

# Personalized Recommendation of Learning Objects Through Bio-inspired Algorithms and Semantic Web Technologies: an Experimental Analysis

Cleon Xavier Pereira Junior<sup>1,2</sup>, Clarivando Francisco Belizário Júnior<sup>1</sup>,  
Rafael Dias Araújo<sup>1</sup>, Fabiano Azevedo Dorça<sup>1</sup>

<sup>1</sup>Universidade Federal de Uberlândia (UFU)  
Uberlândia, MG – Brasil

<sup>2</sup>Instituto Federal Goiano (IF Goiano)  
Iporá, GO – Brasil

{cleonxavier, clarivando, rafael.araujo, fabianodor}@ufu.br

**Abstract.** *The emerging need to explore the Web as a learning source allied with the purpose of providing personalized recommendations is a tough task. Considering this scenario, this work presents an approach that combines Semantic Web technologies and bio-inspired algorithms to perform personalized recommendation of Learning Objects (LOs) using local repositories and Web resources. Web resources are retrieved and structured as LOs. Experiments were performed to verify which bio-inspired evolutionary algorithm would be most appropriate in this context. Also, discussions regarding the quality of recommendations considering local repositories and Web have been made. Initial experiments evaluating the efficiency of the proposed approach have shown promising results.*

## 1. Introduction

Smart Learning Environments (SLE) have been an important contributor to the success of pursuing a personalized teaching and learning process. In this kind of system, the application of Artificial Intelligence (AI) is recurrent and aims, among other goals, to provide more individualized learning paths considering personal necessities [Brusilovsky and Peylo 2003, Colchester et al. 2017]. Observing current SLE, it is possible to conclude that AI also contributes to different stages of the adaptation process, such as automatic student modeling [Bernard et al. 2016] and personalized recommendation of content [Pontes et al. 2014].

Despite the emerging need to explore the Web as a learning source, generating educational content from it in order to provide personalized recommendation is a tough task. Current challenges consist in associating two distinct sources of information (student profile and Web content) to be able to create personalized learning for students in SLE. In this context, Semantic Web technologies can be an interesting way to address this demand. In addition, aiming at facilitating the process of reusing and sharing learning resources, as Learning Objects (LOs), these technologies, which represent an extension to the traditional Web, allow for giving meaning (through metadata) to the content made available online. DBPedia<sup>1</sup>, for example, is an important result of efforts considering Semantic Web. From this perspective, considering metadata standards is essential.

---

<sup>1</sup><https://wiki.dbpedia.org/>

Recommending the best LOs considering individual needs is not a trivial task and may be associated with a specific class of computational problems, called search and optimization, as presented in [Júnior and Dorça 2018]. In this case, it becomes NP-hard [Bernhard and Vygen 2008] and solutions using bio-inspired evolutionary algorithms are usually feasible in these circumstances, as it involves a complex search process.

In this context, this work aims to investigate the personalized recommendation process of LOs through the development of a hybrid approach, which is able to search content from different sources in the Web and to recommend it based on learners individual needs. In addition, we intend to develop an approach that is uncoupled from any specific SLE, allowing its reuse in any learning environment. Considering that the recommendation process is often limited to local repositories of LOs, this work aims to go beyond, dynamically retrieving content from the Web for subsequent personalized recommendations as LOs, allowing a new level of Web content reuse in education.

This paper is structured as follows: Section 2 introduce some fundamental background theories and concepts and a summarized analysis of related work. Section 3 details the proposed approach. Section 4 presents and discusses the experiments and results. Section 5 presents conclusions and future work.

## 2. Background and Related Work

A fundamental concept that constitute the background of this work is LOs and their metadata. Basically, metadata is a set of data that provides summary information about an entity, that is, data about data [Pal et al. 2019]. In the educational field, an entity, known as LO is any digital, or non-digital, resource that can be used in the learning context. In this context, an LO can be a book, film, exercises list, among others. To allow the description and interpretation of LOs, some standards have been created and researchers have been trying to automate the process of creating metadata for LOs. An important metadata standard for LOs is IEEE-LOM (Learning Object Metadata) [IEEE LTSC 2001].

There are different ways to store LOs and their metadata in a repository. Based on their structures, LOs repositories can be classified into four types [Harman and Koohang 2007]:

- *centralized LOs and Metadata*: LOs and metadata are kept on a centralized server;
- *centralized LOs and Distributed Metadata*: in this case, LOs are stored on a central server while their metadata is outside the server;
- *distributed LOs and centralized metadata*: this type maintains metadata on a central server and provides links to external LOs;
- *distributed LOs and metadata*: this is a fully distributed architecture which can connect multiple index servers to multiple databases.

The automatic extraction of metadata from educational content is an effort that has been made by researchers in order to feed LOs repositories [Roy et al. 2008]. Once the LO has its metadata filled in, it becomes feasible to use, reuse and reference it during the learning process as well as exploring the curriculum sequencing [Gasparetti et al. 2018].

There is currently a lot of non-structured content available on the Web. Some of them are stored as videos, texts, audios, among others. Youtube and Wikipedia are examples of Web platforms that have a lot of available content. These platforms provide

APIs (Application Programming Interfaces) which provide access to the content metadata. Unfortunately, automatic metadata access and extraction is not enough to transform the content into an LO. According to [Neven and Duval 2002], in order to have LOs capable of serving their purpose, some educational characteristics are needed, such as granularity and composition. In addition, an automatic process of transforming Web content into LOs requires quality analysis of the content. Regarding metadata standards, it is possible to extend them to add new categories or vocabulary, such as CLEO extensions [CLEO 2003], which is an example of an extension over the IEEE-LOM. It adds additional vocabularies to improve the aggregation level field in the general category, the learning resources types field in the educational category, and the purpose field in the classification category.

With the intention of restructuring and expanding the traditional Web, the Semantic Web is defined by [Berners-Lee et al. 2001] as the Web data described and interconnected to establish context or semantics that adheres to a well-defined language and grammatical rules, allowing machines to understand and search in the Web as humans do. An important component of the Semantic Web is called Ontology, that can be defined as a formal and explicit specification of concepts in a specific domain [Ming and Jie 2002]. It is a concept within the Semantic Web that allows efficiently publishing and retrieving knowledge due to its power of expressiveness. In this way, ontologies formally define domains, their relations and inference rules [Berners-Lee et al. 2001].

The Web Ontology Language (OWL) is an important IEEE standard [Hitzler et al. 2009] that is widely used to develop ontologies. It is an extension of the RDF (Resource Description Framework) which defines the formal structure to make statements about resources in the Web. These statements are composed by a triple in the form <subject><predicate><object> [Decker et al. 2000]. The OWL extends the RDF vocabulary adding more power of expressiveness and allowing logical deductions about the represented knowledge [McGuinness et al. 2004].

SPARQL is a Semantic Web Query Language, similar to SQL (Structured Query Language), that allows for searching in RDF triples. For example, performing a search on DBPedia<sup>2</sup> using the EndPoint provided, with the code presented below, it is possible to retrieve the abstract definition of mitosis. Considering the example query, “dbpedia-owl” and “dbr” are vocabularies used to allow disambiguation of semantically different terms with the same syntax. Vocabularies are fundamental in the Semantic Web. DBPedia links data from Wikipedia and it is already a huge database. Its construction was carried out taking into account the vocabularies already existing on the Semantic Web.

```
PREFIX dbpedia-owl:<http://dbpedia.org/ontology/>
PREFIX dbr:<http://dbpedia.org/resource/>
SELECT * WHERE {
    dbr:Mitosis dbpedia-owl:abstract ?abstract
}
```

Considering the educational context and its inherent diversity of search and optimization problems, the use of bio-inspired evolutionary algorithms to seek for solutions and solve optimization problems is an interesting strategy to support personalized recommendation of LOs [Bhaskaran and Santhi 2017, Kurilovas et al. 2014, Dwivedi et al. 2018]. According to [Krishnanand et al. 2009], bio-inspired evolutionary algorithms are probabilistic search methods that simulate the evolution or the behavior

<sup>2</sup><http://dbpedia.org/sparql>

of biological entities. Such algorithms can be used to obtain near optimal solutions in optimization problems, for which traditional mathematical techniques may fail.

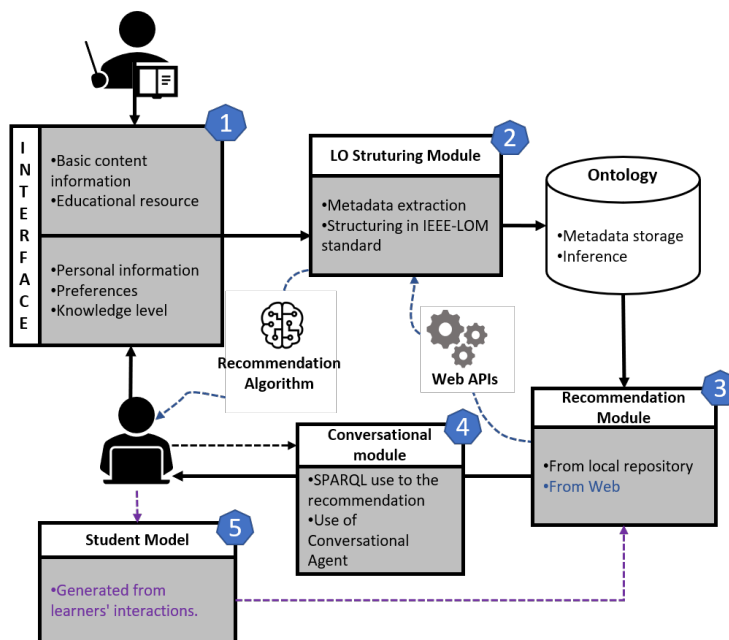
In this context, this work presents an approach that joins Semantic Web technologies and bio-inspired evolutionary algorithms in order to bring new and innovative aspects to address the personalized delivery of content in SLE. Through a literature review and related work analysis, it was possible to notice that this approach brings innovative aspects and advances in the state-of-the-art. Table 1 presents a comparison between the proposed approach and other related approaches.

**Table 1. Analysis and comparison of related work.**

Work	A	B	C	D	E	F	G
[Pal et al. 2019]	x	x				x	
[García-Florianio et al. 2017]		x	x				
[El-Bishouty et al. 2014]			x	x			x
[Kurilovas et al. 2014]				x			x
[Ince et al. 2017]			x	x			x
[Bhaskaran and Santhi 2017]				x			x
[Rastegarmoghadam and Ziarati 2017]			x	x			x
[Dwivedi et al. 2018]				x			x
[Beldjoudi et al. 2018]	x			x		x	x
<b>Proposed Approach</b>	<b>x</b>	<b>x</b>	<b>x</b>	<b>x</b>	<b>x</b>	<b>x</b>	<b>x</b>

A - Retrieving web content; B - Automatic generation of LOs metadata;  
 C - Automatic search in LOs repositories; D - Personalized recommendation;  
 E - Different ways of recommendation; F - Use of Semantic Web Technologies;  
 G - Use of bio-inspired algorithm for recommendation.

### 3. Proposed Approach



**Figure 1. Overview of the proposed approach.**

Figure 1 presents an overview of the proposed approach, which implements a repository considering the distributed LOs and centralized metadata approach, with automation aiming to enrich the process with more educational resources, and supporting

personalized recommendation. Tutors must provide information about the domain to be taught (bullet 1) which is stored in the knowledge base.

The proposed approach uses APIs made available by Wikipedia and Youtube. The Youtube’s API returns JSON (JavaScript Object Notation) [Bassett 2015], while the Wikipedia’s API returns lists, both containing metadata related to the found content. Then, the metadata is transcribed to the IEEE-LOM + CLEO standard, and stored in the knowledge base represented by the ontology (bullet 2). Local content metadata is also stored in this knowledge base. To achieve personalized recommendations (bullet 3), it is necessary to use information about learners (bullet 5), such as cognitive level, individual characteristics and preferences, such as Learning Styles (LS). In addition, information given by tutors is used to know about the domain/concepts that must be taught. A conversational model (bullet 4) helps students in goal-setting, help-seeking and self-evaluation tasks, boosting self-regulated learning behavior in students. The focus of this paper is on the recommendation process that is described next.

Students’ LS allows to filter LOs considering the learning resources types that better fit each individual. Relationship rules between LOs and LS considering the IEEE-LOM + CLEO standard was depicted in [Araújo 2017], in which characteristics of LOs described by the IEEE-LOM extended by the CLEO vocabulary are associated with LS considering the Felder and Silverman LS model [Felder et al. 1988]. An overview of the modeled ontology in this work is presented in Figure 2. It has classes that represent LOs considering their characteristics. In addition, it also represents students’ characteristics (cognitive level, goals, needs and learning styles). With this information, it is possible to infer what we call “ideal LO”, which is a set of features (metadata fields values) that a LO must have in order to properly fit student’s needs.

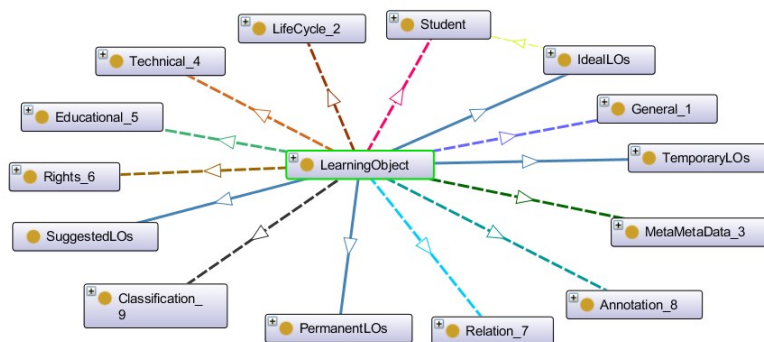


Figure 2. Overview of the Ontology.

The first step of the recommendation process is based on selecting the set of LOs that covers all the desired concepts. Suppose that a specific student needs to learn five concepts represented by the set  $X = \{C1, C2, C3, C4, C5\}$  and the existing LOs in the repository:  $F = \{LO1, LO2, LO3, LO4, LO5, LO6\}$ , where  $LO1 = \{C1, C2\}$ ,  $LO2 = \{C3\}$ ,  $LO3 = \{C4, C5\}$ ,  $LO4 = \{C5\}$ ,  $LO5 = \{C2, C3, C4, C5\}$  and  $LO6 = \{C2, C3\}$ . The cost of each LO is computed based on how similar it is from the ideal LO (inferred in the ontology). Therefore, the lower the cost, the better the LO. For this example, consider the cost vector  $(3,7,1,1,4,2)$  that represents the costs of the LOs in the same order as in F, as shown in Figure 3.

INPUT MATRIX						
	LO <sub>1</sub>	LO <sub>2</sub>	LO <sub>3</sub>	LO <sub>4</sub>	LO <sub>5</sub>	LO <sub>6</sub>
C <sub>1</sub>	1	0	0	0	0	0
C <sub>2</sub>	1	0	0	0	1	1
C <sub>3</sub>	0	1	0	0	1	1
C <sub>4</sub>	0	0	1	0	1	0
C <sub>5</sub>	0	0	1	1	1	0

COST VECTOR					
3	7	1	1	4	2

**Figure 3. Example of a Binary Matrix and its Cost Vector.**

Thus, the goal is to find a subset of LOs, at minimal cost, that covers all concepts. The proposed solution is based on the Set Coverage Problem (SCP) theory [Cormen et al. 2009]. From a binary matrix  $a_{ij}$  with  $m$  lines e  $n$  columns, the SCP tries to cover all lines using a subset of columns at a minimal cost. Given  $x_j = 1$ , if the column  $j$  (with cost  $c_j > 0$ ) is in the solution,  $x_j = 0$  otherwise, then the SCP is defined as:

$$\text{Minimize } \sum_{j=1}^n c_j x_j \tag{1}$$

$$\text{Subject to } \sum_{j=1}^n a_{ij} x_j \geq 1, i = 1, \dots, m \tag{2}$$

$$x_j \in \{0, 1\}, j = 1, \dots, n \tag{3}$$

According to the previous example, the solution {LO1, LO3, LO6} has its cost equals to 6. To each non-satisfied metadata value, the LO gets its cost increased. After calculating the cost of each LO, the next step is to use a bio-inspired algorithm that can solve the SCP. In this paper, we test and compare three different bio-inspired algorithm for the presented recommendation problem: Genetic algorithm (GA) [Whitley 1994], Prey-Predator Algorithm (PPA) [Tilahun and Ong 2015] and, an adaptation of Particle Swarm Optimization (PSO), called Jumping Particle Swarm Optimization (JPSO) [Balaji and Revathi 2016]. Results are presented in next section.

#### 4. Experiments and Results

To analyze the behavior of the three aforementioned bio-inspired evolutionary algorithms for the Set Covering Problem applied to the search and selection of LOs in a given scenario, a large instance was tested – Instance 4.10 available in the OR-library<sup>3</sup> benchmark. This instance is represented by a 200x1000 matrix and its cost vector. It is known from the literature that this instance has an optimal solution equal to 514 as a result of its objective fitness function. Each algorithm was submitted to a base, and had an initial population of 10 individuals. In addition, the test was repeated 30 times.

Figure 4 shows the execution results. Values in the y-axis refer to the average of the 30 executions. The graph also shows the Standard Deviations (SD) of the execution times in Figure 4(a). In order to visualize the results graphically, it was assumed that all iterations used the same execution time as an average of the total time.

<sup>3</sup><http://people.brunel.ac.uk/mastjjb/jeb/orlib/scpinfo.html>

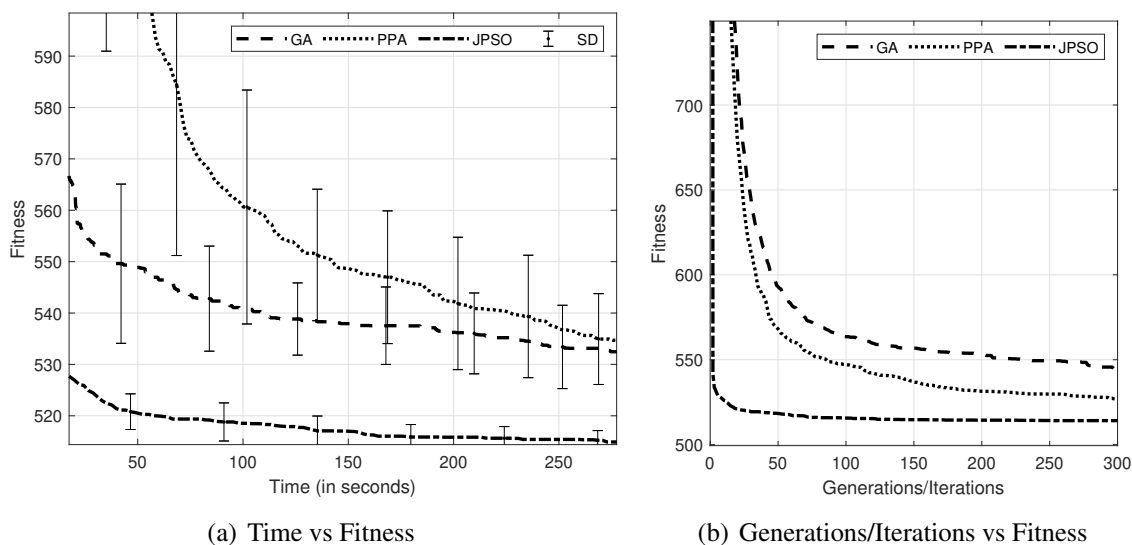


Figure 4. Performance of the algorithms for a 200x1000 size input.

In the two graphs, it is possible to observe that the JPSO algorithm obtained a better performance. Both GA and JPSO converged to the optimal solution at some point in the graph (GA could not be visualized in the graph because it has demanded more generations). The PPA, in thirty executions, only managed to get close to the optimal solution even increasing the number of iterations.

### Content Recommendation

Content to be recommended

Student

#### Recommended Learning Objects

Title	Concept	URL
Mitose vs. Meiose - Divisão Celular - Prof. Paulo Jubilut	mitose,meiose,divisão celular	<a href="https://youtube.com/watch?v=EkRs9kL5Kw4">https://youtube.com/watch?v=EkRs9kL5Kw4</a>

#### Concepts Not Covered

Concept

Não há

#### Related Learning Objects

Title	Concept	URL
Não há		

### Content Recommendation

Content to be recommended

Student

#### Recommended Learning Objects

Title	Concept	URL
Divisão Celular	mitose,prófase,meiose	<a href="#">Link 0002</a>

#### Concepts Not Covered

Concept

divisão celular

#### Related Learning Objects

Title	Concept	URL
Mitose	metáfase,ciclo celular	<a href="#">Link 0003</a>

Figure 5. Recommendation example using the desired knowledge and LS.

A user interface prototype was implemented to allow qualitative experiments through the analysis of recommended LOs considering external repositories and Semantic Web. The prototype allows to set the content to be learned by students and their characteristics. Figure 5 presents a screenshot of a tested scenario that includes the local repository and the Web (Youtube and Wikipedia). Two different sections of the interface show which concepts have not been covered and the related LOs that do not present the concepts of interest. In this test, a maximum of 5 LOs were selected for each concept in each external base (counting up to 45 external digital resources for the example). After performing the inference by the ontology and the search by the algorithm, one YouTube video was selected as the best LO. It is possible to observe that all concepts were covered with one unique LO. Considering the SCP, as explained before, this was the minimal cost solution.

## 5. Conclusion and Future Work

This paper presented an innovative approach to support personalized teaching/learning process in SLE. Besides retrieving Web content as LO, another challenge presented in this work is the personalized recommendation of LOs. Once the search for content has been completed both in the local repository and external sources, the recommendation process takes into account students' preferences and their cognitive level.

In this sense, the problem grows exponentially since students are distinct from each other and the collection of Web content is immeasurable. In this context, an approach supported by bio-inspired evolutionary algorithms and Semantic Web technologies was proposed. Different bio-inspired evolutionary algorithms were tested. Results showed that the proposed approach for personalized recommendation of LOs is promising and contributes to the advance of the state-of-the-art. Experiments in real learning contexts are planned, and tests with real students will allow to improve and develop important features of this approach.

## References

- Araújo, R. D. (2017). *Uma Arquitetura Computacional para Autoria e Personalização de Objetos de Aprendizagem em Ambientes Educacionais Ubíquos*. PhD thesis, Universidade Federal de Uberlândia.
- Balaji, S. and Revathi, N. (2016). A new approach for solving set covering problem using jumping particle swarm optimization method. *Natural Computing*, 15(3):503–517.
- Bassett, L. (2015). *Introduction to JavaScript object notation: a to-the-point guide to JSON*. ” O’Reilly Media, Inc.”.
- Beldjoudi, S., Seridi, H., and Karabadjji, N. E. I. (2018). Recommendation in collaborative e-learning by using linked open data and ant colony optimization. In *Int. Conf. on Intelligent Tutoring Systems*, pages 23–32. Springer.
- Bernard, J., Chang, T.-W., Popescu, E., and Graf, S. (2016). Optimizing pattern weights with a genetic algorithm to improve automatic working memory capacity identification. In *Int. Conf. on Intelligent Tutoring Systems*, pages 334–340. Springer.
- Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The semantic web. *Scientific american*, 284(5):34–43.



- Bernhard, K. and Vygen, J. (2008). Combinatorial optimization: Theory and algorithms. *Springer, Third Edition, 2005*.
- Bhaskaran, S. and Santhi, B. (2017). An efficient personalized trust based hybrid recommendation (tbhr) strategy for e-learning system in cloud computing. *Cluster Computing*, pages 1–13.
- Brusilovsky, P. and Peylo, C. (2003). Adaptive and intelligent web-based educational systems. *Int. J. of Artificial Intelligence in Education*, 13:159–172.
- CLEO (2003). CLEO Extensions to the IEEE Learning Object Metadata. Technical report, CLEO Collaborative Partners (Cisco Systems, Inc., IBM Corporation, Microsoft Corporation, Thomson NETg), Washington, USA.
- Colchester, K., Hagra, H., Alghazzawi, D., and Aldabbagh, G. (2017). A survey of artificial intelligence techniques employed for adaptive educational systems within e-learning platforms. *Journal of Artificial Intelligence and Soft Computing Research*, 7(1):47–64.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2009). *Introduction to algorithms*. MIT press.
- Decker, S., Melnik, S., Van Harmelen, F., Fensel, D., Klein, M., Broekstra, J., Erdmann, M., and Horrocks, I. (2000). The semantic web: The roles of XML and RDF. *IEEE Internet computing*, 4(5):63–73.
- Dwivedi, P., Kant, V., and Bharadwaj, K. K. (2018). Learning path recommendation based on modified variable length genetic algorithm. *Education and Information Technologies*, 23(2):819–836.
- El-Bishouty, M. M., Chang, T.-W., Graf, S., Chen, N.-S., et al. (2014). Smart e-course recommender based on learning styles. *Journal of Computers in Education*, 1(1):99–111.
- Felder, R. M., Silverman, L. K., et al. (1988). Learning and teaching styles in engineering education. *Engineering education*, 78(7):674–681.
- García-Florian, A., Ferreira-Santiago, A., Yáñez-Márquez, C., Camacho-Nieto, O., Aldape-Pérez, M., and Villuendas-Rey, Y. (2017). Social web content enhancement in a distance learning environment: intelligent metadata generation for resources. *International Review of Research in Open and Distributed Learning*, 18(1):161–176.
- Gasparetti, F., De Medio, C., Limongelli, C., Sciarrone, F., and Temperini, M. (2018). Prerequisites between learning objects: Automatic extraction based on a machine learning approach. *Telematics and Informatics*, 35(3):595–610.
- Harman, K. and Koohang, A. (2007). *Learning objects: standards, metadata, repositories, and LCMS*. Informing Science.
- Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P. F., Rudolph, S., et al. (2009). OWL 2 web ontology language primer. *W3C recommendation*, 27(1):123.
- IEEE LTSC (2001). IEEE 1484.12.1 - Draft Standard for Learning Object Metadata. Learning Technology Standards Committee of the IEEE.
- Ince, M., Yiğit, T., and Işık, A. H. (2017). A hybrid ahp-ga method for metadata-based learning object evaluation. *Neural Computing and Applications*, pages 1–11.

- Júnior, C. B. and Dorça, F. (2018). Uma abordagem para a criação e recomendação de objetos de aprendizagem usando um algoritmo genético, tecnologias da web semântica e uma ontologia. In *Brazilian Symposium on Computers in Education*, pages 1533–1542.
- Krishnanand, K., Nayak, S. K., Panigrahi, B. K., and Rout, P. K. (2009). Comparative study of five bio-inspired evolutionary optimization techniques. In *2009 World Congress on Nature & Biologically Inspired Computing*, pages 1231–1236. IEEE.
- Kurilovas, E., Zilinskiene, I., and Dagiene, V. (2014). Recommending suitable learning scenarios according to learners' preferences: An improved swarm based approach. *Computers in Human Behavior*, 30:550–557.
- McGuinness, D. L., Van Harmelen, F., et al. (2004). OWL web ontology language overview. *W3C recommendation*, 10(10):2004.
- Ming, D. Z. T. S. Z. and Jie, Y. D. C. (2002). Overview of ontology [j]. *Acta Scientiarum Naturalium Universitatis Pekinesis*, 5:027.
- Neven, F. and Duval, E. (2002). Reusable learning objects: a survey of lom-based repositories. In *Proc. of the Tenth ACM Int. Conf. on Multimedia*, pages 291–294. ACM.
- Pal, S., Pramanik, P. K. D., Majumdar, T., and Choudhury, P. (2019). A semi-automatic metadata extraction model and method for video-based e-learning contents. *Education and Information Technologies*, 24(6):3243–3268.
- Pontes, W. L., França, R. M., Costa, A. P. M., and Behar, P. (2014). Filtragens de recomendação de objetos de aprendizagem: uma revisão sistemática do cbie. In *Brazilian Symposium on Computers in Education*, pages 549–558.
- Rastegarmoghadam, M. and Ziarati, K. (2017). Improved modeling of intelligent tutoring systems using ant colony optimization. *Education and Information Technologies*, 22(3):1067–1087.
- Roy, D., Sarkar, S., and Ghose, S. (2008). Automatic extraction of pedagogic metadata from learning content. *Int. J. of Artificial Intelligence in Education*, 18(2):97–118.
- Tilahun, S. L. and Ong, H. C. (2015). Prey-predator algorithm: a new metaheuristic algorithm for optimization problems. *Int. J. of Information Technology & Decision Making*, 14(06):1331–1352.
- Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and computing*, 4(2):65–85.