ISSN 2007-9737

Semi-Automatic Creation of Ontologies from Unstructured Pedagogical Texts to Assist in Significant Learning

Yuridiana Alemán, María J. Somodevilla, Darnes Vilariño

Benemérita Universidad Autónoma de Puebla (BUAP),
Faculty of Computer Science,
Mexico

{yuridiana.aleman,mariajsomodevilla,dvilarinoayala}@gmail.com

Abstract. This article shows the advances of an methodology for ontological learning research applied to a pedagogical subdomain. The main proposal focuses on the use of Natural Language Processing and Information Retrieval techniques in all phases of ontological learning process. For analysis, Spanish language, with resources related to significant learning and the personalization of teaching were used. Theoretical elements that are considered important to help in synchronous learning are analyzed: learning styles, types of intelligences and learning strategies. In particular, this work describes the final step (evaluation) through a handcrafted process for building ontologies and the comparison with the obtained results using the semi automatic methodology. Results show a high level of similarity between both process, making the methodology viable to apply to other subdomains of the pedagogical area.

Keywords. Ontology learning, pedagogical domain, significant learning, learning styles, intelligent types, learning strategies.

1 Introduction

An ontology is an explicit formalization, where a text is represented by means of semantic relationships and keywords. There exist several classifications of this structure, the most common being domain ontologies, which are generalizations about specific tasks in a specific area of knowledge. These representations have become one of the most used resources when structuring information. Its use ranges from the formalization of a specific domain to the creation of information retrieval systems for semantic queries.

By analyzing the pedagogical domain, learning is defined as a change in behavior due to experience [4]. Other definitions include elements related to didactics, which are described as the process of acquiring a disposition, relatively durable, to change perception or behavior as a result of an experience [1].

Significant learning is defined as a process that is generated in the human mind when it subsumes new information in a non-arbitrary and substantive way [2]. To address this issue, the personalization of learning is analyzed, where a student learns better with certain schemes and techniques according to their intrinsic characteristics [5].

The general objective of the research consists of the construction of ontologies in a semi-automatic way from unstructured pedagogical texts to express knowledge with clarity and precision. For the study domain, topics related to the implementation of teaching techniques within the classroom are used, this article focuses on the phase of manual creation of the ontologies of mentioned topics, which will be auxiliary in the evaluation of the semi-automatic processes implemented. Also, final results of methodology application were compared with the designs obtained.

2 Related Work

There are some researches about ontologies of a particular subject or topic, in [11] the project

OURAL (Ontologies for the Use of digital learning Resources and semantic Annotations on Line) is proposed, which integrates the disciplines of educational sciences, computer science and cognitive psychology in order to create services for E-learning.

In [7], the educational domain is also analyzed, however, when applied to the Chinese language, they use preprocessing to analyze the characteristics of said language: coupling, relevance and consensus.

In [27], an Internet education system based on ontologies is developed, which implements the exchange and reuse of learning material in different systems. It is a qualitative investigation, where an example is approached with a basic online computer course that describes the modules of the system: learning, interface and resources.

[30] shows the design of learning sequences using formalization through ontologies (OWL language) is presented; This design focuses on individual learning in order to establish a personalized sequence according to the level and characteristics of each student in higher education in virtual environments.

[24] was focused on autonomous online learning, proposing an ontology based on the Internet of Things (IoT) applied to learning within the classroom with the help of technology, taking as a reference the types of students intelligences.

The proposed research work focuses on the shaded area, that is, the construction of the ontology will be as a tool for face-to-face classes. Within this area, [15] he proposes to use an ontological model for the personalization of learning that involves the profile of the students according to Howard Gardner's theory of multiple intelligence, as well as using a domain ontology that helps represent knowledge in virtual learning platforms. This model is used in online education, where the type of intelligence of the students is inferred and relevant content is recommended accordingly. In [21] he presents a reengineering of a learning style ontology in various approaches, which has the objective of supporting the creation, adaptability and use of learning objects.

Recent researches in this topic combine other kinds of techniques, [14] proposes a fuzzy ontology of the academic discipline taking into account the educational material content, its complexity characteristics and studying time. In this research, the individual learning path for a specific discipline was built. In the other hand, [6] analyzes the importance of use ontologies in the design of non-linear learning witch make connections between stuff students already knew and the stuff they didn't, for the active construction of their knowledge. In a general perspective, in [23] an Ontology has been constructed to give a generic view of an Educational Organization and some of its related concerns, the Ontology has some capability to describe a semantic web based on knowledge sharing.

Works analyzed show a strong handcrafted focus, especially in pedagogical domain and the oldest. Thus, this article get importance, which in addition to having an handcrafted approach, is supported by a semi-automatic methodology for the elements extraction for ontologies adapting artificial intelligence techniques. Also, the analyzed papers involve only a principal classes, with out a ontology integration process.

3 Theoretical Concepts

The word Ontology is derived from the Greek ontos (study of being) and logos (word). In computer science, an ontology is defined as a database that describes the concepts in the world or some domain, some of their properties and how the concepts are related to each other [25]. This database is defined from a base corpus; from which they extract the main elements or keywords. Subsequently, the relationships between keywords are inferred from the same text, in this way, a graph structure is created where the node are the keywords and the edges represent the relationship between them. Among the most representative applications of ontologies are the formal representation of knowledge, which facilitates the management and integration of data with different structures. An ontology is created through a process called ontological learning, which is based on unstructured text and is defined as the process of identifying terms, concepts, relationships and, optionally, axioms from textual information and using them to build and maintain an ontology [28].

In an ontology, the concepts represent the basis for the description of the information. This description is made through three components: Terms, attributes and relationships.

Terms are names used to refer to a specific concept that can include a set of synonyms that specify the same concepts. Attributes describe the concept in detail using characteristics, and relationships are used to represent correspondences between different concepts and provide a general structure of the ontology [20]. In [12], the following preliminary criteria are proposed for the design of ontologies: clarity, consistency, extensibility, minimal coding bias and minimal ontological commitment.

3.1 Significant Learning Elements

The theory of significant learning is a theory that, probably because it deals with what happens in the classroom and how to facilitate the learning that is generated in it, has had a profound impact on teachers [19]. To address this issue, the personalization of learning is analyzed, where a student learns better with certain schemes and techniques according to their intrinsic characteristics. [5] defines personalization as adapting the learning experience to each student, all through the analysis of the knowledge, skills and learning preferences of each individual.

Personalization in the teaching-learning process can be conducted following different theories that relate various aspects, given these reflections, to investigate this area, it is proposed to analyze three specific topics in order to detect important concepts that allow establishing the relationship between these. Two of these topics are related to the student (multiple intelligences and learning styles) and one is related to the teachers (teaching-learning strategies). In the following subsections we will analyze each one of these. Figure 1 shows the relationship between the topics selected for study.

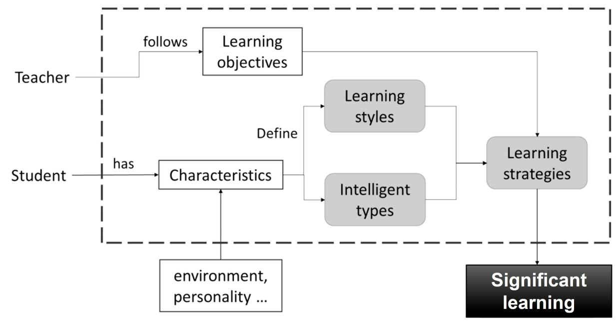


Fig. 1. Roll of selectioned elements in significant learning

— **Learning Styles.** Several theories have been proposed to describe the different types of learning. For this research, the David Kolb [13] model was taken as a reference, in which a learning style is determined using a scale called the Learning Style Inventory (LSI). The theory considers the psychological processes of perception and processing [17]. The method proposes 4 learning styles: active, reflective, pragmatic and theoretical. To determine the predominant learning style, the Honey Alonso Learning Styles Questionnaire (CHAEA) is used, which contains 80 items, 20 related to each style.

— **Multiple Intelligences.** Human beings possess a range of capacities and potentials that can be used in many productive ways, both together and separately, this idea gives rise to Gardner's multiple intelligences [8]. According to the author, intelligence implies the necessary ability to solve a problem or to elaborate products that are important in a cultural context, therefore, he distinguishes eight types of intelligences: logical-mathematical, linguistics, spatial, musical, corporal-kinesthetic, intrapersonal, interpersonal and naturalist. To determine the predominant type of intelligence, the Howard Gardner test is used, which contains 36 items.

— **Teaching-Learning Strategies.** A learning strategy is a set of procedures that a student uses in a conscious, controlled and intentional way as flexible tools to learn and solve problems [3], they can also be defined as behaviors and thoughts that a learner uses during learning with the intention of influencing their coding process [26]. According to [9], learning strategies are behaviors that the student develops during their learning process, influencing their information coding process. Although there are many approaches to classifying learning strategies, [10] defines three main types: cognitive, metacognitives and support strategies.

4 Application of the Methodology

Figure 2 shows the general methodology for the research, which consists of three main phases: Resources compilation, ontology creation and evaluation.

The corpus used for the experiments (first phase) is made up of articles in Spanish languaje related to the main topics selected (types of intelligences, learning strategies and learning styles), in such a way that three initial classes are created. Later, a process of class validation and expansion of an initial corpus were carried out in order to increase the vocabulary richness. Techniques such as classification. Clustering analysis and document filter were implemented, obtaining results higher than 90% in class validation process. These results made it possible to design a relevant method for initial corpus expansion, without affecting the data quality.

For the second phase, the main concepts and the taxonomic and non-taxonomic relationships are obtained. Before these process, a method for get compound terms was needed, according with the data structure, therefore, a process based on the use of conditional probability was implemented to detect collocations. For detecting principal concepts, techniques were focused on textual similarity metrics and part-of-speech analysis. The obtained results were higher for learning styles class, and lower for learning strategies class, this behavior was validated for a domain expert, due to the theoretical content for each class.

Finally, an evaluation module is attached to determine the experiments effectiveness. This process was in two aspects: Information retrieval metrics and comparison with handcrafted made ontologies. For information retrieval metrics, a list of principal concepts (gold standard) was needed. The gold standard was generated manually with the help of the domain expert and contains a number of concepts per class considered important.

Table 1. Elements of the set evaluation per class (extract)

Learning styles	Learning strategies
TIC	Aprendizaje
Reflexivo	Estrategia
Estudio	Conocimiento
Pragmático	Metacognitiva
Académico	Análisis
Intelligence types	**Common**
Persona	Docente
Inteligencia múltiple	Estrategia
Desarrollo	Estudiante
Teoría	Habilidad
AGardner	Individuo

The lists contain 130 items for the types of intelligences, 87 for the learning strategies and 184 for the learning styles, however, these lists have elements in common. Figure 3 shows the Venn diagram with the number of elements that are shared per class.

The three classes share 6 concepts, while the learning styles class is the one that shares a greater number of elements with other classes (14 with learning strategies and 20 with types of intelligences). The learning strategies class, despite having more text than the others, has fewer important concepts and shares few with other classes. Table 1 contains examples of some concepts from the assessment set by class and the that are present in all three sets.

For the semi-automatic results comparison with handcrafted made ontologies, a methodology proposed by [16] was used. The authors uses the following steps:

1. **Determine the Domain and Scope of the Ontology**: It is determined what the ontology is developed for, who will use it and what type of information it will provide.

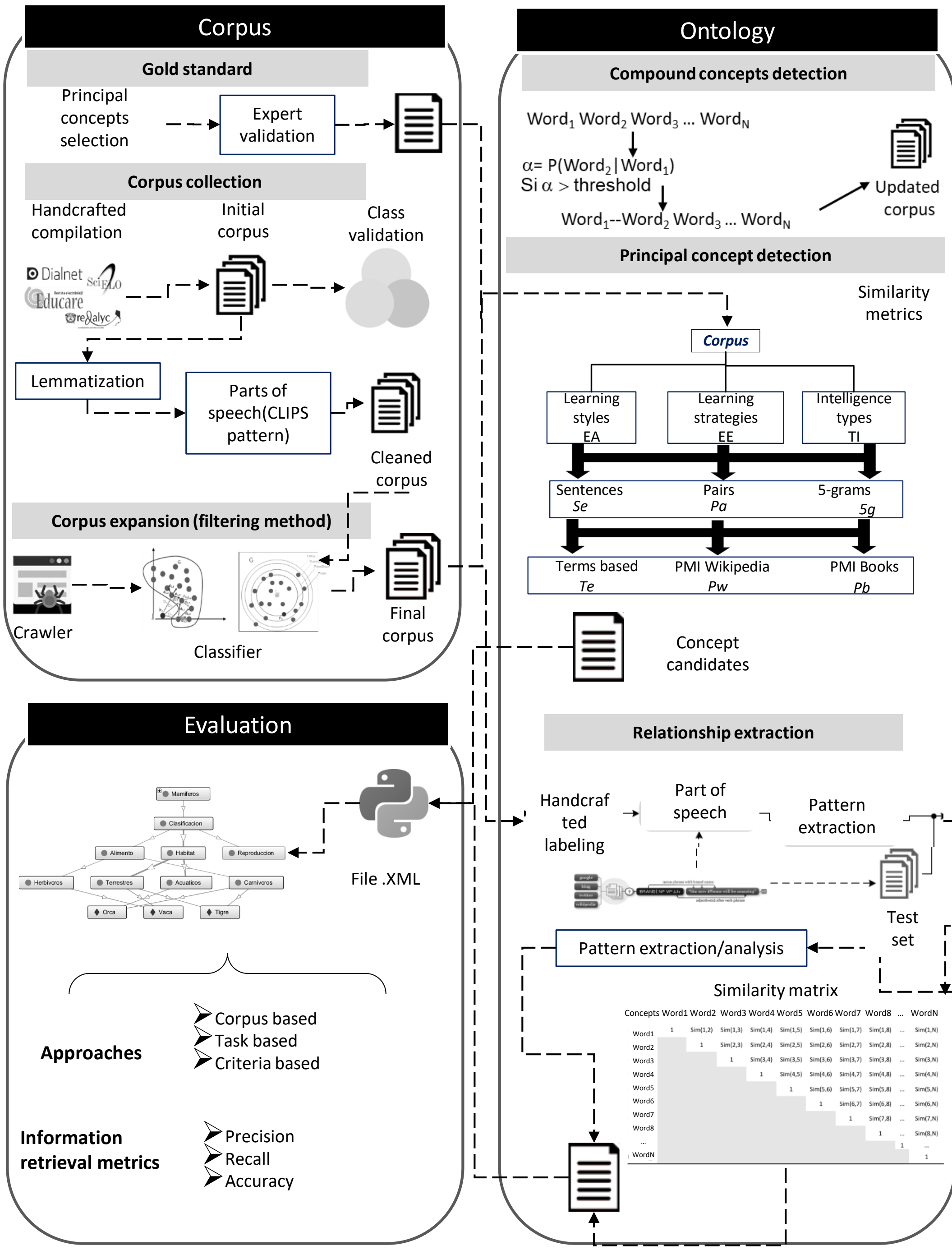


Fig. 2. General proposed methodology

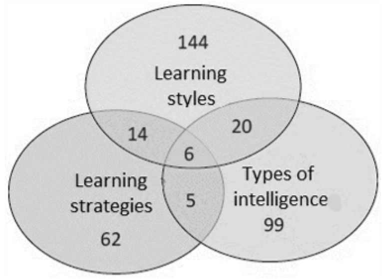


Fig. 3. Number of elements of the set evaluation per class

2. **Consider Reusing Existing Ontologies**: Investigate whether it is possible to extend existing knowledge sources, and which may be useful for mastering the problem.

3. **List Important Terms in the Ontology**: Make a list of the terms provided by the user, indicating the properties of each one, in the most precise and unambiguous way.

4. **Define Classes and Class Hierarchy**: From the list created in the previous step, select those independent terms to constitute the classes.

5. **Define the Properties of the Classes**: Describe the structure of the concepts, the terms that were not selected as classes, are now considered as properties of the class (called slots).

6. **Define the Characteristics of the Slots**: Define the different types of values that describe the slots, for example the type of associated value, cardinality, allowed values, among others.

7. **Create Instances**: Create instances of the hierarchy classes, first a class is selected, an instance is created and the slots are filled with the possible values.

The previous steps result in an ontology with the description of all its elements (classes, relations, inverse relations, instances, class properties). Follow sections contain details about each element.

5 Developing Ontologies

This section analyzes the process of creating the ontologies corresponding to each of the classes. For this analysis, the seven general steps are divided into two sections: the analysis of the theoretical elements (steps 1, 2 and 3) and the final structuring (steps 4 to 7). The first part is analyzed together, that is, the three classes together, while the second part is analyzed class by class.

5.1 Analysis of the Theoretical Elements

As a first step, the domain and scope of each of the ontologies is determined, to obtain this information, the following elements per class were extracted:

— Theoretical approach. For each class, a theoretical approach was selected, according with the size of information available for analysis:

  – Learning styles: Peter Honey and Catalina Alonso approach,
  – Intelligence types: Howard Gardner theory,
  – Learning strategies: Researches of Frida Díaz-Barriga, Gerardo Hernández, María Gonzalez, Javier Turrón among others.

— Purpose. In all classes, is to represent and formalize the domain knowledge.

— Scope. Support to make learning significant through its personalization.

— Target users. Teachers, preferably in college or high school level.

— Knowledge sources. Domain experts and specialized papers on the subject.

The second point is to determine if ontologies will be reused. For the learning styles class, the works of [21] are mentioned, not for reuse of the source file but for help and consultation regarding the structuring of the concepts. The classes of types of intelligences and learning strategies do not have other ontologies or investigations for reference.

After analyzing the collected material and the previously generated list of important concepts,

ISSN 2007-9737

the candidate terms for classes or subclasses are defined. Some of the concepts of each class are described below, in case they have synonyms, these are added between parentheses. The resources are in Spanish language, thus, the concept is in Spanish and English (separating by hyphen), synonyms and description are in English.

Learning styles

- *Actividad*-Activity (Task, Duty). Set of operations or tasks of a person or entity.
- Alonso. Author of the theory Learning Styles
- *Aprendizaje-Learning* (Knowledge). Change in behavior due to experience.
- *Autor*-Author. Person who has produced any scientific, literary or artistic work.
- *Caracterítica*-Characteristic. Said of a quality: That gives character or serves to distinguish someone or something from their peers.
- *EstiloAprendizaje*-LearningStyle (Ea, Style). The cognitive, affective, and psychological traits that serve as relatively stable, of how learners perceive, interrelate and respond to their learning environments.
- CHAEA (CHAEA Questionnaire, Honey Alonso Questionnaire). Questionnaire on Learning Styles that consists of eighty questions (twenty items referring to each of the four Styles) to which it is necessary to respond expressing agreement or disagreement.
- *EstiloActivo*-ActiveStyle. Learning style of people who get involved with new experiences, tend to act first and then think about the consequences.
- *EstiloPragmático*-PragmaticStyle. Learning style that includes people testing their new ideas, theories and techniques, trying to see if they work in practice.
- *EstiloReflexivo*-ReflectiveStyle. Learning style of people who are observers and analyze their experiences from different prospects.
- *EstiloTeórico*-TheoreticalStyle. Learning style that shows within the main characteristics the logic, the method, the objectivity and the structuring in the actions.

Types of Intelligences

- *CorporalKinestética*-CorporalKinesthetic (Kinesthetic Body Intelligence, Kinetics, Body Kinetics, Kinetic Intelligence, Body Kinetic Intelligence). Type of intelligence where the person has the ability to use their own body to carry out activities or solve problems,
- *Discente*-Learner (Pupil, Apprentice, Pupil, Student, Subject). Person receiving education.
- *Espacial*-Spatial (Spatial Intelligence, Visual, VisualSpatial, VisualIntelligence). Type of intelligence that consists of forming a mental model of the world in three dimensions.
- Gardner (HowardGardner). Author of the theory of multiple intelligences
- *Herramienta*-Tool (Technology). Set of theories and techniques that allow the practical use of scientific knowledge.
- *Instrumento*-Instrument (Questionnaire). One who poses a series of questions to extract certain information from a group of people.
- *InteligenciaMúltiple*-MultipleIntelligence (Intelligence, Intelligence Type). It involves the ability to solve a problem or to produce products that are important in a cultural context.
- Interpersonal (Interpersonal Intelligence). Type of intelligence that consists of understanding others.
- Intrapersonal (Intrapersonal Intelligence). Type of intelligence that consists of understanding oneself.
- *Lingüística*-Linguistic (Linguistic Intelligence, Verbal Intelligence). Type of intelligence where people use both hemispheres.

— *LógicaMatemática*-MathematicalLogical (Mathematical Logical Intelligence, Logical Intelligence, Mathematical Logical Intelligence, Mathematical Intelligence). Type of intelligence that corresponds to the thinking model of the logical hemisphere.

— Musical (Musical Intelligence). Type of intelligence that is related to the ability to perceive, discriminate, transform and express oneself through musical forms.

— *Naturalista*-Naturalist (Naturalist Intelligence). Type of intelligence that is used when observing and studying.

Learning strategies

— *Analizar*-Analyze. Submit something to analysis

— *Soporte*-Support (Support, Support Strategy, Regulation Strategy, Resource Regulation). Also called support strategies, they are a series of support strategies that include different types of resources that contribute to the completion of the task.

— *EstrategiaAprendizaje*-StrategicLearning. Those internal processes (cognitive, motivational and emotional) and behaviors that promote effective and efficient learning.

— *Autoevaluación*-SelfEvaluation. Evaluation that someone makes of himself or of some aspect or activity of his own.

— *Cognitiva*-Cognitive (Cognitive Strategy). They refer to the integration of new material with prior knowledge. In this sense, they will be a set of strategies that are used to learn, code, understand and remember information in the service of determined learning goals.

— *Constructivismo*-Constructivism. A pedagogical current created that postulates the need to provide the student with tools that allow them to create their own procedures to solve a problem situation.

— *EstrategiaMetacognitiva*-Metacognitive (Metacognitive Strategy). They refer to the planning, control and evaluation by students of their own cognition. They are a set of strategies that allow the knowledge of mental processes, as well as their control and regulation in order to achieve specific learning goals.

In general, the list of learning styles is more extensive, since it integrates many of the concepts shared by the three classes, for example: theory, student, author, learning, subject, etc. The kinds of learning styles and types of intelligences are the ones that share the most terms, this is largely due to the fact that they are based on already established theories that propose a classification and an instrument to detect it.

Teaching-learning strategies have not yet been fully defined and several authors propose various classifications, although the approach analyzed is one of the most recurrent in the literature, a method generally accepted as in the case of the others has not yet been established in two classes.

5.2 Structuring of Ontologies

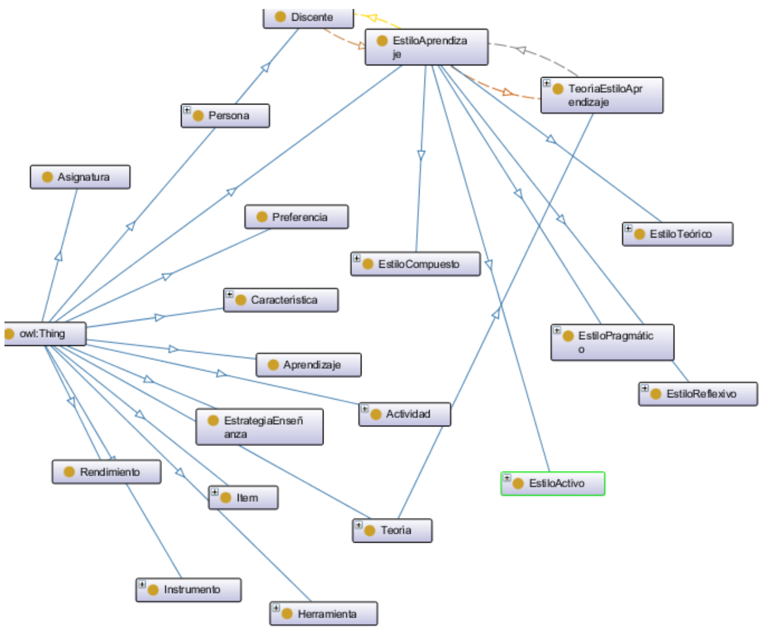
Once the concepts considered important have been selected, the relevant structures for each topic was built analyzed. In the following subsections the most significant elements of each of them are mentioned. Some concepts are in Spanish, due to is the language used for the research.

5.2.1 Learning Styles

Figure 5 shows the process for structuring the ontology to learning styles, according to the methodology used. It starts from point 4, which consists of defining the classes and subclasses.

At this point, the figure shows a three-phase scheme, first an extract of the list of concepts is given, specifying whether they were cataloged as classes or subclasses. The example shows the main classes and subclasses (the types of learning styles), later, these are structured in a small scheme to represent the taxonomic structure. It is precisely these types of structures that allow

ISSN 2007-9737



Metrics

Axiom	776
Logical axiom count	608
Declaration axioms count	168
Class count	150
Object property count	14
Data property count	14
Individual count	4
Annotation Property count	0

Class axioms

SubClassOf	114
EquivalentClasses	42
DisjointClasses	5
GCI count	0
Hidden GCI Count	4

Fig. 4. Graph representing the learning styles ontology (extract)

us to adequately visualize Isa relationships. In the example, IsA (RichTea, LearningStyle), and IsA (LearningStyle, Theory).

Finally, in this same point apart from the IsA relations, other possible relations between different concepts are annexed, for example HasStyle (Student, Pragmatic) and DescribeStyle (Pragmatic, Item2). In this last example, the relationship that each of the items in the CHEA questionnaire has with learning styles is shown.

By handling a nominal scale to answer said questionnaire, each question is focused on determining whether a certain learning style is the dominant one in the students.

The complete process gets the specific relationships between concepts, specifying the domain, range, and inverse relationship. The relationships

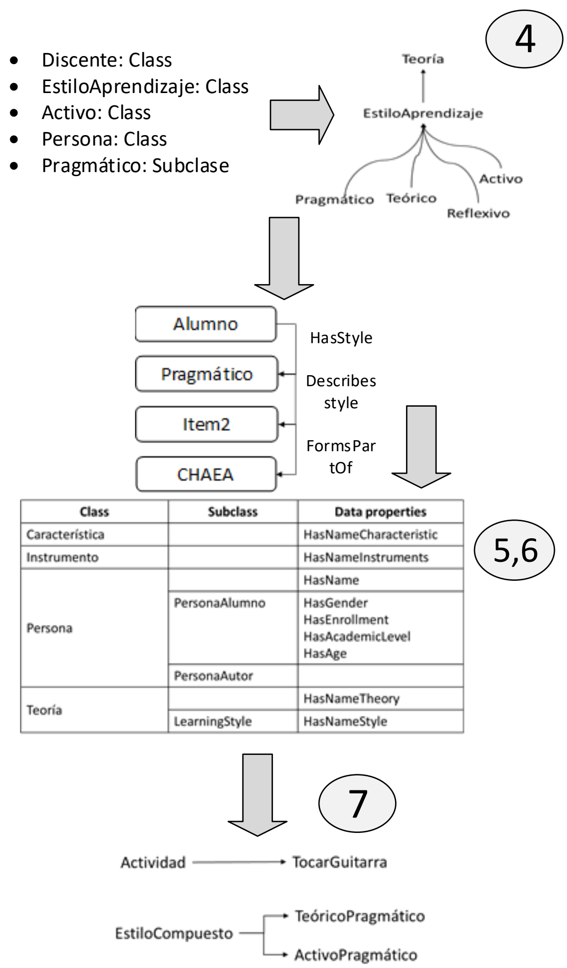


Fig. 5. Creation process of the learning styles ontology

shown mainly analyze the CHAEA instrument, the items that are related to each learning style and the characteristics present in the students of each style.

The HasAuthor relationship is also shown, to relate the creators of the studied approach. In the case of inverse relationships, the range and domain change, for example for the inverse relationship DescribeAStyle, an example of range is ActiveStyle and the domain Spontaneous.

Returning to Figure 5, points 5 and 6 define the properties of the data, specifying the classes and subclasses.

In this fragment, work continues on the classes related to the styles and the analysis of the CHAEA questionnaire.

Finally, in point 7, some examples of instances are given, such as the name of the compound styles and an example of activity (in the ontology, Activity represents what students with a certain learning style prefer to do.

Figure 4 shows some of the integrated classes, in addition to the ontology statistics, among which the number of axioms, classes, instances and subclasses stand out.

5.2.2 Types of Intelligences

Analyzing the types of intelligences, the same process of Figure 5 was applied. In the step corresponding to defining the classes and hierarchies between them (point 4), three elements are added:

— It starts with a list of concepts labeled as classes and subclasses. The pedagogical domain contains some important concepts, so it is common for these concepts to be present in two or all three classes.

— Subsequently, a part of the created taxonomic structure is represented. At this point, the similarity of this class with learning styles is observed. Both start from a theory where there is a classification and each of the sub-classifications has characteristics that describe it. Each student has a predominant characteristic, but it does not imply that he does not have the others in the classification, although to a lesser degree.

— Finally, other possible relationships are attached, also related to the instrument used to detect the type of intelligence in a student (Gardner test). In this scheme, item 2 is related to logical mathematical intelligence.

Table 2. Relationships between concepts in the intelligence type ontology (extract)

Domain	Relationship/Inverse	Range
TeoríaInteligencia	HasAuthor	Gardner
Gardner	IsAuthor	TeoríaInteligencia
TestGarder	DescribesTheory	InteligenciaMúltiple
InteligenciaMúltiple	IsDescribedByInstrument	TestGarder
TestGarder	HasItem	ItemI1
ItemI1	FormsPartOf	TestGarder
ItemI2	DescribesIntelligence	LógicoMatemática
LógicoMatemática	IsDescribedBy	ItemI2
TeoríaGardner	Study	InteligenciaMúltiple
InteligenciaMúltiple	IsStudiedBy	TeoríaGardner
Alumno	HasIntelligence	InteligenciaMúltiple
InteligenciaMúltiple	IsPresentIn	Alumno

Table 3. Relationships between concepts in the learning strategies class ontology (extract)

Domain	Relationship/Inverse	Range
Discente	Has	PerfilCognitivo
PerfilCognitivo	DescribesTo	Discente
EstrategiaEnseñanza	Develops	AprendizajeEstrategico
AprendizajeEstrategico	IsObtainedBy	EstrategiaEnseñanza
Profesor	Implements	EstrategiaEnseñanza
EstrategiaEnseñanza	IsImplementedBy	Profesor
Discente	LearnWith	EstrategiaEnseñanza
EstrategiaEnseñanza	IsMadesBy	Discente
EstrategiaEnseñanza	Encourages	DesempeñoAcademico
DesempeñoAcademico	IsEncouragesBy	EstrategiaEnseñanza
Cognitiva	InvolvesActivity	CentrarAtención
CentrarAtención	IsPartOfStrategies	Cognitiva

Table 2 shows a extract of relationships and inverse relationships raised in this analysis, the same structure of learning styles is followed, relating the theory to the authors, the types of intelligences that make up the theory and the items related to them.

Class properties are defined in the following points. In the extract, characteristics such as the name of the theory, the instrument and a description of the types of intelligence analyzed are added. Finally, in the last point, instances are added to the classes and subclasses, in the example, characteristics of the students are shown, which make them develop a particular type of intelligence.

Figure 6 shows the ontology designed in Protégé in addition to the list of classes and the statistics of said ontology. In general, it has fewer elements than learning styles, since in Gardner's studies, not as many characteristics are taken into account as in the CHAEA questionnaire.

5.2.3 Learning Strategies

In this topic, the concepts used and in general the rich theoretical foundation is not yet as defined as in the other two topics, so when relating them and structuring the ontology, the classes and subclasses found are less than learning styles and intelligence types.

According with the methodology, in point 4 the taxonomy of the classes is structured, first specifying an extract of classes and subclasses, the IsA relations and other relations between concepts. At this point the classification of learning strategies proposed by various authors is mentioned: Cognitive, metacognitive and supportive, as well as an example in the case of two of them. By relating the concepts, in this case the analysis focuses on the impact of learning strategies on the student (learner). Table 3 shows a slightly longer excerpt of these relationships.

The structure presented is different from that of the other two classes, here more relationships are presented focused on the structure of the strategies, not their classification. For example, the elements of a teaching strategy are mentioned, the

ISSN 2007-9737

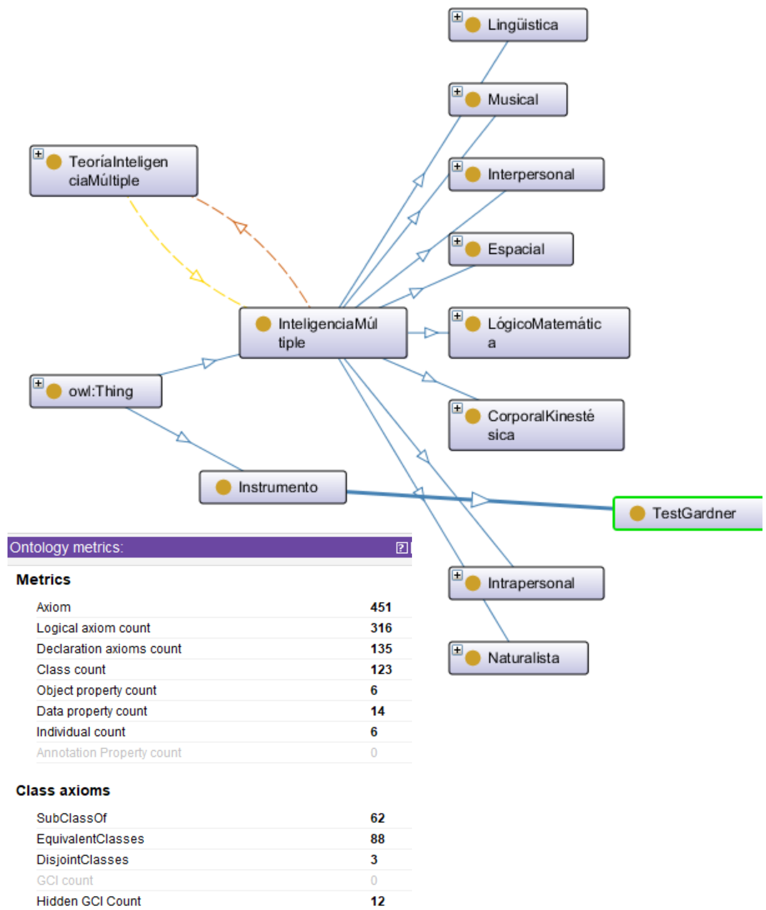


Fig. 6. Graph representing the ontology of styles of intelligence types (extract)

activities they involve. In addition, there are already relationships between the teaching class with the student and the strategies designed.

Figure 7 shows a view of the generated graph. Regarding the number of classes used, it is still less than the other two topics, but this is logical when compared to the initial list of concepts prepared with the help of experts in the domain.

A better comparison between the principal topics is shown in Figure 8, in which the main concepts are integrated, with the *IsA* type relationships

According with the selected theoretical approach and the gold standard designed, the main classes share some concepts such as: *Actividad*, *Aprendizaje*, *EstrategiaEnseñanza*, among others. Applied the semi automatic methodology, the same concepts are retrieved, with others considered not relevant by domain experts, but with high similarity level with the main topics.

Table 4 shows a collection of data related to the ontologies and the validation set created. At the end, the percentage of concepts covered in the design i s appended. The types of intelligences are those with a higher percentage, with 95%, while the learning strategies cover only 68%.

Other data that are integrated in the table is the total of subclasses (IsA relation), equivalent classes (synonyms) and disjoint classes (classes

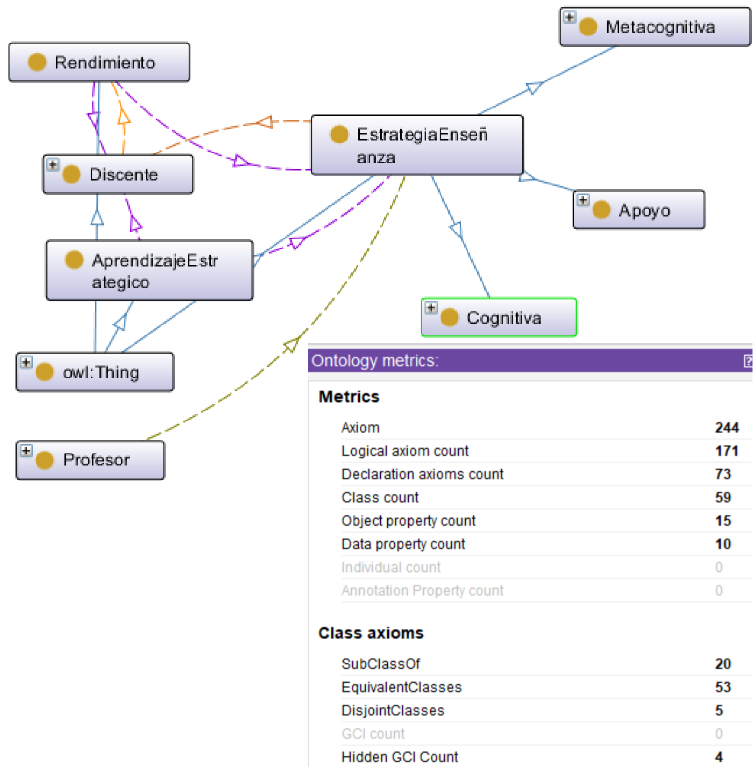


Fig. 7. Graph representing the ontology of teaching learning strategies (extract)

Table 4. Ontology metrics and elements total of the validation set

	Learning styles	Intelligence types	Learning strategies
Axioms	776	451	244
Classes	150	123	59
Subclasses	114	62	20
Equivalent classes	42	88	53
Disjoin classes	5	3	5
Validation set	184	130	87
Coverage	82%	95%	68%

at the same unique taxon level of the ontology). At a general level, there is extensive coverage with respect to the validation set, in addition to the fact that the most important aspects are theoretically covered according to a qualitative analysis.

6 Conclusions and Work in Progress

In this research, the initial objective was obtained, which consists in a semi-automatic methodology for pedagogical domain ontology creation. In general, experiments were carried out for each of the phases of the ontology creation process, using tools and typical procedures of natural language processing area. The used resources were some lemmatizers and lists of words allusive to the studied subdomain. The predominant processes in concepts and relationships detection was the analysis of textual similarity metrics.

Given the characteristics of the research and the proposed objectives, it was necessary to work in the following aspects:

— Choice of main classes: When determining the objective of designing a tool support for

ISSN 2007-9737

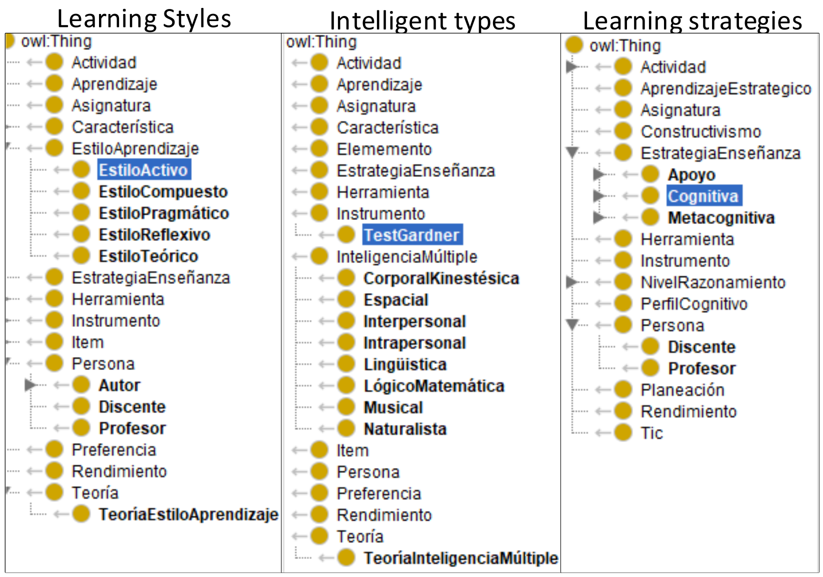


Fig. 8. Classes and subclass on each ontology

significant learning, the choice of topics that were theoretically related to this objective was necessary.

— Using Spanish language, the use of semantic tools is limited, so it is necessary to design resources for processing concepts.

— The pedagogical domain is studied, but not everything, only what is related to significant learning. This brought the need to generate processes that allow analyzing text with a very specific vocabulary, which further limits the external resources that can be use.

Therefore, transversality is an important aspect of this research, although the principal contributions are in computer science area, there are important advances in pedagogy. The lack of linguistic resources for Spanish, especially for the pedagogical domain, allowed a better analysis of the selected topics.

This analysis generated important resources for future researches, such as corpus focused in the principal topics, list of principal concepts, corpus in spanish with other pedagogic topics, and mainly, the handcrafted made ontologies.

As a work in progress, the following activities are being analyzed:

— Experiment with other pedagogy subdomains, in order to further analyze elements that can have an impact on significant learning.

— Expand the techniques used to detect relationships, in order to formalize the experiments done.

— Enrich the ontologies created with a population process, applied to a specific case.

— Apply the created ontologies to pedagogical projects focused on significant learning.

References

1. **Alonso, C., Gallego, D., Honey, P. (2007).** Los estilos de aprendizaje: Procedimientos de diagnóstico y mejora.

2. **Ausubel, D., Novak, J. (1983).** Psicología educativa, un punto de vista cognoscitivo. Trillas.

3. **Barriga, F., Hernández, G. (2004).** Estrategias docentes para un aprendizaje significativo. Una interpretación constructivista. McGraw Hill.

4. **Chance, P. (2001).** Aprendizaje y conducta. Manual moderno.

5. **Cakula, S., Sedleniece, M. (2013).**Development of a personalized e-learning model using methods of ontology. Procedia Computer Science, Vol 26.

6. **El Ouaazizi, M., Akharraz, I. (2019).** Ontology-based model for a linear algebra intelligent tutoring system. International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS).

7. **Fu, J., Jia, K., Xu, J. (2008).** Domain ontology learning for question answering system in network education. 9th International Conference for Young Computer Scientists.

8. **Gardner, H. (2001).** Estructuras de la mente. Fondo de Cultura Económica.

9. **Genovard, C., Gotzens, C. (1990).** Psicología de la instrucción. Santillana.

10. **González, M., Tourón, J. (1992).** Autoconcepto y rendimiento escolar: sus implicaciones en la motivación y en la autorregulación del aprendizaje. Eunsa.

11. **Grandbastien, M., Azouaou, F., Desmoulins, C., Faerber, R., Leclet, D., Quenu-Joiron, C. (2007).** Sharing an ontology in education: Lessons learn from the OURAL project. Seventh IEEE International Conference on Advanced Learning Technologies (ICALT'7).

12. **Gruber, T.R. (1995).** Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies.

13. **Kolb, D. (1976).** Learning style inventory.

14. **Lendyuk, T., Rippa, S., Bodnar, O., Sachenko, A. (2018).** Ontology application in context of mastering the knowledge for students. IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT).

15. **Méndez, N.D.D., Carranza, D.A.O., Ocampo, M.G. (2015).** Representación ontológica de perfiles de estudiantes para la personalización del aprendizaje. Revista Educación en Ingeniería, Vol. 10, No. 19.

16. **Noy, N., McGuinness, D. (2001).** Ontology development 101: A guide to creating your first ontology. Knowledge Systems Laboratory.

17. **Olivos, P., Santos, A., Martín, S., Cañas, M., Gómez E., Maya, Y. (2016).** The relationship between learning styles and motivation to transfer of learning in a vocational training programme. Suma Psicológica, Vol. 23, No. 1.

18. **Esguerra, G., Guerrero, P. (2010).** Estilos de aprendizaje y rendimiento académico en estudiantes de Psicología. Diversitas: Perspectivas en Psicología.

19. **Rodríguez, M. (2011).** La teoría del aprendizaje significativo: una revisión aplicable a la escuela actual. Revista Electrónica de Investigación, Innovación Educativa y Socioeducativa.

20. **Sánchez, S. (2007).** Modelo de indexación de formas en sistemas VIR basado en ontologías. Escuela de ingeniería y Ciencias, Universidad de las Américas Puebla.

21. **Silva, A., Ponce, J. (2013).** Reingeniería de una ontología de estilos de aprendizaje para la creación de objetos de aprendizaje. Eduweb.

22. **Tapia-Leon, M., Rivera, A.C., Chicaiza, J., Luján-Mora, S. (2018).** Application of ontologies in higher education: A systematic mapping study. IEEE Global Engineering Education Conference (EDUCON).

23. **Tanwar, S., Kumar Malik, S. (2018).** Towards blending semantics with an education based ontology using Protege 5.2.0 a revisit. 8th International Conference on Cloud Computing, Data Science & Engineering,

24. **Uskov, A., Pandey, J., Bakken, P., Margapuri, V.S. (2016).** Smart engineering education: The ontology of internet-of-things applications. IEEE Global Engineering Education Conference (EDUCON).

25. **Weigand, H. (1997).** A multilingual ontology-based lexicon for news filtering-the TREVI project. Proceedings of the IJCAI Workshop on Multilingual Ontologies-Nagoya.

26. **Weigand, H.(1986).** The teaching of learning strategies. Innovation Abstracts, Vol. 32, No. 5.

ISSN 2007-9737

27. **Wu, H. (2008).** Research of internet education system based on ontology. Fifth International Conference on Fuzzy Systems and Knowledge Discovery.

28. **Wilson, W., Wei, L., Mohammed, B. (2011).** Ontology learning from text: A look back and into the Future. ACM Computing Surveys-CSUR.

29. **Zhu, F. (2008).** ENGOnto: Integrated multiple English learning ontology for personalized education. International Conference on Computer Science and Software Engineering.

30. **Zhu, F., Yao, N. (2009).** Ontology-based learning activity sequencing in personalized education system. International Conference on Information Technology and Computer Science.

Article received on 31/07/2021; accepted on 29/09/2021.
Corresponding author is Darnes Vilariño.