Sistema de recomendación vocacional en línea

Web-based vocational career recommendation system

Sergio Alcaraz Corona
Universidad Autónoma de Nuevo León, México
sergio.alcarazcrn@uanl.edu.mx

Mario Alejandro Bustos Jiménez
Universidad Autónoma de Nuevo León, México
mariobustosjmz@gmail.com

Resumen

Actualmente elegir la profesión deseada se ha convertido en una de las decisiones más difíciles dada la gran variedad de opciones e información disponibles. el proceso para seleccionar una profesión comienza con la correcta elección del programa educativo que más se ajuste a las habilidades, conocimiento y preferencias del individuo. para apoyar dicho proceso existen diferentes opciones tales como cursos de orientación vocacional así como aplicaciones informáticas, las cuales ayudan a los usuarios en su proceso de selección (katz, 1993); sin embargo, la mayoría de estas aplicaciones tienen un costo alto o no ofrecen resultados confiables. el objetivo de este trabajo es presentar un sistema de recomendación vocacional en línea basado en un algoritmo de asociación y clasificación de información como el principal componente de definición de perfiles vocacionales.

Palabras clave: sistemas de recomendación, orientación vocacional, inteligencia artificial.

Abstract

Now choose the desired profession has become one of the most difficult decisions given the large variety of available options and information. The process to select a profession begins with the right choice of the educational program that suits the skills, knowledge and preferences of the individual. To support this process, there are different options such as courses of vocational guidance as well as computer applications, which help the users in their selection process (katz, 1993); however, most of these applications have a high cost or do not provide reliable results. The objective of this work is to present a recommendation vocational online system based on an algorithm of Association and classification of information as the main component of vocational profiles.

Key words: recommendation, vocational orientation, artificial intelligence systems.

Fecha recepción: Julio 2015 **Fecha aceptación:** Noviembre 2015

Introduction

One of the most important decisions in the life of any person is to select your profession properly. The difficulty of taking such decision has increased due to various factors, mainly economic, technological and social. In a world that changes and moves increasingly faster, the diversity of career choices has become broader and more dynamic because of the emergence of new professions.

Recently changes have been made in the curricula of different careers that offer almost all universities in the world, with the aim of producing professionals better prepared for an increasingly competitive and globalised environment. At the same time, new professional careers have been emerging to meet new demands and labor requirements. The foregoing, together with

other factors, constitutes a large amount of information that can make the process of selection of a career to be a complicated task, especially if there are no elements or support tools.

One of the options that exist to support the process of selection of studies are traditional courses of vocational guidance, drawn up mainly by public and private educational institutions in different formats and contents. One of the essential components of these courses of vocational guidance is clearly the questionnaire with questions aimed at providing the user with information that is useful to your career choice process. Currently it is very common to find such questionnaires as well as other information, by means of an application, or search for information on the internet. However, the majority of such questionnaires and guides are very expensive or do not produce reliable results (Franca, 2012). Why it is essential to have support tools that provide new and better services of vocational orientation (Katz, 1993). The rise of information technologies, mainly everything related to internet, has contributed to the advancement and improvement of a large number of applications in everyday life.

Currently the so-called web applications can cover a wide range of needs, which range from purchasing various items, make a reservation or check general information, among other things. In certain types of web applications, the result of a query can be accompanied by one or more recommendations which provide the user with more information to make a better decision. For example, Amazon.com online shopping portal recommended certain amount of similar to the searched items. These recommendations are obtained using special algorithms employed in so-called applications of artificial intelligence and data mining, which process the information that the user is providing. There is a wide variety of algorithms and methodologies for this purpose, however, a type of these algorithms seeks associate and classify information among themselves according to certain criteria or rules that can make easier the use and exploitation of such information.

The objective of this work is to design and implement a vocational online recommendation system using programming languages and recent information technologies with which the system may be able to produce reliable results and at the same time generate the best study recommendations for each user of the application. Therefore, the recommendation system proposed in this project is a web site of vocational guidance for use mainly of the student community of the Faculty of Mechanical Engineering and Electrical (FIME) of the Autonomous

University of Nuevo León (UANL), whose URL is http://www.fime.vocacionaluanl.com. To do this, system will take as a basis the answers that correspond to questions related to the abilities, skills, interests and tastes that the user registers so they are all processed and classified by the logic of the system and thus determine in each case the most appropriate vocational profile.

DEVELOPMENT

Background

As its name suggests, the main objective of any recommendation system is to provide some information of interest in response to a particular request or search by the user or users (Adomavicius, 2005). Recommender systems are part of an extensive area of computer science that is commonly known as artificial intelligence, whose main objective is the management and administration of knowledge and information through systems or applications with some level of intelligence to solve problems and make decisions within a certain context (Truemper, 2004). Among the most common intelligent systems are expert systems (Truemper, 2004), neural networks (Kasabov, 1996), data mining (Witten & Frank, 2005) and fuzzy logic (Kasabov, 1996), among others.

In a way, recommender systems are considered as a special type of expert system. Additionally there is a lot of applications that recommendation systems have helped improve the processes of search and selection of information. One of the applications most boom has had for the benefits to both users and the organizations that implements it is the e-commerce, mainly for the purchase of commercial products and services (Leavitt, 2006;. Linden et al, 2003).

Although not as extensive as in the commercial field, we found some applications of recommender systems for academic or educational environments (Aher and Wolf, 2013; Sobecki, 2014). In particular, one of these work presents an application of this type of system, which also uses an implementation of the algorithm k Means as part of its mechanism to recommend courses to students under certain criteria or special conditions (Aher y Lobo, 2013).

One of the distinguishing characteristics of recommender systems refers to the logic that follows the procedure that is responsible for producing its recommendations and of which the following categories (Adomavicius, 2005) are derived:

Recommendations based on content. It is the category that requires knowledge of the characteristics related to the content or description or items that have been previously recommended when producing new recommendations for a given user or group of users.

Recommendations based on collaborative filtering. It is another category where only considered the recommendations previously made to users with tastes, interests and / or similar characteristics to generate new recommendations for a given user or group of users, ie, in these cases the content or description of the elements recommended not it is relevant.

Hybrid schemes are those that combine functionality of the previous two categories. In this case, the logic of the system proposed recommendation is the type of the first category since it analyzes the content; specifically, the value of the selected answer for each question as part of the process of generating recommendations.

Definition of Questions

The process of defining questions began with a literature review to understand the most important aspects in the development of questions that correspond to a questionnaire vocational guidance. According to the consultation, it was found that there are several factors that influence the definition of questions in such questionnaires, such as skills, knowledge, preferences, personality and interests of each individual, which are the most important. In general, the above factors have been extensively studied to form the so-called theory of Holland vocational guidance, which found, among other things, that the personality and skills play an important role in any selection process of higher education and referring to what an individual is capable of doing in the future (Betz et al., 1989; Gati et al., 2006; Holland, 1997).

On the other hand, an online survey (Alcaraz et al., 2015), which groups of college students was sent to try to identify the factors that most influenced in their selection process of a career was also created, as well as various situations that led them to make career changes. With the implementation of this survey a sample of approximately 200, which could obtain valuable data for defining the system and thus have more elements to develop a tool for more efficient and reliable vocational guidance was obtained.

From the literature review and the results of the survey, we proceeded to the creation of 50 questions for each race managed by the system, which includes three races of the Faculty of

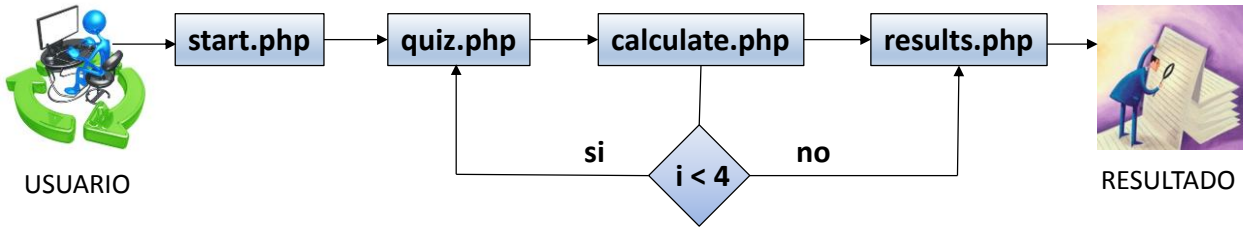
Mechanical and Electrical Engineering of the University in its initial version Nuevo Leon: Software Engineer Technology (ITS), Systems Engineer Manager (IAS), Engineer in Electronics and Communication (IEC). After analyzing the most relevant to the process of defining concepts, it was determined that the previous number of questions can represent the key features and attributes of each race effectively.

The questions are organized within Excel files, starting with the most general to the most specific. Once the questions were sorted, levels (tiers) were planned for them. The levels include 0 to 3, with 3 being the most representative of each career level while 0 is the most basic or general (common core) level. The approximate distribution of questions is proposed level (level, number of questions) (0, 10), (1, 15), (2, 15), (3, 10).

The above distribution may vary depending on several factors, among which are ancient, history, documentation and demand for a particular career. However, when organizing the questions in the above manner is avoided having to implement an additional process to sort them into groups according to the degree or level of representation they have for a particular career. Otherwise, it would have been necessary to implement the algorithm k Means (k-means), which is capable of classifying and grouping a set of n observations, assigning each one of k possible groups (Duda & Hart, 1973). Therefore, the content of each Excel file questions stored in a database following the order mentioned.

Recommendation system structure

Within the increasingly large area of information systems development, there are a variety of options, languages and technology to build this kind of web applications or internet pages dynamic tools. was used mainly PHP (PHP Hypertext Preprocessor) is a programming language commonly used general purpose for developing web applications with dynamic content client-server (Beati, 2011) architectures for this project. The PHP language offers several advantages over other programming languages used to develop applications on the type sought in this project, ie, a dynamic web application with a database. In Figure 1 the diagram of software components that make up the system and are responsible for its operation is shown.



Figura+e 1. Recommendation system structure fime.vocacionaluanl

The start.php module is responsible for establishing the connection to the database and initialized to zero iteration variable (i) with the execution of subsequent modules are subsequently controlled. Once the above function, the module is executed quiz.php whose main function is to display the survey questions starting with the most general level (level 0) for the user you are answered in 15 blocks per screen, as shown in figure 2. Note that the questions are all multiple choice and employing a Likert scale (Likert, 1932) from 1 to 5, where 1 represents the minimum value of association or affinity for a particular question while 5 represents the maximum value . In this case, the variable i represents the level to which belong questions to be displayed at any given time.

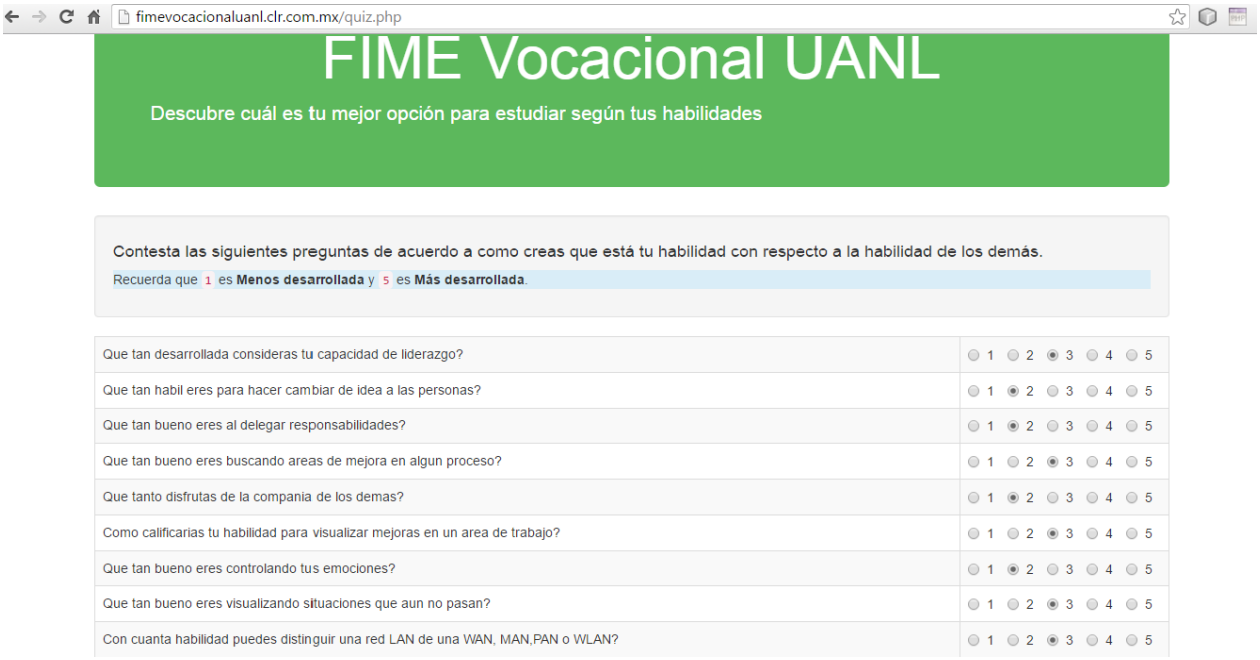


Figure 2. System Operation recommendation

Additionally, questions that are showing are extracted from the database system, which consists of two tables, races and questions. Figure 3 shows the diagram of the database (fime.vocacionaluanlBD) used by the system. On the one hand, the race table stores data for races that support the system starting with a unique identifier for each race as well as the name and date of each race. In this case, the first race (idcarrera = 1) in be stored in the database was IAS, followed by IEC and finally STI. On the other hand, the question table contains all statements of the questions and their respective levels. Fime.vocacionaluanlBD data base was designed and implemented with MySQL which is a management system databases free or open software commonly used in the development of web applications due to its functionality, flexibility and reliability (MySQL, 2016).

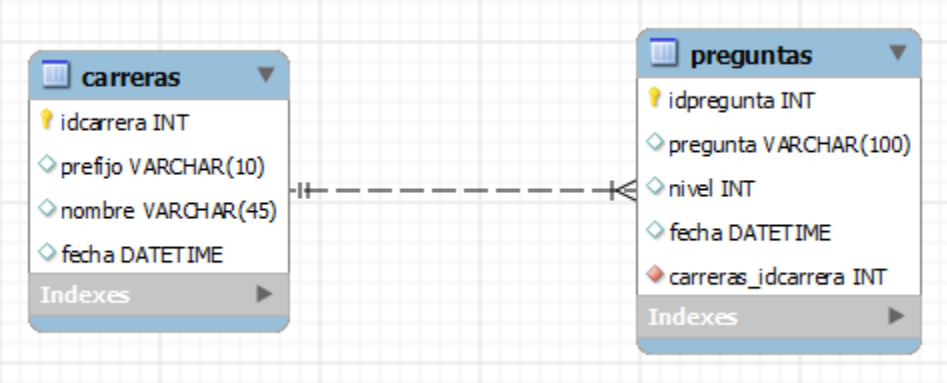


Figure 3. Database schema system

When you finish answering all fifteen questions in a specific screen, quiz.php sends the values set for these questions to the calculate.php module to be used in calculating the weights required and to determine the most appropriate vocational profile the user. Figure 1 shows that if the iteration (i) is less than 4 then control returns to quiz.php to display the next set of questions in a new screen, but not before increasing the value of the iteration variable i for and all come from a higher level. If iteration is not less than 4 then the last results.php module is executed to display the results of the evaluation, which will be explained in detail in the next section.

The aforementioned weights are necessary to determine the percentage of ownership of each question to a specific race, ie a total of 100% for each question percentages membership or association are obtained for each race. As mentioned, the initial version of the system processes three races, then let m = 3 where m represents the number of races that supports the system and n

represents the level. For example, for each question percentages for IAS, IEC and STI careers, ie by calculating the weights it is possible to determine whether IAS, IEC or STIs are a question were allocated.

Table I shows the procedure for calculating the weights according to the level and the number of supported races and equity weights for n = 0 because the level of more general questions, which mostly apply to all races processing system. Instead, from the first level onwards should see greater weight to the race to which a question belongs. For example, for n = 1, the value of the weighting 1 / (3 + 1) = 0.25 is assigned to each of the races that have no relation to the question and the remaining value to complete 1 is assigned to the race associated finally leaving the IAS = 0.25 distribution, IEC = 0.25 and STD = 0.5 for a question STI career. In general, the weights shown in Table I correspond to questions of ITS for each of the different levels, so the highest values found in this column. A big advantage is that the system was implemented so that it is able to fit any number of races dynamically, which in turn facilitates the calculation of new weights without making changes to the system logic every time you need to add a new career and their respective set of questions.

Table I. Calculating weights for each question based on the level and the number of races

Nivel (n)	Ponderación	IAS	IEC	ITS
0	1/(m + n)	0.33	0.33	0.33
1	1/(m + n)	0.25	0.25	0.5
2	1/(m + n)	0.2	0.2	0.6
3	1/(m + n)	0.16	0.16	0.67

Results and discussion

After presenting an overview of the system of vocational fime.vocacionaluanl recommendation, in this section we focus on the end of the operation what are the results and the analysis performed to validate the performance and functionality of the system. Earlier, in Figure 4 we show the sequence diagram that follows the system to be used to visualize the different flows

operating normally carried out between the user, the system and the database. As you can see, there are three main flows of operation: display, and weigh answer questions, which are repeated four times because they are four levels of questions in total (0 to 3). The average time it takes to answer the whole questionnaire (60 questions) is about 10 minutes.

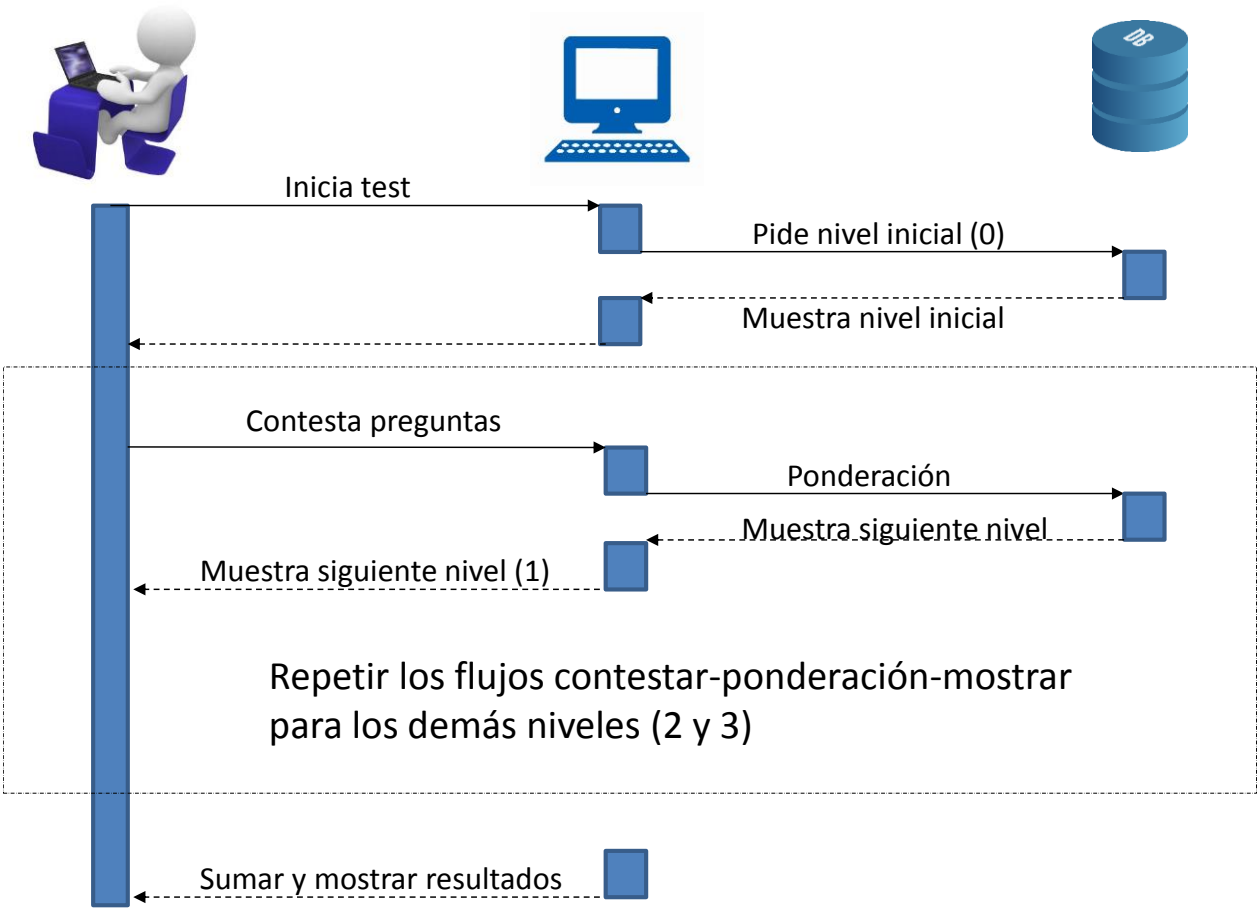


Figure 4. Sequence diagram of system operation fime.vocacionaluanl

The system is available for the moment in a space that was hired on a server to host both system software and its database. In addition, the type of service that would adapt the system hired recommendation to other university faculties using the name of the authority followed by the domain name vocacionaluanl to form your URL. For the graphic design of the system, we decided to use Bootstrap which is a special type of library (framework) Free software designed to facilitate the creation of dynamic web pages. Based on HTML and CSS languages, Bootstrap contains a variety of features to create interactive web applications very complete.

Figure 5 shows an example of how the results appear to finish answering the entire questionnaire. In this case the name of the race that determines the system more compatible with the skills, knowledge and more developed in the user interests appears.

Figure 5. Results Screen System

Finally, it is important to finish presenting an analysis showing that somehow the system has a correct and consistent performance. Therefore, since the questionnaire represents the central part of the functionality of the system recommendation, to test the validity and effectiveness of the analysis with the index or coefficient Cronbach's alpha for a sample of users was obtained respondents was conducted questionnaire. This index is commonly used to determine the consistency and reliability of a measuring instrument consisting of several elements, variables or correlated items (Cronbach, 1951). The range of values that can take this coefficient is between 0 and 1, where 0 represents no 1 reliability while absolute reliability. Typically, a value equal to or greater than 0.8 is considered high reliability.

The sample analyzed amounted to 40 and was taken from a population of students majoring STI and IAS mainly, who were asked to use the system to determine its reliability. However, it is clear that the objective of the system is, on the one hand, guide students who are still undecided

about the race they will study and, on the other, the system is able to cover all engineering careers It offered by the faculty in later versions.

Of the two forms available to estimate the scale Cronbach Alfa, we chose the following:

$$\alpha = \frac{K}{K-1}\left|1 - \frac{\sum V_i}{V_T}\right|$$

Where K represents the number of questions, $\sum V_i$ is the sum of the individual variances for each of the questions and V_T It is the total variance. Figure 6 shows a part of the analysis was performed to estimate the scale Cronbach Alfa based on 40 selected by users who participated in the survey responses. The evaluation of the data by the above formula was an index of Cronbach's alpha 0889, as seen in Fig. Therefore, being a value above 0.8 then it is concluded that the system works correctly and consistently, and generally has an acceptable level of reliability.

USUARIO	MATRICULA	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
1	1582892	4	4	5	4	5	5	4	3	5	5	5	5	3	4	5	5	5	3	4	4	2	1
2	1584080	5	3	5	3	4	1	2	5	4	5	5	4	4	5	5	4	4	5	4	5	5	5
3	1600504	5	4	5	4	4	4	4	3	4	2	4	5	3	5	3	3	4	5	3	4	3	4
4	1591424	4	5	5	4	3	4	2	3	4	3	3	5	3	4	5	5	4	5	4	4	4	3
5	1342626	4	4	3	4	5	5	4	3	3	4	4	5	4	4	3	4	3	5	3	4	3	4
6	1691299	4	4	5	3	4	4	5	5	4	3	3	5	3	4	5	4	3	5	5	4	4	2
7	1525027	5	4	3	5	4	5	4	5	2	4	4	3	5	3	4	4	3	3	5	4	4	5
8	1719272	5	5	5	5	4	4	5	5	3	4	3	4	4	4	4	4	4	3	5	5	4	4
9	1357881	5	4	4	4	5	4	3	4	2	3	3	5	5	3	4	4	3	3	4	4	3	3
10	1485401	4	5	5	3	4	4	4	5	5	5	5	3	5	4	5	5	5	5	5	4	5	5
27	1341162	3	5	3	3	5	3	2	3	3	3	2	5	3	2	2	2	5	3	3	4	5	4
28	8000994	3	2	1	3	3	4	3	2	4	2	3	2	4	1	3	3	4	2	1	1	2	4
29	1562853	3	4	2	3	5	4	2	3	2	4	4	4	5	4	2	3	3	4	5	2	3	3
30	1381517	4	5	4	4	4	4	3	5	4	3	4	5	3	4	4	5	3	5	4	4	5	4
31	1412540	4	5	3	5	4	4	3	4	5	2	4	5	3	4	5	5	5	2	4	5	4	4
32	1483821	3	3	5	4	4	3	3	5	4	5	4	2	3	3	4	5	4	2	4	3	5	4
33	1517117	3	4	2	3	5	4	5	4	3	3	2	3	4	5	5	1	2	5	4	3	3	5
34	1589421	4	5	4	2	3	1	3	5	5	4	3	4	5	5	3	5	5	3	3	5	5	4
35	1691458	3	4	5	4	5	3	1	3	4	3	3	4	3	4	2	5	1	5	4	3	4	4
36	1467182	4	4	2	4	3	3	3	3	3	3	3	2	3	3	3	5	3	2	4	2	3	2
37	1629462	3	5	5	4	3	3	2	4	4	5	5	4	3	2	4	3	2	4	4	3	1	3
38	1509095	5	4	4	4	2	3	5	5	2	3	5	5	4	4	4	5	4	4	4	3	5	5
39	1528465	5	5	4	5	4	4	5	5	3	5	5	2	4	5	5	4	4	5	5	5	2	4
40	1557557	5	3	4	4	5	4	4	4	4	4	3	3	5	2	4	3	4	3	4	4	3	4
	VARIANZAS	0.635	0.892	1.276	0.840	0.897	1.102	1.331	1.208	0.921	1.122	1.118	1.156	0.866	1.251	1.241	1.026	1.281	1.333	1.105	1.046	1.228	1.323
	INDIV.																						

K	60	K/(K-1)	1.017
Σ Vi	71.019	1-(ΣVi/Vt]	0.874
Vt	565.344	Alfa α	**0.889**

Figure 6. Estimated rate Alfa de Cronbach

Either way, the system is available to the public at the URL above, where you can test its performance at any time.

Conclusion

In this paper we have presented a vocational system online recommendation based on an algorithm called artificial intelligence as the main component to determine the best vocational profile for each user. The system aims to become a valuable tool in support of free use for all those students who are considering choosing an engineering career at FIME UANL. The system was designed and implemented with programming languages and tools widely used in this type of dynamic and interactive applications. It could be further validate the consistency and reliability of the system by means of a statistical analysis with data obtained from a sample of 40 users. It is recommended to increase in the future the number of races FIME that can support the system, and even replicate the system so that this process runs other faculties of the UANL. This has to do with research and define the set of questions that correspond to all the races you want to include and arrange them on the levels defined so that everything is stored in the database.

Thanks

We wish to thank the students of the ITS UANL career, Xavier Sanchez Diaz, Alfredo González Garza and Cynthia Reyna Elizabeth Villagomez, for his involvement in the research and development of the first version of the system of vocational recommendation; We also appreciate the support of the Professional Development Program for Teachers (PRODEP) of the Ministry of Education.

Bibliography

Adomavicius, G., y Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6), pp. 734-749.

Aher, S. B., & Lobo, L. M. R. J. (2013). Combination of machine learning algorithms for recommendation of courses in E-Learning system based on historical data. Knowledge Based Systems, 51, pp. 1-14.

Alcaraz, S., Sánchez, X., González, A. (2015). Encuesta orientauanl. FIME UANL. Sitio web: http://bit.ly/orientaUANL

Beati, H. (2011). PHP: Creación de páginas Web dinámicas. Buenos Aires: Alfaomega Grupo Editor Argentino.

Betz, N. E., Fitzgerald, L. F., Hill, R. E. (1989). Trait-factor theories: Traditional cornerstone of career theory. In M. B. Arthur, D. T. Hall, & B. S. Lawrence (Eds.), Handbook of career theory. Cambridge: Cambridge University Press, pp. 26-40.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of test. Psychometrika, 16(3), pp. 297-334.

Duda, R. O., & Hart, P. E. (1973). Pattern classification and scene analysis. John Wiley & Sons, Inc.

Franca, D. (2012). Decide tu carrera, orientación vocacional para jóvenes mexicanos. Relpe. Sitio web: http://www.relpe.org/decide-tu-carrera-orientacion-vocacional-para-jovenes-mexicanos/

Gati, I., Fishman-Nadav, Y., Shiloh, S. (2006). The relations between preferences for using abilities, self-estimated abilities, and measured abilities among career counseling clients. Journal of Vocational Behavior, 68, pp. 24-38.

Holland, J. L. (1997). Making vocational choices: A theory of vocational personalities Uand work environments. Odessa, FL: Psychological Assessment Resources.

Kasabov, N. K. (1996). Foundations of neural networks, fuzzy systems, and knowledge engineering. Cambridge: MIT Press.

Katz, M. R. (1993). Computer-assisted career decision making. Hillsdale, NJ: Erlbaum.

Leavitt, N. (2006). Recommendation Technology: Will It Boost E-Commerce? IEEE Computer Society.

Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, 140, pp. 1-55.

Linden, G., Smith, B., York, J. (2003). Amazon.com Recommendations Item-to-Item Collaborative Filtering, IEEE Internet Computing.

Sobecki, J. (2014). Comparison of selected swarm intelligence algorithms in student courses recommendation application. International Journal of Software Engineering and Knowledge Engineering, 24(1), pp. 91-109.

Truemper, K. (2004). Design of logic-based intelligent systems. Wiley-Interscience.

Witten, I. H., & Frank, E. (2005). Data mining: Practical machine learning tools and techniques. Second Edition, Morgan Kaufmann.