# An Ontology to Trace the Computer Science Student Profile

Ms. Sunny Kelma Oliveira Miranda
*Department of Computer Engineering*
*University of Coimbra*
Coimbra, Portugal
sunnymiranda@dei.uc.pt

Dr. Maria José Marcelino
*Department of Computer Engineering*
*University of Coimbra*
Coimbra, Portugal
zemar@dei.uc.pt

Dr. Paula Alexandra Silva
*Department of Computer Engineering*
*University of Coimbra*
Coimbra, Portugal
paulasilva@dei.uc.pt

***Abstract*—This study proposes an Ontology for Computer Science (CS) students' profiles. This Student model will represent information about the CS student and will be created by extracting constructs and variables from applied questionnaires. The literature presents several instruments that analyze the student's profile and performance in STEM and Computer Science. Some questionnaires were used to determine the profile of CS students to extract constructs and variables such as (a) Sociodemographic and economical; (b) Personality Type; c) Academic performance in secondary education; d) Learning experiences; f) Attitudes toward Programming; g) Science Motivation. h) Parental/Social Support. Our overarching research question of this study is to answer: How can we build an ontology for the computer science student profile that identifies the main factors that motivated him/her to choose a course in this area? This study will build the proposed ontology using the Ontology Development 101 methodology. The proposed ontology will be validated with real-world data and intended to develop case studies to administer internationally validated questionnaires to university students enrolled in Computer Science Major-related courses in public universities. The expected results are to know and trace the profile of CS students (female and male) since secondary education to promote better actions that increase the interest of secondary school students, mainly girls, in the IT area. Thus, this work intends to build an Ontology of CS Students as part of a recommendation system that will make suggestions for actions that encourage more students to pursue a career in Computer Science and then, be able to increase the supply of IT professionals in the market: workforce, especially women.**

***Keywords—Computer Science Student, Computer Science Major, Secondary School Students, Ontology***

## I. Introduction

The motivation for this study is that women are underrepresented in the IT field. Whilst when the number of job opportunities in IT has never been greater, only 22% of all technology roles in European IT companies are held by women [1]. The European Institute for Gender Equality [2] has affirmed that attracting more women to Science, Technology, Engineering and Mathematics (STEM) jobs would boost the job market where labor shortages are envisaged. The McKinsey technical report [1] says that if the percentage of women in technology jobs were to double by 2027, the GDP of the European Union could increase by up to 600 billion euros. However, the percentage of women graduating in STEM subjects is decreasing, where only 19% of bachelor's degree students in ICT are women [1]. Therefore, attracting, retaining, and developing female talent in the IT area is a business imperative.

The principal objective of this study is to create an ontology for the Computer Science (CS) student profile, identifying the factors that most influenced the choice of course in the CS area. This ontology is the first part of a more extensive work that intends to use this ontology to develop a recommendation system to encourage more students, especially girls, to choose CS in higher education as their professional area. For that, initially, it is crucial to determine the motivations and factors that can lead a student to choose a CS degree and persist in this area. Then, it is essential to implement actions that could effectively trigger a secondary student's interest in pursuing a CS degree and a career in IT.

According to [30], two primary questions must be answered while constructing a student model: What information about students must be represented in a student model, and What approaches must be adopted for modeling that information? Thus, the overarching research question of this study is to answer: How can we build an ontology for the computer science student profile that identifies the main factors that motivated him/her to choose a course in this area? For this, specific research questions were identified as follows).

*1) RQ1:* What factors lead a student to choose CS as a higher education course?

*2) RQ2:* How do students' performance and learning experiences during secondary school influence his/her choice?

*3) RQ3:* Is it possible to identify similar characteristics between female and male students who choose a course in the CS area.

*a) Is it possible to identify profile patterns?*

*b) Is there a relationship between the profile and the course choice?*

Therefore, this study intends to determine the CS student (female and male) profile based on data collection through the application of validated questionnaires that will identify her/his interests in the IT area and her/his learning experiences since pre-university education, her/his personality traits and learning style. Thus, these data will serve to validate the proposed ontology.

This paper is structured in four sections, as follows. The background in session II provides an overview of the factors that motivate this study. Session III presents the ontology proposal. Session IV reports the brief conclusions and future work.

## II. Background

This section presents an overview of the main topics that support this study.

### A. Student Motivation and Educational Aspirations

In investigating students' motivation, we studied two main theories of Psychology that explain the motivation for learning and the student's choice of higher education and career: the

2023 IEEE Frontiers in Education Conference (FIE) | 979-8-3503-3642-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/FIE58773.2023.10342999

Social Cognitive Theory – SCT [3] and the Social Cognitive Career Theory - SCCT [4]. From the SCT of Bandura [5], a relevant construct emerges - human agency - that comprises the individual interacting with the social environment, where personal, behavioral, and environmental factors interact reciprocally, promoting changes and being modified by the individual. Self-efficacy and self-regulation are two further essential concepts from Bandura's Theory. Self-efficacy refers to an individual's belief in his/her capacity to execute behavior's necessary to produce specific performance attainments. Self-regulation allows the subject to voluntarily control feelings, thoughts, and actions, aiming to achieve established goals. All these are key factors in students' motivation. The SCCT derives from Bandura's Theory, and it focuses on personal, cognitive and contextual factors for career/academic choices. Thus, these theories will be considered to build the CS student ontology.

### B. *Factors that Encourage the Choice of STEM/CS*

STEM education in schools develops 21st-century skills like creativity, critical thinking, collaboration, and innovation [6]. The level of STEM career knowledge a student has will directly influence a student's interest in pursuing a future STEM career [7]. The main factors influencing a STEM career are learning experiences, motivation beliefs, and positive outcome expectations [8]. Moreover, the factors influencing the plan of a Computer Science Career are listed as prior experiences in Computing, interest in the IT field, promising career, and high salary [9]. To encourage girls' participation in Computer Science, parental authority and female role models/social support are some of the factors most cited as influencing the choice [10]. It is, therefore, important to capture these factors when mapping the CS student profile.

### C. *Assessing Student Profiles in STEM and Computer Science Areas.*

Previous research has produced instruments to analyze students' profiles and STEM and Computer Science performance. Internationally validated questionnaires were found that extracted constructs and variables to identify motivations for choosing a CS course, academic performance, learning experiences, attitudes toward programming, personality type, learning styles and parental/social support, as shown below.

*1) Questionnaire 1 - Motivation:* The instruments measure the motivating factors to start and continue studying IT [11]. The questionnaire uses a 5-point Likert-type scale (5 - very important to 1 - not at all important) to assess the importance of each item in influencing student motivation to study IT. Questionnaire items are divided into three categories: 1) intrinsic motivation (a person does something because it is interesting or enjoyable); 2) extrinsic motivation (a person does something because it leads to a separable result); and 3) unspecified [11].

*2) Questionnaire 2 - Learning style:* N-ILS (New Index of Learning Styles) [12] - This is a new and reduced version of the Felder and Silverman ILS [13] test to identify learning styles. The Felder and Silverman model is composed of four dimensions, which represent the stages of learning, wherein each of them the student tends to a pole: 1) Perception (sensitive or intuitive); 2) Input (visual or verbal); 3) Processing (active or reflective); 4) Understanding (sequential or global). The N-ILS consists of 20 questions in which each of the four bipolar dimensions has five questions related to it. The calculation of preferences is done as Felder and Soloman model. Each dimension has two response options, with the participant being asked to choose between the letter "a" or "b" to indicate their response to each question [28].

*3) Questionnaire 3 - Personality types:* Ten Item Personality Inventory (TIPI) [14] - It is a brief personality assessment scale by Gosling et al. [15], whose Portuguese version - Ten Items Personality Inventory by Lima and Castro [14] - is based on the Five Great Factors model by Costa and McCrae [26], consisting of 10 items that are assessed using a scale 7-point Likert type, ranging from (1) "Strongly Disagree" to (7) "Strongly Agree". This assessment measure contains two items for each of the five factors: Openness to Experience, Agreeableness, Conscientiousness, Emotional Stability, and Extroversion. [27].

*4) Questionnaire 4 - Interview:* The instrument was a semi-structured interview with questions pre-planned and constructed using Social Cognitive Career Theory as a foundation to identify and understand reasons why students choose to study Computer Science at university [16].

Thus, the proposed ontology in this study will be built considering the questionnaires presented above.

### D. *Ontology and Recommender System*

An ontology is an explicit formal description of concepts in a domain that consists of objects, classes, properties, relationships, rules and constraints [17]. The construction of ontologies is a complex and expensive process, but there are already tools that propose to overcome this problem by automatically creating ontologies [18]. An ontology representation language is used to formally describe concepts and properties between them. The most common languages are Web Ontology Language (OWL), Resource Description Framework (RDF), DARPA Agent Markup Language (DAML) and Ontology Inference Layer (OIL) [17]. These coding languages often include predefined semantics and reasoning rules that support the processing of this knowledge [18].

Ontologies are used in different domains and for different purposes. An ontology can be classified as a domain ontology representing knowledge about a specific part of the world [18]. The recommender systems can use domain ontology to establish relationships between users and their preferences about the recommendation subject. For example, according to Tarus et al. [18], most knowledge-based e-learning recommender systems use ontologies to represent knowledge about the learner and learning resources. In this case, the ontology establishes the relationship between students and their preferences about learning resources. Since ontology-based recommenders rely more on domain knowledge than classifications, they do not experience problems such as cold-start, rating sparsity and overspecialization [18].

The According to García et al. [19] ontologies have been shown to be the most suitable structure to represent knowledge models. In addition to ratings and similarity of users or items [20], educational domain recommender systems should consider additional student information such as prior knowledge, background, and learning style during the recommendation process [18].

Although this work is not directed to e-learning, we tried to study how ontologies were applied in the context of an educational recommendation system.

### E. *Creating a Computational Ontology*

Computational ontology formally represents knowledge in a particular domain, typically created using computer-based tools and methods. Creating a computational ontology involves identifying the key concepts, the definition of the categories, properties, and relationships between the concepts, data, and entities that substantiate one within a domain, defining them in a formal language such as RDF or OWL, and then using tools such as reasoning engines and query languages to manipulate and use the ontology.

The literature presents ontology construction methodologies and methods from the mid-1990s to the present; however, Ontology Development 101 [23] stands out as one of the most used [29]. Ontology Development 101 has an iterative nature, and its script advocates a method supported by fundamental rules that support the decisions that permeate the construction work. The development stages contemplate the whole process, including defining class hierarchies, properties, and individuals [23].

This work will adopt the Ontology Development 101 methodology, and the development process will be presented in the next section.

## III. Ontology Proposal

This section presents the construction of an ontology to represent the domain knowledge involving CS student profiles. The proposal followed the Ontology Development 101 methodology [23] for its construction according to the seven steps specified below.

### A. *Step One: Determine the Domain and Scope of the Ontology*

This phase must contain information about the domain, scope, and general purpose. Thus, the CS Student Ontology will structure a formal representation of an undergraduate student model of a course related to Computer Science. The ontology intends to be comprehensive in the classes but not exhaustive in the student's personal and academic data attributes. This ontology will be part of a knowledge-based recommendation system to be developed later.

### B. *Step Two: Consider Ontology Reuse*

At this phase, the reuse of ontologies should be considered; however, no ontologies published and ready for import precisely like the proposal in this work were found. In any case, references to ontologies already built and described in academic articles were used to create the proposed model. In addition, the OWL content of ontologies that will support this proposal will be imported using the Protégé 5.6.1 ontology editor [25] to start the specific ontology development for CS students.

### C. *Step Three: List Important Ontology Terms*

This step involves selecting domain reference materials and adopting knowledge extraction methods. For extracting knowledge and identification of domain concepts, this ontology model will be created from constructs and crucial variables extracted from the questionnaires, as mentioned above. The identification of domain concepts (see Figure 1) and some properties are listed below.

- Personal identity (Birth date; Gender; Nationality; City of birth; City of Residence; Father's profession; Mother's profession).
- Course (Faculty, Course Name; Course year).
- Learning style (Active / Reflective; Sensitive / Intuitive; Sequential / Global; Visual / Verbal).
- Personality type (Extraversion; Agreeableness; Conscientiousness; Emotional Stability; Openness to Experiences).
- Motivations and Interest in IT area (Interest in computers; Prior contact with IT; Knowledge in the IT field; good salary in IT field; Good prospects in the job market; ICT teacher; IT hobby or competitions).
- Learning experiences (Programming Experience Level; Robotic Experience Level; Math self-assessment; Science self-assessment; TIC self-assessment; Art self-assessment).

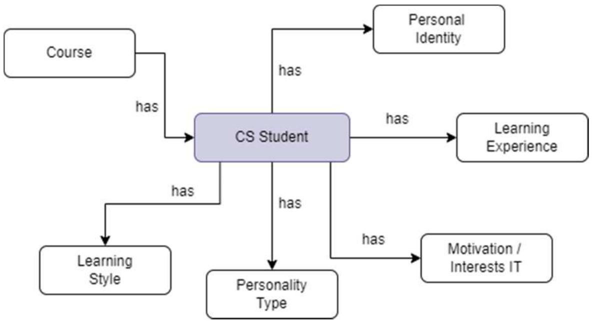


Figure 1: Conceptual Map of the Ontology Proposed at High Level

### D. *Step Four: Define the Classes and Their Hierarchy*

For the definition of classes, combined development was chosen. According to Ontology Development 101 [29], a combined development is the combination of top-down and bottom-up. In this form of development, the most outstanding concepts are defined first, which are then generalized and specialized appropriately.

In this first iteration, the following concepts were identified and modeled as classes using Protégé 5.6.1 ontology editor, as shown in Figure 2.

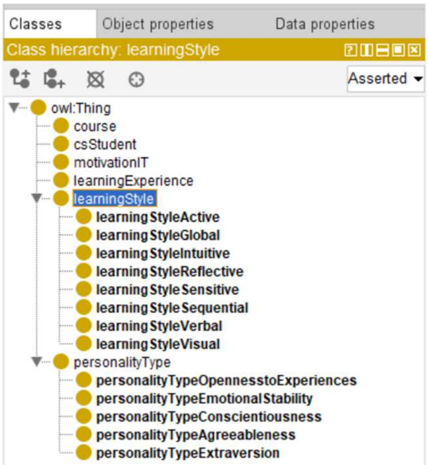


Figure 2: Class Hierarchy in the Protégé of the Proposed Ontology

### E. *Step Five: Define the Properties*

In OWL, properties represent relationships between two individuals. There are two types of properties: Object Property and Data Property. Object Property connects two individuals (two classes instances). The Data Property, in turn, connects an individual to a value with a well-defined data type (numeric, strings) [25].

Properties are considered independent entities, and they are linked to classes after their creation. Figure 3 presents examples of data properties linked to *CS Student* and *course* classes; and object properties linked to *CS Student class* with *Learning Style, Personality Type* and *Course* classes.

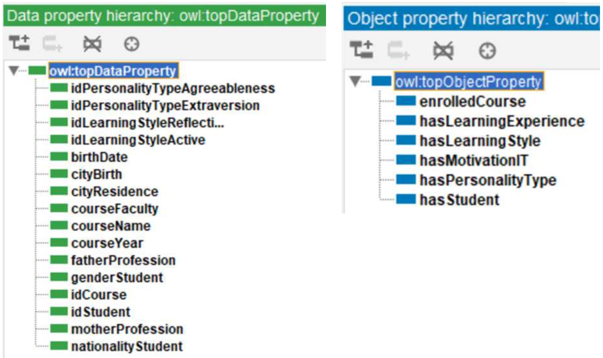


Figure 3: Example of Data and Object Properties of the Proposed Ontology

## F. Step Six: Detail and Describe the Classes, Types of values, Quantities, Axioms and Rules

In the sixth step of the script, the types of values are defined, values that are allowed, the cardinality of the properties and other characteristics that the properties can have [23]. Definition of domain and range of object properties. When a semantic model in OWL is built, it can define, for a given property, its domain (domain) and range (range), as shown in Figure 4.

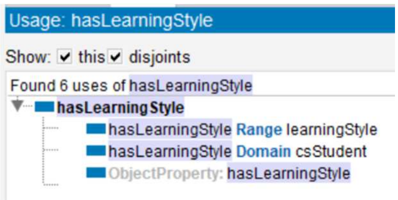


Figure 4: Example the definition of domain and range to "hasLearningStyle" property

## G. Step Seven: Create Instances

Defining individual instances of classes is a process that requires: (1) choosing a class; (2) creating an individual instance for the class, and (3) populating data property values and establishing object property connections.

Instances were created for all classes in this ontology and relationships were also created to object properties. After that, tests were carried out with fictitious data to validate the construction rules. Below are examples of questions asked to Reasoner through Protégé Software's DL Query. See Figure 5 with queries results.

*Q1:* What are the motivations for choosing an IT course?
*Q2:* Who responded with the motivation "great jobs"?
*Q3:* Who answered liking programming as motivation and being female?
*Q4:* Who has learning experience in robotics?
*Q5:* What are the identified learning styles?
*Q6:* What are the identified personality types?

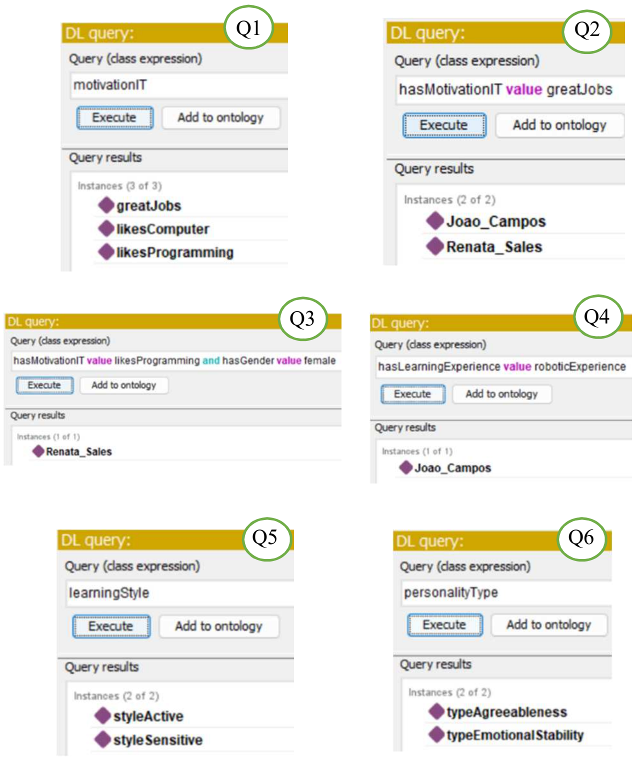


Figure 5: Examples of Tests with Instances to Validate Ontology Rules.

Further on, the goal is to carry out validation with actual data collected through internationally validated questionnaires that will be administered to university students enrolled in Computer Science Major-related courses in Portugal public universities.

## IV. Conclusion and Future Work

This work addresses the problem of the need for more women in the IT field. The literature presents the reasons and factors that lead a student to choose a Computer Science course and follow a professional career, as well as the barriers that discourage this choice. Thus, intending to attract more girls and boys to the area of computing, the present study proposes an ontology to trace the Computer Science student profile covering his/her academic pathway to university, motivation, and influences for choosing a CS course, personality type and other characteristics, as the first step to developing an ontology-based recommender system. By better understanding the profiles of CS students, it will be possible to recommend better interventions in secondary education to encourage more students, especially girls, to choose a CS course.

As this work is still in progress, the next steps are 1) perform the tests with real-world data applying the questionnaires mentioned in this work during the case studies, and, finally, 2) build the knowledge-based recommender system using the proposed ontology.

The expected result of the system implementation can contribute to guiding public policies to increase the supply of computer skills to secondary school students. Moreover, it may lead to the reduction of gender inequality and the achievement of a more economically balanced, innovative, and sustainable society.

## REFERENCES


[1] McKinsey & Company (2022). Women in tech: The best bet to solve Europe's talent shortage. Sven Blumberg, Melanie Krawina, Elina Mäkelä, and Henning Soller. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/women-in-tech-the-best-bet-to-solve-europes-talent-shortage

[2] EIGE - Women and men in ICT: a chance for better work–life balance. 2018. DOI: 10.2839/310959

[3] Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ: Prentice Hall.

[4] Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. Journal of vocational behavior, 45(1), 79-122

[5] Bandura A. (2006). Toward a Psychology of Human Agency. Perspectives on Psychological Science. 2006;1(2):164-180. doi:10.1111/j.1745-6916.2006.00011.x

[6] Kaufman, K. J. (2013). 21 ways to 21st century skills: why students need them and ideas for practical implementation. Kappa Delta Pi Record, 49(2), 78-83.

[7] Razali, F. (2021). Exploring Crucial Factors of an Interest in STEM Career Model among Secondary School Students. International Journal of Instruction, 14(2), 385-404.

[8] Wang, X. (2013). Why students choose STEM majors: Motivation, high school learning, and postsecondary context of support. American Educational Research Journal, 50(5), 1081-1121.

[9] Sax, L. J., Newhouse, K. N., Goode, J., Nakajima, T. M., Skorodinsky, M., & Sendowski, M. (2022). Can Computing be diversified on "principles" alone? Exploring the role of AP Computer Science courses in students' major and career intentions. ACM Transactions on Computing Education (TOCE), 22(2), 1-26.

[10] Sun, C., & Clarke-Midura, J. (2022). Recruiting K-12 youth into Computer Science: insights from a multi-year research project. ACM Inroads, 13(2), 22-29.

[11] Kori, K., Pedaste, M., Altin, H., Tõnisson, E., & Palts, T. (2016). Factors that influence students' motivation to start and to continue studying information technology in Estonia. IEEE Transactions on Education, 59(4), 255-262.

[12] Vieira Junior, N. (2012) Planejamento de um ambiente virtual de aprendizagem baseado em interfaces dinâmicas e uma aplicação ao estudo de potência elétrica. 2012. 234 p. Tese (Doutorado em Engenharia Elétrica) - Universidade Estadual Paulista, Ilha Solteira.

[13] Soloman, B. A., & Felder, R. M. (2005). Index of learning styles questionnaire. NC State University. Available online at: https://www.webtools.ncsu.edu/learningstyles/

[14] Gosling, S. D., Rentfrow, P. J., & Swann J.W. B. (2003). A very brief measure of the Big-Five personality domains. Journal of Research in Personality, 37, 504-528. doi:10.1016/S0092-566(03)00046-1

[15] Lima, C., & Castro, S. (2009). Ten-Item Personality Inventory, Portuguese version. Retrieved 10 abril 2023, http:/homepage.psy.utexas.edu/gosling/scales_we.htm.

[16] Alshahrani, A., Ross, I., & Wood, M. I. (2018, August). Using social cognitive career theory to understand why students choose to study Computer Science. In Proceedings of the 2018 ACM conference on international Computing education research (pp. 205-214).

[17] Poelmans, J., Ignatov, D. I., Kuznetsov, S. O., & Dedene, G. (2013). Formal concept analysis in knowledge processing: A survey on applications. Expert systems with applications, 40(16), 6538-6560.

[18] Tarus, J. K., Niu, Z., & Mustafa, G. (2018). Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. Artificial intelligence review, 50(1), 21-48.

[19] García, I., Benavides, C., Alaiz, H., & Alonso, A. (2013). A study of the use of ontologies for building computer-aided control engineering self-learning educational software. Journal of Science Education and Technology, 22(4), 589-601.

[20] Zhao, X., Niu, Z., Chen, W., Shi, C., Niu, K., & Liu, D. (2015). A hybrid approach of topic model and matrix factorization based on two-step recommendation framework. Journal of Intelligent Information Systems, 44(3), 335-353.

[21] Mendonça, F.M. & Soares, A.L.. (2017). Bulding ontologies with ontoforinfoscience methodology: A detailed approach to the activities of ontology development. Ciencia da Informacao. 46. 43-59. https://revista.ibict.br/ciinf/article/view/4013.

[22] Gómez-Pérez, A. and Fernández, M. and Vicente, A. de (1996). Towards a Method to Conceptualize Domain Ontologies. In: "12th European Conference on Artificial Intelligence (ECAI'96)", 13/08/1996, Budapest, Rumanía.

[23] Noy, N. & Mcguinness, Deborah. (2001). Ontology Development 101: A Guide to Creating Your First Ontology. Knowledge Systems Laboratory. 32.

[24] Suárez-Figueroa, M. C. D5. 4.1. NeOn Methodology for Building Contextualized Ontology Networks. NeOn project. Avaiable in http://www.neon-project.org

[25] Horridge, Matthew. (2011). A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools Edition 1.3.

[26] McCrae, Robert & Costa, Paul. (1987). Validation of the five factor model of personality across instruments and observers. Journal of personality and social psychology. 52. 81-90. 10.1037/0022-3514.52.1.81.

[27] Brito-Costa, Sonia & Bem-haja, Pedro & Moisao, Ana & Alberty, Afonso & Castro, Florencio & Almeida, Hugo. (2016). Pychometric properties of portuguese version of big five inventory (bfi). International Journal of Developmental and Educational Psychology. Revista INFAD de Psicología.. 1. 83. 10.17060/ijodaep.2015.n2.v1.325.

[28] A. C. O. Duarte and D. L. do . Nascimento, "Using the New Index of Learning Styles (ILS) to determine the Learning Styles of High School students and teachers", RSD, vol. 10, no. 13, p. e285101321209, Oct. 2021.

[29] Uma ontologia de perfil holístico para estudantes de graduação. Claiton Custódio da Silva. Brasília. 2021

[30] AISSAOUI, Ouafae EL e Lahcen OUGHDIR: A literature review on student modeling purposes. 2020. https://www.researchgate.net/publication/341091023_A_literature_Review_on_Student_Modeling_purposes