VRSIL: A Framework for Video Recommendation Integrating Semantic Intelligence with a Large Language Model

Anubrat Bora[1] and Gerard Deepak[2]

[1]Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India
[2]B.M.S Institute of Technology and Management, Yelahanka, Bangalore, India
[2]gerard.deepak.christuni@gmail.com

Abstract. This framework presents a strategic model for video recommendation in the Web 3.0 era, integrating hybrid machine intelligence, generative AI, and semantic artificial intelligence through advanced semantic reasoning and quantitative semantic measures. The model dynamically generates ontologies from preprocessed query words, which are then used to select features based on linked similarity. These features are employed to classify the dataset from the perspective of the query using a logistic regression classifier. Semantics-oriented reasoning is achieved by computing the Normalized Pointwise Mutual Information (NPMI) to determine quantitative thresholds, and the Jian-Konrad Index is utilized as a criterion function for optimization. This optimization is carried out using the Elephant Optimization algorithm, which is executed only once to maintain the diversity and number of recommended entities. This approach ensures that the recommendations are both relevant and varied, aligning with the dynamic nature of Web 3.0.

Keywords: Semantic Intelligence, Ontology Generation, Semantic Similarity

1 Introduction

The rapid evolution of Web 3.0 and the Semantic Web has ushered in a new era of information retrieval and recommendation systems, but the dynamic nature of these technologies poses significant challenges, particularly in video recommendation. Traditional systems relying on collaborative filtering and content-based approaches struggle to keep up with the complex, evolving relationships within Web 3.0. This creates a critical need for techniques that can effectively navigate and leverage the rich semantic structures of this new web paradigm. Quantitative semantic reasoning becomes essential to bridge the gap between user queries and the vast array of video content available online, allowing recommendation systems to move beyond surface-level associations and provide more contextually relevant and accurate suggestions. Ontologies play a pivotal role by organizing and representing knowledge domains, enabling precise feature selection and alignment with user intent. This enhances the effectiveness of machine learning models, particularly when integrated with semantic intelligence. Semantic similarity measures help these models understand and interpret the nuances of user queries, leading to more refined and personalized recommendations. As the demand for intelligent content delivery in Web 3.0 grows, it is crucial to develop recommendation systems that are both computationally efficient and deeply informed by Semantic Web principles. This paper addresses these challenges and presents a novel approach that utilizes dynamic ontology generation and machine learning to improve semantic reasoning and recommendation accuracy.

Motivation: The primary motivation for the development of a video recommendation system in the Web 3.0 era stems from the need for strategies that are fully compliant with the evolving standards and capabilities of Web 3.0. Current video recommendation systems either fail to meet these standards or, if they do, are still in a nascent stage of development. What is needed is a comprehensive, hybrid intelligence-based framework that optimizes video recommendations through a knowledge-centric approach. Such a framework must be semantically aligned to meet the demands of the dynamically growing and increasingly complex Semantic Web. This method offers a strong response to the problems presented by this new digital environment by guaranteeing that recommendations are both pertinent and flexible enough to accommodate Web 3.0's constant change.

Contribution: The proposed framework makes several significant contributions to video recommendation systems in the Web 3.0 era. It features dynamic ontology adjustment and generation based on preprocessed query words and user clicks, ensuring that the system's recommendations evolve to better meet user needs. This iterative approach, which continues until user clicks align with recommendations, reflects a high degree of user-centric refinement. Additionally, the framework employs dynamically generated ontologies for classifying video datasets using a logistic regression classifier, leveraging augmented, contextually relevant knowledge to enhance recommendation accuracy. The incorporation of Capsense generation using LMA (Large Language Model Architecture) further enriches the system with advanced semantic intelligence, making the framework both innovative and effective in addressing the complexities of video recommendation.

Organization: The paper's structure is as follows: A thorough list of books relevant to this paper is provided in Section 2. An overview of the suggested system design is given in Section 3. The outcomes produced by the suggested model are listed in Section 4. The paper is finally concluded in Section 5.

2 Related Works

2.1 Video Recommendation Systems

Video recommendation systems have evolved significantly over the years, with traditional methods like collaborative filtering and content-based filtering playing a prominent role. Early works by Sarwar et al. [1] focused on collaborative filtering, leveraging user-item interaction matrices to predict user preferences for unseen items. Similarly, content-based filtering, as discussed by Lops et al. [2], recommended videos based on the textual content associated with video metadata, such as titles, descriptions, and tags. While effective to some extent, these methods often struggled with scalability and capturing the nuanced context of user preferences in dynamic environments like Web 3.0.

2.2 Semantic Intelligence in Video Recommendations

The integration of semantic intelligence into video recommendation systems marks a significant improvement over traditional method. A key study by Middleton et al. [3] introduced the use of ontologies to enhance video recommendations by capturing the semantic relationships between concepts, which allowed for more accurate and context-aware suggestions. Similarly, Cantador et al. [4] explored how semantic reasoning could be applied to improve the relevance of recommendations by incorporating user profiles and preferences into a semantically rich framework. These approaches laid the groundwork for more sophisticated models that integrate semantic intelligence with other advanced technologies.

2.3 Large Language Models (LLMs) in Recommendation Systems

Prominent Large Language Models (LLMs) with human-like text comprehension and generation capabilities include GPT-3 and BERT. Brown et al. [5] introduced GPT-3, which could generate coherent, contextually relevant text based on a vast amount of training data, making it a powerful tool for enhancing recommendation systems. Furthermore, BERT, as explored by Devlin et al. [6], enabled the deep understanding of context within text, which can be applied to improve the precision of video recommendations by better interpreting user queries and preferences. These LLMs serve as a foundation for integrating language understanding with recommendation algorithms.

2.4 Ontology-Based Frameworks for Knowledge Representation

Ontologies have been widely used to represent knowledge in a structured manner, facilitating the organization of information and improving the accuracy of recommendations. Gruber [7] defined ontologies as explicit formal specifications of terms in a domain and relations among them, which are crucial for building intelligent systems. Building on this, Staab et al. [8] developed ontology-based frameworks that enabled better knowledge management and retrieval, leading to more effective recommendations. These frameworks are essential for understanding how semantic relationships between concepts can be leveraged in video recommendation systems.

2.5 Hybrid Models Integrating Semantic Intelligence and Machine Learning

Hybrid models that combine semantic intelligence with machine learning techniques have shown promise in enhancing recommendation systems. Burke [9] discussed the benefits of hybrid recommender systems, which blend multiple recommendation strategies to overcome the limitations of individual approaches. More recently, Sun et al. [10] introduced a hybrid model that integrated semantic information with deep learning techniques, resulting in significant improvements in recommendation accuracy and user satisfaction. These models exemplify the potential of combining semantic intelligence with machine learning to create more robust and adaptable recommendation systems.

2.6 Optimization Algorithms in Recommender Systems

Optimization algorithms play a critical role in fine-tuning recommender systems to achieve better performance and scalability. The Elephant Optimization Algorithm, introduced by Wang et al. [11], is an example of a nature-inspired metaheuristic that has been effectively applied to optimize complex systems, including video recommendation frameworks. Additionally, genetic algorithms, as explored by Holland [12], have been widely used for optimizing the feature selection process, ensuring that the most relevant features are utilized in the recommendation process. These algorithms are vital for enhancing the efficiency and effectiveness of recommender systems, particularly when dealing with large and diverse datasets.

3. Proposed Methodology

Fig. 1. depicts the video recommendation framework's proposed system architecture.

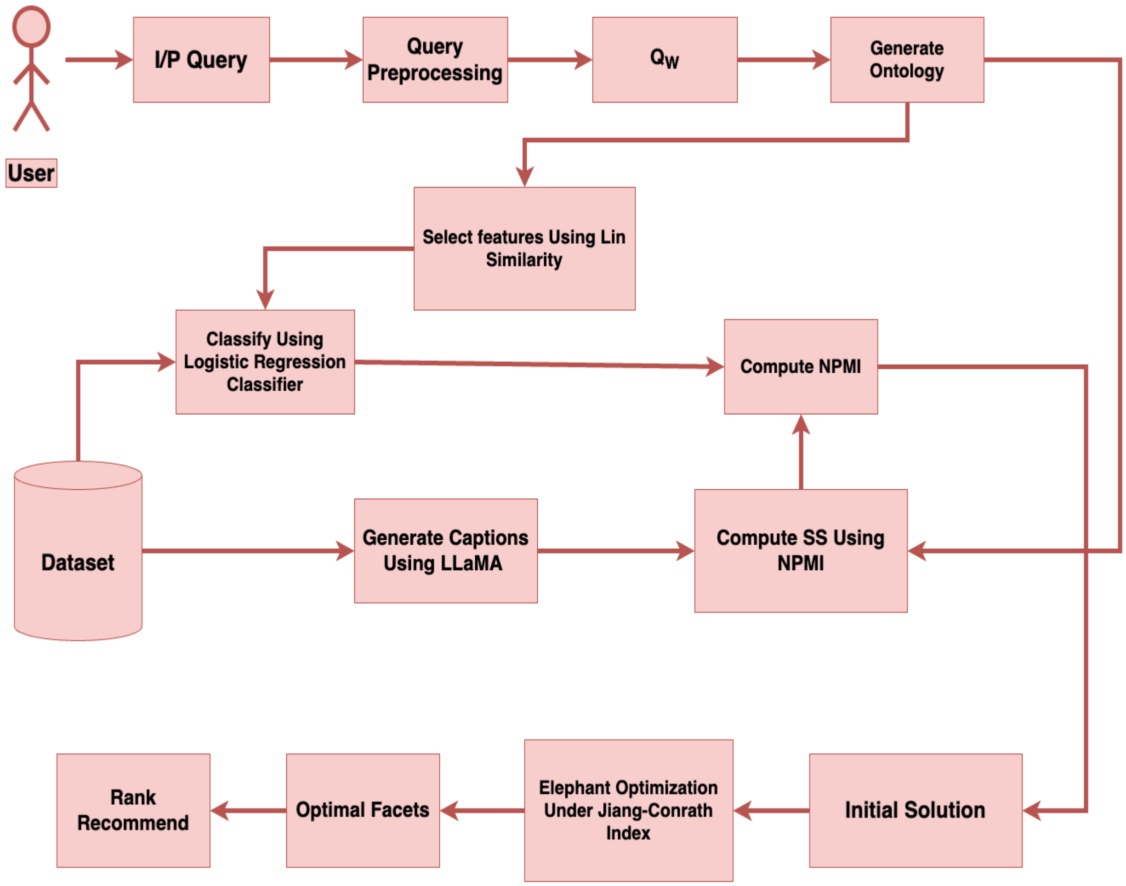

Fig. 1. Proposed Architecture

The proposed system begins with query pre-processing to refine the user's input. This stage involves several key steps: tokenization splits the query into individual words, while lemmatization converts these words to their base forms, ensuring that they are in their non-inflectional states. Stop word removal then filters out commonly used words that do not contribute significant meaning to the query. Additionally, Named-entity Recognition (NER) identifies and categorizes specific entities within the query, such as names, locations, and organizations. This pre-processing ensures that the query is in an optimal form for further analysis. Following pre-processing, the system proceeds to ontology generation. The refined query words are fed into Ontocolab, the chosen tool for generating ontologies. To maintain the integrity of the ontology and ensure it stays true to the domain of the query, certain constraints are applied: the ontology is limited to 24 levels of concept-subconcept hierarchy, and only two levels of individuals (instances) are allowed. These restrictions help preserve the essence of the domain, preventing the ontology from becoming overly complex or deviating from its intended focus. The last phase of the procedure involves integrating the generated ontologies and the video dataset labels into LLAMA (Large Language Model Architecture). LLAMA, as a generative AI framework, offers several advantages. Its capabilities as a large language model (LLM) enable it to generate human-like text, making it effective for producing coherent and contextually relevant responses. By leveraging LLAMA, the system can process the input data more effectively, leading to more accurate and meaningful results. This integration of ontologies with LLAMA ensures that the system can handle complex queries and generate responses that are well-aligned with the user's intent.

3.1 Pointwise Mutual Information (PMI)

A statistical metric called Pointwise Mutual Information (PMI) is used to assess the relationship between two events, x and y. It measures the extent to which these occurrences' co-occurrence is greater or less than one would anticipate if they were independent.

$$PMI(x, y) = \log\left(\frac{P(x,y)}{P(x).P(y)}\right) \qquad (1)$$

where $P(x, y)$ is the joint probability distribution and $P(x)$ and $P(y)$ are individual probabilities.

3.2 Normalized Pointwise Mutual Information (NPMI)

By normalizing PMI, Normalized Pointwise Mutual Information (NPMI) improves it and makes sure the resultant number falls into a defined range, usually between -1 and 1. When examining relationships between various datasets or contexts, this standardization is very crucial.

$$NPMI(x, y) = \frac{PMI(x,y)}{-\log\big(P(x,y)\big)} \qquad (2)$$

The negative value of logarithm serves to balance the measure, making it more interpretable. Positive NPMI values suggest that the events co-occur more frequently than expected under independence, whereas negative values indicate the opposite.

3.3 LaMA

LLaMA is a noteworthy development in the field of generative AI, having the capacity to efficiently handle and interpret large amounts of textual input. As a cutting-edge language model, LLaMA is excellent at comprehending challenging queries and producing text that is logical and pertinent to the situation. It is very efficient for a variety of natural language processing applications since its architecture is designed to facilitate large-scale learning and adaptability. These covers activities including semantic analysis, comprehension, and text production. Because of its adaptability and scalability, LLaMA works well in a variety of applications, such as content recommendation systems and conversational bots. It is an effective tool for improving user interactions and providing personalized information because of its capacity to learn from large datasets and produce complex responses.

3.4 Roche Deletion Classifier

A specialized tool used mostly in genomic research to detect and categorize genetic deletions is the Roche Deletion Classifier. To identify regions where deletions have occurred, this classifier analyzes genomic sequences. This information can be crucial for comprehending genetic illnesses and variants. Through the identification of these deletions, scientists can learn more about the ways in which genetic modifications affect health and illness. In precision medicine, where treatment decisions are guided by an awareness of an individual's genetic profile, the Roche Deletion Classifier is especially useful.

3.5 Elephant Optimization

A specialized tool used mostly in genomic research to detect and categorize genetic deletions is the Roche Deletion Classifier. To identify regions where deletions have occurred, this classifier analyzes genomic sequences. This information can be crucial for comprehending genetic illnesses and variants. Through the identification of these deletions, scientists can learn more about the ways in which genetic modifications affect health and illness. In precision medicine, where treatment decisions are guided by an awareness of an individual's genetic profile, the Roche Deletion Classifier is especially useful.

3.6 GN-Contract Index

In contract theory, the GN-Contract Index is a metric used to assess how well a contract aligns the incentives of the parties. It aids in evaluating how well the agreement addresses information asymmetry, controls risks, and promotes cooperation. When several parties must balance their incentives and interests to obtain mutual gains, the index is especially useful in complex contracts.

$$C_{GN} = \frac{u_A.p_A + u_B.p_B}{r_A + r_B + (I_A - I_B)^2} \quad (3)$$

- $u_A.p_A + u_B.p_B$: These terms represent the expected utilities for parties **A** and **B**, respectively. They consider the likelihood that each party will fulfill their contractual obligations. Expected utility is a measure of the benefit or satisfaction each party anticipates from the contract, factoring in the probability of various outcomes.
- $r_A + r_B$: This term represents the total risk assumed by both parties under the contract. It captures the combined risk exposure of parties **A** and **B**. A higher total risk indicates greater uncertainty and potential difficulty in maintaining the contract's effectiveness, as the parties may face significant challenges in meeting their obligations.
- $(I_A - I_B)^2$: This term accounts for the information asymmetry between the two parties. It measures the difference in the amount of information available to each party. A significant discrepancy in information can lead to an imbalance, making the contract less effective. Squaring the difference amplifies the impact of severe information asymmetry, highlighting how a major imbalance in information can affect the contract's overall efficacy.

4 Performance Evaluation and Results

A Standard Dedicated Single Large Dataset was developerd by formalizing four distinct datasets: the Video Recommendations Based on Visual Features Extracted with Deep Learning by Kvifte [16]; the WeChat Short Video Recommendation Dataset by Voler [17]; the YouTube Recommended Videos Network from Eduardo Velho [18]; and the 2021 DIGIX Video Recommendation Dataset by Voler [19]. These datasets were hybridized by incorporating newer annotations through a customized annotation generator. Following this, the annotations were rearranged and reprioritized, with annotation population achieved using specialized tag annotation generators to establish dedicated categories and terms.

Table 1. Performance Comparision

Model	Average Precision %	Average Recall %	Average Accuracy %	F-Measure %	FDR
ACTR [13]	90.22	91.87	91.00	91.04	0.10
SBRR [14]	91.28	92.19	91.74	91.73	0.09
DNNYR [15]	92.70	93.09	92.89	92.89	0.08
Proposed VRSIL	96.08	97.09	96.59	96.58	0.04

From Table 1, it is evident that the proposed VRSIL framework significantly outperforms models such as ACTR [13], SBRR [14], and DNNYR [15] across all evaluated metrics. With an average precision of 96.08%, VRSIL demonstrates superior accuracy in retrieving relevant content. Its recall rate of 97.09% indicates a thorough capture of relevant items, minimizing omissions. Additionally, VRSIL achieves the highest accuracy at 96.59%, reflecting its overall effectiveness in classification tasks. The F-Measure of 96.58% further underscores its balanced performance in both precision and recall. Notably, VRSIL has the lowest false discovery rate (FDR) at 0.04, significantly lower than ACTR (0.10), SBRR (0.09), and DNNYR (0.08), emphasizing its capability to minimize false positives.

The VRSIL framework's strength lies in its strategic design, incorporating dynamic ontology generation based on user input, a hybrid approach that integrates a large language model (LLaMA) for caption generation, and a logistic regression classifier. This classifier leverages features derived from the ontologies, applying Lin similarity with a strategic threshold to ensure accurate classification. Furthermore, VRSIL's robust ecosystem for semantic relatedness computation, which includes normalized quantization measures, the GN-Contract Index, and the elephant optimization algorithm, guarantees optimal solutions and a fair scaling of potential entities within the framework. In contrast, models like ACTR [13], SBRR [14], and DNNYR [15] exhibit notable limitations. ACTR, which relies on adaptive collaborative topic regression, often suffers from overfitting due to its simplistic two-stage process, leading to sparse results and reduced precision. SBRR, while addressing sampling bias, lacks the complexity needed for dynamic ontology generation and LLM integration, limiting its adaptability to diverse data sources. Similarly, DNNYR, though employing deep neural networks for candidate generation and ranking, is hindered by the dichotomy of these processes. This separation, while potentially enhancing learning, increases the risk of overfitting, especially when relying on a singular dataset, and fails to achieve the adaptability and precision offered by the VRSIL framework. Most importantly, VRSIL excels as a knowledge-centric, semantically reasoned system that dynamically generates and integrates auxiliary knowledge through query words and ontologies. This dynamic process continues until user interactions indicate satisfaction, at which point the LLM generates quantified auxiliary knowledge based on the gathered data, aligning the results with user needs. The framework's efficiency and accuracy are further enhanced by semantic reasoning through multi-computation, combined with advanced optimization algorithms.

In summary, the results clearly demonstrate that the VRSIL framework not only outperforms the DNNYR model [15] but also represents a noteworthy development in video recommendation. The DNNYR model, with its reliance on deep learning stages for candidate generation and ranking, is prone to overfitting and sparse results. In contrast, VRSIL's hybrid approach, incorporating dynamic ontology generation, lightweight machine learning classification, and LLM-based captioning, addresses these issues effectively. The integration of semantic reasoning with optimized thresholds and criteria functions, such as the GN-Contract Index, positions VRSIL as the best-in-class model. Given the growing demand for more accurate, context-aware, and semantically intelligent video recommendations, a model like VRSIL is not just advantageous but essential. Its robust solution, combining advanced semantic reasoning with optimized algorithms, ensures the delivery of highly relevant and user-satisfactory results, meeting the needs of modern Web 3.0 environments.

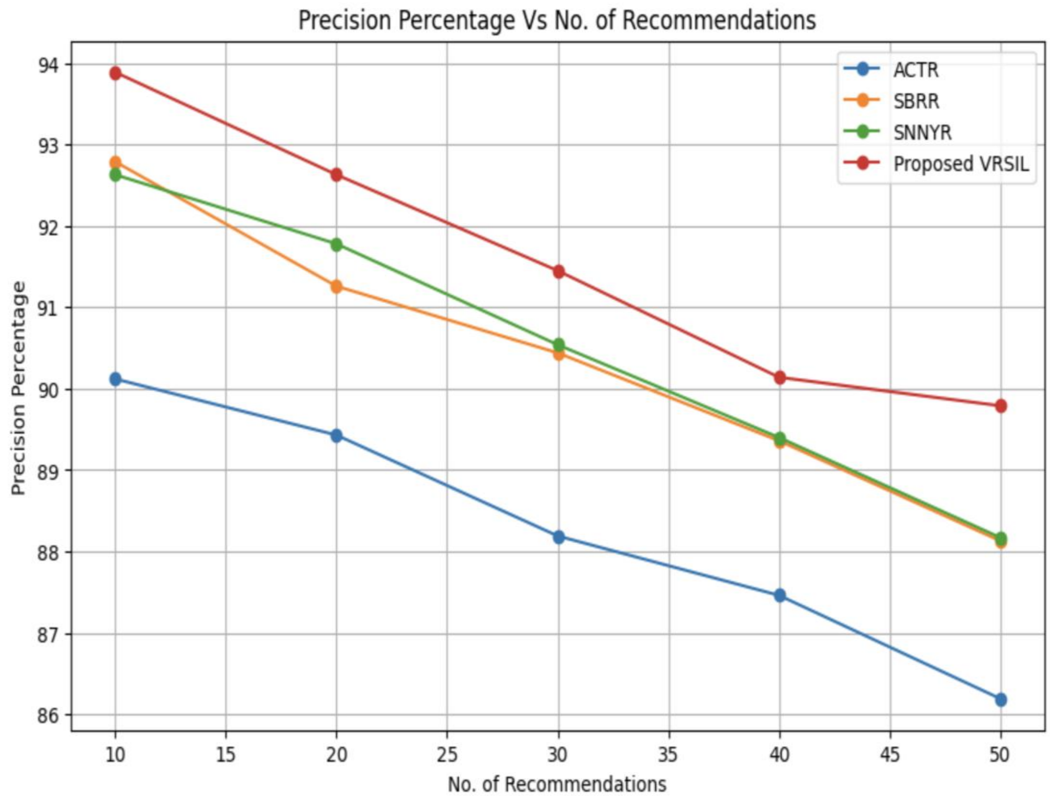

Fig 2. Precision Percentage Vs. Number of Recommendations

Fig 2. illustrates the recommendation distribution, where VRSIL occupies the highest portion of the hierarchy due to its comprehensive and effective approach. The ACTR model ranks lowest [13], followed by SBRR [14] and DNNYR [15], which occupy intermediate positions. The VRSIL framework's superior performance is attributed to its advanced integration of semantic reasoning, quantitative semantic reasoning, and LLM-based techniques, making it a more robust and effective model for video recommendations in the Web 3.0 context.

5 Conclusion

This paper proposes a novel strategic model for video recommendation in the Web 3.0 environment. The model encompasses dynamic ontology generation based on the perspective of query words, followed by feature selection on the generated ontology. This process is used to classify video datasets using a lightweight logistic regression classifier. The approach is designed to enhance knowledge-centric machine learning by utilizing a computationally efficient and feasible lightweight classifier, thereby ensuring the system operates effectively without overloading computing resources. This balance between efficiency and effectiveness makes the model particularly suitable for Web 3.0 applications.

6 References

1. Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. *Proceedings of the 10th International Conference on World Wide Web*.
2. Lops, P., De Gemmis, M., & Semeraro, G. (2011). Content-based Recommender Systems: State of the Art and Trends. *Recommender Systems Handbook*.
3. Middleton, S. E., Shadbolt, N. R., & De Roure, D. C. (2009). Ontological User Profiling in Recommender Systems. *ACM Transactions on Information Systems*.
4. Cantador, I., Fernández-Tobías, I., Berkovsky, S., & Cremonesi, P. (2011). Cross-domain recommender systems. *Proceedings of the 5th ACM Conference on Recommender Systems*.
5. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.
6. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL-HLT*.
7. Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*.
8. Staab, S., & Studer, R. (2001). Handbook on Ontologies. *Springer*.
9. Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*.
10. Sun, Y., Wang, S., & Zhang, Y. (2019). A Hybrid Recommendation System for Personalized Learning Resources. *IEEE Access*.
11. Wang, G.-G., Deb, S., & Coelho, L. S. (2015). Elephant herding optimization. *Neural Computing and Applications*.
12. Holland, J. H. (1992). Adaptation in natural and artificial systems. *MIT Press*.
13. Chen, Jingyuan, et al. "Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention." *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*. 2017.
14. Hidasi, Balázs, et al. "Session-based recommendations with recurrent neural networks." *arXiv preprint arXiv:1511.06939* (2015).
15. Covington, Paul, Jay Adams, and Emre Sargin. "Deep neural networks for youtube recommendations." *Proceedings of the 10th ACM conference on recommender systems*. 2016.
16. Kvifte, Tord (2021). Video Recommendations Based on Visual Features Extracted with Deep Learning [Dataset]. https://data.niaid.nih.gov/resources?id=ZENODO_4889728
17. Voler (2021). WeChat Short Video Recommendation [Dataset]. https://www.kaggle.com/voler2333/wechat-short-video-recommendation
18. Velho, Eduardo (2021). YouTube Recommended Videos Network from the Olavo de Carvalho video "Globalismo e Poder" [Dataset]. http://doi.org/10.5281/zenodo.5557833
19. Voler (2021). 2021 DIGIX Video Recommendation [Dataset]. https://www.kaggle.com/voler2333/2021-digix-video-recommendation