



# Faculty Day: Thesis Expo

Wednesday 16th December 2020

Master in Computer Science

## Thesis proposals A.Y. 2020/2021

<a href="#"><u>Runtime verification and monitoring</u></a>	<a href="#"><u>2</u></a>
<a href="#"><u>Design and foundations of programming languages</u></a>	<a href="#"><u>4</u></a>
<a href="#"><u>Virtual, Mixed and Augmented Reality</u></a>	<a href="#"><u>6</u></a>
<a href="#"><u>Computer Vision</u></a>	<a href="#"><u>9</u></a>
<a href="#"><u>Computer Graphics</u></a>	<a href="#"><u>13</u></a>
<a href="#"><u>Network Analysis</u></a>	<a href="#"><u>15</u></a>
<a href="#"><u>Bioinformatics</u></a>	<a href="#"><u>15</u></a>
<a href="#"><u>Computational and Statistical Learning</u></a>	<a href="#"><u>16</u></a>
<a href="#"><u>Data Management</u></a>	<a href="#"><u>17</u></a>
<a href="#"><u>High-Quality Software Development</u></a>	<a href="#"><u>20</u></a>
<a href="#"><u>Software Testing</u></a>	<a href="#"><u>23</u></a>
<a href="#"><u>Collaborations with external companies</u></a>	<a href="#"><u>25</u></a>



## Thesis proposals in “Runtime verification and monitoring”

Titolo: **Runtime monitoring and verification in RML**

Keywords:

Runtime verification and monitoring, domain specific languages, pervasive computing, ubiquitous computing, multi-agent systems, formal methods

Tutors: Davide Ancona, Viviana Mascardi, Angelo Ferrando

Abstract:

Runtime verification (RV) [1,2] is an approach to verification consisting in dynamically checking that the event traces generated by single runs of a system under scrutiny (SUS) are compliant with the formal specification of its expected correct behavior.

RV is complementary to other verification methods and integrates well with both formal verification [3] and software testing [4].

**RML (Runtime Monitoring Language)** [5,6,7]

is a system agnostic RV DSL which allows users to define formal specifications independently of code instrumentation; its key features are simplicity, expressive power, supported by parametric and generic specifications, a general event model based on JavaScript object literals, a formal semantics, a fully fledged Kotlin compiler able to generate from the specifications SWI-Prolog monitors that support real-time verification, through several protocols, of the events emitted by the instrumented SUS in the data-interchange format JSON.

The main aim of this research theme is to further advance the theory and practice of RML; different separate directions can be considered, depending on the skills and interests of the prospective PhD student.

- **Theory**: there exist several interesting unresolved theoretical issues concerning the formal semantics [9] of RML, decidability and expressive power.
- **Language design and implementation**: several directions for extending RML can be considered; different, but not necessarily mutually exclusive, aspects may include:
  - Adding support for data aggregation, and constraint checking over data series;

- extending RML to define and generate Internet of Things rule engines able to react to specific actions triggered by events;
  - integrating RML with ontology matching, to manage event types in a more flexible way;
  - providing a better support to RV error reporting, and static checking of RML specifications, possibly integrated with an IDE, to favor language usability and specification development.
- **Applications:** assessment on the usability of RML calls for more practical and challenging experiments with RML in interesting application domains, including Internet of Things [8] and Multi-agent systems [10,11].

#### References:

- [1] Y. Falcone, K. Havelund, G. Reger, A Tutorial on Runtime Verification, in: Engineering Dependable Software Systems, 141–175, 2013.
- [2] E. Bartocci, Y. Falcone, A. Francalanza, G. Reger, Introduction to Runtime Verification, in: Lectures on Runtime Verification - Introductory and Advanced Topics, 1–33, 2018.
- [3] W. Ahrendt, J. M. Chimento, G. J. Pace, G. Schneider, Verifying data- and control-oriented properties combining static and runtime verification: theory and tools, Formal Methods in System Design 51 (1) (2017) 200–265.
- [4] M. Leotta, D. Clerissi, L. Franceschini, D. Olinas, D. Ancona, F. Ricca, M. Ribaudo, Comparing Testing and Runtime Verification of IoT Systems: A Preliminary Evaluation based on a Case Study. ENASE 2019: 434–441
- [5] <https://rmlatdibris.github.io/>
- [6] L. Franceschini, RML: Runtime Monitoring Language, Ph.D. thesis, DIBRIS - University of Genova, URL <http://hdl.handle.net/11567/1001856>, March 2020
- [7] D. Ancona, L. Franceschini, A. Ferrando, V. Mascardi, A SWI-Prolog based implementation of RML, Workshop on Trends, Extensions, Applications and Semantics of Logic Programming, ETAPS 2020, <https://www.coalg.org/tease-lp/2020/a-swi-prolog-based-implementation-of-rml>
- [8] D. Ancona, L. Franceschini, G. Delzanno, M. Leotta, M. Ribaudo, F. Ricca, Towards Runtime Monitoring of Node.js and Its Application to the Internet of Things. ALP4IoT@iFM 2017: 27–42
- [9] D. Ancona, A. Ferrando, V. Mascardi, Can determinism and compositionality coexist in RML? EXPRESS/SOS 2020, <https://arxiv.org/abs/2009.00391>
- [10] A. Ferrando, D. Ancona, V. Mascardi, Decentralizing MAS Monitoring with DecAMon, AAMAS 2017
- [11] D. Ancona, A. Ferrando, V. Mascardi, Parametric Runtime Verification of Multiagent Systems, AAMAS 2017

## Thesis proposals in “Design and foundations of programming languages”

### *Line of research: Secure and Reliable Systems*

Reference course: Principles and Paradigms of Programming Languages

People: Davide Ancona, Elena Zucca + PhD students: Francesco Dagnino (3rd year), Pietro Barbieri (2nd year), Riccardo Bianchini (1st year)

Overall topic: Formal/mathematical tools to model and reason about programming languages, notably execution models (operational semantics) and type systems enforcing absence of (some kind of) errors. The thesis work can be study/experiment with such formal models, and/or provide their implementation.

### **Specific themes**

(1)

Title: **Deconfined intersection types in Java**

Key-words: foundations of programming languages, Java, type systems

Tutors: Elena Zucca, Paola Giannini (UniPo)

Abstract: Recent work proposes an extension of Java where intersection types are “deconfined”, that is, they can be freely used as types. The extension is formalized by a translation from an extended to a core Java calculus, which is proved to preserve typing and semantics. The thesis work consists in studying and experimenting the proposal, and implementing the two calculi, the translation, and the type system.

(2)

Title: **Union types with union-elimination in Java**

Tutors: Elena Zucca, Francesco Dagnino

Key-words: foundations of programming languages, Java, type systems

Abstract: Union types can be added to Java-like languages, and union-elimination rules allow to derive a type for expressions which have subterms of union types. The thesis work consists in investigating and implementing a Java-like calculus with union-elimination rules.

(3)

Title: **CoFJ at work**

Key-words: foundations of programming languages, Java, coinduction

Tutors: Davide Ancona, Elena Zucca, Francesco Dagnino

Abstract: Corecursive Featherweight Java (coFJ) is a Java-like calculus where objects can be cyclic and the programmer can specify the behaviour of methods on such objects. There is a prototype implementation via translation in coinductive logic programming. The thesis work could consist in providing a full implementation (including a user interface and dealing with optimization issues) and/or designing a type system.

(4)

Title: **Encoding inference systems in Agda**

Key-words: induction and coinduction, proof assistants

Tutors: Elena Zucca, Francesco Dagnino

Abstract: Agda is one of the nowadays most popular proof assistants. That is, it is a dependently typed language where the programmer can also write theorems and their proofs, whose correctness is statically ensured by the type checker. Recently, an Agda library has been developed for inference systems, a widely-used framework to express inductive and coinductive definitions, together with the associated proof principles. The thesis work consists in designing a user-friendly language to write inference systems, and developing a tool generating the corresponding Agda code.

## Thesis proposals in “Virtual, Mixed and Augmented Reality”

(1)

**Title: Making virtual and real environments semantically consistent via ontologies**

**Tutors:** Manuela Chessa, Viviana Mascardi

**Keywords:** Virtual Reality, Knowledge Representation, Ontologies

**Abstract:** This thesis aims at developing a system to create immersive virtual environments consistent with the real world surrounding the user. Once the objects in the real room are detected and the corresponding bounding box created, the virtual scenes should be composed on-the-fly, and in a (semi-)automatic way, by inserting objects of the same shape, position and spatial properties of the corresponding real ones. These virtual objects should be semantically consistent with the virtual scene.

For example, if the virtual scene represents a bush, and a real object has been detected with a parallelepipedal bounding box starting from the ground and 70cm tall, it might be substituted, in the virtual scene, by a shrub. If the height is 1.70cm, it might be substituted by a tree.

If the virtual scene represents a shore, those two objects might be substituted by a chair and a beach umbrella, respectively, and if it is a road in a urban environment, they might be a trash bin and a pole with a road sign.

In order to associate virtual objects with real objects in a semantically consistent way, we need to represent the knowledge about which virtual objects make sense into a virtual scene, also depending on their geometric features, via one or more ontologies. The design, implementation, and experimentation of the ontology are the main goals of the thesis.

(2)

**Title: Visualizing human view**

**Tutors:** Manuela Chessa, Enrico Puppo, Fabio Solari

**Keywords:** Eye tracking, Data visualization

**Abstract:** By means of an eye tracker, build and visualize a 3D mental map of an observed scene. The map can be used to visualize data, accordingly to the strategies used to explore the scene and by exploiting space variant representation. Build psychophysical experiments to assess the validity of the model, in the context of data visualization.

(3)

**Title: Interaction techniques in Mixed Reality**

**Tutors:** Manuela Chessa, Fabio Solari

**Keywords:** Virtual Reality, Human-Computer Interaction

**Abstract:** Develop a mixed reality setup, where people can interact by using several interaction techniques (vision-based, controllers, VR gloves) with virtual objects characterized by different physical properties (textures, weight, material,...). Devise, design and implement experiments to understand the differences of the considered techniques.

(4)

**Title: Mixed Reality Environments for first aid simulation**

**Tutors:** Manuela Chessa, Fabio Solari

**Keywords:** Virtual Reality, Human-Computer Interaction, Computer Graphics

**Abstract:** Develop a Mixed Reality system, composed of both virtual environments and physical tools, to be used for the simulation and training of first-aid procedures. In particular, the aim of the thesis is to develop virtual scenarios to replicate specific real-world situations, tracking the most important elements of the real scene (e.g. objects and instruments) and maintaining the co-localization between real and virtual elements.

This thesis is within the joint lab DIBRIS-SIMAV.

(5)

**Title: Exergames based on Virtual Reality**

**Tutors:** Manuela Chessa, Fabio Solari

**Keywords:** Virtual Reality, Human-Computer Interaction

**Abstract:** Develop exergames based on Virtual Reality, both with immersive and not immersive techniques, to be used for the cognitive and physical assessment and rehabilitation. The exergames will adapt to the individual characteristics of each user and on his/her actual status and level of improvements, and they will replicate specific real-world activities. In particular, web-based techniques will be exploited to cope with the need of maintaining the physical distance especially with fragile people.

(6)

**Title: Virtual/augmented reality experiments and economic decision-making**

**Tutors:** Manuela Chessa, Michela Chessa (CNRS and University of Nice)

**Keywords:** Virtual Reality, Game Theory

Abstract:

This thesis aims at developing virtual reality and/or augmented reality environments and situations to test strategic and cooperative behaviors and attitude toward risk in economic decision-making.

The thesis will consist in: 1) designing and implementing a game, e.g. a shape-sorter game, which aims at reproducing a game-theoretic scenario 2) performing experiments with subjects 3) analyzing the experimental data

The thesis will be conducted in collaboration with the GREDEG Laboratory, CNRS at the LEEN (Laboratoire d'Économie Expérimentale de Nice).



## Thesis proposals in “Computer Vision”

(1)

**Title:** Video-based posture estimation at the desk

**Tutors:** Nicoletta Noceti, Francesca Odone, Gaurvi Goyal

**Keywords:** Computer Vision, Machine Learning, pose estimation

### Abstract

Wrong sitting posture may cause temporary or permanent damages to sedentary young people. The thesis will explore the use of pose estimation algorithms to estimate the pose of a user sitting in front of a computer, analysing the pose deviation with respect to a correct posture.

### References

<https://link.springer.com/article/10.1007/s00138-019-01033-9>

<https://arxiv.org/abs/1804.03142v1>

(2)

**Title:** Efficient real-time vehicle recognition system operating in a low resources regime

**Tutors:** Francesca Odone, Nicoletta Noceti, Giorgio Cantarini (in collaboration with Imavis srl)

**Keywords:** Computer Vision, Machine Learning, Deep learning, Object Detection, Few-Shot Learning

### Abstract:

In this thesis we will design and validate different techniques to the problem of vehicle recognition: from the detection of the vehicles and the plates, exploiting fast object-detection models such as SSD, to the classification of the color and the recognition of the model, using machine learning and few-shot learning methods; you will have the opportunity to study and test the state-of-the-art models with an eye to efficiency.

The work will be firstly evaluated on benchmark datasets and then to real data acquired from real systems. The whole study will take into account technical specifications obtained from real traffic monitoring systems.

### References:

<https://arxiv.org/abs/1702.01721>

<https://arxiv.org/pdf/1512.02325.pdf>

<https://openreview.net/pdf?id=HkxLXnAcFQ>

[https://ai.stanford.edu/~jkrause/cars/car\\_dataset.html](https://ai.stanford.edu/~jkrause/cars/car_dataset.html)

(3)

**Title: Implementation of reverse engineering methods for neural network interpretation for video analysis in medical applications**

**Tutor:** Vito Paolo Pastore, Francesca Odone

**Keywords:** Computer Vision, Machine Learning, Motion analysis

**Abstract**

Over the past decade, deep learning has demonstrated superior performances in solving many problems in various medical applications compared with other machine learning methods. Despite the high accuracy in prediction, neural networks are black boxes, with no information easily accessible to understand how the model actually processes the input (e.g., images, videos) to predict a specific output. The proposed work will investigate on algorithms to “unbox” a neural network, with specific focus on reverse engineering to provide interpretability to deep learning predictions. The designed interpretability algorithms will be applied to automatic assessment of epileptic seizures in videos and the analysis of general movements in premature infants.

(4)

**Title: Improving detection performances using motion recognition in surveillance scenarios**

**Tutors:** Nicoletta Noceti, Francesca Odone, Giorgio Cantarini

**Keywords:** Computer Vision, Machine Learning

**Abstract:**

Object detection has achieved remarkable performances, although in specific application domains it is still struggling for the intrinsic difference between the large dimensional benchmark datasets used to train the algorithms, and the data obtained in the field. Video-surveillance is one of such scenarios. The goal of the thesis is to combine a classical motion-based approach to video surveillance with CNN-based object detection. In particular the thesis will explore the possibility of deriving motion patterns to classify entities by the way they move, with the goal of solving detection ambiguities.

**References:**

<https://www.frontiersin.org/articles/10.3389/frobt.2017.00014/full>

(5)

**Title:** **Efficient Pose estimation**

**Tutors:** Francesca Odone, Giorgio Cantarini, Vito Paolo Pastore

**Keywords:** Computer Vision, Machine Learning, Deep learning

**Abstract:**

The proposed work consists in designing and validating neural network models for pose estimation with a specific focus on computational efficiency.

The thesis will explore two main fields that are getting more and more attention from the ML community: network pruning and optimal neural architecture search.

The first one aims at reducing the number of nodes and connections of a neural network (i.e., the number of parameters) to decrease the time needed for inference and the memory usage, with only a marginal or absent drop in accuracy. The optimal neural network search aims to find the best architecture (e.g. number of layers, type of convolution, etc.) for a given task (i.e., classification or regression), in order to maximize the prediction accuracy while minimizing the time needed for training and prediction.

The candidate will have the opportunity to study, implement and test state-of-the-art models and to practise the use of DL framework (such as Tensorflow, Keras,...).

**References:**

<https://arxiv.org/pdf/1812.08008v2.pdf>

<https://arxiv.org/pdf/2003.03033v1.pdf>

<https://arxiv.org/pdf/1707.07012v4.pdf>

<https://arxiv.org/pdf/1904.07850v2.pdf>

(6)

**Title:** **Graph-based Motion Analysis**

**Tutors:** Nicoletta Noceti, Francesca Odone

**Keywords:** Computer Vision, Machine Learning

**Abstract:**

Graph models provide an effective way of representing structured data, and in recent years their effectiveness in representing the human body has been well assessed. In this thesis we explore their use to the purpose of representing and understanding human actions, with particular attention to upper body movements, and leveraging deep learning methodologies, where their use is still largely unexplored.

**References**

<https://arxiv.org/abs/2003.08802>

(7)

**Title: Efficient methods for motion segmentation**

**Tutors:** Elena Nicora, Nicoletta Noceti

**Keywords:** Computer Vision, Motion detection, Deep Learning

**Abstract:**

Change Detection (CD) has a long history, and it has evolved alongside the areas of image processing and computer vision. CD is a well defined classification problem: it deals with the automatic detection of “significant” changes occurring in a scene by the elaboration of a video sequence. In this thesis we will investigate the combination of an efficient filtering strategy (based on the use of Grey-Code Kernels) and of Deep Learning methods for change detection in Human Motion Analysis

**References:**

<https://ieeexplore.ieee.org/document/4069256>

## Thesis proposals in “Computer Graphics”

(1)

**Title: Weighted Averages on Meshes**

**Keywords:** geometric meshes, surfaces, mapping, weighted averages.

**Tutors:** Prof. E. Puppo, C. Mancinelli.

**Abstract:**

Weighted averages provide mathematical tools for resolving a large number of problems in a variety of domains. In Computer Graphics, they are at the basis of the control of spline curves and surfaces, as well as texture mapping, shape morphing, definition of barycentric coordinates, etc. Computing weighted averages in a non-Euclidean space, such as the manifold corresponding to the surface of a solid object, is a difficult and expensive task, requiring geodesic computations. The purpose of this thesis is to implement an innovative efficient approach to the computation of weighted averages on surface meshes. The work starts from a basic layer of software already developed by the advisors and follows a clear path that has already been designed. The student is asked to possess good skills in C++ programming and basic notions of geometric data structures.

(2)

**Title: Gradient Fields Estimation on Meshes**

**Keywords:** geometric meshes, gradient fields estimation.

**Tutors:** Prof. E. Puppo, C. Mancinelli.

**Abstract:**

The estimation of the differential properties of a function sampled at the vertices of a discretized domain is at the basis of many applied sciences. The accuracy of such estimation heavily depends on the nature of the signal defined on the domain, and on the connectivity of the latter. In the literature, several algorithms that address such problem have been proposed, each one with pros and cons depending on the experimental setup. The purpose of the thesis is the implementation and experimentation of one or more of these methods, and, possibly, the design of a new method. The work starts from a basic layer of software already developed by the advisors. The student is asked to possess good skills in C++ programming and basic notions of geometric data structures.

(3)

**Title: Affine Connections on Meshes**

**Keywords:** geometric meshes, affine connection, discrete differential geometry.

Tutors: Prof. E. Puppo, C. Mancinelli.

Abstract:

Affine connections are mathematical tools that provide basic notions to define a metric on a manifold surface. This is necessary, for instance, to measure lengths, angles and areas, and develop differential geometry in the intrinsic domain of a manifold. There exist several approaches to the definition of an affine connection on a geometric mesh. The purpose of the thesis is the implementation and experimentation of at least one such method. The work starts from known methods in the literature and it is open to possible innovative approaches. Implementation shall be developed in the context of a geometric library providing all the necessary basic tools and data structures. The student is asked to possess basic notions of differential geometry of surfaces (e.g., from the course in Geometric Modeling), good skills in C++ programming and basic notions of geometric data structures.

## Thesis proposal in “Network Analysis”

**Title:** Graphlet-based representation for the understanding of multiplex networks

**Tutors:** Annalisa Barla, Davide Garbarino

**Keywords:** Network features, multiplex networks, Machine Learning

### Abstract:

The main aim of this thesis project is to find the best representation of complex (textual) information when limited prior knowledge is available. This project is composed of two parts: devise the optimal NLP+machine learning combination to infer binary network structures and define a graphlet-based metric to evaluate network similarity

The methods will be applied to real-world data - acquired ad hoc - consisting of large-scale websites (such as those of complex institutions or public administrations) or to legal corpus of laws.

### References:

Pržulj, Nataša. "Biological network comparison using graphlet degree distribution." *Bioinformatics* 23.2 (2007): e177-e183.

Berry, Michael W. "Survey of text mining." *Computing Reviews* 45.9 (2004): 548.

=====

## Thesis proposal in “Bioinformatics”

**Title:** ATAC-seq (Assay of Transposase Chromatine sequencing) data analysis project within a Multiple Sclerosis project

**Tutors:** Annalisa Barla, Margherita Squillario

**Topics:** Bioinformatics, Machine Learning, Sparsity, Immune System

### **Abstract**

This project consists in the development of a high-throughput data analysis pipeline (few samples and large number of variables) to understand the effect of the presence or absence of certain metabolites on the transcriptome of immune system cells

## Thesis proposals in “Computational and Statistical Learning”

(1)

Title: **Fast and accurate randomised algorithms for large scale machine learning**

Tutor: Lorenzo Rosasco

Keywords: Machine Learning, Large-Scale Data

Abstract:

This project aims at developing efficient algorithmic and software solutions to process massive amounts of data. The idea is to combine randomised algorithms with fast recursive approaches to obtain scalable solutions. The algorithmic efforts are complemented with the development of software libraries taking advantage of GPUs.

(2)

Title: **Machine learning for robotics**

Keywords: Machine Learning, Robotics

Abstract:

Perceptual and motor skills tasks in robotics can naturally be tackled with machine learning techniques.

In this setting, efficiency in terms of data and computations is of paramount importance, and a number of weak forms of supervision can be exploited for increased efficiency. Different projects are available, with the possibility to work closely with roboticists.

(3)

Title: **Machine learning for time series prediction with applications**

Keywords: Machine Learning, Time series

Abstract:

Time varying data are ubiquitous in applications and offer both opportunity and challenges for machine learning. Several projects are possible ranging from algorithms design to applications in weather or energy market prediction.

(4)

Title: **Machine learning for physics**

Keywords: Machine Learning, Physics

Abstract:

Machine learning is rapidly changing the way physics model are constructed, by replacing deterministic with statistical modelling. Several projects are possible, ranging from the analysis of high energy physics data, to using reinforcement learning to understand navigation.



## Thesis proposals in “Data Management”

(1)

**Title: Location inference from social media data**

**Tutors:** Giovanna Guerrini, Federico Dassereto, Michela Bertolotto (UCD, Dublin)

### Abstract:

The multifaceted nature of user-generated data, along with its geographic component, is nowadays being exploited to better understand social dynamics and propagation of information. Social media activities can be associated with both an explicit (e.g., in Twitter, metadata like the user profile location and the GPS coordinates of the device from which the activity is performed) and an implicit (i.e., inferred from data, with a variable degree of confidence) geographic information component. Unfortunately, explicit tagging is used only in a small percentage of tweets, due to the fact that location services of mobile devices are often disabled or switched off to save battery, thus inferring additional tweeting position is an important issue to make microblog analysis more effective.

In the geospatial context, the geographical component is no more enough to assess connections between elements. In particular, geographical objects can be aggregated by position, category, density, and many other dimensions.

Embeddings allow representing the content of geographical objects in terms of vectors in a high dimensional space. In such a space, the distance reflects the “semantic distance” that holds among objects. This novel representation opens the door to the integration of geospatial ontologies into machine learning algorithms. The aim of this project is on inferring microblog messages location (i.e., the position of the user when the tweet was sent) rather than the user’s home location, investigating the coherence between geographical objects and embeddings. We claim that using embeddings to retrieve the semantic closest terms allows us to geolocate at sub-city level microblog messages. To this aim, we rely on several information, including toponyms contained in texts, social relationships and interactions between users, and text content.

Along with applications, we work on the theoretical side to deeply understand embeddings and proposing alternative algorithms to embed data.

(2)

**Title: Representing Geospatial Ontologies via Embeddings**

**Tutors:** Giovanna Guerrini, Federico Dassereto, Michela Bertolotto (UCD, Dublin)

### Abstract:

Assessing similarity among data items becomes harder and harder, as the amount of semantics that it is daily added to the Web is continuously increasing. In terms

of machine understanding, processing and interpreting this information requires novel fast and scalable data representations.

In the geospatial context, the geographical component is no more enough to assess connections between elements. In particular, geographical objects can be aggregated by position, category, density, and many other dimensions.

Embeddings allow representing the content of geographical objects in terms of vectors in a high dimensional space. In such a space, the distance reflects the “semantic distance” that holds among objects. This novel representation opens the door to the integration of geospatial ontologies into machine learning algorithms. The aim of this project is to investigate the coherence between geographical objects and embeddings, with specific attention to real applications (e.g., microblogs geolocation). Along with applications, we work on the theoretical side to deeply understand embeddings and proposing alternative algorithms to embed data.

(3)

**Title: Augmenting Tables in Data Lakes**

**Tutor:** Giovanna Guerrini, Federico Dassereto

**Abstract:**

Querying data lakes containing a large amount of unstructured data is a difficult task, accentuated by the facility of producing data due to the rise of the World Wide Web. A traditional operation on data lakes is to search for joinable and unionable tables, to find tables related to specific user requests. Existing approaches search for columns with the best overlap possible, returning columns from tables in data lakes with high joining or unionability rating. These approaches fail to capture the situation in which a data scientist wants to find related tables related to a particular query table when she is trying to build a machine learning model to predict a column of her query table. The reason for which existing approaches fail is that they do not consider the impact that a joining table has on the predicting tasks, due to the lack of semantic surrounding the data lake. The goal of the thesis is to define an automatic table augmentation framework, to improve the performance of machine learning models in predicting target columns of a query table. The idea is to start by studying the augmentation problem as a problem on data lakes only, with measures and concepts derived from information theory and functional dependencies, and then integration embeddings in the process. Embeddings will be integrated by mapping tables to knowledge bases and exploiting their semantics to automatically produce the augmentation. More precisely, the thesis will first propose a framework for automatic horizontal table augmentation in data lakes, allowing a data scientist to add relevant features to her query table from a ranking of joining tables.

Furthermore, an extension comprehending a mapping of tables to knowledge bases and exploiting knowledge bases embeddings will be developed, to study the role of embeddings in automatic tables augmentation. Finally, to sum up, the previous results, an ad-hoc embedding algorithm for knowledge bases will be developed to emphasize the relevant properties a knowledge base should have to produce a relevant augmentation.

(4)

**Title:** **Ethic-by-design Data Preparation**

**Tutors:** Barbara Catania, Chiara Accinelli, Giovanna Guerrini

**Abstract:**

Nowadays, large-scale technologies for the management and the analysis of big data have a relevant and positive impact: they can improve people's lives, accelerate scientific discovery and innovation, and bring about positive societal change. At the same time, it becomes increasingly important to understand the nature of these impacts at the social level and to take responsibility for them, especially when they deal with human-related data.

Properties like fairness, diversity, serendipity, or coverage have been recently studied at the level of some specific data processing systems, like recommendation systems, as additional dimensions that complement basic accuracy measures with the goal of improving user satisfaction.

Due to the above-mentioned social relevance and to the fact that ethical need to take responsibility is also made mandatory by the recent General Data Protection Regulation of the European Union, nowadays the development of solutions satisfying - by design - non-discriminating requirements is currently one of the main challenges in data processing and is becoming increasingly crucial when dealing with any data processing stage, including data management stages. Our actual interest is to consider non-discriminating requirements, like fairness and diversity, during the query processing pipeline. The development of technological solutions satisfying such requirements is currently one of the main challenges in data management and it has been investigated for many front-end data processes (e.g., ranking, set selection, etc.). We are mainly interested in non-discriminating approaches for data preparation, focusing on back-end strategies: an unfair data preparation process might have a relevant impact on front-end analysis.

More precisely, the aim of our research is to design, implement, and evaluate ad hoc query processing techniques to automatically enforce specific beyond-accuracy properties, with a special reference to fairness and diversity, during the data preparation steps.

## Thesis proposals in “High-Quality Software Development”

(1)

**Title: Software Engineering for IoT Systems**

**Tutors:** Gianna Reggio, Maurizio Leotta

**Keywords:** IoT requirement capture & specification, IoT design, IoT testing

**Abstract:**

- Deriving acceptance test cases from IoT-Req requirement specifications
- Modelling methods for (designs of) IoT-systems
  - Not just for documentation, but also simulation, rapid prototyping, automatic generation of the implementation
  - For deriving IoT-unit test cases (i.e. test cases for edge/cloud-software, web/mobile apps composing the system)
  - Investigate the suitability of existing notations: UML, ThingML, Modelica and Node-Red as modelling notation

(2)

**Title: Frameworks & Software Engineering**

**Tutors:** Gianna Reggio, Maurizio Leotta, Filippo Ricca

**Keywords:** Software engineering, framework

**Abstract:**

- SE for software development based on frameworks
  - Which are the problems, if any ? What is needed?
  - Should application frameworks and software frameworks considered differently?
- SE for developing frameworks
  - Which are the problems, if any ? What is needed?
  - What is the role of the supporting tools (e.g. wizard) ?
  - Framework instantiations using ad hoc languages/notations versus model driven development

(3)

**Title: UML Modelling to Improve Business Process Maturity**

**Tutor:** Gianna Reggio

**Keywords:** business process modelling, UML

**Abstract:**

- Modelling the UNIGE business processes (with Prof. Annalisa Barla)
  - ongoing initiative (UniGE federate website initiative)

- introducing the use of UML
- exploiting business process models for further goals/uses (e.g. (semi) automatic documentation generation)
- Modelling the DIBRIS business processes, a micro enterprise
  - cooperative effort to discover the hidden processes and to model them
  - analysing the effects

A lot of interactions with Italian-speaking domain experts is required.

(4)

Title: **Easy/Smart Beginner Programming Lab**

Tutor: Gianna Reggio

Keywords: smart system development, e-learning

Abstract:

Providing a smart software support to run beginner programming lab course.

Many different needs:

- **Instructors:** managing the assignments, partly automatizing the students' submissions correction, smart analysis of students' submissions (e.g. to detect plagiarisms, discovering misunderstood concepts)
- **Students:** simple ad hoc IDE covering only the basic concepts, with smart detections and explanations of errors (at compile and run time, bad smells and style issues defined by the instructors)
- **Both:** high usability, should run on the mainstream platforms (Windows, Linux and MacOS)

(5)

Title: **Smart Cities**

Tutors: Maura Cerioli, Marina Ribaudò

Keywords: participation, blockchain, mobile

Abstract:

Projects helping organizations, PA to become smart

- promoting usage of state of the art technologies
- enhancing the participation of administered people

Examples:

- ZenaClean, a system allowing citizens to notify problems, with a reward system powered by Ethereum
- Open source alternative to Municipium (an application to support all aspects of management for public administration)

(6)

**Title: Innovative Microsoft technologies from theory to practice**

**Tutors:** Maura Cerioli, Marina Ribaudo, Maurizio Leotta, Gianna Reggio, in collaboration with **Edisoftware**

**Keywords:** .Net, Azure, Blazor, Cloud, Containers, Dockers, Web

**Abstract:**

Exploratory projects to understand the industrial potential of some of the most recently introduced Microsoft technologies.

Examples:

- Workload tests with containers in MS Azure Cloud
- Development web application using MS-Blazor technology
- Microservice architecture based on cloud serverless functions

(7)

**Title: Programming as a Card Game**

**Tutors:** Maura Cerioli, Gianna Reggio

**Keywords:** gamification, user experience, e-learning

**Abstract:**

Learning how to program is difficult and often frustrating.

Key idea: a gaming platform where beginners can play programming poker :

- each card represents a statement/declaration
- players use them to solve given problems
- the first to complete the program wins ...but the others may challenge the solution

Main challenges:

- devising the game rules
- implementing an enticing game platform

## Thesis proposal in “Software Testing”

“Expose and remove defects as early as possible”.

The increasing complexity of software systems requires automated testing approaches. Our research group aims at devising automated testing approaches and tool to reach an important goal of software engineers: testing more and better.

We focus on the following kind of software systems:

- Web
- Mobile
- IoT
- Machine Learning

(1)

**Title: Testing Framework Evaluation**

**Tutors:** Maura Cerioli, Maurizio Leotta, Filippo Ricca

**Keywords:** testing, QA, assertions, empirical SE

**Abstract:**

Projects and experiments to compare usability of different testing frameworks

- analysis of current programming practice
- comparisons w.r.t. understandability, efficacy, productivity...
- automatic analysis of repos and code
- experiments with human subjects and statistical analysis

(2)

**Title: E2E Software Testing**

**Tutors:** Maurizio Leotta, Filippo Ricca

**Keywords:** testing, QA, e2e testing, empirical SE, Web, Mobile

**Abstract:**

- E2E Test Scripts Generation for Web / Mobile applications
  - Improving approaches based on a Navigational Model of the Applications (paths gen)
  - Understanding how to automatically generate effective Assertions (able to detect bugs)
- E2E Test Dependencies Detection for Web / Mobile applications
  - Improving existing approaches for detecting dependencies among automated test cases

- Understanding how to minimize the whole test suite execution time (i.e., the number of executed tests) while, at the same time, maintaining a high coverage of the SUT
- How AI/ML can improve the Effectiveness of Test Automation
  - generating E2E tests by learning behavioral patterns of the application users
  - making tests more robust to app changes by learning which are the most reliable features (hooks)



## Thesis proposals in collaboration with external companies

The four proposals below are the result of very recent interactions with the two involved companies, **SPX Lab** (<https://www.spxlab.com/>) and **Ovunque s.r.l.** (<http://www.ovunque-si.it/comuni/noi.php>), and for this reason they are still incomplete: the full and more focused proposals are under development. Nevertheless, the social impact of some themes and the exciting novelty and multi-disciplinarity of others, pushed us to share them, even in draft form.

(1)

**Title: Exploiting Natural Language Processing and Machine Learning for Risk Assessment of Gender-Based Violence**

**Key-words:** Machine Learning, Natural Language Processing, Gender-Based Violence

**Tutors:** Alessandra Giorgi (external, SPX); Viviana Mascardi; Francesca Odone

**Abstract:** The thesis will be inserted in the SPX Social Innovation thread of activities, consisting in the design and development of an algorithm for the risk factor assessment of gender-based violence. The algorithm will compute the risk factor based on the answers to a thirty questions test. The results of the algorithms based on the answers to the test will be compared with the evaluation of the same answers made by professionals from anti-violence centers; the data by professionals will be re-entered into the system as a feedback to train the algorithm and to make it more and more efficient over time. The thirty questions test is currently available in Italian, and a version in other languages is foreseen. The activity includes collaboration within an interdisciplinary team and on the field experimentation.

External links:

[www.savethewoman.com](http://www.savethewoman.com)

<https://vimeo.com/388797807/fe21b57cc7>

(2)

**Title: The Intelligent Avatar in the Mirror**

**Key-words:** Virtual and Mixed Reality, Cognitive Agents, Human-Computer Interaction

**Tutors:** Alessandra Giorgi (external, SPX); Massimiliano Margarone (external, SPX); Viviana Mascardi; Manuela Chessa

**Abstract:** Mirò is a mirror whose reflection capabilities are enhanced by technology: besides reflecting the image, it can augment it with virtual elements;

also, thanks to a camera hidden in the frame, details of the reflected image can be zoomed in and manipulated in different ways. So far, Mirò has been conceived by SPX as an artwork for retail, entertainment and personal care sectors. If equipped with cognitive and natural language abilities, it might also be exploited in healthcare for rehabilitation, in cultural heritage preservation, in training and education. The technology exploited by Mirò is QT (<https://www.qt.io/>).

**We propose to extend Mirò in three different ways, by adding**

- Cognitive capabilities to it according to the "Theory of Mind", namely the ability to attribute mental states -- beliefs, intents, desires, emotions, knowledge, etc. -- to oneself and to others [PW78]  
(thesis related with the contents of the MAS course);
- Natural language interaction capabilities to it, with a special attention to the proper understanding of the user's sentiments and emotions thanks to the use of ontologies of sentiments, and with support to multi-language interaction (thesis related with the contents of the NLP course);
- Interaction with an avatar, which could be a virtual companion, e.g. a storyteller for museum applications or a personal trainer for physical training and/or rehabilitation. The avatar could also be a virtual representation of the user him/herself with possible applications for both fashion (e.g. a magic mirror for testing new dresses or hairstyle) and psychological training (e.g. self-perception)

All the extensions will be carefully engineered to ensure them to perform properly even with limited computational resources. No degradation of the performances should be experienced by the user.

External links:

[www.mirointerior.com](http://www.mirointerior.com)

<https://vimeo.com/388803192/394d9873bd>

References:

[PW78] D. Premack, G. Woodruff, Does the chimpanzee have a theory of mind?, Behavioral and Brain Sciences 1 (1978) 515–526.  
doi:10.1017/S0140525X00076512

(3) / (4)

**Title: Integration of a Chatbot and of a Gamification module into the "I Love Guido" platform**

Key-words: Natural Language Processing, Conversational Agent, Gamification, Virtual Touristic Guides, Augmented Reality

Tutors: Mirko Melis & Ombretta Pavarino (external: Ovunque s.r.l.), Viviana Mascardi, Manuela Chessa

Abstract:

The aim of these theses is

1. To integrate a conversational interface into the "I Love Guido" platform, available at [www.iloveguido.it](http://www.iloveguido.it), that helps tourists in finding a guided tour led by a professional guide. By sharing the tour with other participants, all the participants have an economic gain, and share the fun. The conversational interface should be integrated into the search module of the platform, in order to help tourists in improving their search via natural language interaction. In addition, the conversational interface should access a semantic categorization of the contents (i.e. guided tours description) and, in case, of the related blog site.
2. To increase the tourist engagement by adding a gamification extension to I Love Guido based on conversational agents and augmented reality. Engaging stories will be designed and the platform will propose, in predefined points of the tour, the visualization of images consisting in clues useful to the tourist to collect elements for his character's growth, eventually to conquer the "treasure" that will lead the user to the next level of the game, or to win some bonus.

External links:

[www.iloveguido.it](http://www.iloveguido.it)

(5)

Title: **Personalized diet for weight management in patients**

Tutors: Professors Francesca Odone, Alessandro Verri (UniGe) and Dr Laura Toni (UCL), in collaboration with **Camelot Biomedical Systems**

Abstract:

Weight management cover an important role in cancer patients. For example, in breast cancer patients, pre- diagnosis and post-diagnosis obesity is associated with higher breast cancer-specific and overall mortality, while overweight is associated with higher overall mortality. The long-term goal of this project is to optimize patients' food intake and nutritional status, involving them in a dietary real-life self-monitoring process with nutritional "real world" data.

This project is aimed at posing the first steps toward this long-term goal direction. Specifically, the first step is to design a computer vision based architecture able

to identify the food intake from patients images. The project will start with existing open dataset (e.g., [1]) and implement and validate a food-recognition platform. Should dataset would be limiting, the second task of the student will be to augment existing dataset, developing techniques of webscraping that are able to build a rich enough annotated dataset. Finally, should the time allows it, data from the food intake analysis from patients, enriched by immunological data will be analyzed to infer possible patterns in real-world data.

References:

[1] <https://github.com/phiresky/pic2kcal>

(6)

Title: **Thesis opportunities with Esaote**

Keywords: Machine Learning, Image Processing, medical image analysis

Internal Supervisor: Francesca Odone

Abstract:

Esaote is one of the world's leading producers of medical diagnostic systems. The projects below are proposed by our ultrasounds R&D department with the common purpose of improving the user experience. An ultrasound exam is a challenging task due to the simultaneous interaction with patient, device and probes. Poor signal-to-noise ratio inherently associated with ultrasound imaging makes the task ever harder. An improvement of device usability allows the physician to focus on image interpretation and diagnosis, instead of ultrasound system correct usage.

The projects we propose are:

- **Machine learning methods for standard scanplanes detection in images and videos** to help the physician to follow all the required steps to fully analyze the area of interest and to comply to standard procedures ([luca.zini@esaote.com](mailto:luca.zini@esaote.com))
- **Automatic acquisition and refinement of 3D volumes** ([luca.zini@esaote.com](mailto:luca.zini@esaote.com))
- **Ultrasound image denoising and enhancement with traditional image processing or deep learning methods** ([paolo.nicolardi@esaote.com](mailto:paolo.nicolardi@esaote.com))

(7)

Title: **Anomaly detection in image analysis in the visible spectrum and thermographies**

Tutors: Francesca Odone, Alessandro Verri, in collaboration with **AlgoWatt spa**

Keywords: Image Processing, Machine Learning

Abstract:

The thesis addresses the problem of detecting anomalies in image sequences acquired by a drone flying over photovoltaic (PV) plants. The main objectives of the work will include single panel detection in PV plants and functional anomalies detection in each panel.

The activity will include data gathering and data cleaning and a feasibility study which will compare different approaches based on image processing and machine learning.