

INTELLIGENT DECISION SUPPORT FOR UNIVERSITY APPLICATION USING RIASEC CODES

João Pedro Silva¹, Filipe Portela², Manuel Filipe Santos² and Maria do Céu Taveira³

¹*Information System Department, University of Minho, Azurém, Guimarães, Portugal*

²*Algoritmi Center, Information System Department, University of Minho, Azurém, Guimarães, Portugal*

³*Psychology Research Center, University of Minho, Braga, Portugal*

Abstract

A study performed recently in Portugal showed that there are currently about 320,000 students attending high school. Typically, 51 % of them didn't know concretely which course to select, five months before the application date. A significant number (about 160,000) of students needed to be guided and informed about the educational provision at the national level. In addition, 25% of the approximately 309,000 students attending Portuguese Public Higher Education (about 77,000) have changed or thought about changing course during their academic career due to some dissatisfaction related to the current course. In order to minimize these difficulties a decision model was outlined. The solution is based on the construction of a tool that, through a carefully prepared questionnaire, will identify what are the best alternatives for students' application. Key information has been collected from the analysis of several variables in various contexts (e.g. social, economic, personal, and psychological), from scientific studies and from real facts. The resulting model is a weighted model where variables can be set by the user in order to guide the decision making process. Therefore the model is able to adapt to the intrinsic characteristics of each user. This paper focuses in the psychometric capability of the solution (which is unique in this environment) adopting the RIASEC Codes (domains) as the vocational component of the decision models.

Keywords: Higher Education, University application process, Decision Models, Psychometric Tests, RIASEC Codes

1 INTRODUCTION

The main goal of this project is to develop of a Decision Model based on a comprehensive study of the Portuguese Higher education application process. In a depth level this paper is focused in the analysis of RIASEC codes in order to help the students in the University application process and how it can be crossed with the university areas (e.g. science, health, technologies, agriculture and natural resources and architecture). During an experimental phase (first stage) in order to understand the viability of the project it was introduced a Decision Support System (DSS) called C.U.R.S.O. (Universal Centre Collection Tips Oriented). This DSS assisted the students who are in transition year between secondary and higher education, in order to choose the most suitable course for them according their profile. Based on answers provided by the user, the DSS used a data model containing information about the courses of public higher education in Portugal to present to the student a list of courses that best fit their profile. These models help the user analysing what are the most relevant questions for him selecting the weight that each question (variable) has in the final solution. In the decision model process the information was represented through variables from various contexts (social, economic, personal, vocational...) and the system displays as output an ordered list of the most suitable courses for the user. The success of this first version using only a few numbers of variables motivated the improvement of the decision models. To support the second part of the work (understand the problem reality and evaluate the viability of the solution) were performed questionnaires among possible users of the system (students attending both systems High School and College). Two separate questionnaires were developed. This exercise took place during the month of February 2013, about five months before the College application and it was answered by 2000 students. For the two segments mentioned above, the questionnaires had as main goals to prove that there is a large number of secondary school students who do not know yet which course to choose, and that there is a large number of higher education students, who despite having entered a course, after some time eventually changed the course thus proving that their initial expectation was wrong [1]. The attained results were fundamental to develop the second stage.

In this second stage the decision model was improved and adapted to the Portugal reality using the RIASEC codes and DGES areas. In parallel a set of other variables were studied in order to be included (e.g. University Quality, Personal, Social, Academic). The results of this work (model) can be adapted to other countries. The work was developed crossing knowledge from different research areas: information systems, education and psychology.

This paper addresses the development of the decision models mentioned above, focusing the psychometric variables of the system through the RIASEC Codes allowing for a solution that considers the user's vocational attributes in its analysis. The document is divided in seven chapters. The first one explains the context where this project appeared, chapter two handles the background that defines the scope of this work, similar services, techniques used to portray information about the target audience and also the description of the development methodology that was used. In chapter three are presented the variables that were studied and analyzed, in particular their explanation, association with the model and their data source. In this chapter it is also explained in detail the psychometric component presenting the RIASEC Codes association with the Decision Models. In chapter four the informational model is presented and in chapter five it is presented a logical block of the decision model, more concretely the Vocation block. Finally, in the sixth chapter the results are discussed and then in the seven chapter are analyzed the conclusions and proposed future actions to this work.

2 BACKGROUND

In the beginning of this work it was accomplished a market analysis for this context. During this phase some questionnaires were developed in order to understand the context [1], similar services (related work) were searched and a methodology to support the development of the decision models was defined.

2.1 Related Work

A work of this size, in addition to be useful, must necessarily be something innovator in the context in which it appears. So, it was carried out a comprehensive search of similar services that are available to the market segment. Regarding to the specific exercise intervention in this work, after an exhaustive survey, none system has been found with similar characteristics.

At national level, two online services were found that allow users to do a search of all courses of higher education (research), where they could establish a set of criteria in order to filter the answers that will appear (form). The services are supported by the site of the Directorate General of Higher Education (DGES) and the Office of Higher Education. In both cases the user can survey courses in higher education and filter by location, type of institution, field of study and the specific ingress therein. At international level, there are several systems similar to those that can be found in Portugal. In addition, in some systems the user fills out the variables of a short questionnaire and then the best options for their profile are presented. Table 1 present the services found characterizing the types of service (search and form). As can be observed none of the services permit the association of weights, i.e., the user / student cannot configure the importance of choices according their tastes.

Table 1. Similar Services

Institution	Country	Search	Form	Weight
Empresário	México		X	
Univafu	México		X	
Universidades.com	Argentina	X		
UK <i>Course Finder</i>	United Kingdom	X		
Go2Uni	United States	X		
Guia do Estudante	Brazil	X		
<i>Good Universities Guide</i>	Australia	X		
Guia da Carreira	Brazil		X	
Gabinete do Ensino Superior	Portugal	X		
DGES	Portugal	X		

2.2 Psychometric Tests

Although there isn't any decision support system to help students in order to find the most appropriate choice to Higher Education course, there are other tools with many years of use in this community. The most common are the psychometric tests that, through a series of questions and mental exercises, and with the help of a psychologist, put the student in situations related to professional areas, in order to find what area makes him feel more comfortable. However these tests do not compete directly with the DSS in development, since they point to future jobs. Though intended for the same market segment, they are for different situations.

2.3 Methodology for the development of Decision Models

To develop this project a methodology has been followed in order to drive all the work. The methodology chosen is based in the principles of the decision process presented by Herbert A. Simon [2]. Recently, the methodology received contributions from Turban [3]. The resultant model is divided into 5 phases:

- 1) Intelligence: Gathering the information inherent in the process for expectations about the answers that the model will represent;
- 2) Design: Drawing of the decision model, consisting of flow charts, decision trees and other tools for structuring the model;
- 3) Choice: Select the best model and consequent configuration of the data structure that will be presented in the Model;
- 4) Implementation: Deployment of Decision Models in a platform (IDSS);
- 5) Monitoring: Monitoring of the system after its implementation [4].

With this approach the models are able to operate in an Intelligence Decision Support System (IDSS), an Artificial Intelligence structure that it is capable to support the decision that the users need to execute, in this case the support on the College application.

3 MODEL VARIABLES

3.1 Data Source

To design a complete and transversal decision model it was carried out a deep study on the variables typically used in this context and their sources. The variables represented in the decision model were defined through a study based on scientific sources to ensure the reliability of the information. This point was essential to find information corresponding to the variables treated. Table 2 present the main variables considered by the models and the corresponding data source, i.e., the local from where the information was collected.

Table 2. Variables Data Source

Variable	Data Source
Admission exams	Public Guide for College Admission [5] DGES website
High School Areas	DGES website
Scholarship assignment importance	Direção Geral de Estatísticas da Educação e Ciência (DGEEC) entitled "Inscritos no ano letivo 2011-2012 por NUTs [6]; DGES website
City life quality importance	"Os Municípios e a qualidade de vida", dos autores José Pires Manso, António de Matos e Fátima Gonçalves no ano de 2012 (Manso & Simões, 2012)
Academic life importance	Institutional Websites of Colleges and Student Associations
Cultural life importance	Institutional Websites of Colleges and Student Associations
Sports prestige importance	Federação Académica de Desporto Universitário (FADU)
Rankings	Webometrics [7]
College courses areas	DGES website

During the study was detected the importance of developing some indicators in order to achieve a better comprehension of which variables are more important to the user (student). The following table (table 3) resumes these indicators.

Table 3 – Variables Indicators

Variable	Indicator
Scholarship assignment importance	(total of assigned scholarships / total of submitted scholarships) per establishment
City life quality importance	Association of the quality of life of municipalities index here there are establishments of higher education to the same
Academic life importance	Percentage of activities related to the academic life occurring in a particular establishment
Cultural life importance	Percentage of activities related to the cultural life occurring in a particular establishment
Sports prestige importance	(total medals won / total modalities practiced) per establishment
Rankings	Division of 5 variables for 5 distinct rankings
DGES areas	Relation with high school areas through a comprehensive analysis of curricular plans

3.2 Vocational Variables (RIASEC Codes)

RIASEC is characterized by a specific action within the psychometric tests performed in a completely different way when compared with the traditional approaches. The differences are more evident in terms of how to involve the user and how to display the results [10]. The user answers to a set of questions related with work situations. At the end of the exercise the user is faced with a RIASEC code: **R**ealistic, **I**nvestigative, **A**rtistic, **S**ocial, **E**nterprising and **C**onventional. In the last set (class) of questions the objective is try to find the association between user answers and a number of vocational variables. As a result it is expected that this block can "steer" the choice more for certain courses and less for others. To frame this class in the model it was necessary to use an expert authority in the area that could, not only facilitate access to information and other contents but also formalize how to fit all these data in Decision Models. To this end, and as described in the introductory section, it was contacted an expert in Psychology area in order to make a bridge between the system and the vocational variables.

This part of the work was based on the RIASEC framework and had as objective to find which vocational variables should be included in Decision Models. The first concern of this part of the work it was ensuring that when this information is processed in the models and when it is represented through the questions (in the implementation phase), the results continue to make reference to an intact variables association and according what was projected in this study. At this point there are numerous forms of match the RIASEC with university courses. With the help of psychology school it was possible define a real, practicable and valid path. To perform this task it was necessary analyze the courses, evidence of access, and other elements of the training areas, from more pedagogical criteria, in order achieve a final match. Briefly, the intervention of psychology professor in this project had always in the mind a point, the fact of these variables cannot fully condition the decision of the user, i.e., a user which chosen one set of vocational variables, cannot be fully excluded from a course that is not related to the same. This way we can guarantee that the informative nature of this class are not totally decisive, being the results dependent of other variables. This class must be performed in a characteristic way, differing from the others classes. The user must choose three of the variables and defining the preferably order i.e., he should indicate what is the area that more identifies. Finally and after having defined the decision process, the knowledge areas were described according RIASEC codes (domains), then and having the objective to find a match to the courses it was defined an order that the domains should appear.

- **Realistic (R)** - activities and environments that often involve preference for working with objects, develop knowledge and mechanical skills, solving problems and practical solutions, physical work and athletic activities. Involves often deal with plants, animals, natural materials, tools, machines and technologies. Includes work or study environments that, by comparison with other imply more working outdoors and less bureaucratic work, finally it is embodiment of life values related to tradition, freedom and independence.
- **Investigative (I)** - activities and environments that often involve work with objects and ideas. Frequently involve dealing with abstract ideas, thought, solving mentally problems, collect and analyze data and allow more often achieve life values, such as independence, logic and personal conducting (school, home and professional).
- **Artistic (A)** - activities and environments that often involves working with ideas and people. Often involve using imagination, self-expression, creativity, and working with forms, designs

and patterns, and open tasks without much initial structuring or strict rules to follow. Allow achieve life values as aesthetics, self-expression, imagination and non-conformism.

- **Social (S)** - activities and environments that involve preferential working with people. Involving communication and interaction with others, and the provision of services in order to help, support, educate, educate and advise, allowing often realize related life values, for example, altruism, ethics, and equality.
- **Entrepreneurs (E)** - activities and environments that often involve preference by working with data and people. Involves the use of persuasion and decision-making, leadership, management, administration, and other behavioral skills such influence as well as the organization of people, work, and events. This domain also involves some risk taking and business deals. Allow more often, achieve life values such as tradition, economic achievement, and ambition.
- **Conventional (C)** - better known as the interests of organizational or bureaucratic problems, can be developed in almost all environments of school and professional work, though in some of them being more evident. Involves a preference for working with data and objects, rather than with people or ideas. Activities like how to organize, sort, be detailed, follow instructions maintain routines and standard procedures, meets this domain. Typically is associated to life values such as tradition, ambition, obedience, economic achievement and comfort.

The ultimate goal of this component was to make an association between the domains and vocational areas:

- **Science** - must appear when the
 - First domain chosen interests is the Investigative (I);
 - Second field of interest chosen is realistic (R) or Conventional (C);
 - Even when the third domain of interest for the conventional (C) or the realistic (R).
- **Health** - should appear when the
 - First domain chosen interests is the Investigative (I);
 - Second domain chosen interests is the Investigative (I), or social (S);
 - Third chosen domain is the Investigative (I) Conventional (C) or realistic (R).
- **Technologies** - must appear when the
 - First domain chosen interests is realistic (R);
 - Second domain chosen interests is the Investigative (I);
 - Third chosen domain is the Conventional (C) enterprising (E) or Social (S).
- **Agriculture and Natural Resources** - should appear when the
 - First field of interests chosen is the Investigative (I);
 - Second domain chosen is realistic (R);
 - Third chosen domain is the entrepreneur (E) Conventional (C) or Social (S).
- **Architecture, Fine Arts and Design** - should appear when the
 - First domain chosen interests is the Artistic (A);
 - Second field of interest chosen for the Investigative (I), Social (S) or realistic (R);
 - Third area chosen is realistic (R) or Social (S).
- **Educational Sciences and Teacher Education** - should appear when the
 - First field of interest chosen for the Social (S);
 - Second domain selected interests is Artistic (A), Conventional (C) or Investigative (I);
 - Third is chosen domain Artistic (A), Enterprising (E) or Investigative (I).
- **Law, Social and Human Sciences and Services** - should appear when the

- First domain chosen interests is Social (S);
- Second field of interest chosen for the Entrepreneur (E) or conventional (C);
- Third chosen for the Enterprising (E), Conventional (C) or Investigative field (I).
- **Economics, Management and Accounting** - should appear when the
 - First field of interests chosen is the Entrepreneur (E);
 - Second field of interest chosen for the Social (S) or Conventional (C);
 - Third field is the chosen Social (S), conventional (C), or Investigative (I).
- **Humanities, Secretarial and translation** - should appear when
 - First domain chosen interests is Social (S);
 - Second field chosen for the Conventional (C), Artistic (A) or Entrepreneur (E);
 - Third area chosen is the conventional (C), Artistic (A) or Entrepreneur (E).
- **Physical Education, Sport and Performing arts** - should appear when
 - First domain chosen interests is the Artistic (A) or realistic (R);
 - Second field of interests chosen is Artistic (A), realistic (R), Conventional (C), Investigative (I), Entrepreneurs (E), or social (S);
 - Third field chosen is Conventional (C), Investigative (I), Enterprising (E) or social (S).

After completed the association phase this information was converted into a format capable of being integrated into Decision Models. It was used a scale from 0 to 5 to relate the variables with vocational areas where 5 would be a strong relationship and 0 unrelated. To ensure consistency with the assumptions stipulated by the area of psychology: "The model shouldn't close paths to the user, i.e., the system cannot say to a student who have most vocation to arts that cannot come to attend a technology course because the system cannot take responsibility for preventing the access to information in a certain area."

The system do not replaces the student opinion, only is responsible to give more information to the decision-making process, presenting the better choices to a determinate profile. In this case it was opted by not assign the values 0 and 1 to any relationship of vocational variable. This exception is different from all the others variables, since, more than any other variable, the user may not be conditioned by its listing under these 6 categories (an enterprising individual can perfectly follow the area of science, however this area has this variable as less related). Table 4 present the relation made between RIASEC codes and vocational areas based in the principles presented in this sub-chapter.

Table 4 – RIASEC vs Vocational Areas

	Science	Health	Technology	Agriculture and natural resources	Architecture, Arts and Design	Education and Teaching	Law, Social and Human Sciences and Services	Economy, Management and Accounting	Humanities, Secretary and Translation	Sports and Entertainment Arts
Realistic	4	3	5	4	3	2	2	2	2	5
Investigative	5	5	4	5	4	4	3	3	2	3
Artistic	2	2	2	2	5	4	2	2	4	5
Social	2	4	3	3	3	5	5	4	5	4
Enterpriser	2	2	3	3	2	3	4	5	4	3
Conventional	4	3	3	3	2	4	4	4	4	3

3.3 Model Blocks

Defined the structure of decision models it was created a logical structure to aggregate variables so that could be presented to the user in order to frame it in the context. To this end, were defined logical blocks variables that that are filled by the data presented in the database tables according the relational model designed. The decision model is composed by 10 blocks that correspond to the aggregation of variables into classes by logical predicates. Each one is composed by a set of attributes resulting in a total of 27 variables analyzed [8]. For each block it was necessary found the indicators that best represent each variable. In this context it was performed a statistical analysis in order to create classes to the model variables. It was decided to create a five data grouping classes, instead of using a rule of Frequency Distribution, Sturges Law [9]. This procedure was performed for all variables where it was necessary to organize variables in groups / classes. For this it was necessary to analyze the statistical assumption based on the law of power base 9, allowing for the identification of how many classes can be created. In some variables of the models were excluded the value of 0 in order to mitigate the difference with the records that have higher values. In other cases also was removed the value 1 when the area that addressed this point is irrelevant to the project.

Combining all the variables present in the 10 blocks, it is possible to find a strong relation between the variables and the courses. This relationship is a demonstrative example of how one might relate vocational variables with courses in Higher Education. Table 5 present the 10 logic blocks and the variables associated to each one. Each block is corresponding to one of the components that are treated in the decision models. The variables will be represented in the IDSS by a question / form.

Table 5 – Model Blocks Variables

Id	Name	Description (Logic component)	Variables
1	High School	User variables corresponding to High School	School, Conclusion Average, Studies Area, If still attending High School, Required Exams
2	College	User variables corresponding to all College courses	Average, Duration, Regime
3	DGES Areas	User variables corresponding to DGES Areas	All the 10 institutional areas created by DGES
4	Social	User variables corresponding to all Social details	Scholarships, City quality of life
5	Academic	Treats all variables corresponding to the user's academic oriented responses	8 College Academic Events
6	Cultural	Treats all variables corresponding to the user's cultural oriented responses	7 College Cultural Entities
7	Sports	Treats all variables corresponding to the user's sports oriented responses	College Sports Prestige
8	Prestige	Treats all variables corresponding to the user's responses related to prestigious establishments of Higher Education	Visibility, Impact, Opening, Excellence and Global Rankings
9	Vocational	Treats all variables corresponding to the user-oriented vocational answers	Realist, Researcher, Artistic, Social, Enterpriser, Conventional
10	Personal	Treats all corresponding to the user's personal nature responses variables	Satisfaction with High School Area, Will to study away from home

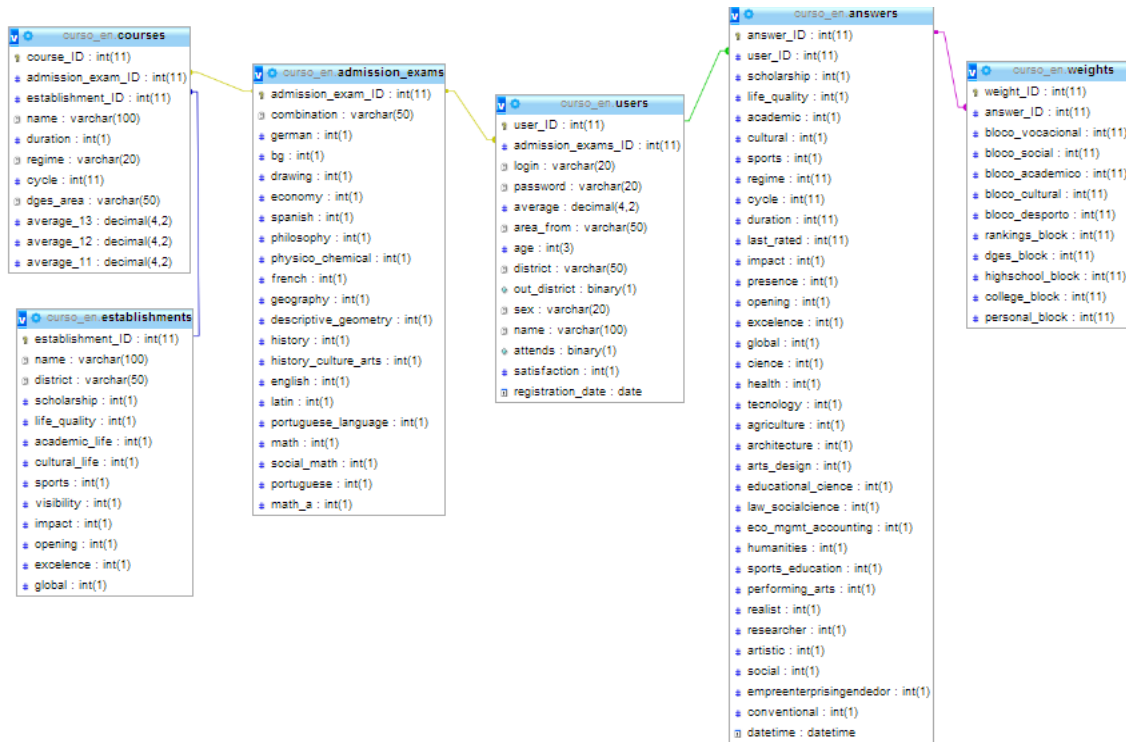
This paper focus on vocational logical block (9) that contains the information regarding the vocational profile of the user.

4 INFORMATION MODEL

After being studied the variables that compose the decision models it was possible to structure a relational model. This model gave origin to a database containing all the information related with Higher Education courses. During the choice phase they were designed four different decision models structures. After evaluate all the models it was choosen the model which best meets the needs. Figure 1 shows the relational model chosen, taking into account that the structure could be configured in various ways, this model is one that best met the requirements and the objectives . In this figure it is possible observe the variables grouped by tables: course, establishments, exams, users, answers and weights. In this tables it is possible see which variables are evaluated by each

group. For example to the establishment it is used: name, district, scholarship, life quality, academic life, cultural life, sports, visibility, impact, opening excellence and global appreciation.

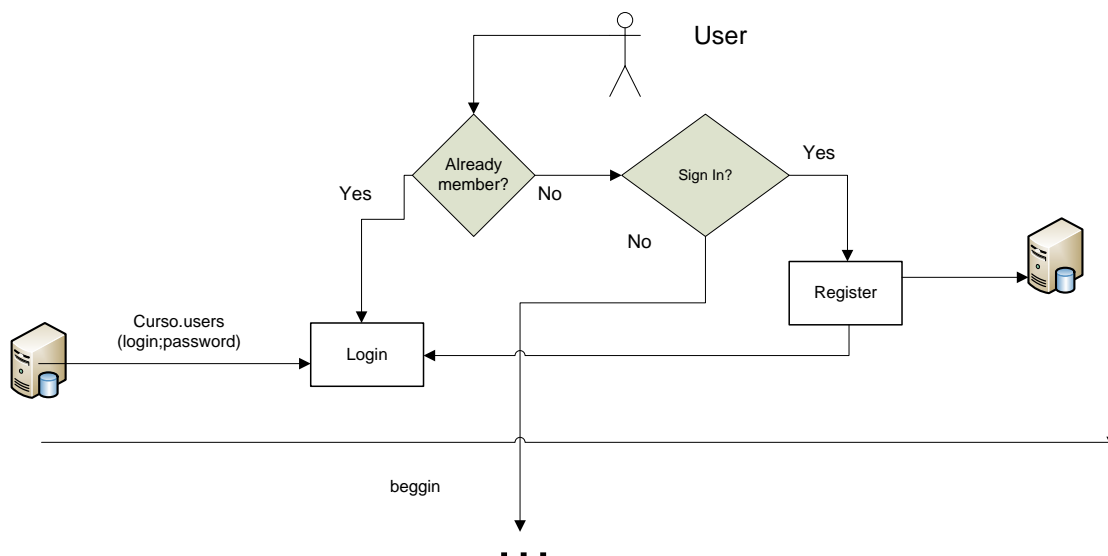
Fig. 1. Relational Model

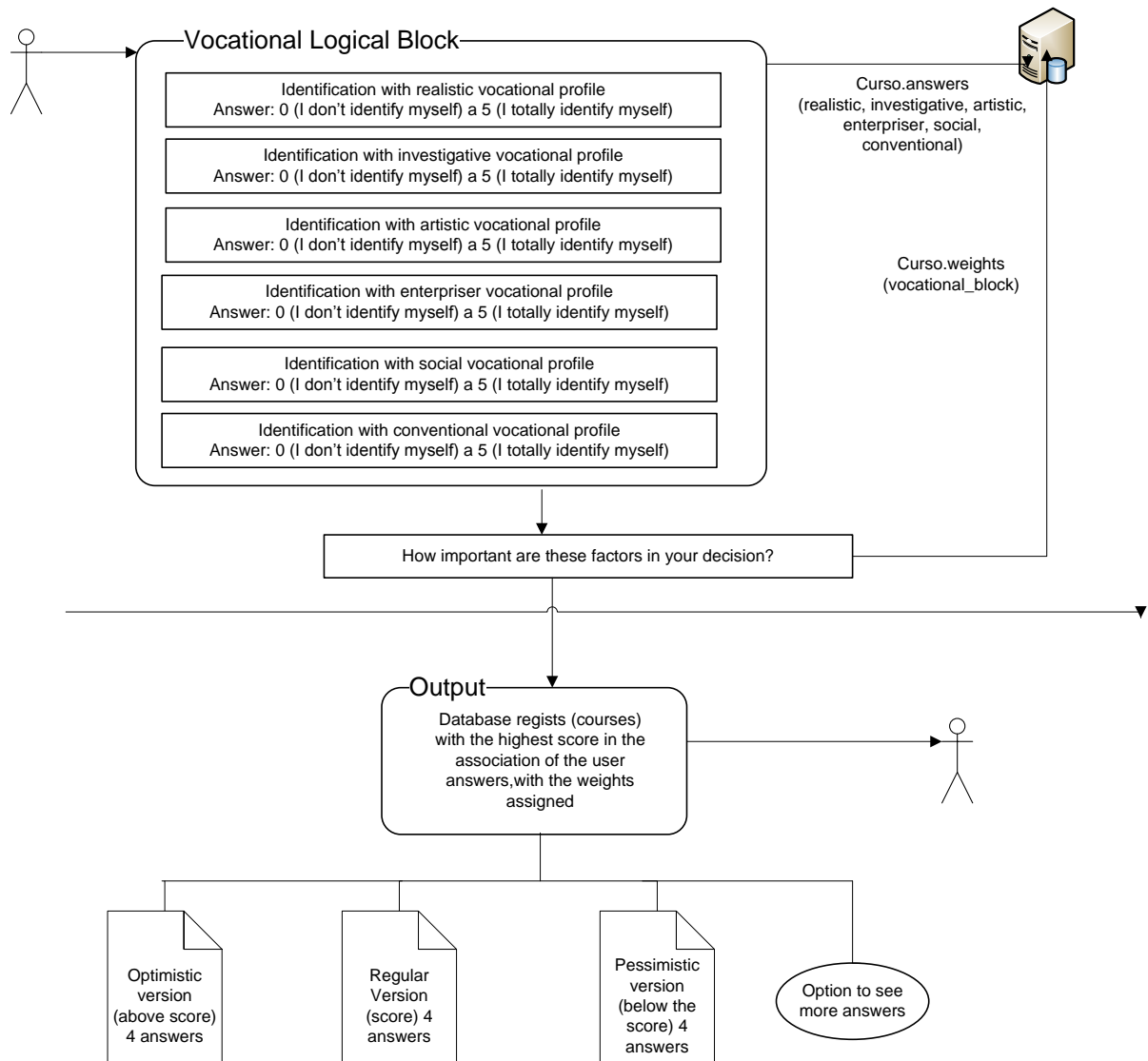


5 VOCATIONAL BLOCK

After the definition of the data model it was created the logical structure to support the logical blocks defined. This structure is defined considering the interaction of the user with the decision models. Since this paper is about the vocational component of de Decision Models, in Figure 2 it is possible to observe the execution of this structure. In the beginning the user is identified, then there are a set of interaction with all the logical blocks (focus on the vocational block). For example, as can be observed in the Vocational logical block, the user has to answer to six questions related with each one of the domains and set the respective weight to this block. Finally and after all the answers be compiled by the model the obtained results are presented to the user.

Fig. 2. Logical block of Vocation variables





6 DISCUSSION

This work presented the tool developed and outlined the feedback received in the experimental phase. To prove the pertinence of developing an IDSS like this a set of questionnaires were developed [1]. This resulted in a model that makes use of 10 logical blocks containing 27 variables.

A block including vocational variables was introduced based in a decision model that combines RIASEC codes with other variables object of study (High School, College, DGES Areas, Social, Academic, Cultural, Sports, Prestige, Vocational, and Personal). As mentioned before, RIASEC codes allow students select three domains and, based in their choices, present a set of areas. For instance, a course from technologies area is presented if the student has indicated realistic (R) as first domain interests, Investigative (I) in second domain interests and Conventional (C), Enterprising (E) or Social (S) for the third domain. It was based in this presupposition that the information was converted into a format capable to be associated with Decision Models, by using a scale from 0 to 5 (table 4). However, even the student choice indicates that there are strong correlations with technologies courses it can be presented courses from other area. This matching gives the opportunity to the student to have a better idea of which domain is more adequate to their vocational profile.

The final result depends on the choices that students made relatively to nine different blocks and the weights associated to each one. This model represents a new way to help the students during the application process. Due to its completeness the model can be considered universal (country independent). Making an overview of decision models developed, it is possible verify the viability of combining a set of variables and different areas in order to create new knowledge at the time of decision.

7 CONCLUSION AND FUTURE WORK

This paper presented an overview of the global decision model focusing in the vocational variables. RIASEC codes proved to be useful when combined with other variables in order to support the university application process. The combination of three distinct areas of knowledge (information systems, education and psychology) was the key of success.

The main contribution of this work is the explanation of how RIASEC codes can be used as vocational variables and the model generated that can be universal and correlate six domains with vocational areas. The proposed model is unique in this context and contemplates the introduction of a weighting system for these variables. The weights can be set by the user and make the model able to adapt to the intrinsic characteristics of each student. As the output of the model it is presented a set of possible choices adequate to the student profile.

Future work includes the deployment of an Intelligent Decision Support System making use of the models presented in this work. At the same time a monitoring process will be incorporated to receive the users (students) feedback. This feedback can be used to enrich the models making the IDSS adaptive.

ACKNOWLEDGMENTS

This work has been supported by FCT – Fundação para a Ciência e Tecnologia in the scope of the project: PEst-OE/EEI/UI0319/2014.

REFERENCES

- [1] Silva, João Pedro; Portela, Carlos Filipe; Santos, Manuel Filipe (2013). "A Decision Support System for Portuguese Higher Education course selection – First Round". KMIS 2013 - International Conference on Knowledge Management and Information Sharing. Vilamoura, Portugal
- [2] Simon, H. A. (1978). Rational Decision-Making in Business.
- [3] Simon, H. A. (1978). Decision Support Systems, 50(3), Turban, E., Sharda, R., & Delen, D. (2011). Decision Support and Business Intelligence Systems. (S. Yagan & E. Svendsen, Eds.) (p. 715). Prentice Hall.
- [4] Turban, E., Sharda, R., & Delen, D. (2011). *Decision Support and Business Intelligence Systems*. (S. Yagan & E. Svendsen, Eds.) (p. 715). Prentice Hall.
- [5] Direção Geral do Ensino Superior (2013), "Guia das Provas de Ingresso". http://www.dges.mctes.pt/NR/rdonlyres/C17C6105-9379-40E9-AEF1-470F0DEF08E6/7074/GuiadasProvasdeIngressoPC3BAblico2013_20130520.pdf (Accessed: 21st February 2013)
- [6] Direção Geral de Estatísticas da Educação e Ciência (2013), "Vagas e Inscritos". <http://www.dgeec.mec.pt/np4/84> (Accessed: 21st February 2013)
- [7] Cybermetrics Lab (2013), "Ranking Web of Universities" www.webometrics.info (Accessed: 21st February 2013)
- [8] Silva, João Pedro, Portela, Filipe, Santos, Manuel Filipe (2014). "University Application Process - Decision Support Models". WorldCIST'14 - World Conference on Information Systems and Technologies. Madeira, Portugal
- [9] Stable A (2012). The Choice of a Class Interval Author(s): Herbert A. Sturges Reviewd Work(s).
- [10] Gardiner V. & Gardiner G. (1979). Analysis of Frequency Distribution.