



**Software and Service Engineering Group**



Database Technologies and Information Management Group

# **TFG-TFM proposals**

<https://gessi.upc.edu/en/students/pizza-day/propostestfgs-pizzaday2018.pdf>

# GESI Pizza Day 2018

The GESI<sup>1</sup> and DTIM<sup>2</sup> research groups of the Universitat Politècnica de Catalunya (UPC) conduct research in many fields of software engineering and data and knowledge management.

Currently, we have several tasks to carry on (see description below) related to the following subjects:

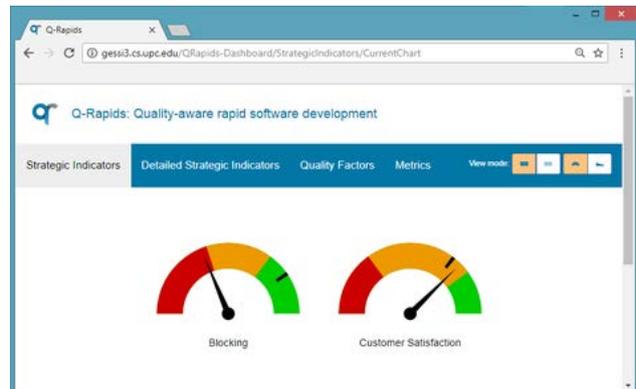
- Quality-aware Software Development
- Requirements Recommendation and Reuse
- Smart APIs Generation
- Data Analytics
- Others

If you are interested to collaborate in them, just contact us. Also, feel free to propose us any other topic you may be interested in.

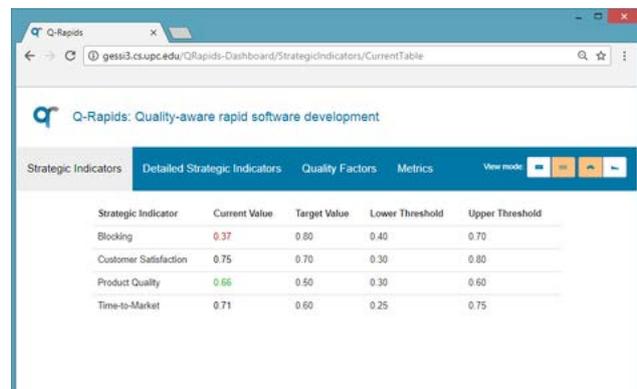
## Quality-aware Software Development

**Title: Developing compatibility features for a strategic dashboard**

**Description:** Q-Rapids is a research and innovation project funded by the European Commission (H2020) coordinated by the GESI group at the UPC. The consortium involves universities and IT companies. Companies are: Nokia, Bittium (Finland), Softeam (France) and iTTi (Poland). The project aims at providing tool support to decision-makers for quality management in agile software development processes. This tool support includes the construction of a strategic dashboard to be presented to decision-makers. This strategic dashboard needs to offer a clear view of several strategic indicators like time to market, product quality, expected revenue, risk, team performance, etc.



The dashboard should include some **export features** to allow decision-makers using the dashboard results in their internal meetings and reports. The work would include **exporting the textual data (e.g. excel files)** or **graphically (e.g. power point, jpg)** or producing **customisable reports (e.g. pdf)**. As part of the TFG, the student should analyse some existing tools that provide report facilities to be integrated in the dashboard (e.g. **BIRT, JasperSoft Community, ...**). We need to use existing technologies, therefore the work will also include the analysis of existing



The screenshot shows a web browser window displaying the Q-Rapids dashboard. The dashboard has a navigation bar with tabs for 'Strategic Indicators', 'Detailed Strategic Indicators', 'Quality Factors', and 'Metrics'. Below the navigation bar, there is a table with the following data:

Strategic Indicator	Current Value	Target Value	Lower Threshold	Upper Threshold
Blocking	0.37	0.80	0.40	0.70
Customer Satisfaction	0.75	0.70	0.30	0.80
Product Quality	0.66	0.50	0.30	0.60
Time-to-Market	0.71	0.60	0.25	0.75

<sup>1</sup> <https://gessi.upc.edu/en>

<sup>2</sup> <http://www.essi.upc.edu/dtim/>

dashboards and or specific visualisation techniques, synthesising their pros and cons.

**Degree:** GEI

**Requirements:** Software Engineering specialization of GEI

**Technologies:** java, Spring, AngularJS, postgresSQL. The Q-Rapids Dashboard is a web application

**Contact:** Lidia López i Xavier Franch

**e-mail:** llopez@essi.upc.edu, franch@essi.upc.edu

**Title: Extending a Strategic Dashboard to compute Strategic Indicators using Bayesian Networks**

**Description:** Q-Rapids is a research and innovation project funded by the European Commission (H2020) coordinated by the GESSI group at the UPC. The consortium involves universities and IT companies. Companies are: Nokia, Bittium (Finland), Softeam (France) and iTTi (Poland). The project aims at providing tool support to decision-makers for quality management in agile software development processes. This tool support includes the construction of a strategic dashboard to be presented to decision-makers.

This strategic dashboard needs to offer a clear view of several strategic indicators like time to market, product quality, expected revenue, risk, team performance, etc. The student should extend the strategic dashboard to include a functionality to introduce the information required to create Bayesian networks that will be used to compute strategic indicators.

**Degree:** GEI

**Requirements:** Software Engineering specialization of GEI

**Technologies:** java, Spring, AngularJS, postgresSQL. The Q-Rapids Dashboard is a web application.

**Contact:** Martí Manzano i Cristina Gómez

**e-mail:** mmanzano@essi.upc.edu, cristina@essi.upc.edu

**Title: Developing Quality Alerts management system for a strategic dashboard**

**Description:** Q-Rapids is a research and innovation project funded by the European Commission (H2020) coordinated by the GESSI group at the UPC. The consortium involves universities and IT companies. Companies are: Nokia, Bittium (Finland), Softeam (France) and iTTi (Poland). The project aims at providing tool support to decision-makers for quality management in agile software development processes. This tool support includes the construction of a strategic dashboard to be presented to decision-makers.

This strategic dashboard needs to handle the reception, filtering and management of Quality Alerts (Quality Alerts are triggered when the quality of an aspect of the software is below a specific threshold). As part of the TFG, the student should **design and develop** the part of the dashboard related to Quality Alerts.

**Degree:** GEI

**Requirements:** Software Engineering Specialization of GEI

**Technologies:** Spring, Java, postgresSQL, AngularJS

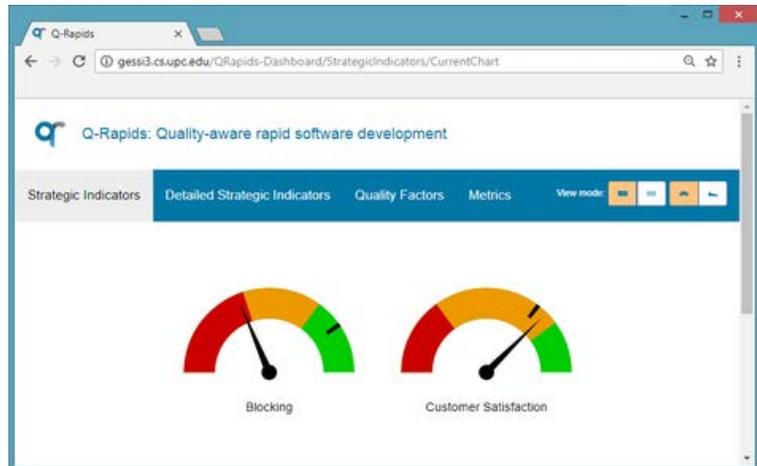
**Contact:** Marc Oriol, Carles Farré

**e-mail:** moriol@essi.upc.edu, farre@essi.upc.edu

## Title: Integrating Quality Requirements Catalogue in an strategic dashboard

**Description:** Q-Rapids is a research and innovation project funded by the European Commission (H2020) coordinated by the GESSI group at the UPC. The consortium involves universities and IT companies. Companies are: Nokia, Bittium (Finland), Softeam (France) and iTTi (Poland). The project aims at providing tool support to decision-makers for quality management in agile software development processes. This tool support includes the construction of a strategic dashboard to be presented to decision-makers. This strategic dashboard needs to offer a clear view of several strategic indicators like time to market, product quality, expected revenue, risk, team performance, etc.

One of the features of the dashboard requires the use of a **Quality Requirement Patterns catalogue**. We need to include the adequate views to allow the user to navigate through this catalogue in order to use the patterns to analyse the impact of including concrete requirements in the strategic indicators. As part of the



TFG, the student should analyse the needs for this feature and design and develop the appropriated views to be integrated in the dashboard web application.

The following images (up) you can see the current layout of the Q-Rapids Strategic Dashboard and a screenshot of a PABRE-PROJ (down) tool to give an idea on what kind of view can be designed (the tree view on the left, titled Patterns) .

**Degree:** GEI

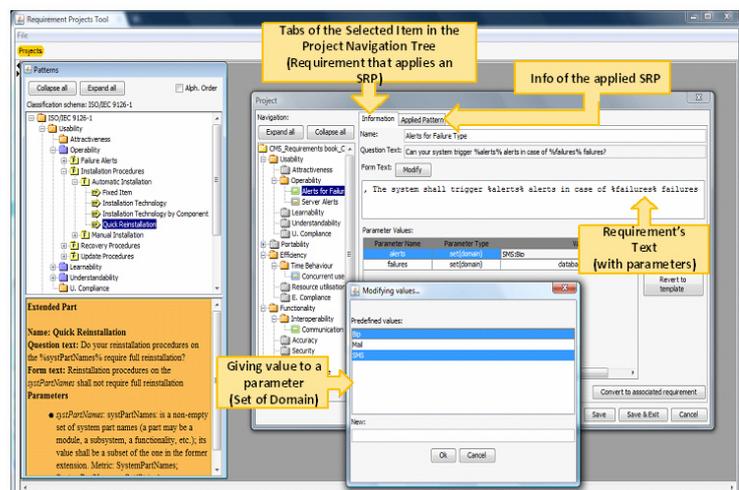
**Requirements:** Software Engineering specialization of GEI

**Technologies:** java, Spring, AngularJS.

The Q-Rapids Dashboard is a web application

**Contact:** Lidia López i Xavier Franch

**e-mail:** llopez@essi.upc.edu, franch@essi.upc.edu



## Requirements Recommendation and Reuse

### **Title: Automatic Improvement of Requirements Quality**

**Description:** One of the goals of the OpenReq (<http://openreq.eu/>) European project is to help users to write better requirements by automatically suggesting tips to the users on how they can improve their requirements. The goal of this project is implement some of the tips that can help to improve such quality, such as:

- Checking the conformance to templates of natural language requirements
- Suggesting to the user that some information is missing in its requirements (e.g., you are writing a bug in an open source project tool, but you have not entered the data to reproduce such bug)
- Checking the use of the terms defined in a glossary by the user

**Degree:** GEI, MIA

**Requirements:** Computing or Software Engineering specialization of GEI

**Technologies:** Java, Spring, MySQL, OpenNLP, DKPro, Lucene

**Contact:** Cristina Palomares

**e-mail:** [cpalomares@essi.upc.edu](mailto:cpalomares@essi.upc.edu)

### **Title: Recommender systems for Requirements Engineering**

**Description:** As part of the OpenReq (<http://openreq.eu/>) European project, we aim to apply recommendation techniques (as the ones used in Amazon or Netflix) to the Requirements Engineering field. The goal of this project is to predict some properties of requirements (e.g., priority or risk) by matching requirements at hand with those that have already been defined (in the same or previous requirements projects). To do the matching of the requirement at hand with already defined requirements, we think that clustering/classification algorithms might help to do it. Different components (e.g., Weka, Mahout, TensorFlow) and algorithms (e.g., Naïve Bayes, Neural Networks) for clustering and classification already exist. In addition, some algorithms might need certain parameterizations (e.g., batch size for Naïve Bayes).

**Degree:** GEI, MIA

**Requirements:** Computing or Software Engineering specialization of GEI

**Technologies:** Java, Spring, MySQL, Weka, Mahout, TensorFlow

**Contact:** Cristina Palomares

**e-mail:** [cpalomares@essi.upc.edu](mailto:cpalomares@essi.upc.edu)

### **Title: Recommendation of Users for Requirements**

**Description:** As part of the OpenReq (<http://openreq.eu/>) European project, we aim to apply recommendation techniques (as the ones used in Amazon or Netflix) to the Requirements Engineering (RE) field. The success of a software project crucially depends on identifying the “right” stakeholders who should cooperate in the RE process. This is even more crucial when thousands of requirements and stakeholders are involved in the process (think for instance of huge open source projects such as Eclipse or Qt). The goal if this TFM/TFG is to construct an API containing services

that are able to create and update the profile of a user based on the observed context, and the recommendation of users to contribute to a specific requirement. The idea is to construct a profile of a user (e.g., interests and availability) by analysing the history of the user in the OpenReq system (e.g., what kind of projects and requirements he has visited and contributed in the past, the projects and requirements he has visited and contributed recently, etc.). Then, with that profile, users will be recommended for an specific requirement (e.g., by extracting the topics that a requirement is about and matching them with the profile of the users).

**Degree:** GEI, MIA

**Requirements:** Computing or Software Engineering specialization of GEI

**Technologies:** Java, Spring, MySQL, Weka, Mahout, TensorFlow

**Contact:** Cristina Palomares

**e-mail:** cpalomares@essi.upc.edu

#### **Title: Context Analysis to Create a User Profile**

**Description:** As part of the OpenReq (<http://openreq.eu/>) European project, we aim to apply recommendation techniques (as the ones used in Amazon or Netflix) to the Requirements Engineering field. The goal of this project to construct the profile of a user of the OpenReq system by analysing the context. The idea is to construct a profile of a user (e.g., interests and availability) by analysing the history of the user in the system (e.g., what kind of projects and requirements he has visited and contributed in the past, the projects and requirements he has visited and contributed recently, etc.).

**Degree:** GEI, MIA

**Requirements:** Computing or Software Engineering specialization of GEI

**Technologies:** Java, Spring, MySQL, Restful

**Contact:** Cristina Palomares

**e-mail:** cpalomares@essi.upc.edu

#### **Title: Software Requirement Patterns Management and Search**

**Description:** The PABRE system (<http://www.upc.edu/gessi/PABRE/Tools.html>) allows the definition and management of software requirement patterns stored in a Relational Database Management System (RDBMS). However, software requirement patterns have conditions that do not require some properties that such type of database system has, and other that maybe could be provided in a better way for a NoSQL system. In this project the idea is to explore the advantages of having the software requirement patterns catalogue in a NoSQL system.

**Degree:** GEI

**Requirements:** Software Engineering specialization

**Technologies:** Elasticsearch, REST web services, Javascript, AngularJS

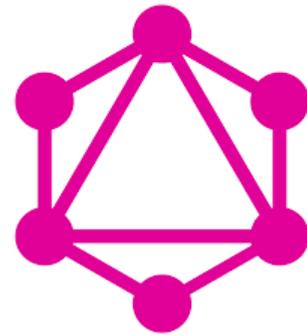
**Contact:** Carme Quer

**e-mail:** cquer@essi.upc.edu

## Smart APIs Generation

### Title: Interactive tool to generate GraphQL & REST APIs

**Description:** GENESIS (GENERation and Evolution of Smart APIs) is a project funded by the National Spanish Program for Research Aimed at the Challenges of Society 2016. GENESIS aims to provide techniques and tools to collect information, metadata, and behavior of the functionalities to be exposed; generate a repository of monitorable APIs with their documentation, metadata and implementation; automatically manage data flows and data repositories to implement the APIs; propose and evaluate API evolutions; and facilitate the automatic evolution of APIs. The main goal of this project is to develop an interactive tool to help the generation of GraphQL and REST APIs that provide an integrated, uniform and transparent access to a set of heterogeneous data sources and services.



**Degree:** GEI

**Requirements:** Software Engineering Specialization

**Technologies:** GraphQL, REST, JavaScript, Java, RDF, SPARQL

**Contact:** Carles Farré

**e-mail:** farre@essi.upc.edu

### Title: APIs Monitoring

**Description:** The market of Application Programming Interfaces (APIs) has experienced a significant usage growth in the last recent years. However, current techniques for the maintenance and evolution of these APIs are still done ad-hoc, with little automated support. For example, understanding how the users interact with the APIs, the Quality of Service (QoS), and how the APIs have evolved so far, may help developers in their maintenance and evolution activities.



This project consists on developing a second version of a monitoring system for RESTful APIs to gather (1) the Quality of Service (QoS), (2) the interaction of users with the API, and (3) changes in the evolution of the API. Such monitoring system will be developed using RESTful service technologies and will include a web-interface to manage the monitoring system and show the monitoring results.

**Degree:** GEI

**Requirements:** Software Engineering specialization

**Technologies:** RESTful services, JSON, Java, Javascript, web-technologies

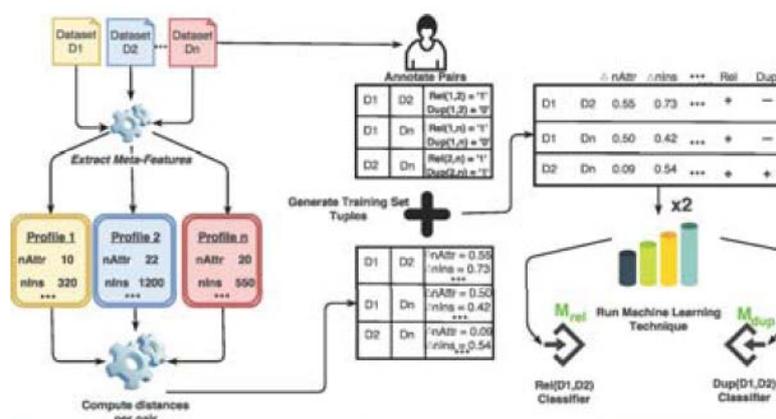
**Contact:** Marc Oriol, Jordi Marco

**e-mail:** moriol@essi.upc.edu, jmarco@cs.upc.edu

## Data Analytics

### Title: Unsupervised Clustering for Efficient Schema Matching

**Description:** With the emergence of Big Data stored in large-scale data repositories consisting of raw data in their original structures, commonly called Data Lakes (DL), there is a need to understand the information stored in such data. The goal of this research project is to investigate effective and efficient data



profiling techniques which can support in understanding the information inside datasets (mainly CSV and relational DB data, possibly JSON and XML). This includes the following activities:

- Automatic extraction of schema metadata using profiling techniques, including attribute data types (e.g. numeric or nominal attribute types), attribute names, and attribute data distributions.
- Applying unsupervised clustering data mining techniques to group related schemata into coherent clusters.
- Using the profiling metadata to improve the efficiency of schema matching which aims to find common information between different schemata.
- Using clustering techniques for near-duplicate schemata and outliers detection.

**Degree:** Bachelor project

**Requirements:** English-language thesis / graduation project

**Technologies:** Java, SQL, familiar with clustering techniques (e.g. k-means)

**Contact:** Ayman Alserafi

**e-mail:** alserafi@essi.upc.edu

### Title: Efficient Schema Matching using Profiling Techniques

**Description:** With the emergence of Big Data stored in large-scale data repositories consisting of raw data in their original structures, commonly called Data Lakes (DL), there is a need to understand the information stored in such data. The goal of this research project is to investigate effective and efficient data profiling techniques which can support in understanding the information inside datasets (mainly CSV and relational DB data, possibly JSON and XML). This includes the following activities:



- Automatic extraction of schema metadata using profiling techniques, including attribute data types (e.g. numeric or nominal attribute types), attribute names, and attribute data distributions.
- Designing efficient and effective algorithms which are capable of profiling big multi-structured datasets including: extraction of structure, computing statistical data distributions of attributes stored in the datasets, and matching the data values stored in attributes.
- Implementing similarity measures between different datasets based on syntactic measures (string distances), TF-IDF similarity scores and semantic measures (term dictionaries and thesaurus like the WordNet API) using the extracted profiles from the previous step

**Degree:** Master's degree project

**Requirements:** English-language master thesis project, familiar with basic statistics (e.g. simple data distribution analysis)

**Technologies:** Java, SQL, can work with online APIs, familiar with data mining techniques and tools (e.g. WEKA)

**Contact:** Ayman Alserafi

**e-mail:** alserafi@essi.upc.edu

## Other Subjects

**Title:** Comparison of different optimization algorithms for Release Planning

**Description:** Software Release Planning (SRP) solves the problem of finding the best combination of features or requirements to implement in a sequence of releases. SRP seeks to maximize business value and stakeholder satisfaction without neglecting the constraints imposed by the availability of adequate resources and the existence of dependencies between features, among other constraints. Many strategies and algorithms have been proposed in the literature to tackle this NP-hard problem. In this project, we want to compare our current genetic-algorithm implementation with other approaches such as linear programming and branch & bound.



**Degree:** GEI

**Requirements:** Software Engineering or Computer Science specialization

**Technologies:** Java, CPLEX

**Contact:** Carles Farré

**e-mail:** farre@essi.upc.edu

**Title:** Detecting dependencies in Agile software development

**Description:** The DATAMASE (<https://www.ca.com/us/company/innovation/research-and-development/ca-strategic-research/projects/datamase.html>) project is an industrial research collaboration between GESSI UPC and a global software company CA Technologies. DATAMASE is focused on exploring data-driven mechanisms to improve Agile software



development processes in large organizations. By looking at Agile software development projects over time and by using Artificial Intelligence to analyze historical development data (such as backlogs and the behavior of the different stakeholders involved in the software development process), we plan to do deeper process optimization to improve the cost effectiveness and speed of the development teams. Therefore, our goal is to identify patterns, exhibited by teams, and analyze their impact on work performance and team collaboration.

In this project we leverage telemetric data from an Application Lifecycle Management (ALM) tool, which captures the time-based nature of user actions — this has rarely been modelled by researchers in the past. Combined with information from other data sources, we hope to discover the user's high level behaviour (i.e. structured navigation) by exploring and interpreting sequences of low level interaction events.

One of the goals of the project is to transfer event-driven data to a process mining model that would specify unique characteristics of the trace and identify long-term dependencies. By analyzing traces of an end-to-end execution of a process (executed by a user) we hope to predict users' future steps (based on groups they belong to) and identify implicit dependencies between tasks. We plan to employ neural networks, Bayesian networks, relational dependency networks or other machine learning methods to model dependencies between different factors (e.g. sprint progress or sprint planning).

**Degree:** GEI

**Requirements:** Software Engineering or Computer Science specialization

**Technologies:** Python, PyTorch, Keras, Spark SQL, Pandas, NumPy, Scipy, Matplotlib, REST web services

**Contact:** Katarzyna Biesialska, Xavier Franch

**e-mail:** katarzyna.biesialska@ca.com, franch@essi.upc.edu

### **Title: Tool support for documenting NFRs in Agile**

**Description:** The GESSI group has developed a proposal for documenting non-functional requirements (NFRs) in Agile Software Development (ASD). The variability of NFRs both in scope and detail makes it necessary to provide several representation artefacts to cope with all of them. Therefore the proposal uses epic, user stories or user stories acceptance criteria to represent NFRs and provides guidelines on how to select the most adequate representation and how to use it depending both on the scope and detail of the NFRs.



In order to provide tool support for documenting NFRs in ASD, it is necessary to select an adequate requirements tool (Jira, etc) and extend/configure it to give support to the proposal.

**Degree:** GEI, MEI

**Requirements:** Requirements Engineering, Software Engineering

**Technologies:** To determine as part of the project (Java, XML...)

**Contact:** Dolores Costal, Xavier Franch

**e-mail:** dolores@essi.upc.edu, franch@essi.upc.edu