure AI



Computer Graphics

- 3D geometry processing and fabrication
  - meshing, parametrization, segmentation
  - computational design of objects based on customizable materials
  - 3D fabrication technologies for small-scale series production
- **Impact** on industry and manufacturing, towards a market for unique, personalized designs





J. Panetta et al.: *Elastic Textures for Additive Fabrication*. ACM Transactions on Graphics 34, 2015

L. Maolomo et al.: *FlexMolds: Automatic design of flexible shells for molding*. ACM Transactions on Graphics 35, 2016

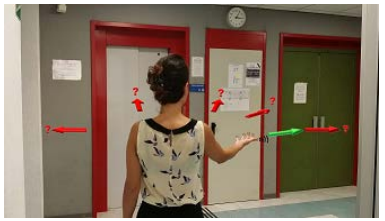
N. Pietroni et al.: *Position-based tensegrity design*. ACM Transactions on Graphics 36, 2017

VISUAL COMPUTING LABORATORY

ISTI

Line 8 – Multimodal Interaction and Accessibility

Multimodal Interaction [HIIS Lab]

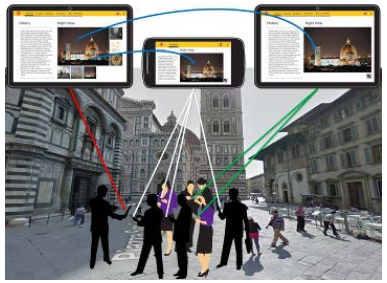


Context-dependent User Interfaces Methods and techniques for enabling user interfaces to adapt to the context of use. We consider various contextual aspects related to users, technologies, environments and social relations. The technologies employed range from sensor-based solutions for detecting the actual contextual parameters to high level tools for defining UIs adaptive behaviors on a rule-based manner.

Tools for Accessibility Evaluation We work on tool support for usability and/or accessibility evaluation in order to provide evaluators and developers with information that can be helpful to improve their interactive applications. We support various techniques ranging from intelligent analysis of logs of user interactions detected on the users' devices to code inspection in order to check consistency with relevant accessibility guidelines.

http://mauve.isti.cnr.it

MultiModal User Interfaces We design and develop methods, languages, tools, and applications that exploit multimodal user interfaces, which are able to interact with a system by exploiting multiple human senses, in order to improve the user experience. For this purpose we consider various modalities (exploiting different physiological parameters) such as graphics, voice, gesture, vibro-tactile feedback, gaze, brain activity, .., which can be combined in different ways depending on the desired effect.



Cross-device User Interfaces We have designed and implemented frameworks for cross-device applications with user interface distributed among public displays and personal devices to allow users to interact through cross-device mid-air and touch gestures and cross-device interactions in which input in one device produces some output in another one.

IMPACT and CHALLENGES

- Better exploit opportunities offered by multi-user interactions, by researching and developing technologies augmenting human interaction in groups in both professional and private contexts.
- Developing future interactive systems offering higher quality experiences, for instance through systems which are mobile, support additional senses, have higher accuracy or incorporate bio or environmental sensors.

Human Interfaces In Information Systems Laboratory
The Human-Computer Interaction Group

ISTI

Line x – according to the White Paper

ISTI CONTRIBUTION

- Bullet points
- Bullet points

- Insert Pictures/Videos representing the current activity and the main results

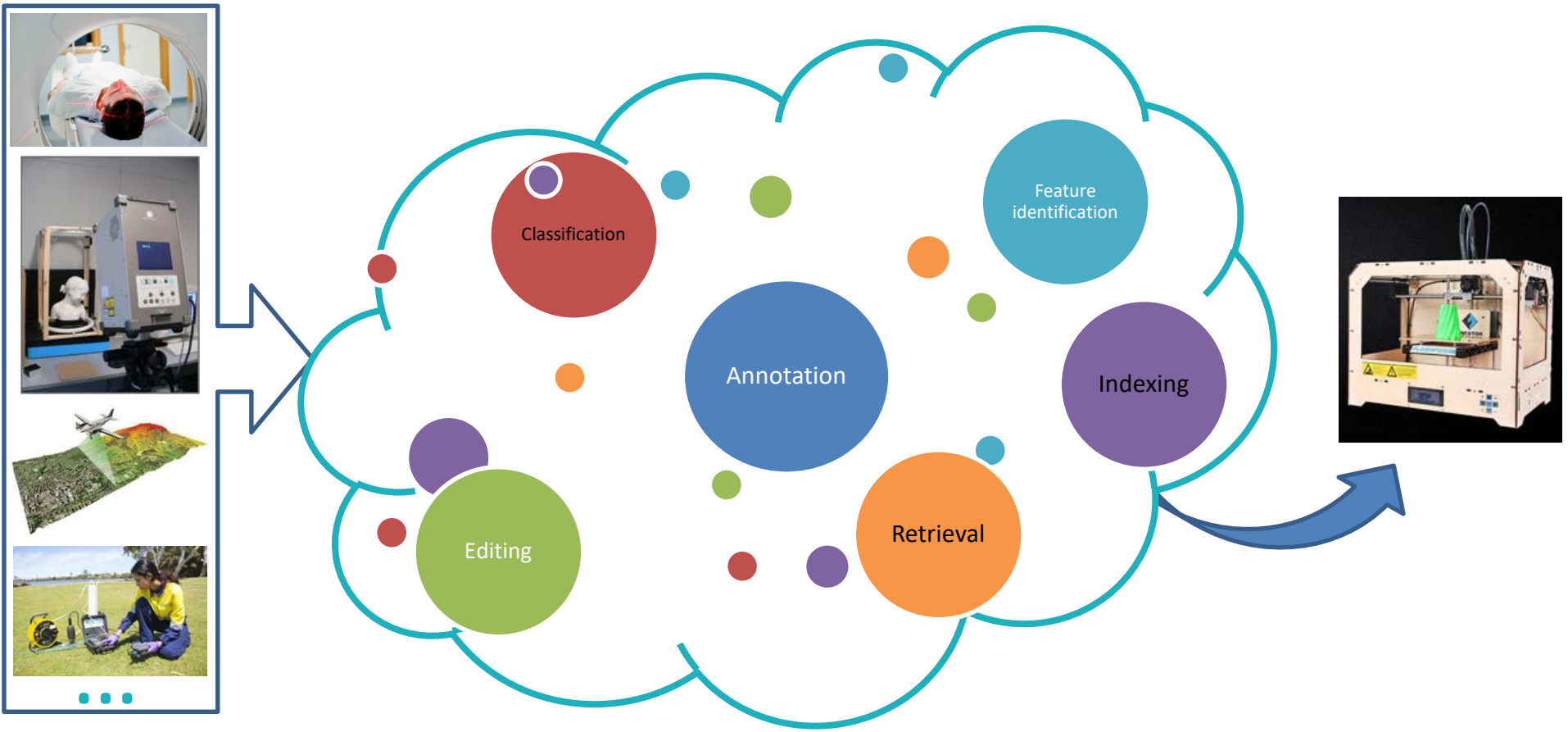
IMPACT and CHALLENGES

- To specify the impact of consolidated research and/or challenges and future work

CONSIGLIO NAZIONALE DELLE RICERCHE

CNR-IMATI

Shape and Semantics research

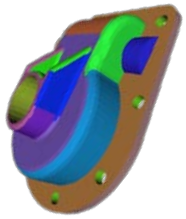
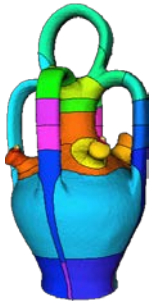


Shape and Semantics research

- Methods and algorithms for the representation, approximation, manipulation of the geometry
  - mesh processing, geometric modelling, implicit modelling

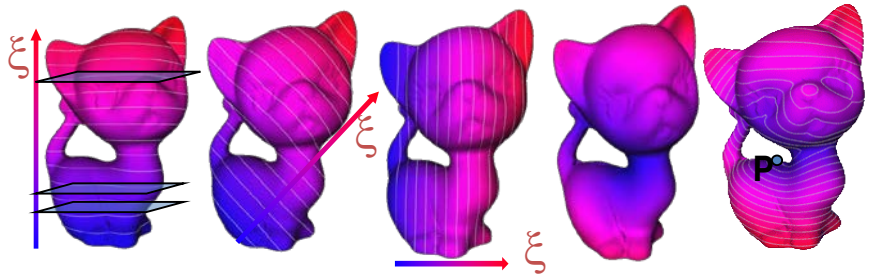
- Methods and algorithms for the analysis, structuring and understanding of shape data
  - geometric reasoning, statistical methods, spectral analysis, skeletonisation, finite element analysis

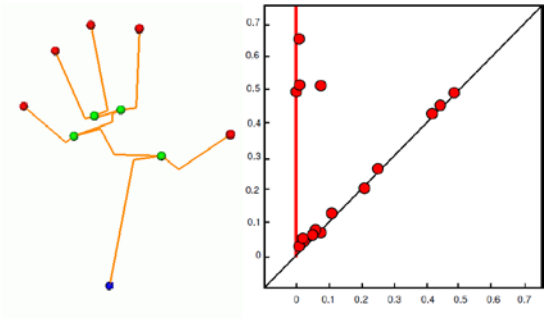


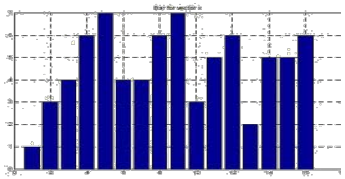
Shape and Semantics research

- Computational Geometry and Topology
  - Morse theory, homology and persistence analysis
  - Reeb graphs, persistent topology



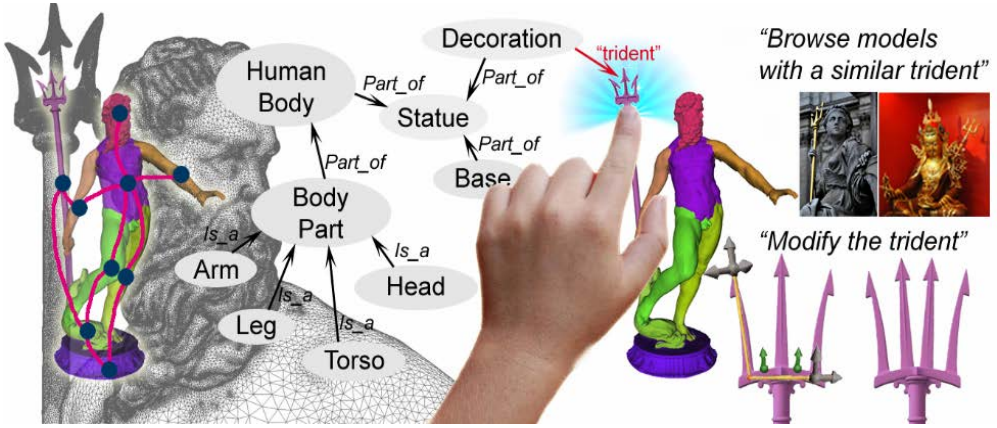


- Shape similarity



Shape and Semantics research

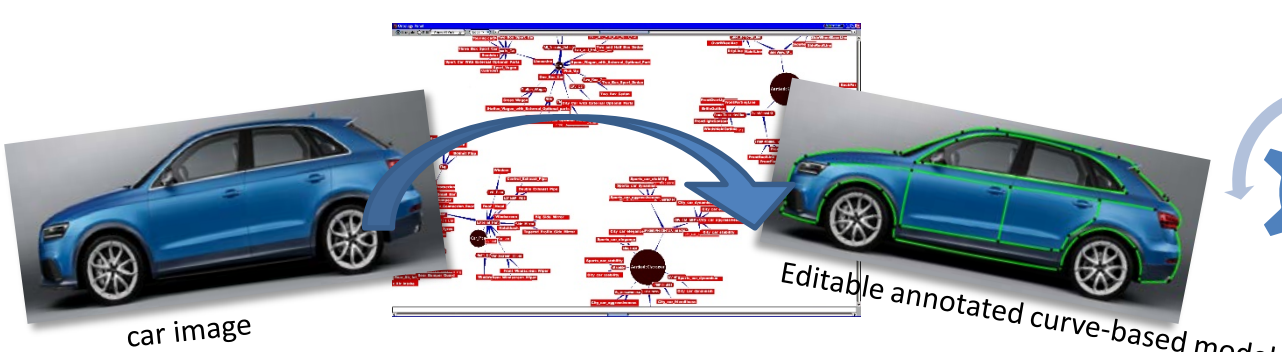
- Coding, exploring and analysing knowledge with distinctive focus on
  - issues arising in applications dealing with 3D content
  - data discovery by semantic similarity and granularity
- Semantic annotation and interoperability
- Linked Data Publishing and Consumption
- Information Visualization

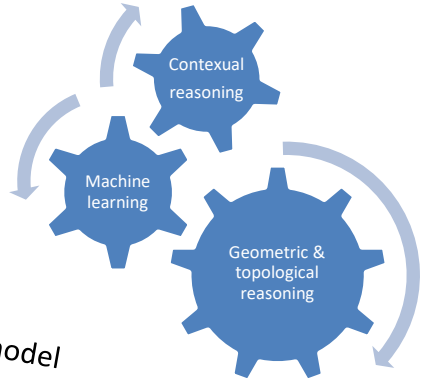


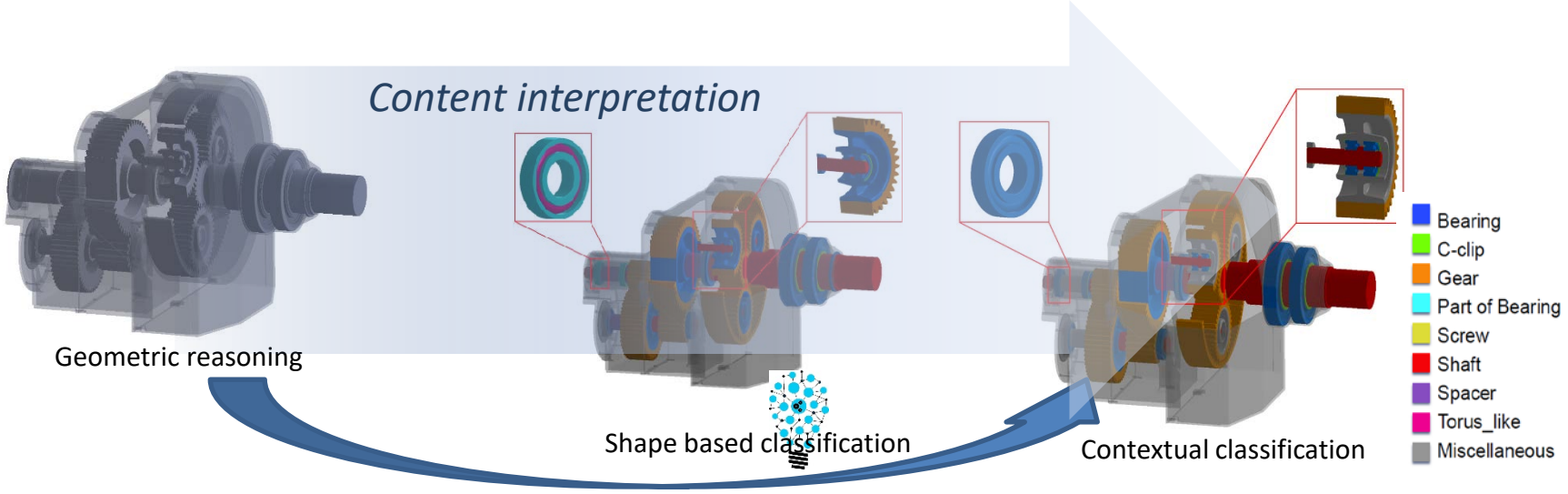
Visual Content understanding

From geometric description to semantic interpretation through a combination of

- geometric reasoning processes for implicit information extraction
- machine learning techniques
- context-specific knowledge formalization and exploitation for content interpretation

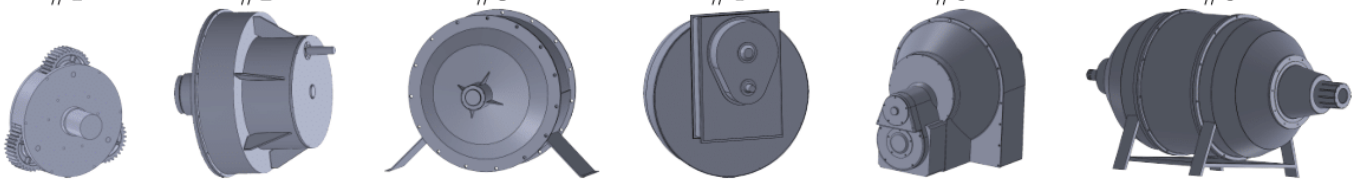
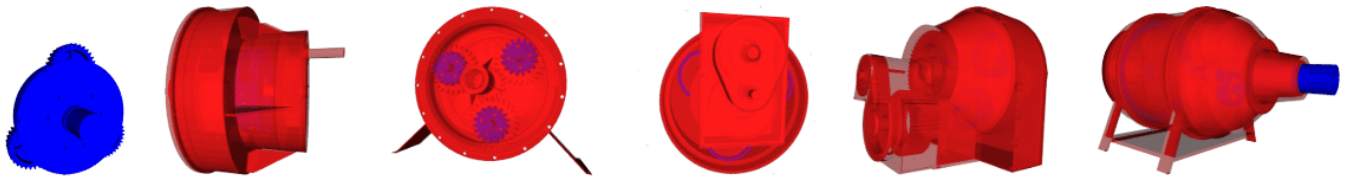






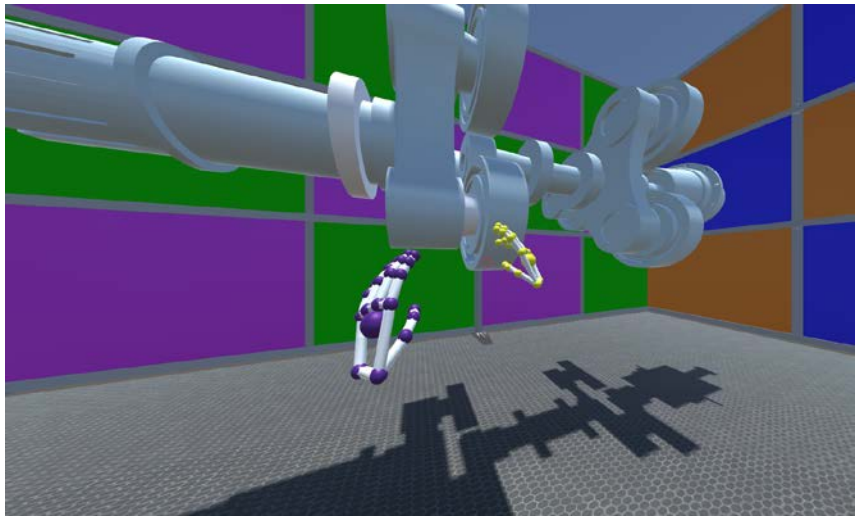
Multi-criteria 3D shape retrieval

- Criteria for CAD assemblies retrieval: components' shape, position, joints, types

	#1	#2	#3	#4	#5	#6
CAD models						
Matched parts						
Shape similarity	1.00	1.00	0.78	0.66	0.76	0.74
Joint similarity	1.00	1.00	0.00	0.00	0.00	0.00
Position similarity	1.00	1.00	1.00	1.00	0.80	0.61
Local similarity	1.00	1.00	0.59	0.55	0.52	0.45
Partial similarity	1.00	1.00	0.44	0.42	0.39	0.34
Global similarity	1.00	0.18	0.06	0.04	0.06	0.10

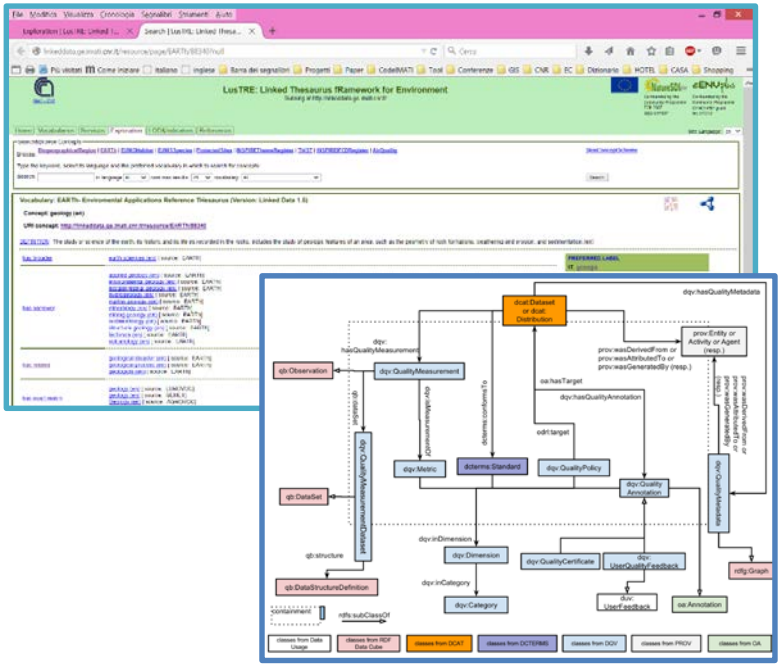
MultiModal Interfaces in VR/AR

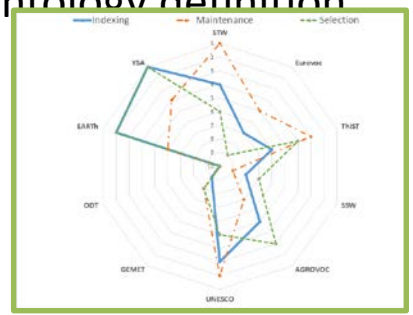
- Methods for visual 3D shape analysis and navigation in 3D repositories
- Study and understanding of
  - natural interaction in virtual environments
  - human behaviour
  - complex 3D scenes

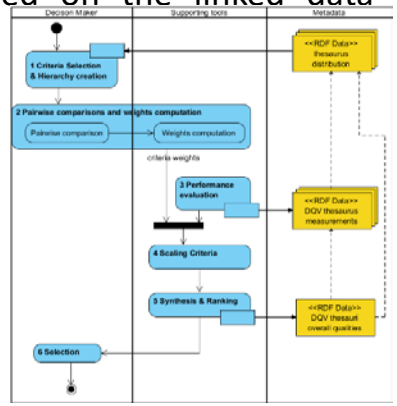


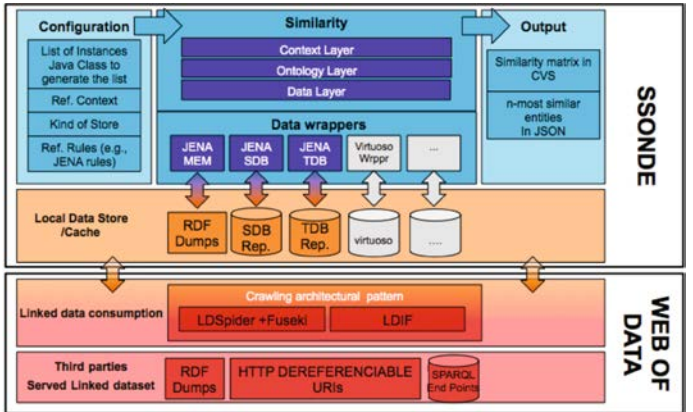
Knowledge extraction and management

- Definition and extension of metadata models to ensure the full characterisation of the pipeline of data production and data quality.
- Definition of quality metrics for Open (Linked) Data and data enrichment:
  - ✓ Overall quality measure based on Multi Criteria Decision Making methodology.
  - ✓ Linkset quality for data enrichment
- Participation in standardization bodies for interoperability and information discovery, e.g. W3C working groups on Data on the Web Best practice (DWBP) and Data Exchange (DXWG).
  - Data Quality Vocabulary (https://www.w3.org/TR/vocab-dqv/)
- Publication and consumption of linked data:
  - multilingual Linked Thesaurus fRamework for the Environment (LusTRE) to support data discovery http://linkeddata.ge.imati.cnr.it/.
  - SSONDE, a context-dependent and asymmetric similarity tool for analysing entities exposed on the linked data http://purl.oclc.org/NET/SSONDE
- Domain ontology definition



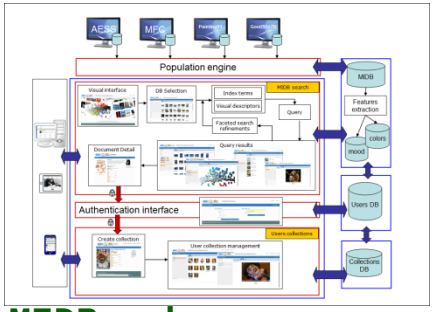




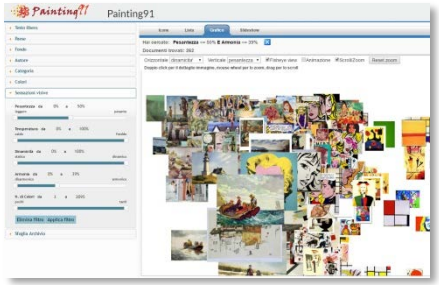


MIDB || Multi Image DataBase Search

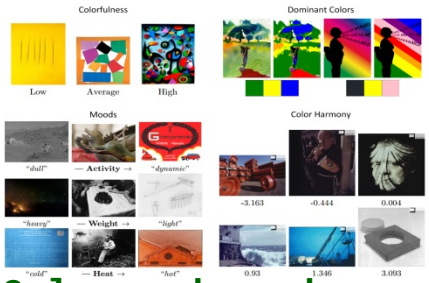
http://arm.mi.imati.cnr.it/midb



MIDB schema



Results on the graph



Color and mood descriptors

MIDB portal has been designed and implemented to navigate and search image databases using keywords, colors and moods and is characterized by the following features:

- Different image collections related to social or cultural heritage (other archives on different subjects can be added)
- Multimodal search, both on keywords and on pictorial, emotional and aesthetic characteristics automatically extracted from text and images;
- Multiple data display modes: list and thumbnail, or mapped on a Cartesian plane, in which the user dynamically chooses the attributes from dynamicity, colorfulness, mood features, …

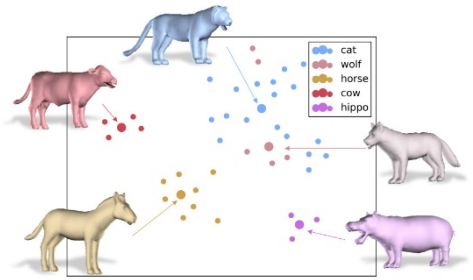
- Developed in collaboration with the DISCo Department of Milano Bicocca University

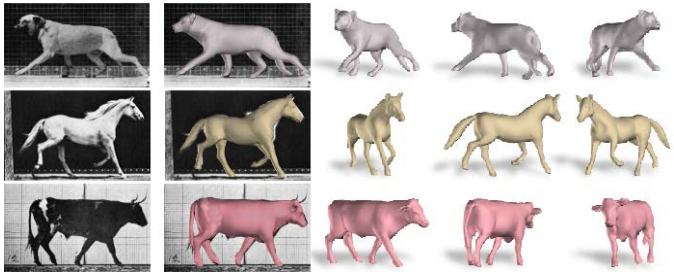
imati

Inference in probabilistic graphical models and 3D statistical shape models of articulated objects

D-PMP: Particle-based Belief Propagation for multi-mode inference in graphical models with continuous variables (Zuffi et al., ICML 2014)

The Stitched Puppet Model: a part-based model of the human body learned from 3D scans (Zuffi et al., CVPR 2015)





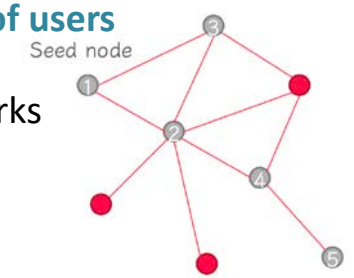
SMAL: Skinned Multi-Animal Linear model learned from 3D scans, can represent animals of different species in a common shape space (Zuffi et al., CVPR 2017)

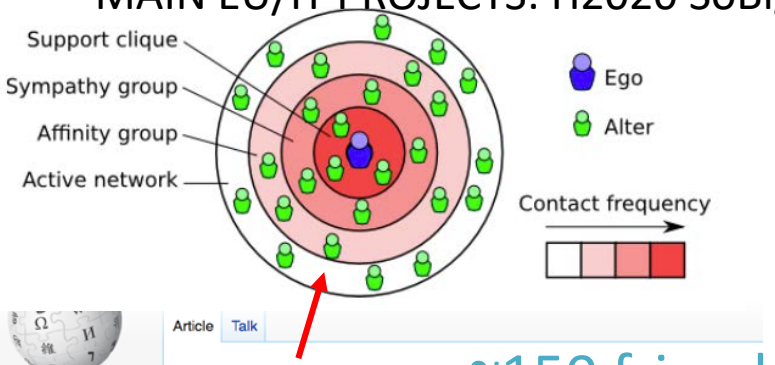
SMALR: Capturing 3D articulated shape and appearance of animals from images (Zuffi et al., CVPR 2018)

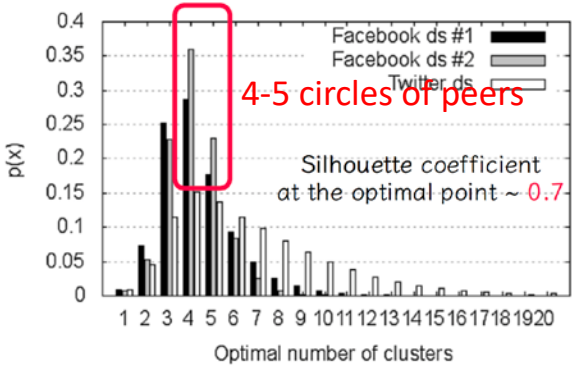
CNR-IIT

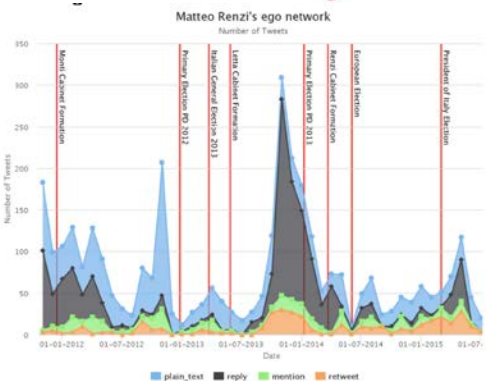
Online Social Networks as a microscope for studying human behaviour

- MAIN CONTRIBUTION
  - Considering the structural properties of **ego networks** (graph-based abstraction representing a person's personal network) and the limits imposed by **cognitive constraints**, it is possible to accurately model **information diffusion** both over individual social links, as well at the entire network level, i.e., it is possible to accurately model information "cascades".
  - When information diffusion is driven by trust, the average length of shortest paths is more than twice the famous one (about 6 or less) obtained when all social links can be used for dissemination.
- IMPACT
  - Ego networks analysis empowers users in their decision-making process, and aids social, economic and political decision-makers, as well as marketing designer, in steering their communication strategies in order to optimize information diffusion.
- CHALLENGES
  - understand the **interplay** between **information diffusion** and **network structures**
  - characterise the **dynamic properties** of personal ego networks for specific **classes of users**
  - build OSN **services** based on **ego-network context information**.
  - understand the **mixed** role of OSN as both **information-** and **social-oriented** networks
- MAIN EU/IT PROJECTS: H2020 SoBigData



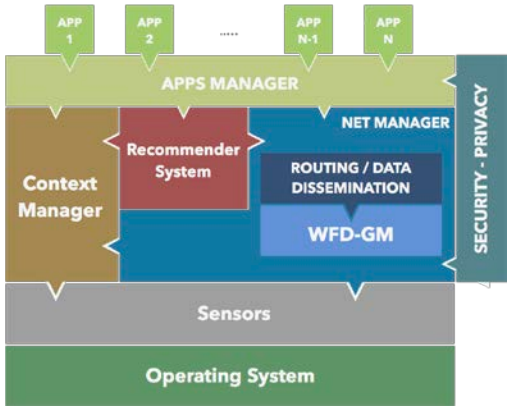


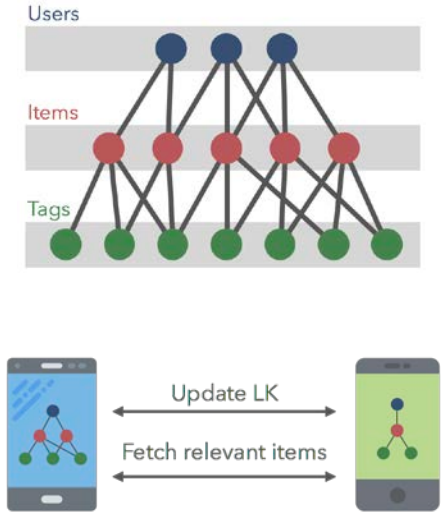


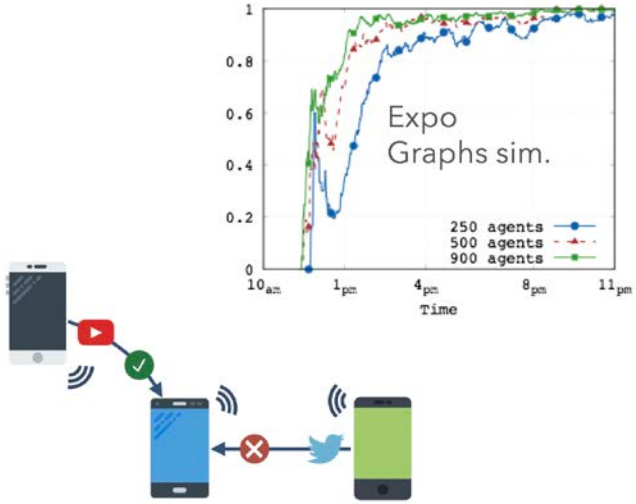


Recommender Systems for Mobile and Online Social Networks

- MAIN CONTRIBUTION
  - definition and evaluation of a new context-aware RS for OSN and MSN based on the analysis of **heterogeneous context information**, derived both from the **physical devices** of the users (sensing information) and their **social context** (e.g., physical proximity and OSN interactions)
  - RS for MSN rely on a partial and dynamic knowledge of the network and the available data, and this requires additional optimization of the algorithms.
- IMPACT
  - This activity can contribute to empower different users categories in the decision-making process, at social, political and economical levels.
- CHALLENGES
  - The analysis of heterogeneous context information, both in online and mobile environments, can further optimize the application of RS in several application domains, from **smart city** to **health** and **well-being**, improving the **service personalization**.







Social Media

MAIN CONTRIBUTION

- Social media crawling, crowdsensing and crowdsourcing
- User reliability and Fake content detection on Social Media
- Multimedia and multimodal data analysis

IMPACT

- Effective systems for the detection and the prevention of spread of fake news and hoaxes.

CHALLENGES

- integration of effective fact-checking methods from knowledge bases or crowdsourcing; mitigation of false information spreading

FUTURE WORK

- developing effective tools to contrast the spreading of false information and to boost fact-checking functionalities.

MAIN EU/IT PROJECTS

- smartnews - Regione Toscana (FAR FAS), 200k€

INVOLVED CNR INSTITUTES

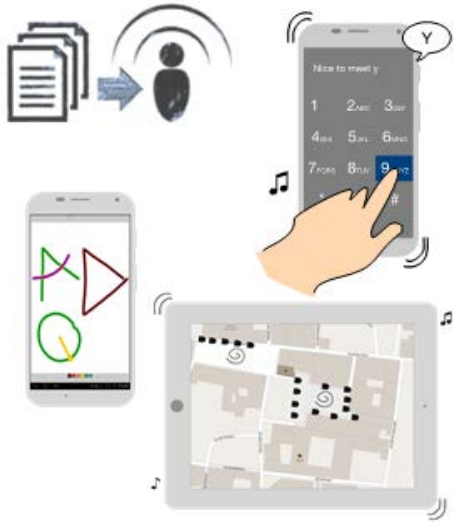
- IIT, ISTI

SmartNews

Multimodal Interaction and Accessibility

- MAIN CONTRIBUTION:
  - Inclusive design for interaction with mobile and electronic devices; Innovative HCI methods to provide most suitable and natural way of interaction also using wearable/specialized devices.
- IMPACT:
  - Empowering individuals: accessible design empowers and is valuable for all users.
- CHALLENGES:
  - New emerging challenges include personalization, system adaptation, seamless integration of IoT devices, intelligent data processing, and also standards definition.
- MAIN EU/IT PROJECTS:
  - Città educante (PON 2007--2013 MIUR)
  - Suoniamo
- INVOLVED CNR INSTITUTES: IIT

cittàeducante



CONSIGLIO NAZIONALE DELLE RICERCHE

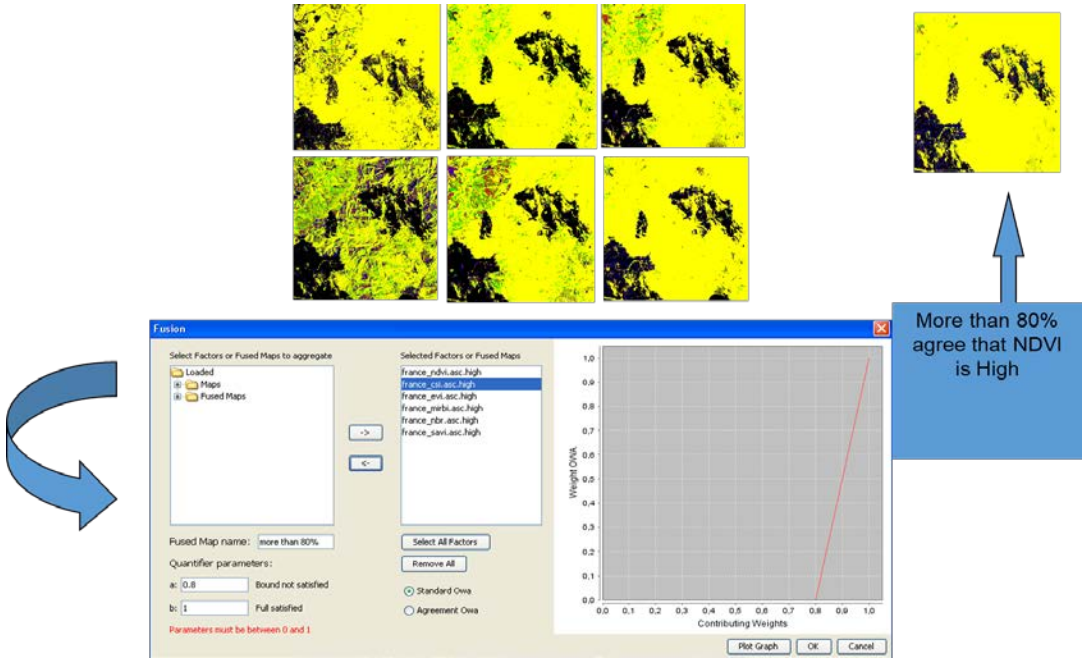
CNR-IREA

Soft computing for Geo information

- MAIN CONTRIBUTION:
  - **Mining implicit information:** geoinformation extraction, spatio temporal analysis, (unmixing) classification
  - **Quality assurance and assessment:** quality indicators**,** fitness for use assessment,
  - **Multi source geo data synthesis:** flexible decision support to target needs tolerating partial domain knowledge and uncertainty /imprecision on data
- IMPACT:
  - **semantic interoperability** and re-use of georeferenced user generated content from multiple sources
  - **Geo data synthesis** modeling complementary/redundant geo information and distinct decision attitudes.
- CHALLENGES:
  - **Semantic Interoperability**: knowledge formalization by linguistic descriptors and (fuzzy) ontologies to improve semantic understanding of data and awareble re-use
  - **Personalization and customization of geo data synthesis:** modelling distinct scenarios based on distinct decision attitudes (optimistic, pessimistic, ..), majority opinion, data driven; Linguistic summarization of geo data synthesis
- FUTURE WORK: quality-based discovery service, downstream service, mapping geospatial natural and artificial phenomena
- MAIN IT PROJECTS: URBAN GEO_BIG DATA, STRESS, SIMULATOR-ADS
- INVOLVED CNR INSTITUTES: IREA

Flexible Synthesis of Geo Information

- ✓Representation of partial knowledge
  - ✓Propagation of imprecision
  - ✓synthesis in two steps:
    - ✓Soft constraints to identify contributing factors
    - ✓Linguistic quantifier guided aggregation to generate a synthetic map

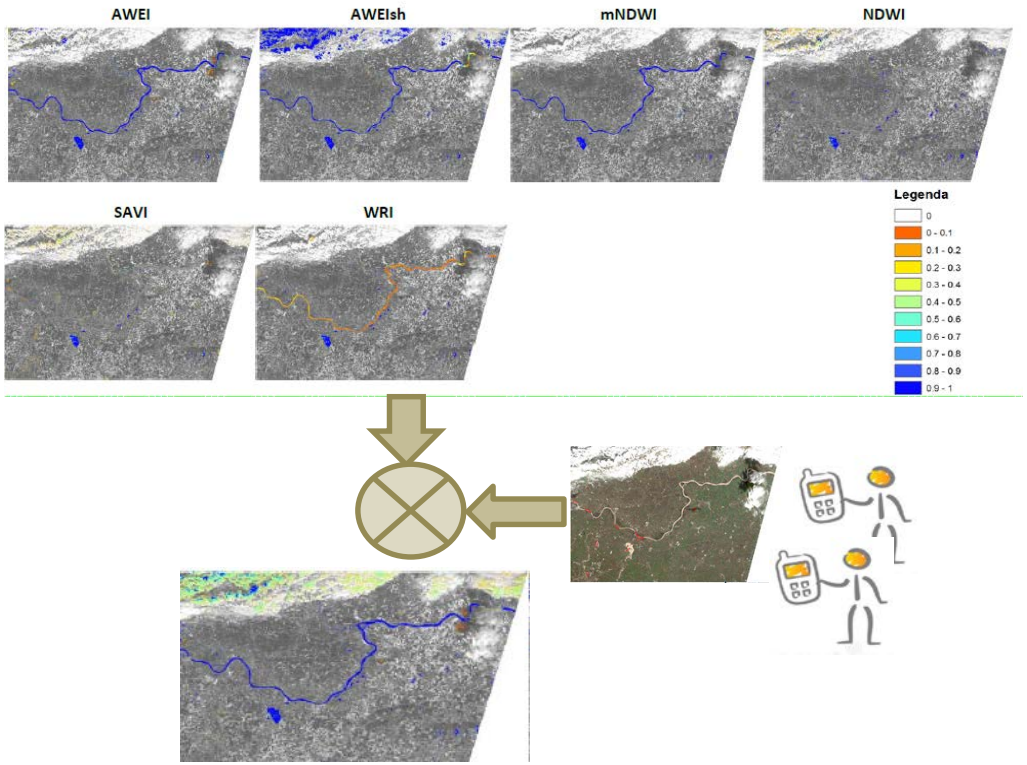


Flexible Synthesis of Geo Information

Fondazione CARIPLO

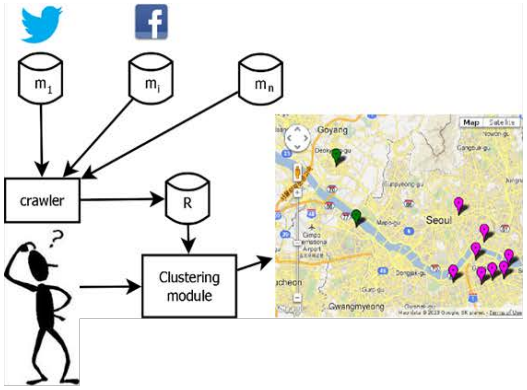
STRESS Resilient Smart Societies

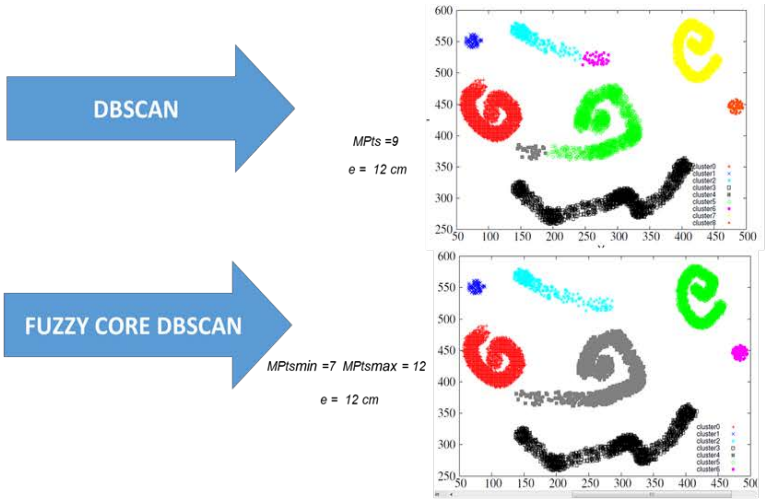
Flexible Synthesis of multiple remote sensing images by fuzzy aggregation & machine learning optimization based on in situ observations (VGI) to map flooded areas



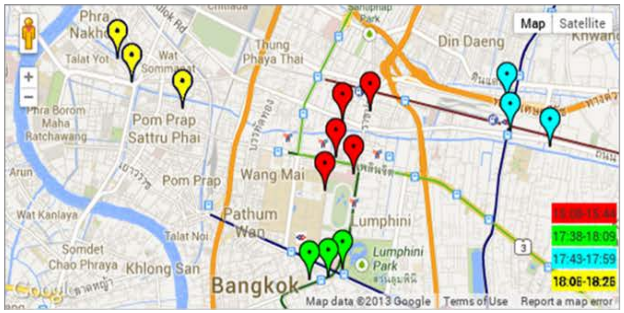
Flexible Synthesis of Geo Information

Periodic/aperiodic event detection in Social networks by geo-temporal density based clustering





Most frequent Time periods and locations of daily recurring traffic jams in Bangkok during 6-12 2013 from Twitter

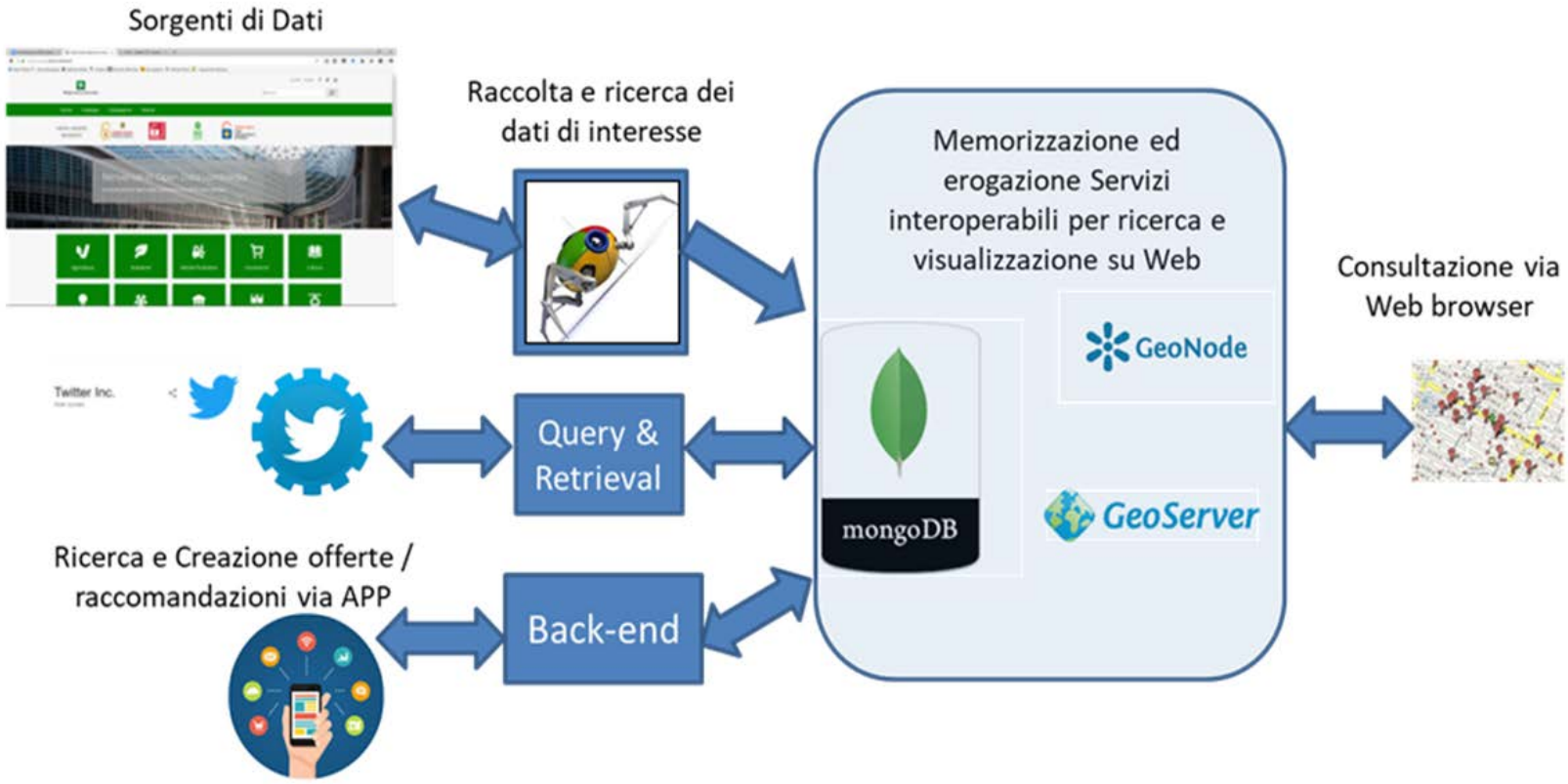


Most frequent Time periods of daily recurring traffic jams at global scale during 6-12 2013 from Twitter

Integration of Multisource Geo Information

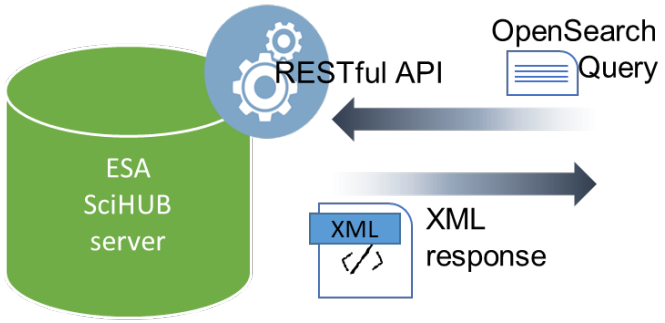
FHfFC project

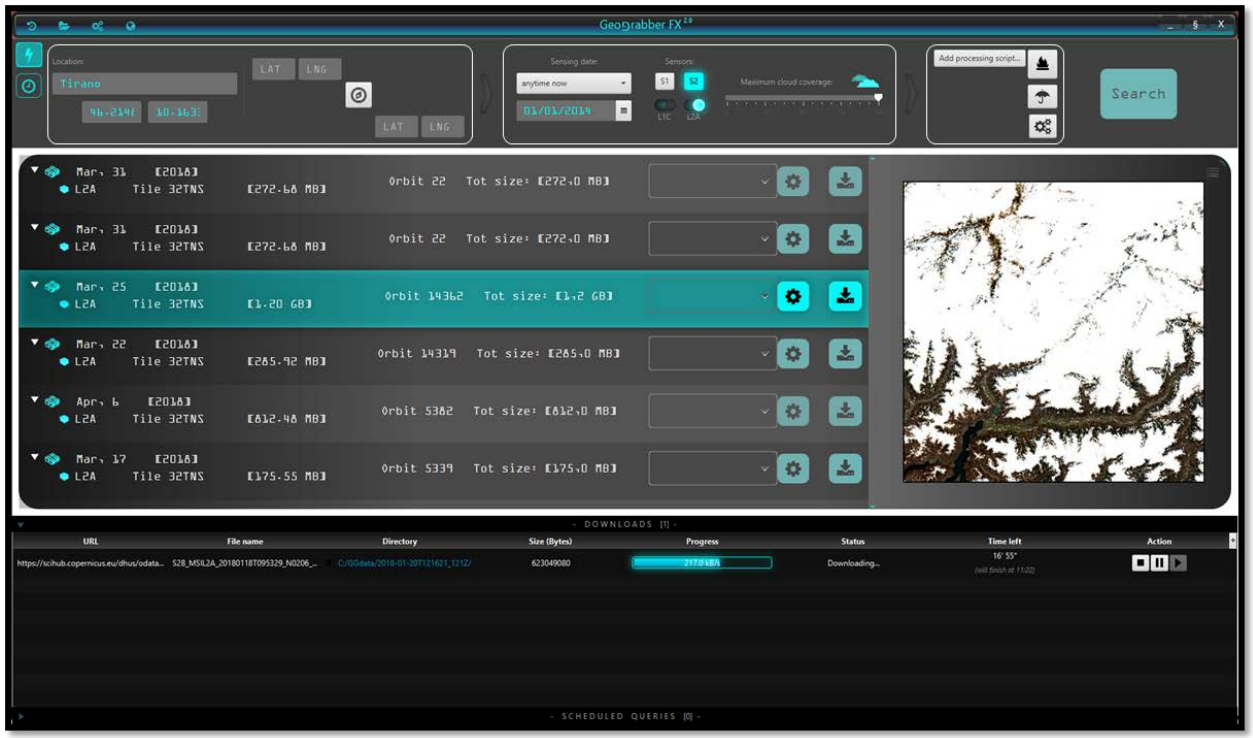
Crawling, georeferencing and interoperable Web sharing of Open Data and related Social network messages



Retrieval and Processing of Geo Information

Enabling not experts to deploy periodic Downstream services



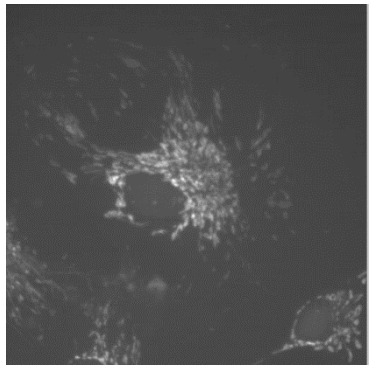


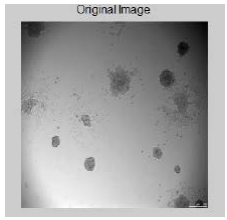
CONSIGLIO NAZIONALE DELLE RICERCHE

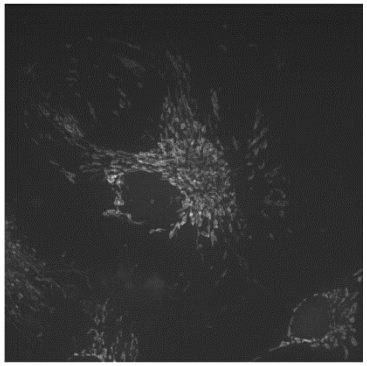
CNR-IAC

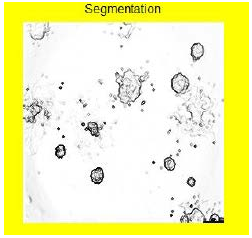
Multimedia and Multidimensional data: Image and Video analysis

1) Multiresolution analysis and non-linear approximation: time-frequency/time-scale transforms, regularization and source separation methods with sparseness constraints, regularization with visual perceptual constraints, functional approximation using metrics for images quality, non-linear relaxation methods for variational image segmentation.





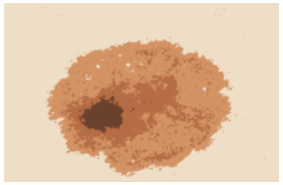


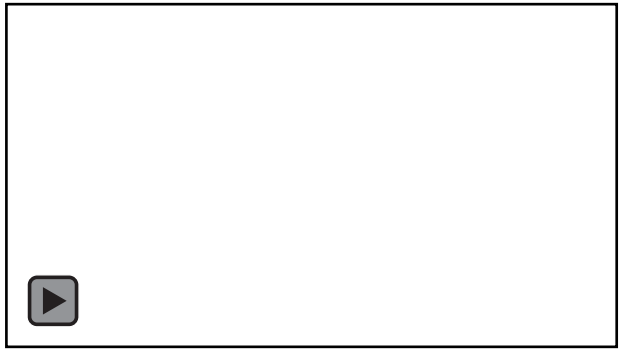


Objectives:

- ✓ denoising, deblurring, segmentation, compression and video motion estimation;
- ✓ edge detection and features extraction in high oscillating signals;
- ✓ image processing and integration in a safety monitoring system of the territory;
- ✓ digital restoration of archives;
- ✓ colour quantization in dermoscopic images for skin marks;
- ✓ target tracking (video surveillance) and iris tracking (biometry).

Italian patent with international extension PCT Pub. No.: WO/2016/199025 A1
title: "Microscopio confocale e relativo procedimento di acquisizione ed elaborazione di immagini" - Cotitolarità: ***CNR-Crestoptics srl****.*
Inventors: ***D. Vitulano****, V. Bruni, V. Ricco, A. Santinelli*

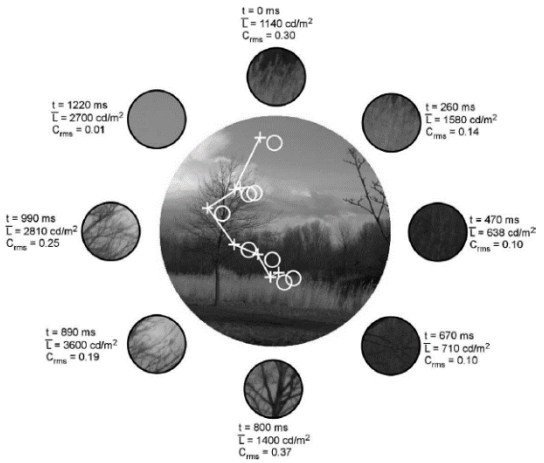




2) Information and perception theory: metrics for the distance between probability densities, relative entropy and mutual information, asymptotic equipartition property, channel coding theorem.

Objectives:

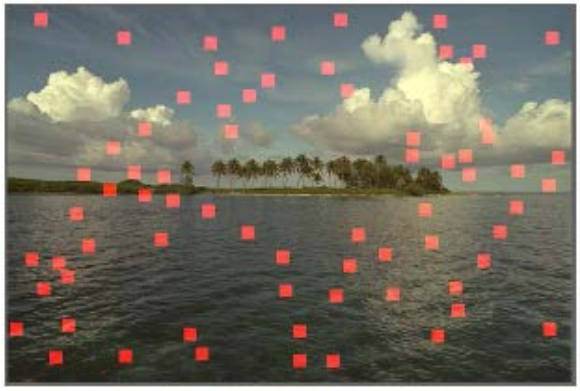
- ✓ new measures to evaluate the image quality according to visual perception rules by characterization of observation points in pre-attentive phase.
- ✓ definition of formal relations between human visual perception and fundamental theorems of information theory to determine the *just noticeable detection threshold*.

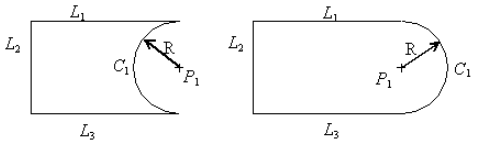


3) Geometry and digital topology: methods based on the usage of geometric and topological characteristics for digital images.

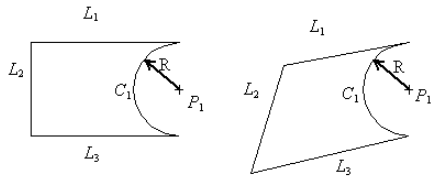
Objectives:

- ✓ low-level image analysis, digital geometry, digital topology, computational geometry, geometric modeling, visual geometry and modeling, shape representation, shape description, color image segmentation.

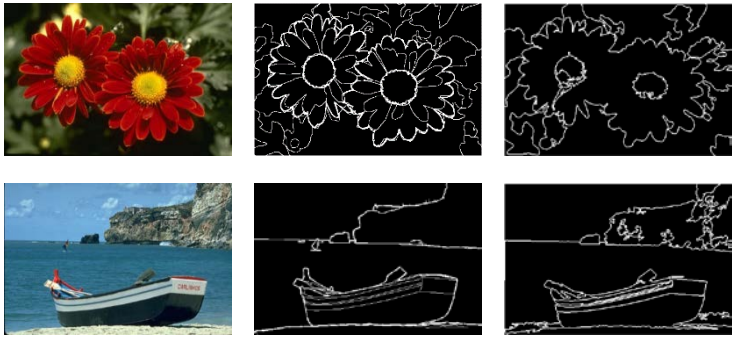




(a) Same geometry but different topology



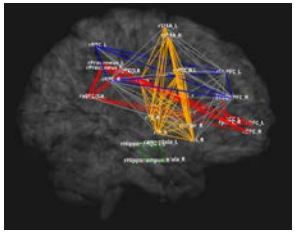
(b) Same topology but different geometry

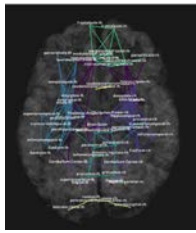


Input image | Human ground truths | Color segmentation

Computational and Statistical Tools for Knowledge Discovery and Data mining (KDD)

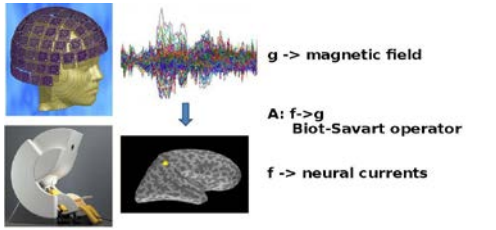
1) Statistics for high dimensional data analysis: convex and folded concave penalization, group selection, partial linear hybrid additive models, high dimensional cox regression, functional data analysis, functional data clustering, geo-referencing and geo-spatial statistics, Bayesian tracking, particle Filter, optimization, Galois lattices induced by graphs and binary relations, bootstrap algebraic multigrid.

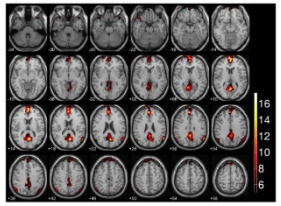


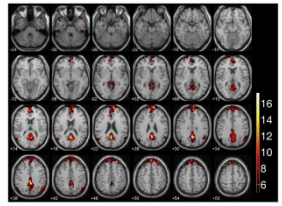


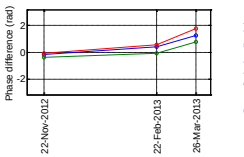
Objectives:

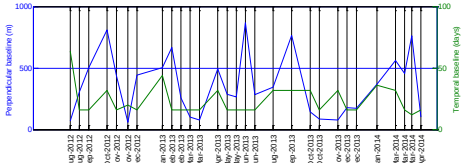
- ✓open source software, NeuroPython (nipype-based) for fMRI, MEG, EEG, iEEG signals;
- ✓open source software (R-based) for residential energy consumption prediction;
- ✓multispectral remote sensing images analysis;
- ✓x-ray data analysis;
- ✓open source software (R-based) for biomedical, biomolecular and clinical data integration;
- ✓characterization of agri-food products on morphometric, biochemical and genetic base;
- ✓spectral clustering of sparse networks.

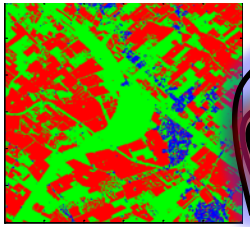


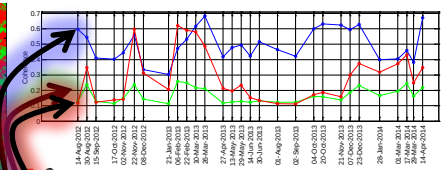


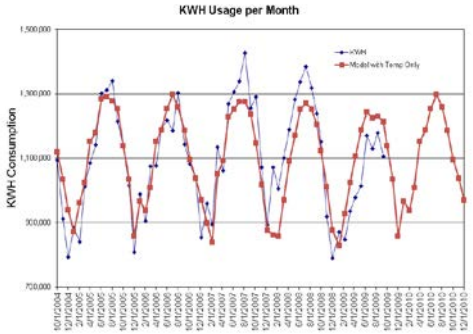


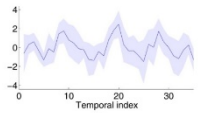


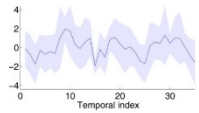


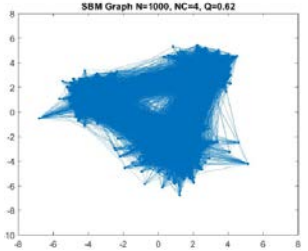


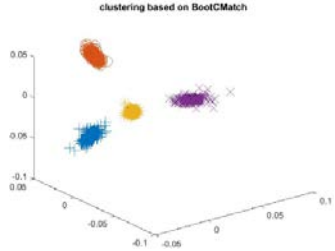








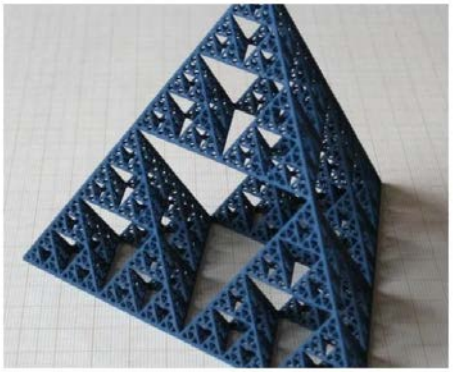




2) Advanced techniques to accelerate 3D printing with FDM

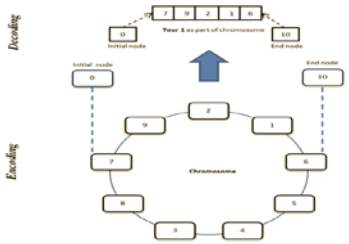
3D printer: WASP Delta 20x40, 1 extruder, plastic material PLA.

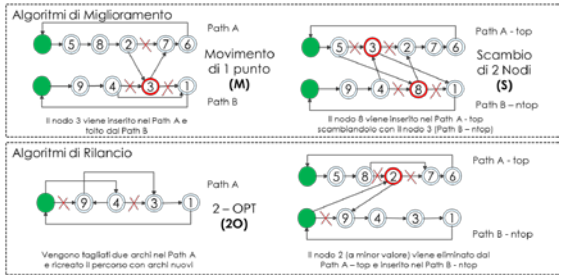
Objectives: shape optimization.



3) Team orienting problem

Objectives: decision support system in tourism.





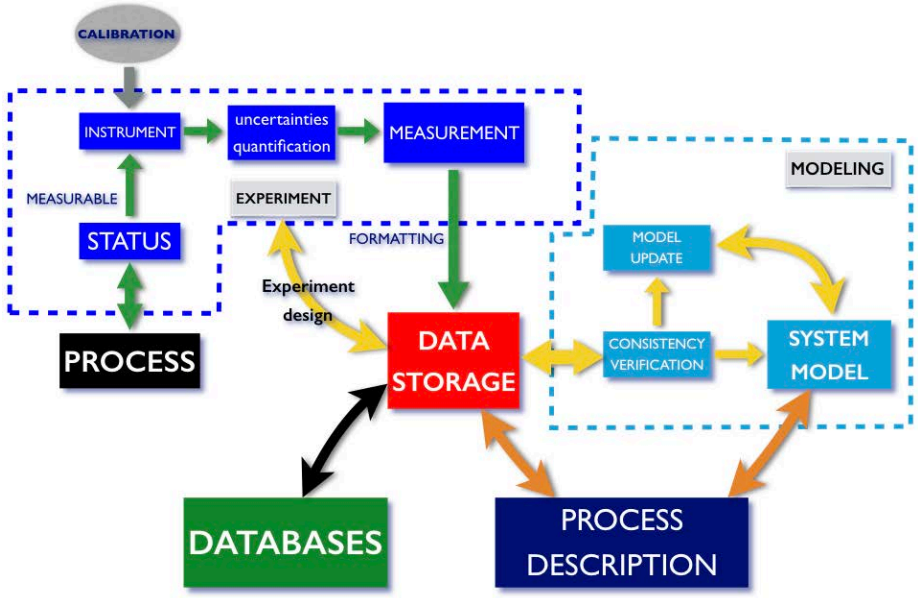
CONSIGLIO NAZIONALE DELLE RICERCHE

CNR-IRC

Virtual Chemistry of Smart Energy Carriers

- PURPOSE: Collecting and cataloging how data is used within the European combustion community with the goal of promoting efficient data exchange, management and usage.
- IMPACT: Harmonization of the existing combustion chemical kinetics databases will allow for the integration of the European research efforts in the field and a substantial advancement in the Smart Energy Carriers science and technologies development.
- CHALLENGE: Identify effective ontologies of main elementary reactors used in chemical kinetics studies and propose corresponding standard descriptors to be implemented in existing and future chemical data repositories.

- FUTURE WORK: preparation of a white-paper on the topic, definition of descriptors and cataloguing of main experimental combustion reactors by the end of 2018.
- PROJECT: COST Action CM1404 - SMARTCATs (Smart Energy Carriers Chemistry and Technologies)
- INVOLVED INSTITUTE: IRC-CNR



CONSIGLIO NAZIONALE DELLE RICERCHE

CNR-ICAR

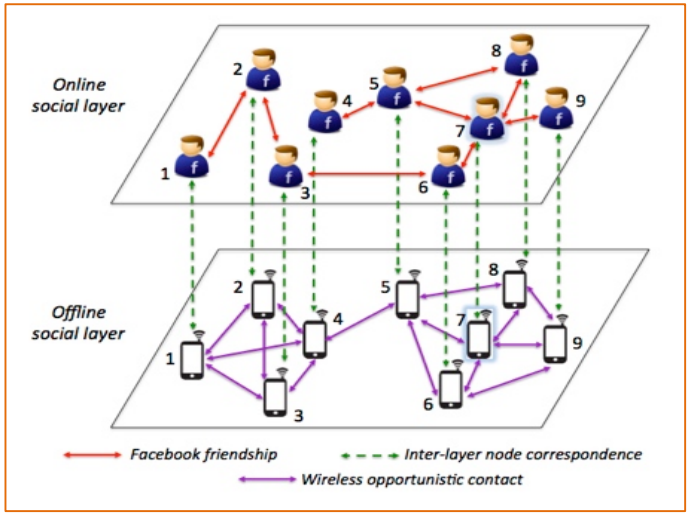
Data Analitcs

- MAIN CONTRIBUTION:
  - Complex Networks analysis and mining
  - Behavior analitcs
  - Query-Answering to natural questions.
- IMPACT:
  - organization models of networks
  - information propagation
  - Recommendation
  - viral marketing
  - automatic business processes
- INVOLVED CNR INSTITUTES : ICAR

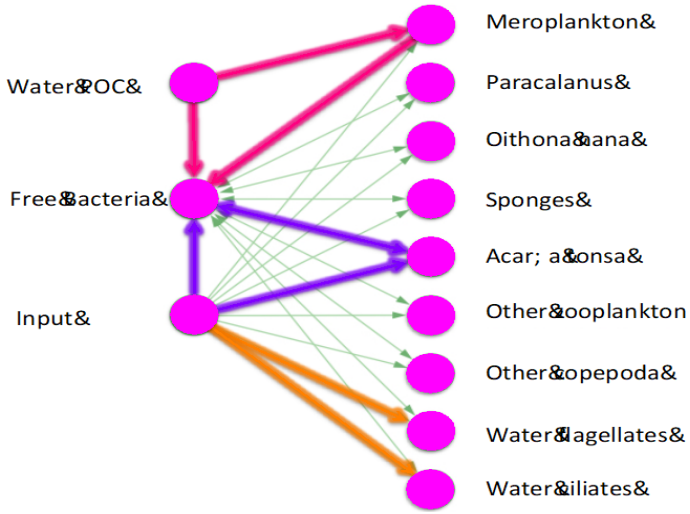
Community Detection in Multi-Layer and Attributed Networks

- In social networking environment a user usually has multiple relationships on different online social networking services. Facebook, Twitter, LinkedIn, are just some examples of the social dimensions an online user may have. Multi-layer are valid formalism to model the multiple types of social connections between users.
- Motifs are important structures that can give insight into the functional mechanisms of the analyzed system, and the division of a network based on diverse motifs can reveal different organization models of the same network.
- The compositional dimension contained in many real world networks has been recognized fundamental to find network divisions which better reflect group organization.

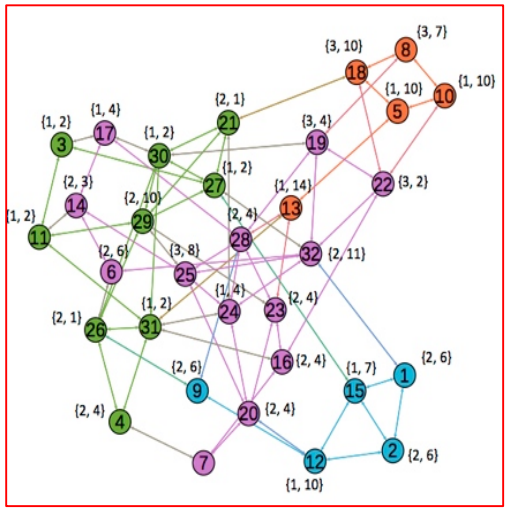
Multi-layer network



Motif-based Communities: Florida Bay food web network with three types of 3 nodes-motifs



Attributed Networks



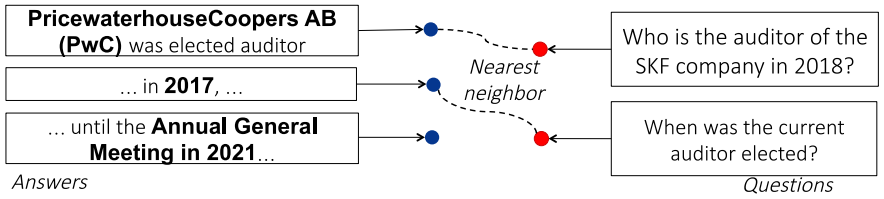
Artificial Intelligence for Question Answering

- **Question Answering (QA)** has the goal to automatically provide pertinent answers to natural language questions
- It is a complex task that requires contextual **natural language understanding (NLU)**, **deep learning** and **reasoning** abilities.
- Almost all **Natural Language Processing** (NLP) tasks can be seen as a QA problem (e.g. **entity extraction**, **sentiment analysis**, **machine translation**).

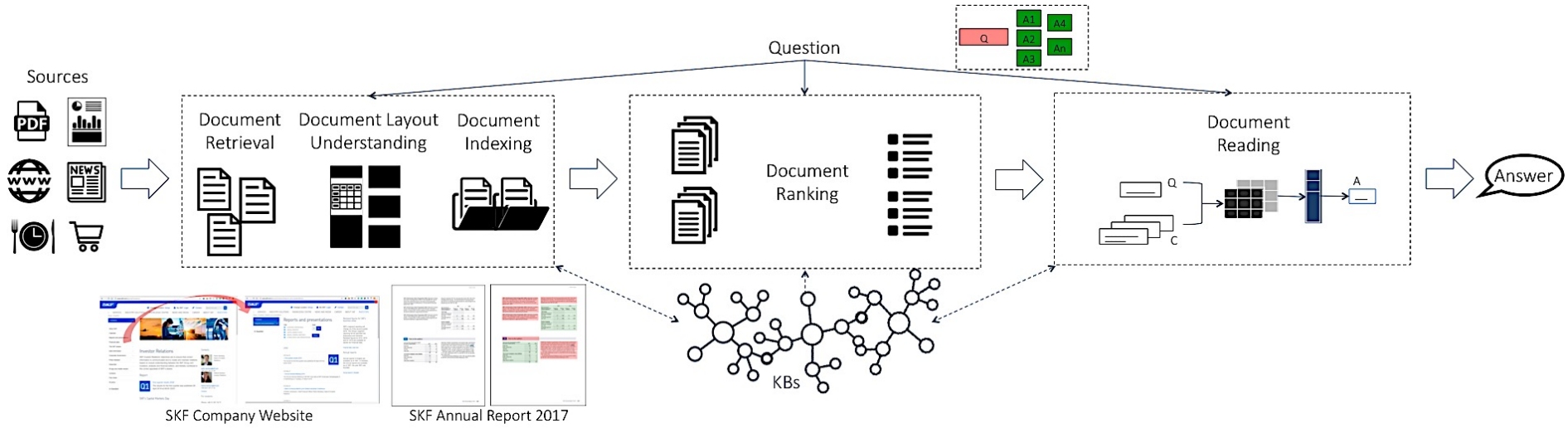
Example

"At the Annual General Meeting in 2017, PricewaterhouseCoopers AB (PwC) was elected auditor for AB SKF until the Annual General Meeting in 2021."

annual-report-skf-group-2017

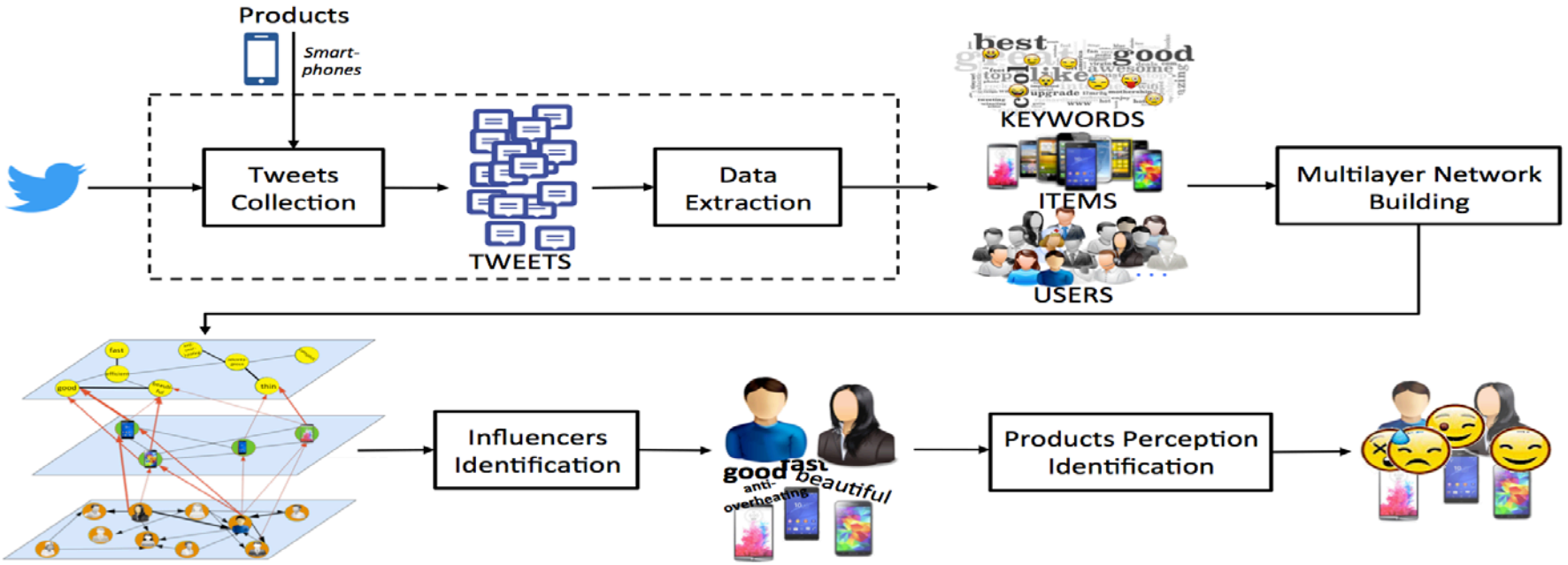


Architecture and Models



Social Network Influencers

- **The identification of social network key influencers with their products perception and preferences is crucial to enable marketers to apply effective techniques of viral marketing and recommendation.**
- Methodology able to identify the most influential twitterers of specific topics or products, along with their perceptions and opinions about them, that combines topological and contextual information.



Behavior Computing

- Modeling and analysis of behaviors and relationships of heterogeneous entities within complex environments

Social Network Analysis

Wearable data analysis

Medical diagnosis

Process optimization

Consumer behavior

Smart cities and mobility

Predictive maintenance

CONSIGLIO NAZIONALE DELLE RICERCHE

CNR-IASI

Big Data Analytics

- MAIN CONTRIBUTION
  - Algorithmic approaches for providing accurate answers to joint queries from multiple aggregate databases
  - Automatic extraction of words associations from texts based on the joint distribution of words occurrences
- IMPACT
  - Performances of strategic queries in data warehousing
  - Efficacy in comparison between trends, business predictions and decision making
- CHALLENGES
  - Time series analysis to study the spans of life trajectories (based on demographic characteristics, education and other lifestyle variables) in the form of sequence analysis
  - Clustering
- INVOLVED CNR INSTITUTES: IASI

Knowledge Representation, Reasoning and Engineering

- MAIN CONTRIBUTION
  - Ontology Engineering
  - Semantic similarity (**SemSim** method)
  - Semantic relatedness
- IMPACT
  - Knowledge organization and formalization
  - Knowledge interoperability
  - Semantic search and retrieval
- CHALLENGES
  - Exploratory search
- INVOLVED CNR INSTITUTES: IASI

SemSim: Research Objectives

Semantic Search and Retrieval, based on a Reference Ontology and Similarity of Concepts

To this end, we need to:

- Build a **Reference Ontology** (Weighted Reference Ontology, WRO)
- Use the WRO to **annotate** the **digital resources**
- Formulate a **Request**, in terms of concepts in the WRO
- Searching target resources by computing **similarity reasoning**

Then

- **Search results** will be a list of resources, ranked according to their conceptual similarity wrt the Request

Semsim: Major Outcomes

- **Method** for weighting concepts in the ***WRO***
  - **Concept weight**: probability that a DR is characterised by the concept

Weighted Reference Ontology

- **Method** for computing the similarity of 2 concepts:
  - **Concept similarity**: based on their weight and the structure of the ontology

CONSIM

- **Method** for the Semantic Similarity between two Conceptual Vectors (Request Vector & Resource Vector)

SEMSIM

Concept Similarity: consim

Concept Similarity (consim) [LIN] probabilistic approach based on information content notion

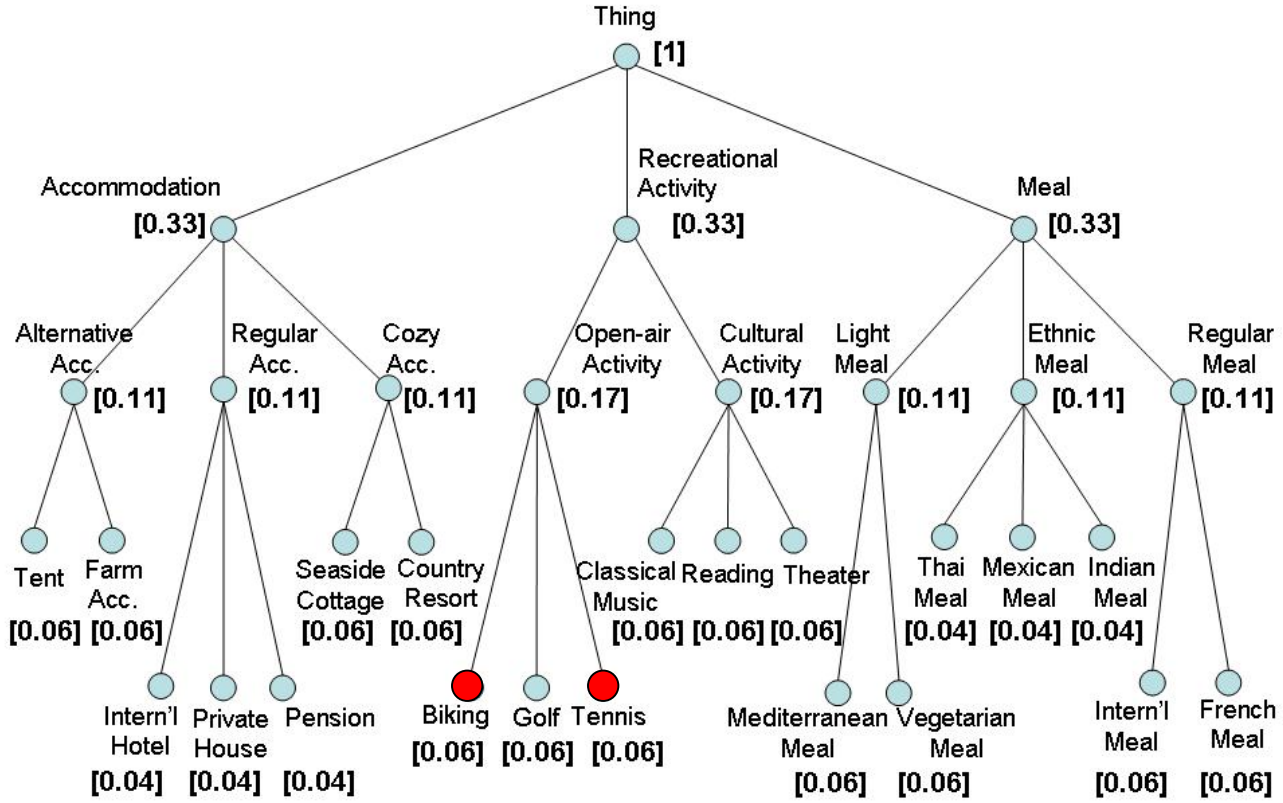
given two concepts c_i *and* c_j *in WRO*

$$consim(c_i,c_j)=2 \log w(lub(c_i,c_j)) / (\log w(c_i)+\log w(c_j))$$

- log w(c_i) is the information content carried by c_i
- log w(lub(c_i,c_j)) is the maximum information content shared by c_i *and* c_j in the WRO

[Lin] An Information-Theoretic Definition of Similarity.
Proc. of 15th the International Conference on Machine Learning (1998)

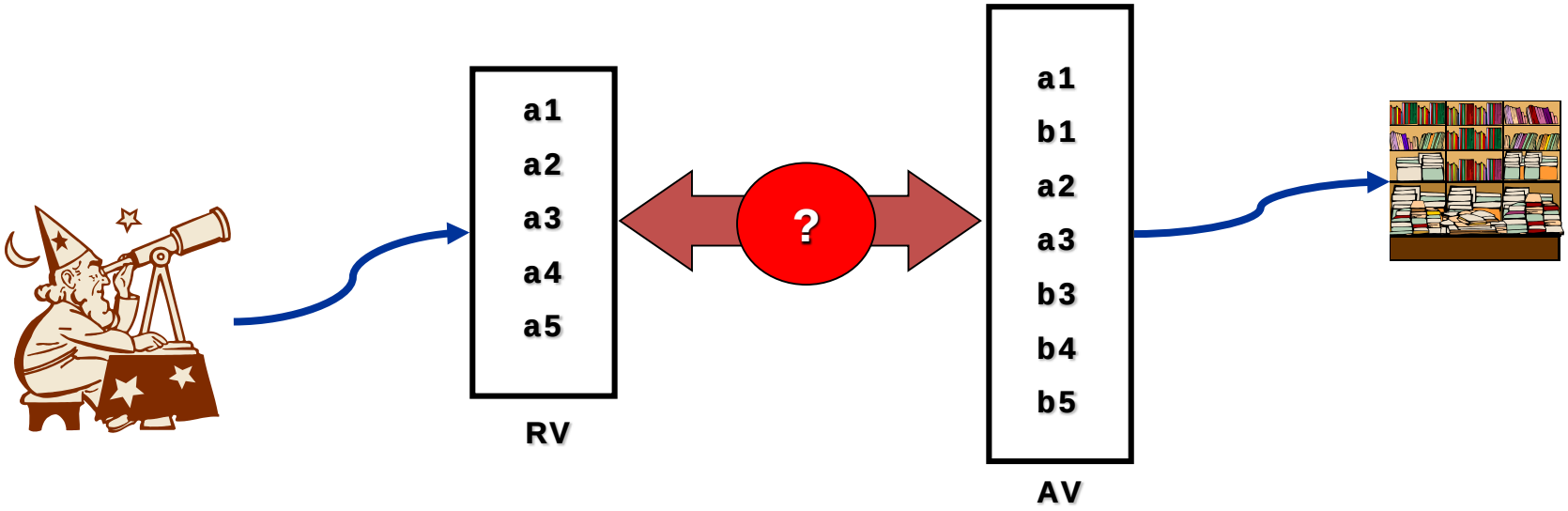
Consim: an example



consim(Biking, Tennis) = 2 log w(Open-airActivity) / (log w(Biking)+ log w(Tennis))=0.63

The Semsim method

- Allows similarity degree computation between a Request Vector (RV) and an Annotation Vector (AV)



Semantic Similarity: semsim

Semantic Similarity (semsim) based on a wedding approach (Hungarian Algorithm)

given a rv and a ofv

$$\textbf{semsim(rv,ofv)} = \textbf{max}(\sum \textbf{consim}(c_i,c_j)) \,/\, \textbf{max}(n,m)$$

where $c_i \in rv$, $c_j \in ofv$ *and*

$n = |rv|$ *and* $m = |ofv|$

Each concepts can participate in one pair exclusively

CONSIGLIO NAZIONALE DELLE RICERCHE

CNR-STIMA

Ontology-based Virtual Factory/Environments

MAIN CONTRIBUTION

- Ontology-based interoperability and solutions for applications in manufacturing, domotics, health-care, and posture recognition domains.
- Knowledge formalization in the manufacturing domain.
- Augmented/Virtual Reality (AR/VR) to support a user-centered applications with visual, auditory, tactile feedback and interaction.

IMPACT

- The use of ontologies is a way to create a formal reference knowledge.
- In manufacturing, the integration of heterogeneous digital tools (e.g. DES, 3D layout design in VR environment, optimization based on AI techniques and mathematical modelling) will have an impact on KPIs such as throughput, costs, sustainability.
- In industry, VR-based training or AR-based support while performing a certain tasks may result in a save of time and costs.
- In rehabilitation, the provision of innovative rehabilitation training – also at home – can help in reducing the social and economic burden of chronic pathologies.

CHALLENGES

- Usability and scalability of ontology-based software tools in industrial applications.
- Standardization of industrial ontologies to boost knowledge and data sharing.
- Integration and interoperability of continuously updated AR/VR technologies.
- Development of VR/AR applications that are not only effective, but also well designed and accepted by their final users.

FUTURE WORK: ontology engineering for industrial applications, VR/AR applications for manufacturing industry

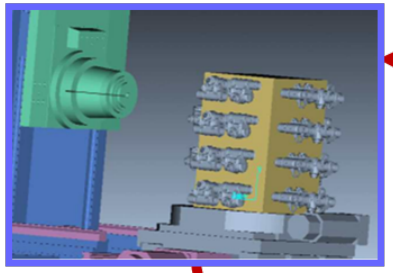
PROJECTS: Ansaldo Energia Lighthouse Plant; W3C Linked Building Data Community Group; Industrial Ontology Foundry (IOF)

INVOLVED CNR INSTITUTES: STIIMA (ITIA)

Ontology-based Virtual Factory/Environments

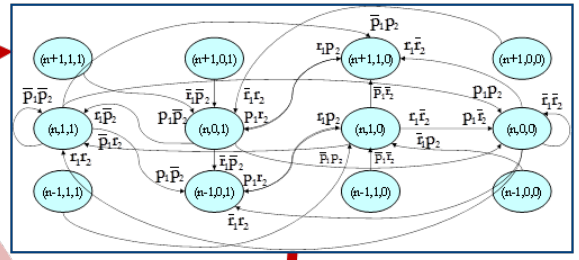
How to guarantee the coherence of the Models?

Kinematics model



Production Process Design and Simulation

Mathematical model



Design and Performance Evaluation

MANAGEMENT

PROCESSES

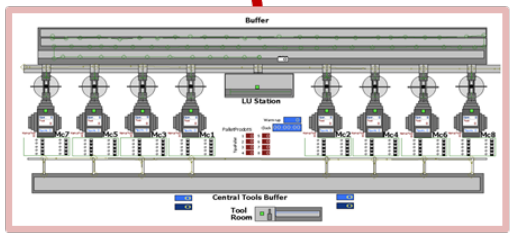
RESOURCES

(Descriptive) Model of the Factory as a whole

Visual model

Layout Design

Simulation model



Performance Evaluation

Meta-model of the Factory as a whole

Ontology-based approach (OWL as meta-metamodel)

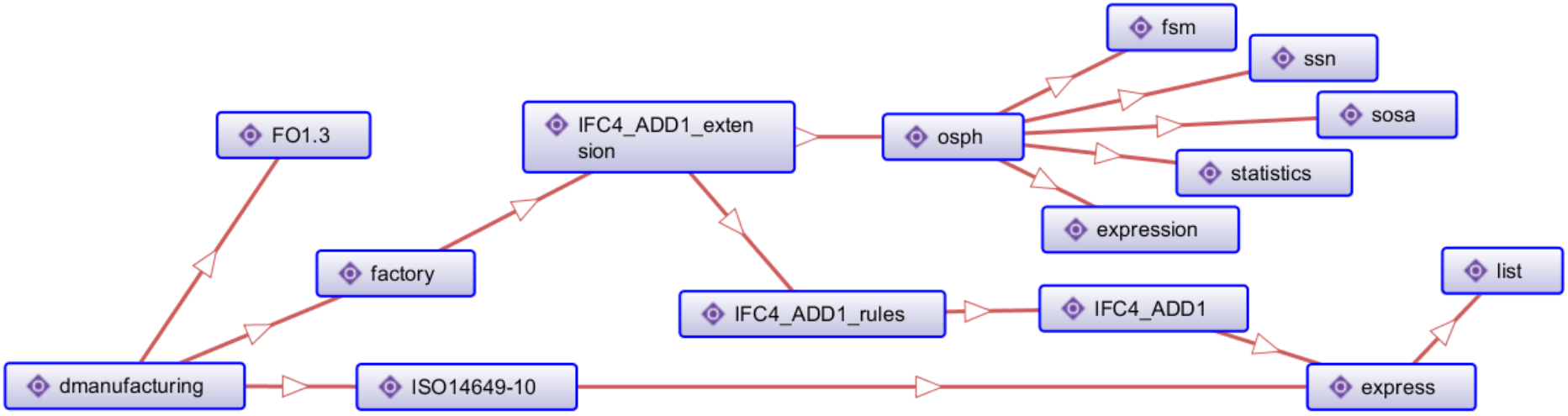
1)Specific meta-models have an ontology interface to the meta-model of the Factory as a whole (T-box)

2)The specific models implements an interface to the model of the Factory as a whole (A-box)

Ontology-based Virtual Factory/Environments

Virtual Factory Data Model (VFDM): a common language (i.e. a Data Model) to support System/Factory design and management.

Virtual Factory Data Model (VFDM) – a possible integrated ontology

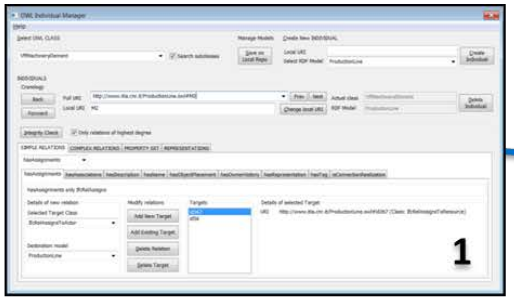


Ontology-based Virtual Factory/Environments

Industrial Case example: Design of Production Line

Design of a manufacturing line dedicated to the production of **short-block assembly:**

1. Define the process plan and assign each process step to a type of machine tool
2. Design the layout of the plant, defining the number of machines
3. Evaluate the system performance using discrete event simulation
4. Analyse the results of the simulation



OntoGUI

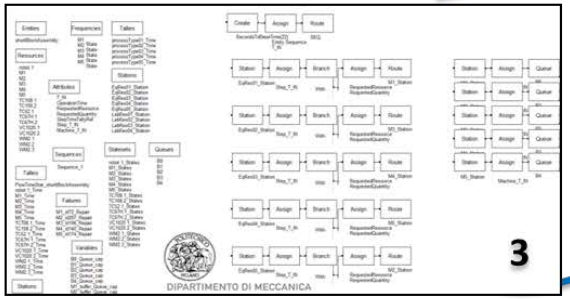
catalog of equipment, process plans

Production system layout; process assignments

Factory Model as a whole

GIOVE-VF – Design mode

Process plans, demand, production system layout



ARENA by Rockwell + plugin

System performance, simulation log

System performance, simulation log

GIOVE-VF – Animation mode

Ontology-based Virtual Factory/Environments

Industrial Case: Design of Production Line

Layout planning

Dynamic animation

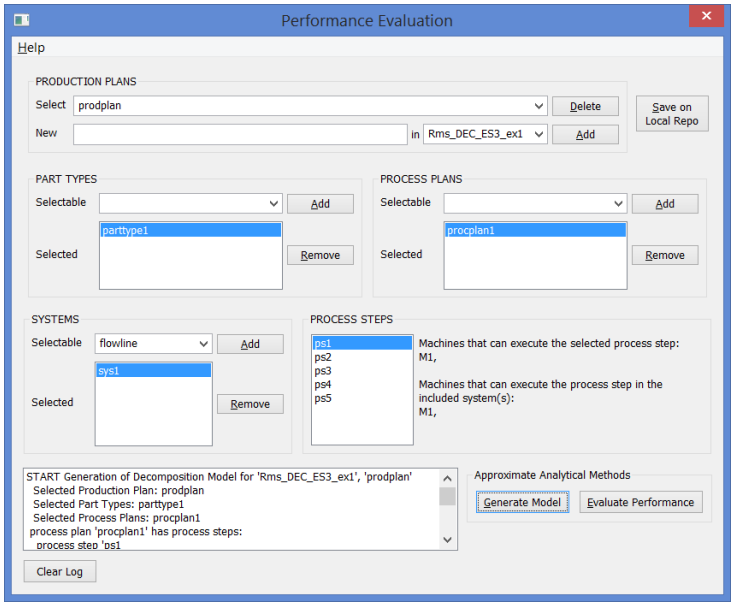
GIOVE-VF is a virtual reality collaborative environment to support the factory layout design. Developed by ITIA-CNR onto the C++ library GIOVE (Graphics and Interaction for OpenGL-based Virtual Environments).

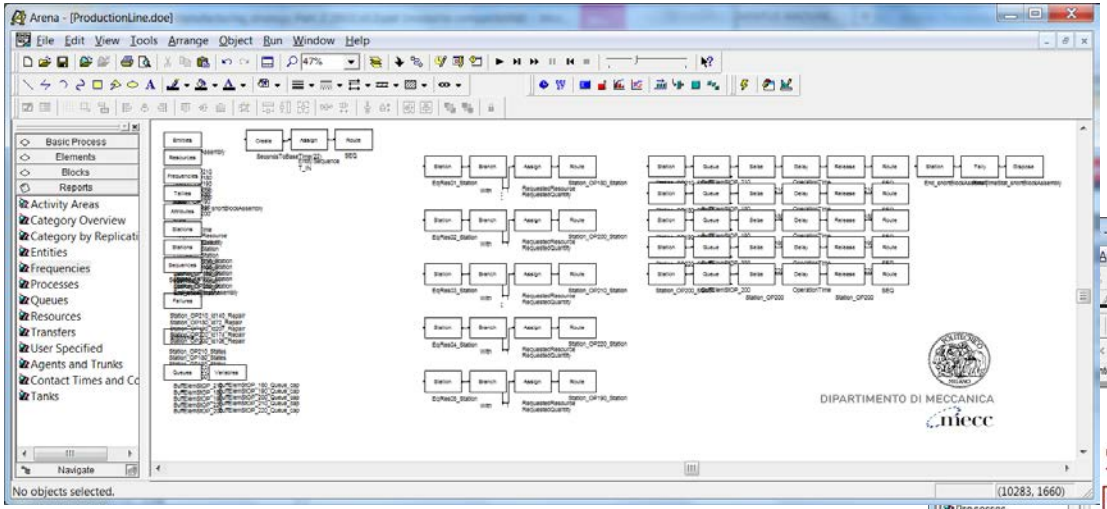
Ontology-based Virtual Factory/Environments

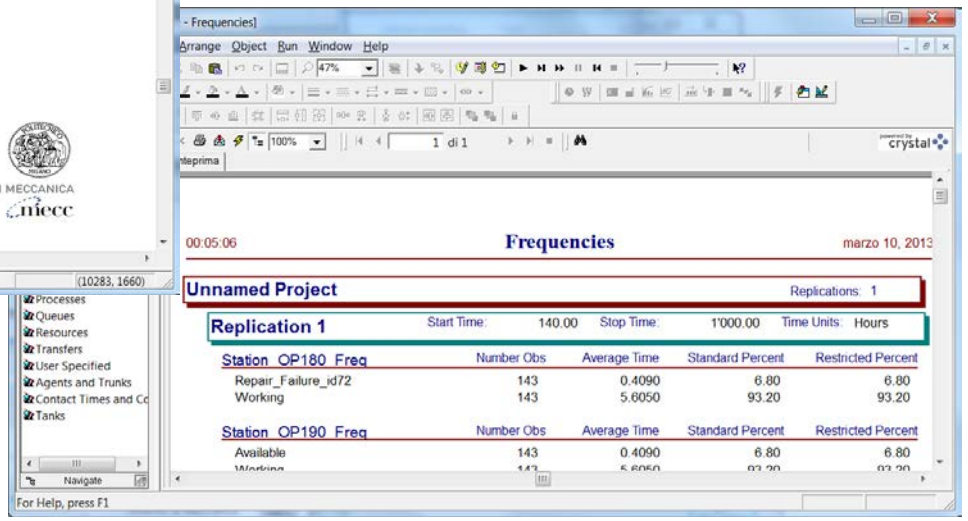
Industrial Case: Design of Production Line

Perfomance evaluation

Automatic ontology-based generation of models of ***approximate analytical methods*** or ***discrete event simulation*** (e.g. Rockwell ARENA, Plant Simulation) to evaluate the performance of production systems.







CONSIGLIO NAZIONALE DELLE RICERCHE

CNR-IGI

- **MAIN CONTRIBUTION**
  - **Management of data from large scientific experiments**: data acquisition, remote data access, data sharing in scientific communities
  - **Data representation and ontologies for nuclear fusion research**: methods and tools for developing agreed data visions in research communities
- **IMPACT**
  - Data management is of paramount importance in new generation physics experiments, in particular Nuclear Fusion research.
  - The new large nuclear fusion experiments such as ITER will rely more and more on the availability of large amount of simulation and experimental data in order to maximize the predictability of the experiments and to reduce cost and risk in experimental activity.
- **CHALLENGES**
  - An unprecedented amount of experimental and simulation data must be handled in order to advance scientific and technical knowledge required to reach the final target of replicating fusion on earth
- **FUTURE WORK**
  - Integration of recent big data technologies from other domains
- **MAIN EU/IT PROJECTS**
  - **MDSplus**: de facto standard system for data management in world-wide fusion research
  - ITER Integrated Modelling & Analysis Suite (**IMAS**)
- **INVOLVED CNR INSTITUTES**
  - **IGI**