Decision Support System for IST Degree Coordination

Gonçalo Lopes[1]

Instituto Superior Técnico, Lisbon, Portugal
goncalo.paulos.lopes@tecnico.ulisboa.pt

Abstract

The coordinator of a higher education degree must make strategic decisions to ensure its quality. The decision-making of the coordinator needs to be supported by data relating to the main areas of the degree. In many cases, the data is collected and processed manually, which requires time and effort and is prone to human error. We propose a decision support system that automatically performs the tasks of collecting and storing data in a format suitable for analysis. The data is used to create dashboards, that indicate the performance a degree's different areas, ultimately providing useful insights that will enhance the decision-making of the coordinator.

Contents

1 Introduction

A degree coordinator's decision-making impacts all parties involved in the degree, mainly on students and instructors. Therefore, it is fundamental for the coordinator to have access to the correct data on how the degree works, to understand which areas need to be improved.

Currently, at *Instituto Superior Técnico*[1] (IST), collecting relevant data about students and their performance is mostly a manual, laborious and time consuming process, that the degree coordinators have to perform at the end of each semester.

Particularly, the coordinator of the Degree in Computer Science and Engineering (*Licenciatura em Engenharia Informática e de Computadores - Taguspark*, LEIC-T[2] in portuguese) organizes the data collected in Microsoft Excel files, which he uses to obtain relevant data through the application of various formulas. This data is used to generate semiannual reports, describing the overall performance of students and courses.

Given the repetitive nature of gathering and processing data, and its reliance on manual human labor, these tasks are prone to error. Their automation would mitigate the error rate and increase the overall efficiency, while providing the coordinator with accurate data for decision-making. As such, we propose a a software application, known as *IST Degree Coordination Decision Support System* (*Sistema de Apoio à Decisão para Coordenação de Curso do IST*, SAD-CCIST in portuguese), to automatically gather, structure and store data for analytical purposes.

The main contributions of this work are (i) designing and implementing a data warehouse, guided by Kimball's design methodologies, to store data related to the performance of the LEIC degree, (ii) designing and implementing extract-transform-load processes, that automate the current manual process of obtaining and processing data into a suitable format and (iii) designing and implementing three interactive dashboards, for student activity, student graduation and course related metrics, used for enhancing the LEIC-T Coordinator's decision-making.

2 Basic Concepts

This section describes concepts related to higher education, applicable to IST, focusing on student activity and performance.

At the beginning of an academic year[3], students apply for a degree. A student that is accepted by the university is called an *admitted student*. The set of admitted students of a given degree and academic year, that enrolled to all the first semester courses are called a *student generation*.

An academic degree comprises several courses that students must enroll to and complete, so they can graduate. Completing a course provides students with a number of credits. These are part of the *European Credit Transfer and Accumulation System* (ECTS), that defines the courses' workload and estimates the time students must dedicate to them. Each degree year[4] is composed of 60 ECTS (i.e., 30 ECTS per semester), divided by a set of courses.

In terms of enrollment in a given course, students may be classified as (i) *evaluated* if he/she delivered all mandatory evaluation elements of the course, or (ii) *non-evaluated* if he/she failed

[1] https://tecnico.ulisboa.pt

[2] https://fenix.tecnico.ulisboa.pt/cursos/leic-t

[3] Academic year is the period in which the academic tasks take place, typically beginning in September of a given calendar year and ending in July of the following calendar year

[4] Degree year is the portion of a degree's curricular plan, that must be undertaken by the students over the course of an academic year.

to deliver at least one of the mandatory evaluation elements of the course. Evaluated students may be further classified as (i) *approved* if he/she obtained a positive final grade (i.e., 10-20), having successfully completed the course, or (ii) *failed* if he/she obtained a negative final grade (i.e., 1-9), having failed the course.

The students' activity in a semester is determined based on their enrollments and evaluations. According to these criteria, a student may be classified as (i) *active*, if he/she has been evaluated in at least one course, in a given semester, or (ii) *inactive*, if he/she has either not enrolled to any course, or has not been evaluated in any course, in a given semester.

A modification of a student's activity status from one semester to the next determines two important measures: (i) *comebacks*, that occur when a student previously inactive becomes active in the following semester, or (ii) *withdrawals*, that occur when a student becomes inactive after being active in the previous semester.

3 Related Work

In this section, we discuss relevant works describing decision support systems in the context of higher education institutions, as well as their shortcomings.

3.1 A Decision Support System for IST Academic Information

The Decision Support System for Academic Information [3] (SADIA in Portuguese) was a decision support system proposed in 2003 for managing the IST academic information. Its goal was to provide current and historical data, organized in terms of *Key Performance Indicators* (KPIs), to enhance decision-making.

The authors proposed a solution based on DW, designed according to the *Business Dimensional Lifecycle* methodology proposed by Kimball [5]. Using this methodology, the business requirements were defined in an interview-oriented way, that led to the identification of the user analysis queries the system must address.

To define the dimensional model, the authors used a DW Bus Architecture [4], to enable modeling each process individually. The authors identified the main business processes: (i) IST Student Admission process, (ii) Undergraduate Degree Performance Evaluation process, (iii) Course Performance Evaluation process and the Student Performance Evaluation process. The common dimensions identified were: (i) Time, (ii) Student, (iii) Admission, (iv) Geography, (v) Course, (vi) Degree and (vii) Department. The Academic Year dimension was modeled separately from the Time dimension, as most user queries intended to analyze data for a particular academic year and semester.

The authors propose a validation of the ability to respond to each of the user analysis queries. In that sense, the authors created validity matrices, to determine the various dimensions and factual measures that helped responding to each user query. However, for organizational reasons, the SADIA system was never implemented.

3.2 Design of a Data Warehouse Model for Decision Support at Higher Education: A Case Study

The authors present a case study of the implementation of a DW for the University of Business and Technology (UBT) [1] in Jeddah, Saudi Arabia.

First, the authors monitored the business processes at UBT and ensured that a DW system was, in fact, needed. This was followed by the business requirements specification,combining interviews, business process observation and document examination.

Kimball's bottom-up approach was used, in the logical design step, meaning that the business processes modeled would be merged to form an enterprise-wide DW for UBT. Using Kimball's four-step dimensional design process, and considering the business requirements identified, two business processes were selected: Course Registration and Academic Performance. The two business processes resulted in two data marts.

The validation of the implemented models was performed using reports, to answer a series of strategic questions provided by the executive managers at UBT, regarding the registration of students by major per year. Several queries were constructed to answer the questions in the form of a report, containing a table with the years and majors as rows and columns respectively, as well as the respective number of registrations.

3.3 Discussion

From the aforementioned articles, it is possible to observe that the DW design was mostly guided by Kimball's methodologies (i.e., Bus Architecture in [3] and the Four-Step Dimensional Design Process in [3, 1]).

Logical design decisions are detailed in [3]: to enhance the performance of user queries that refer to a specific academic year, the authors introduced an academic year dimension, separated from the conventional time dimension, which is relevant for our work, as the LEIC-T Coordinator's analysis often focuses on specific academic years and semesters.

In terms of validation methodology, the articles propose different approaches. In [3], the authors propose a verification of the models' compliance with the user analysis queries. In [1], the validation of the model was performed by generating reports to provide answers to strategic questions randomly asked by the executive managers of UBT.

4 Business Requirements

By analysing the available input data and through interviews with the LEIC-T Coordinator, we determined that the system should allow monitoring the performance of two particular areas: courses and student generations. We identified several business questions related to both areas, that the system should be able to answer through several KPIs.

For the courses, the business questions are: (i) "What are the approval rates in a given semester?", (ii) "How have the approval rates changed throughout the years?". For the student generations, the business questions are: (i) "How did a generation of students perform in a specific year / semester?", (ii) "Did students perform better when enrolling in fewer courses?", (iii) "Is there a most difficult course for a generation of students?", (iv) "How did the students perform after the duration of their degree?" and (v) "Are the withdrawals significant in a generation of students?".

5 Implementation

This sections describes the implementation of SAD-CCIST. Our system integrates academic data from a set of input Excel files gathered throughout the years, obtains KPIs related to the performance of students in their degree and displays them in the form of dashboards.

5.1 System Architecture

The layers that compose the architecture of our DSS are: the *Data Sources*, the *Data Staging Area* (DSA), the *Data Warehouse* (DW), the *OLAP Server* and the *Dashboards*. The overall architecture of our system is presented in Figure 1.

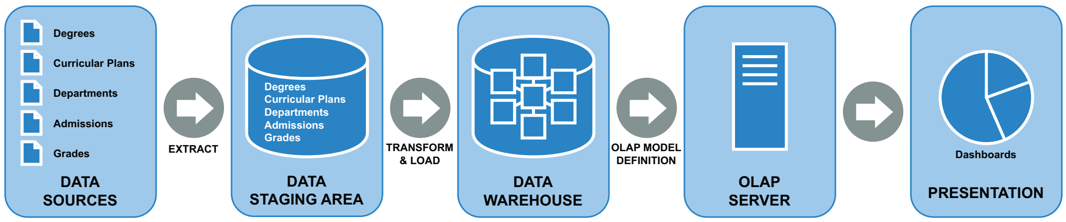

Figure 1: Architecture of the developed system

5.2 Data Staging Area

We created the DSA as relational database using MySQL. We create a table for each group of Excel files (i.e., degrees, curricular plans, departments, admissions and grades), replicating the structure of the files, to enable a direct extraction of data.

5.3 Dimensional Model

The dimensional model of the DW was designed using Kimball's DW design methodologies. We identified the main business processes (i.e., student admission, student evaluation, student activity and student graduation) and conformed dimensions (i.e., *degree*, *course* - including the *scientific area* and *department* hierarchies, *student* and *time* - classified as *admission time* or *evaluation time*), as well as their logical relationships, which are expressed in the Bus Architecture Matrix [4] presented in Figure 2.

BUSINESS PROCESSES	COMMON DIMENSIONS				
	Degree	Course	Student	Admission Time	Evaluation Time
Student Admission	X		X	X	
Student Evaluation		X	X	X	X
Student Activity	X		X	X	X
Student Graduation	X		X	X	X

Figure 2: Proposed bus architecture matrix

The logical design was guided by the Kimball Four-Step Dimensional Design Process methodology [5], which consisted of tackling each business process in the bus architecture matrix individually, for which we identified the granularity, dimensions and factual measures. Using this methodology we created several dimension and fact tables, that together form our dimensional model, presented in Figure 3. The DW dimensional model was implemented as relational database using MySQL.

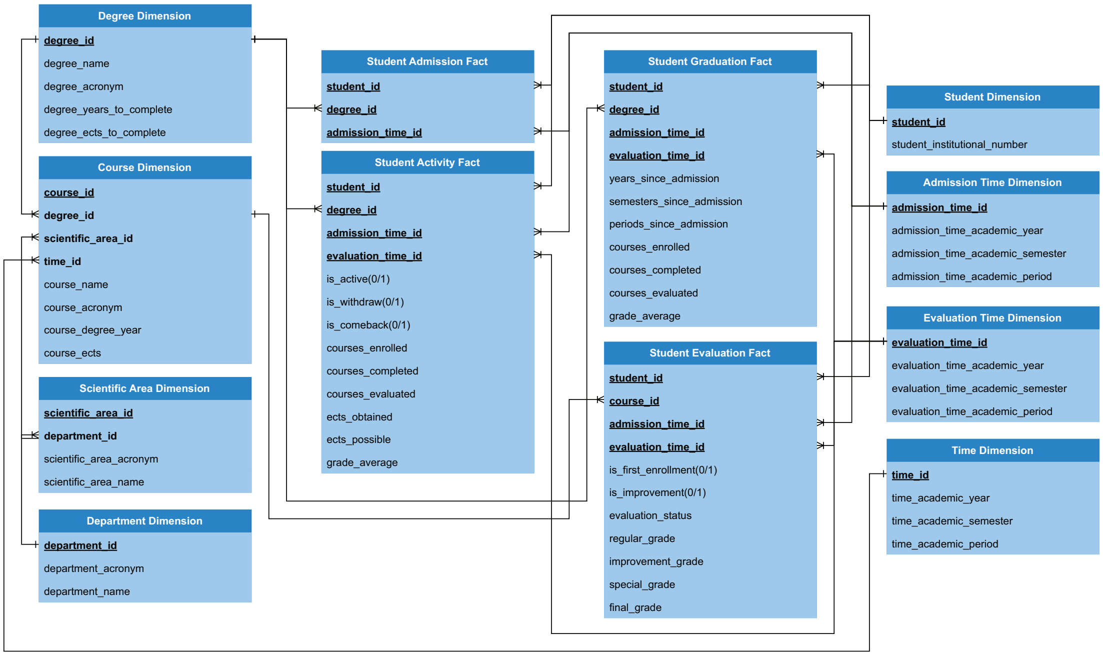

Figure 3: Proposed dimensional model

5.4 Extract-Transform-Load Processes

Since we included a DSA, the ETL was split into two main processes: (i) *extraction* process, that reads all the data in the Excel files and stores it in the tables of the DSA and (ii) *transform-load* process, which takes the data from the DSA and structures it into a suitable format for storing it in the DW. These processes were created using Pentaho Data Integration (PDI).

5.5 Online Analytical Processing

To increase the efficiency of analytical queries, we introduced an OLAP server. Using Pentaho Schema Workbench (PSW) and the Mondrian OLAP engine, we designed and implemented an OLAP model over the model of the DW. creating OLAP cubes using the fact and dimension tables. The proposed model failed to obtain all the required KPIs, which led to its exclusion.

5.6 Dashboards

Our DSS presents the data in the form of interactive dashboards. We used the Pentaho Business Analytics (BA) platform to access the DW and use the data obtained from it to create three dashboards. With these dashboards, the LEIC-T Coordinator is able to obtain insights about the performances in terms of *student activity*, *student graduation* and *course evolution*.

The first dashboard is related to student activity. This dashboard is based on the analysis of the activity of a student generation performed by the LEIC-T Coordinator every semester, which focuses on the activity status of students, withdraws/comebacks, grade averages, ECTS obtained, approval rates and courses completed. Users can use the dashboard filters to view these measures for a specific generation of students and semester. The layout of this dashboard is presented in Figure 4.

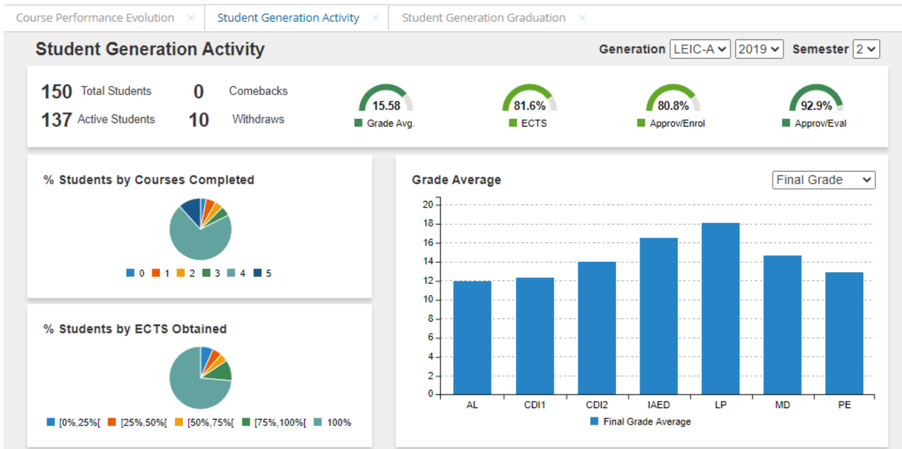

Figure 4: Student activity dashboard

The second dashboard is related to student graduation. This dashboard focuses on the number of graduations, enrollments, withdraws/comebacks and grade average of a generation of students. Users can use the dashboard filters to view these measures for a specific generation of students.

The third dashboard is related to course performance. This dashboard focuses on the approval rates of a course. Users can view these measures for each course, by selecting the desired option on the dashboard filters.

6 Evaluation

6.1 Dimensional Model Validation

To determine whether the various dimensions and facts involved in the dimensional model can obtain the required KPIs, we created validity matrices. A *validity matrix* is a tabular structure that indicates which dimensions and/or facts are involved in obtaining a certain KPI.

The KPIs are divided into *course* and *student generation* KPIs. The course KPIs can be obtained using the measures from the student evaluation fact and its linked dimensions (i.e., student, course, admission time and evaluation time). The student generation KPIs can be obtained using the measures from the student activity and student graduation facts and their linked dimensions (i.e., student, degree, admission time and evaluation time).

Having created the validity matrices, we have verified that the dimensional model created is able to provide answers to all the business questions identified.

6.2 Data Integrity Validation

A fundamental aspect of a decision support system is the integrity of the data within it. To guarantee that the data is accurate, the data stored in our DW was validated against the data contained in the Excel files created by the LEIC-T Coordinator.

For visualization purposes, the LEIC-T Coordinator possesses another group of Excel files, with various metrics related to the performance of the generations of students. These files contained the final grades of all students and metrics related to their semestral activity (i.e., courses enrolled and approved, ECTS obtained and possible).

The LEIC-T Coordinator pointed out two possible issues with these files: the grades data sometimes did not include the special grade values and the set of students considered as part of each generation could be slightly different from ours, as they were manually selected. This implies that deviations might occur when comparing this data with the one stored in our DW.

To produce trustworthy comparisons, these issues should be corrected. Correcting them must be done manually, which made us select only a portion of the available data. As such, we selected only the files related to the LEIC-T degree, which include data from 2007 to 2019.

To perform the comparison, we created a Python script that obtains data from both data sources, for each academic year, then compares all the records obtained and outputs a percentage of matching records. We compared the data stored in our data warehouse against the original Excel files and against the Excel files upon correcting their input data (i.e., updating the grades data and updating the set of students). The results yielded by these comparisons are presented in Table 1.

	Student Evaluation	**Student Activity**	**Overall**
Original Files	96,26%	77,88%	90,05%
Files with updated Grades	96,65%	79,10%	90,72%
Files with updated Grades and Students	100,00%	100,00%	100,00%

Table 1: Results of comparing data from SAD-CCIST against LEIC-T Coordinator Excel files

The difference in terms of the sets of students being considered in the Excel files and in the DW caused a significant difference in terms of accuracy. A more in-depth analysis determined that the set of students considered in the Excel files was composed of *1223* students, whereas our system, considered only *1051* students. The Excel files included several students that had cancelled their admission, which were causing mismatches in terms of student activity. After removing all students that did not fulfill the criteria, we were left with the 1051 students that matched the ones in our system, and the comparison yielded *100%* of matches.

6.3 Usability Validation

We invited several users to evaluate our system in terms of usability. The set of users considered comprised of instructors of IST (i.e., some are current or former degree coordinators) and administrative staff of the IST Department of Computer Science and Engineering (DEI). We considered these users the most representative possible, since they are the most likely to benefit from our system.

In total, there were 14 users participating in the usability tests. 11 users are instructors of IST and 3 belong to the administrative services of DEI. A total of 7 users (i.e., 50%) are/were degree coordinators.

The users were introduced to a set of tasks, for each dashboard (i.e., student activity, student graduation and course evolution dashboards), that consisted of finding specific information within them. The users were encouraged to use the dashboards on their own, as much as possible, though they may require assistance to complete the tasks. We opted by not letting the users fail the tasks, no matter how long they may take, and instead of measuring their success through task completion, we decided to record the following measures: (i) *elapsed time* from the start until the end of the task, (ii) *number of errors* registered in the task and (iii) *number of assistances* given to the users while performing the task.

Upon completing the tasks for each dashboard, users were asked a *single ease question* [7], to assess the difficulty of the tasks, on a scale of 1 to 7 (i.e., very difficult, very easy).

Figure 5 presents the average of all the measures recorded for each task. The three tasks were considered relatively easy by the users, with an average of *5.64*, *5.86* and *4.93*, which are all values above average. Despite this fact, the third task, that consisted of using the course evolution dashboard, presented less desirable results than the first two. Users took more time performing this task, with an average of *03:38* minutes spent. This task also had the highest average number of errors and assistances (i.e., *0.43* and *1.07* respectively), which, unsurprisingly, led to being considered the most difficult of the three (i.e., *4.93*).

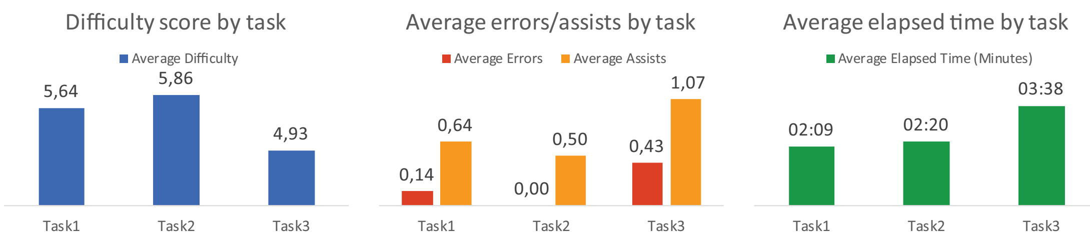

Figure 5: Performance measures of each set of tasks

After completing all the tasks, the users were asked to evaluate the overall usability of the dashboards. We used the *System Usability Scale* (SUS), which is a 10 item usability questionnaire, with response options that range from 1 to 5 (i.e., strongly disagree to strongly agree) [2]. Based on the users' answers to the questionnaire, we computed the average, standard deviation, maximum and minimum of the SUS scores, which are presented in Table 2.

Average	**Std. Deviation**	**Maximum**	**Minimum**
85.18	17.74	100.00	42.50

Table 2: SUS average, standard deviation, maximum and minimum score values

The average SUS score of *85.18*, indicates an overall *excellent* usability score according to this scale [6]. The standard deviation value of *17.74* is slightly high, which indicates that, despite most users giving the system a very good score, a few users did not consider the system as good, as also suggested by the maximum and minimum scores of *100* and *42.5*.

7 Conclusions

In this paper we proposed and developed a DSS for IST degree coordination, known as SAD-CCIST. This system arose from the need for automating the current manual process performed by the LEIC-T Coordinator, that consists of extracting data related to the performance of students and courses of a degree, processing the data and generating visualizations from it.

We specified the business requirements and created a DW dimensional model guided by Kimball's methodologies, as well as relevant works on higher education DSSs. The DW was populated through ETL processes. An OLAP model was introduced, to increase the efficiency of analytical queries, though its inability to obtain all the required KPIs led to its exclusion from the final solution. Finally, we created dashboards that access the DW and display KPIs of different areas of the degree, to provide useful insights to the end users.

To evaluate our solution, we used validity matrices to determine the proposed dimensional model's capability of obtaining the required KPIs. To assess the quality of the data data stored in the DW we compared it against the Excel files produced by the LEIC-T Coordintor. Additionally, we performed usability tests, to determine the usability of the dashboards created and to understand how they could be improved.

8 Future Work

The LEIC-T Coordinator expressed interest in having SAD-CCIST including new data, related to the the IST Course Unit Quality (QUC [5] in portuguese).

The ETL should perform incremental loads as opposed to the current full loads, for efficiency purposes. Additionally, the ETL will need to accommodate a curricular plan modification that will take place in the academic year of 2020/2021, which will introduce courses taught on a quarterly basis (i.e., academic periods), rather than being taught only on a semester basis.

The OLAP model introduced must be redesigned to be capable of obtaining all KPIs, which will enable more efficient analytical queries.

More dashboards can be created using the data currently stored our the system. For instance, the LEIC-T expressed interest in a dashboard for observing the evolution of the number of ECTS obtained by all generations of students, in each semester and degree years.

9 Acknowledgments

To Professors Helena Galhardas and Nuno Mamede, for their guidance on this work and to all participants of the usability test sessions, for the time they spent on such an indispensable part of this project.

References

[1] Isam M Aljawarneh. Design of a data warehouse model for decision support at higher education: A case study. *Information Development*, 32(5), 2016.

[2] John Brooke. Sus: a "quick and dirty'usability. *Usability evaluation in industry*, page 189, 1996.

[3] Elsa Cardoso, Helena Galhardas, Maria Trigueiros, and António Rito Silva. A decision support system for IST academic information. *Informatica (Ljubljana)*, (3), 2003.

[4] Ralph Kimball and Margy Ross. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. John Wiley & Sons, Inc., 3rd edition, 2013.

[5] Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, and Bob Becker. *The Data Warehouse Lifecycle Toolkit*. John Wiley & Sons, Inc., 2nd edition, 2008.

[6] Jeff Sauro. *A practical guide to the system usability scale: Background, benchmarks & best practices*. Measuring Usability LLC, 2011.

[7] Jeff Sauro and Joseph S Dumas. Comparison of three one-question, post-task usability questionnaires. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1599–1608, 2009.

[5] https://quc.tecnico.ulisboa.pt/en/