ASEAN Engineering Journal

Full Paper

ONTOLOGY CONSTRUCTION: BIOINSPIRED IMPROVED SEA LION OPTIMIZATION MODEL FOR SEMANTIC INFORMATION RETRIEVAL

Shital Kakad[a*], Sudhir Dhage[b]

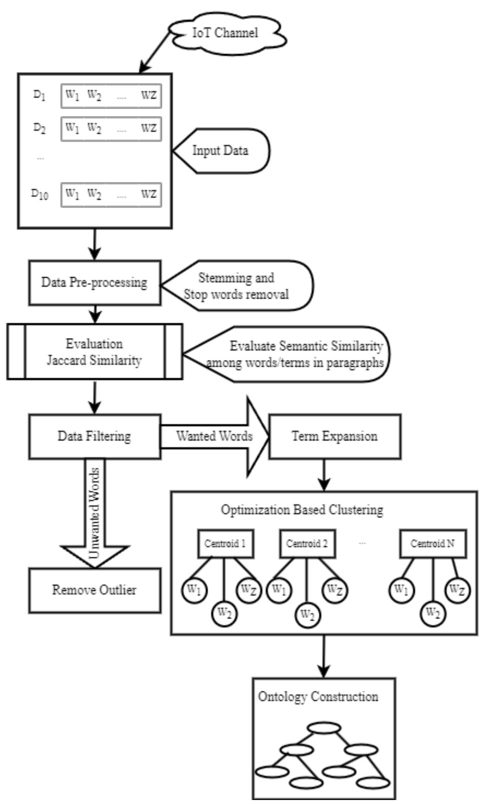
[a]Faculty of Engineering, Vishwakarma Institute of Technology, Upper Indira Nagar, Bibwewadi, Pune, India

[b]Faculty of Engineering, Sardar Patel Institute of Technology, Andheri, Mumbai, India

Article history
Received
18 October 2023
Received in revised form
04 February 2024
Accepted
06 February 2024
Published online
30 November 2024

*Corresponding author
Shital.kakad@vit.edu

Graphical abstract



Abstract

The search effectiveness and efficiency completely depend on the ontology design. The objective of information retrieval (IR) is search and retrieves precise and accurate data in response to a user's query. Notably, the existing search engine relies on conventional keywords to search information. It purely compares the user's query with the database and retrieves outcomes without understanding the intended meaning behind the user's query. Therefore, a significant proportion of the outcomes contain unrelated information. Further, designing a new ontology from scratch and evaluating it are challenging tasks. This research work is divided into two decision-making processes (i) data filtering and (ii) data annotation. In this paper, the steps of ontology construction as follows: a) Pre-processing of data b) Implementation of Proposed Jaccard Similarity Evaluation to evaluate the similarity of data c) Data filtering and outlier detection and final step d) Semantic annotation and cluster. The data is filtered by using the evaluated similarity function. Then, the data is grouped separately into wanted data and unwanted data. The unwanted words are called outliers. Based on the semantics, the data annotation is performed and the process of clustering is evaluated for developing the precise cross-domain applications-based ontology. Moreover, the clustering is done based on the similarity evaluation under multiple dictionaries. In the clustering procedure, the optimal centroid selection is considered a challenging crisis. Hence, for solving this issue, this research work widened with the introduction of an Improved Sea Lion (ISnLO), which is the improved version of the Sea Lion Optimization algorithm.

Keywords: Ontology, Semantic Web, Data Filtering, Data Annotation, Cross-Domain network, Optimization.

1.0 INTRODUCTION

Generally, the data on the Internet is stored in heterogeneous format. It is really difficult to retrieve accurate information from a huge database. Every day heterogeneous data is collected from different sources. Ontology represents a relationship of information/data. It represents the semantic of information. It is useful to retrieve accurate data. The semantic web makes the user query machine-understandable. Web 3.0 is called a semantic web. Ontology plays the main role in offering a global reference view [1-4]. Domain knowledge is important to build any ontology. Domain expert involvement plays a prime role to build relationships of data and build ontology. Creating relationships of data manually is a very complicated process. It takes a lot of time of researchers to build a new ontology from scratch. The machine can easily understand and reply to user queries using data annotation. The generation of the unified view was made by incorporating the heterogeneous data from diverse sources within one ontology instance. This in turn offers a viable solution for information sharing, reusing, and data integration [7-10], which is the groundwork of semantic webs information exchanging, reusing and sharing. There are a number of ontologies constructed in various domains such as environment, health, education, industry, etc. It exploits the development of smart manufacturing systems development via rule reasoning and semantic matching [5] [6]. The main purpose of ontology construction is sharing, reusing, and exchanging cross domain knowledge.

Even though, in large-scale ontology, the efficiency requirement is the major challenge that exists because of the continuous maintenance of concepts/data and their relations and difficulty of automatic construction of ontology [11] [12]. Typically, the domain experts construct the ontology manually and are inefficient [13]. Hence, there is a need for new solutions that can create automatic ontology from databases. In contrast, the semantics and formats of enterprise data gathered from databases can be associated with intelligent manufacturing systems usage [14-18]; this, in turn, might alter the particular business circumstances. In recent times, there is an apparent need for cross-domain applications [19] [20] which might enclose everyday life's multiple constraints. On considering two diverse cross-domain applications, it is obvious that the utilization of multiple source superiority through collaborative learning methods [21-24] is essential for achieving superior knowledge among multiple domains. Though, few of the unified cross-domain collaborative learning frameworks have been proposed in the literature [25-27], since the media data pose multi-modal, multi-domain, supervised, and sparse properties. The author constructed cross-domain based ontology model using hybrid whale optimization algorithm and rider optimization algorithm [28-31]. The cross-domain applications data sets are considered for ontology construction. Cross domain data can be cross platform data or cross network data. In particular, most of the diverse domain's media data are heterogeneous, and as well may complement one another [32] [33]. Still far as well poses domain discrepancy.

The research work contribution is as follows:
The abundant applications using recent technology are increasing exponentially on the Web. This expansion of information creates difficulties and complications, such as searching and retrieving appropriate information. The user spends most of the time on the Web to browse information [1]. The searching and retrieving accurate information effectively from the web are difficult tasks. The current search engine is not capable to understand the meaning of user query and mostly retrieves irrelevant information. Specifically, the data on the web is heterogeneous, which is collected from different sources. The main problem of the current web is data interoperability, data integration, data annotation and data filtering [2] [3]. In this context, cross domain ontology design and development plays a vital role to search and retrieve meaningful information using semantic web technology.

There is need to improve the information retrieval metrics such as precision and recall as well as need to reduce ontology construction time [7] and user query execution time. More research needed in construction of cross-domain ontology [15] [16] in terms of intelligent information retrieval and to solve data interoperability and data integration by adding cross-domain data. An appropriate optimization-based clustering algorithm can be discovered to integrate heterogeneous data [10] [17]. The data filtering, data annotation, and semantic similarity search [8] [18] is a motivating challenge in this research.

1. Multi-dictionaries-based similarity evaluation is performed for clustering to design the cross-domain based ontology.
2. Ontology constructions steps as follows: Pre-processing of data, Proposed Jaccard Similarity Evaluation, Data filtering and Outlier (unwanted words) Detection, and Semantic annotation and clustering of data.
3. Optimal centroid selection in the cluster by introducing a new algorithm termed Improved SnLO for solving the optimization issue is the main contribution of this research work to retrieve accurate and precise information.
4. The ontology construction time and execution time is evaluated to compare the implemented model with state of art models by changing training percentage and cluster size.
5. The performance of the Improved SnLO is measured using information retrieval metrics such as precision and recall. The overall performance of the proposed model is improved in terms of precision and recall.

The arrangement of this paper is explicated as follows: Section 2 shows the literature survey of the ontology construction model. Section 3 describes the cross-domain based ontology construction with a short description of process execution. Next, Section 4 depicts the semantic annotation and proposed clustering approach using a multi-dictionary similarity evaluation. Section 5 explicates the proposed improved sea Lion algorithm with solution encoding and objective function. Section 6 portrays the results and discussions. Finally, conclude the paper in section 7.

2.0 LITERATURE REVIEW

The construction of new applications using modern technology is facing exponential growth on the internet. This rapid expansion of information generates challenges and difficulties, particularly in tasks like searching for and retrieving relevant data. The existing search engines often struggle to understand the user queries, resulting in the retrieval of largely irrelevant information. Ontology enables the reuse, sharing, and exchange of domain knowledge. Currently, a vital focus of research pertains to organizing knowledge across multiple domains. The internet is comprised of diverse domain-specific data, making the management of such multi domain data a complex work.

The construction of a cross-domain ontology necessitates the incorporation of cross-domain knowledge. To address this, a real-time ontology (RTO) [1] is constructed using an adaptive filtering approach that responds to user requirements. The process involves the extraction of ontologies and sub-ontologies based on user-defined needs. These chosen ontologies are subsequently compared and seamlessly merged to give rise to a novel ontology referred to as the real-time ontology. The efficiency of this method is achieved through the implementation of a relevancy decay function and an adaptive threshold. The effort to construct a significantly large-scale ontology through the matching and integration of distinct domain-specific ontologies, employing both direct and indirect methodologies, poses considerable challenges and demands substantial time investment. In terms of recall, accuracy, correctness, and runtime, the performance of the real-time ontology betters that of super-large ontology (SLO) and top-level ontology (TLO).

Currently, the e-commerce sector is experiencing rapid growth in online shopping sector. Customers now eagerly deliver feedback on products, sharing their thoughts through textual input, original images, or videos of products. These online reviews play a crucial role in supporting decision-making during the process of purchasing products online. In this, a cross-domain ontology [2] is developed by exactly analysing a large volume of customer reviews in diverse domains. These online reviews mainly consist of unstructured, textual content. The construction of cross domain ontology model involves several stages: corpus preparation, ontology construction, and ontology alignment. The performance is achieved in terms of performance metrics, encompassing precision, recall, and F-measure, which measures the effectiveness of information retrieval.

Semi-automatic ontology construction [3] employs a hybrid approach that combines both manual and automated techniques in the making of an ontology. The objective is to join the advantages of each approach while modifying their individual limitations. In this process, once data/information is collected, automated methods like text mining or machine learning are used to extract concepts and relationships from the collected data. The concepts and relationships thus extracted may show potential incompleteness, inconsistency, or redundancy. Consequently, manual curation becomes crucial to refine and effortlessly integrate the extracted information into the existing knowledge framework. Following this integration, the structure of the ontology is developed through manual integration. This approach significantly reduces the time and effort required for ontology development as compared to fully manual construction methods. Moreover, it enhances the accuracy and consistency of the resultant ontology when compared with depend on just on automated methods. Additionally, this semi-automatic approach enhances the scalability and flexibility of ontology construction compared to fully manual methodologies.

Researchers have devoted significant time towards automating the process of ontology construction, aiming to enhance Information Retrieval (IR) and ease the burdens associated with construction of new ontology. Base domain ontology [4], which contributes to refining the precision of information retrieval. Subsequently, the base domain ontology is exactly formed using the reference collection, which is comprised of precisely labelled human-generated information.

In this, a comprehensive overview of the progress in ontology research and its practical utilization is provided. As ontology technology continues to grow, there is a growing emphasis on subject-focused web information retrieval techniques that integrate ontology. An inventive approach that merges semantic web technology with conventional information retrieval methods [5] is introduced. Moreover, it presents a corresponding algorithm rooted in ontology to measure relevance across varying subject matters. The presented information retrieval system mode demonstrates the capability to successfully prevent the loss of valuable information. This achievement is appreciated by semantically expanding the user's retrieval criteria. Furthermore, the initial retrieved documents undergo thorough filtration via a document analyzer, resulting in a remarkably precise alignment with the user's retrieval requirements.

The Internet of Things (IoT) creates a network where objects and devices interact independently, communicating without human involvement. A significant challenge in the IoT arises from semantic interoperability issues due to the different data formats of devices. To address this, ontologies are employed to embed meaning into raw data, thereby standardizing data representation. Despite the diverse applications of IoT, current models remain segregated in distinct vertical groups, each employing its unique ontology. A new cross-domain ontology, known as CDOnto, [6] is proposed. It functions as a versatile framework, adaptable with domain-specific ontologies. The model accepts a contextual strategy, effectively categorizing and arranging combined domain representations (contexts).

The review discusses the difficulty of retrieving information/data from the Internet due to its various formats of data. The main importance is on improving information retrieval using text feature extraction. A novel approach is introduced by using utilizing feature extraction and selection techniques [7] to enhance retrieval. The T-Order algorithm is implemented to reduce dimension complexity. The approach is applied to the frequency analysis of the BBC news dataset, resulting in improved accuracy and classification outcomes. Hybrid optimization model is constructed by combining whale optimization algorithm and rider optimization algorithm [28]. The main goal of the paper is to make the information retrieval system more efficient using the semantic web. This optimization-based ontology is implemented using the semantic web. In this Decision Support System for Alzheimer's disease Diagnosis [32] has been implemented using different existing ontology. This ontology helps to diagnose Alzheimer's disease patients.

3.0 PROPOSED ONTOLOGY CONSTRUCTION MODEL

The Proposed Ontology Construction model for cross domain applications stages is as (i) Pre-processing of data, (ii) Proposed Jaccard Similarity Evaluation, (iii) Data filtering and Outlier (unwanted words) detection, and (iv) Optimization-based cluster.

3.1 Pre-Processing of Data

In this, the cross-domain applications dataset is taken through the IoT channel to construct ontology. The initial and starting phase is pre-processing of data. In which the normalization is carried out using stop-word removal and stemming. A brief explanation of these two techniques is given below:

Stop-word removal [27]: This is the first step of data pre-processing. Most of the words in some texts are preferred to connect the sentence. Hence, the stop-word removal technique is applied to remove these connected words. The stop words such as “the” or “and”.

Stemming [27]: In this, the process of dropping unwanted words to their root form by removing unnecessary characters, typically a suffix, is referred to as stemming. Snowball and Porter are existing stemming methods.

The Jaccard similarity is evaluated after data pre-processing. Once this semantic evaluation is complete, the data filtering process relies on evaluated semantic similarity. The unwanted data are also called outlier. Subsequently, the clustering process is evaluated with which the centroid (semantics) acts as the prime part.

Though the centroid selection is appended as a promising issue, and for this reason, the optimization tactic is evolved in this paper. The optimization concept is the improved version of the SLnO algorithm, named ISLnO, which aim is accurate selection of optimal centroid. Figure 1 explicates the architectural presentation of the proposed cross domain ontology construction. The data is taken through the IoT channel.

3.2 Evaluation of Proposed Jaccard Similarity

The [28] explication of this implemented work is made using the exemplary illustration of data with auto and motorbikes. We have downloaded a data set from Kaggle. The cross-domain applications data are heterogeneous in nature. Let us assume, auto data Ho involves 10 paragraphs/data as $\{P_1^{Ho}, P_2^{Ho} ...., P_{10}^{Ho}\}$ and the motorbike Ba with 6 paragraphs as $\{P_1^{Ba}, P_2^{Ba} ...., P_6^{Ba}\}$. Apparently, the evaluation of ontology construction is moved from level 3 to level 1, as depicted as per Figure 2. The different count of words in each paragraph is shown in Eq. (1). In this $C = 1,2,....C_w$, C_w represents the number of words. The motorbike data is as well explained as per the similar notation and is shown in Eq. (2). Table 1 represents mathematical symbols and Description.

$$\begin{aligned} &P_1^{Ho} \Rightarrow w_1^{Ho} = \{w_{11}, w_{12} ...., w_{1C}) \\ &P_2^{Ho} \Rightarrow w_2^{Ho} = \{w_{21}, w_{22} ...., w_{2C}) \\ &.... \\ &P_{10}^{Ho} \Rightarrow w_{10}^{Ho} = \{w_{101}, w_{102} ...., w_{10C}) \end{aligned} \tag{1}$$

$$\begin{aligned} &P_1^{Ba} \Rightarrow w_1^{Ba} = \{w_{11}, w_{12} ...., w_{1C}) \\ &P_2^{Ba} \Rightarrow w_2^{Ba} = \{w_{21}, w_{22} ...., w_{2C}) \\ &.... \\ &P_6^{Ba} \Rightarrow w_6^{Ba} = \{w_{61}, w_{62} ...., w_{6C}) \end{aligned} \tag{2}$$

Table 1 Mathematical Symbols and Description

Mathematical Symbols	Description
C_w	Number of words
C_{Ho}	Count of Paragraphs
$w_z^{selected\ Ho}$	Total count of chosen words
TS	Count of term set
w_m	Word inside the cluster
cl	Cluster
K_{cl}	centroid word of the cluster cl
$\hat{i}$	Total count of clusters.
$\overrightarrow{Ds}$	The distance between the sea lion and the target prey
$\vec{S}(t)$	Vector position of sea lion.
$\vec{T}(t)$	Vector position of target prey.
t	Present iteration.
$\vec{L}$	Random vector.
$\overrightarrow{S_{leader}}$	The speed of sea lion leader’s sound
$\overrightarrow{C_1}$	The sound speed in water
$\overrightarrow{C_2}$	The sound speed in air
$\vec{H}$	Value gets decreased from 2 to 0 over the following iteration.
q	Random number

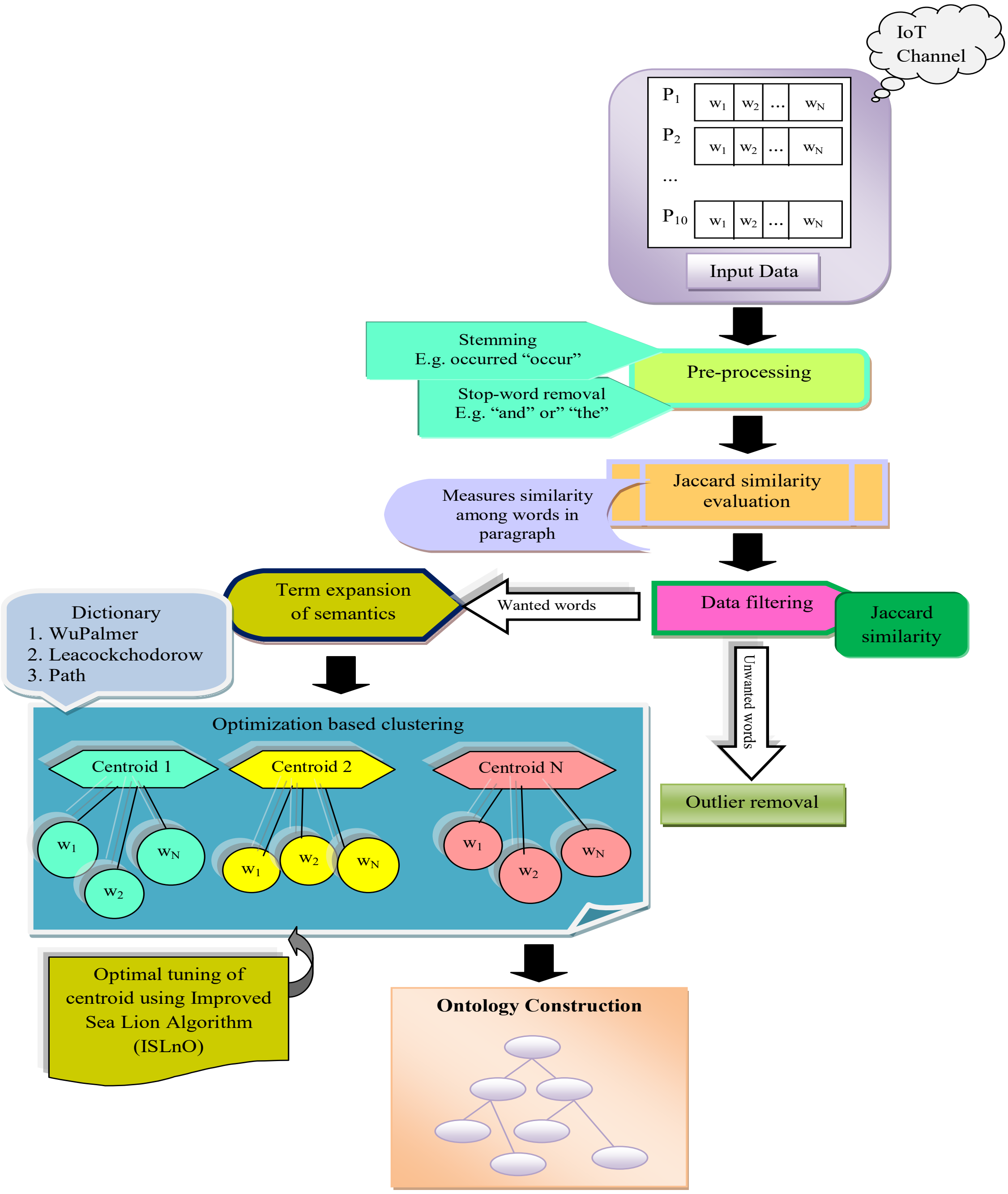


Figure 1 Architectural representation of proposed cross domain Ontology construction

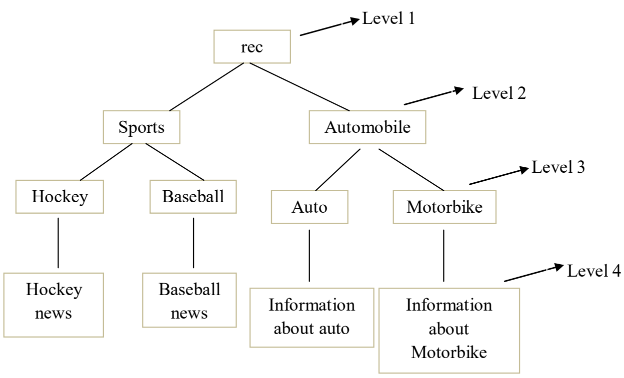


Figure 2 Hierarchy of data

Figure 2 shows the hierarchy of sports data and automobile data. In this, we can retrieve the data related to hockey news and baseball news. Similarly, we can retrieve information about auto and motorbikes. In this, user submit query such as- which is best motorbike? Next, the data will be filtered based on user query. The tree is constructed for selected words, which are relevant to user query by evaluating Jaccard similarity. The remaining words (unwanted words) are outlier. Data is consistently retrieved from level 3 down to level 1 by evaluating Jaccard Similarity. Part of the data is illustrated in Figure 2. A detailed explanation of the data set is given in the experimental setup.

Eq. (3) illustrates the evaluation of the Jaccard similarity among the words of each paragraph in auto. The evaluation of the Jaccard similarity coefficient for finite sample sets involves comparing the size of the intersection with the size of the union of the sets. Eq. (4) demonstrates the finalized Jaccard similarity, which is the mean of entire the Jaccard similarity [28]. In which, count of paragraphs is denoted by C_{Ho} in Ho based data. Ho represents motorbike data. Moreover, the Jaccard similarity of the motorbike is estimated as well.

$$J(P_x^{Ho}, P_y^{Ho}) = \frac{P_x^{Ho} \cap P_y^{Ho}}{P_x^{Ho} \cup P_y^{Ho}} \tag{3}$$

$$\mu J_x^{Ho} = \frac{1}{C-1} \sum_{\substack{y=1 \\ y \neq x}}^{C_{Ho}} J(P_x^{Ho}, P_y^{Ho}) \tag{4}$$

3.3 Data Filtering and Outlier Detection

The subsequent step is data filtering and outlier detection after executing similarity evaluation. The similarity data involves both wanted and unwanted data. Jaccard similarity evaluation is used to distinguish wanted and unwanted data from the complete data. The $C\left|w_1^{Ho}\right| + C\left|w_2^{Ho}\right| + \ldots + C\left|w_{TS}^{Ho}\right|$ indicates total word count [28]. The count of term sets is represented by TS, representing the selected terms. and, C shows the overall word count. The selected or wanted words are indicated as per Eq. (5), in which $w_z^{selected\ Ho}$ shows the entire count of chosen words. It is derived from Jaccard similarity and is expressed using Eq. (6).

$$C_w = C \times \sum_{z=1}^{TS} \left\{ w_z^{selected\ Ho} \right\} \tag{5}$$

$$w_z^{selected\ Ho} = \left\{ w_x^{Ho} \right\} \qquad \forall x : \mu J_x^{Ho} > \mu J \tag{6}$$

The unwanted words are outliers. In specific, the selection of outliers depends on the condition $\mu J_x^{Ho} > \mu J$.

Example: Consider a scenario where the first paragraph contains a total word count of 10. During the data filtering stage, the selected words can be $w_{11}^{Ho}, w_{13}^{Ho}, w_{15}^{Ho}, w_{16}^{Ho}, w_{17}^{Ho}, w_{18}^{Ho}, w_{19}^{Ho}$, using condition $\mu J_x^{Ho} > \mu J$, and the outliers (not selected words) can be achieved as $w_{12}^{Ho}, w_{14}^{Ho}, w_{110}^{Ho}$.

4.0 SEMANTIC ANNOTATION: IMPROVED SEA LION ALGORITHM

4.1 Semantic Annotation

Moreover, the processing of semantic annotation is very essential. Semantic annotation means term expansion. Every word contains its own semantics. Let us consider the chosen words $w_{11}^{Ho}, w_{13}^{Ho}, w_{15}^{Ho}, w_{16}^{Ho}, w_{17}^{Ho}, w_{18}^{Ho}, w_{19}^{Ho}$. The word w_{11}^{Ho} semantics is evaluated in Eq. (7).

$$w_{11}^{Ho} = \begin{Bmatrix} w_{11,1,1}^{Ho} & w_{11,1,2}^{Ho} & \cdots & w_{11,1,C}^{Ho} \\ w_{11,2,1}^{Ho} & w_{11,2,2}^{Ho} & \cdots & w_{11,2,C}^{Ho} \\ \cdots\cdot & & & \\ w_{11,10,1}^{Ho} & w_{11,10,2}^{Ho} & \cdots & w_{11,10,C}^{Ho} \end{Bmatrix} \tag{7}$$

Therefore, altogether, there is a necessity to create clusters from groups of semantically related words. Clustering in semantic information retrieval enhances the organization, search, and exploration of information/data. It provides users with more meaningful and efficient access to relevant information based on user query. It enhances relevant information retrieval experience and allows users to quickly locate relevant data.

In terms of clustering, a centroid is a central point within a cluster. It is used as a summary of the data points in that cluster. The centroid's location is determined based on the attributes or features of the data points it represents. Centroid is the best solution for semantic information retrieval from large databases. It plays a main role in the clustering process by defining the center around which data points are grouped based on similarity. The selection of centroid in the clustering process is a complicated task. Hence for the accurate attainment of the optimal centroid from the cluster, the optimization algorithm with improved strategies is established. For this reason, a new Improved SLnO algorithm is developed in this paper, which is the improvement of SLnO Algorithm.

4.2 Optimization based Clustering

Furthermore, the selection of centroid is optimally made based on the distance evaluation with multi-dictionary similarity formulation. Different dictionaries are used to find similarity of words. Each word has semantic (more than one meaning). Eg. Car can be denoted by motor or automobile or machine. As this work involves three dictionaries like Wu Palmer, Leacock Chodorow, and Path, the evaluation of distance is made for every individual cluster, and is given as per Eq. (8), (9) and (10), respectively for three dictionaries. It enhances the relevant information retrieval based on user query. In which, w_m is a word inside the cluster. Here cl represents the cluster, the centroid word of the cluster cl is delineated as K_{cl} and $\hat{i}$ is the entire count of clusters.

$$\left[d_{cl_{\hat{i}}}^{Ho}\right]^1 = \frac{1}{\left\| d_{cl_{\hat{i}}}^{Ho} \right\|} \sum_{m=1}^{d_{cl}^{Ho}} sim_1(w_m, K_{cl_{\hat{i}}}) : cl = 1,2,\ldots\hat{i} \tag{8}$$

$$\left[d^{Ho}_{cl_{\hat{i}}}\right]^2 = \frac{1}{\left\|d^{Ho}_{cl_{\hat{i}}}\right\|}\sum_{m=1}^{d^{Ho}_{cl}} sim_2(w_m, K_{cl_{\hat{i}}}) : cl = 1,2,...\hat{i} \quad (9)$$

$$\left[d^{Ho}_{cl_{\hat{i}}}\right]^3 = \frac{1}{\left\|d^{Ho}_{cl_{\hat{i}}}\right\|}\sum_{m=1}^{d^{Ho}_{cl}} sim_3(w_m, K_{cl_{\hat{i}}}) : cl = 1,2,...\hat{i} \quad (10)$$

The overall similarity is achieved as per Eq. (11), in which wt_1, wt_2 and wt_3 are the weights that are selected randomly.

$$Sim = wt_1 \times \left[d^{Ho}_{cl_{\hat{i}}}\right]^1 + wt_2 \times \left[d^{Ho}_{cl_{\hat{i}}}\right]^2 + wt_3 \times \left[d^{Ho}_{cl_{\hat{i}}}\right]^3 \quad (11)$$

The standard deviation measures the amount of variation in the set of words and is computed as per Eq. (12). A low standard deviation indicates that the values tend to be close to the mean of the set of words.

$$\sigma = std\left(d^{Ho}_{cl_{\hat{i}}}, \mu d^{Ho}_{cl_{\hat{i}}}\right) \quad (12)$$

The selection of centroid is optimally made based on the distance evaluation, in which the standard deviation (σ) of the distances should be minimum, such that, the optimal centroid word K_{cl} will be selected for every cluster.

5.0 PROPOSED IMPROVED SEA LION ALGORITHM

5.1 Objective Function and Solution Encoding

The solution that is offered as the input to the proposed ISLnO algorithm for optimized based clustering is illustrated in Figure 3, in which the total centroid count in the clusters is represented by C. Moreover, the objective (*obj*) defined in this research work is given in Eq. (13), in which, the standard deviation σ is given in Eq. (12).

$$obj = \min(\sigma) \quad (13)$$

K_{cl1}	K_{cl2}	....	$K_{cl,C}$

Figure 3 Solution Encoding

5.2 Sea Lion Optimization Algorithm

SLnO [26] algorithm is the nature-inspired metaheuristic optimization algorithm. It is developed by mimicking the hunting behaviour of Sea Lions. The sea lions have posed a few fascinating characteristics. They can move faster. Sea lions have a clear vision and superior property of hunting the prey.

Sea lions naturally possess inherently highly sensitive characteristics called whiskers, which supports them in accurately identifying the location of their target prey. These whiskers are also utilized to determine the position, shape, and dimensions of the prey. In terms of their hunting behaviour, sea lions undergo distinct phases such as

- Sea lions utilize their whiskers to trace and hunt the prey.
- Call other members (Subgroup members) to pursing and encircling the target prey.
- Attack on the target prey.

Mathematical Modelling:

The SLnO algorithm is mathematically exploited through four distinct stages referred to as:

a. Prey tracking b. Social Hierarchy c. Prey attack d. Prey encirclement

Detecting and tracking phase: The whiskers are used to sense the existing prey as well as to detect prey position. The water waves direction matter. Sea lions can sense prey. The whiskers direction should be opposite to the direction of water waves to sense the prey. However, the movement of the sea lion's whiskers is minimal while aligning with the current orientation of water waves.

Sea lion locate the prey location. They call other members to join its subgroup. The purpose is to collectively track and capture the target prey. The leader is the sea lion which directs other members. Sea lion leader continually update the position of target prey. Other members follow the leader position, moving towards the prey [26]. Sea lion algorithm assumes the target prey as the closer one to the optimal solution (best solution). This is denoted in Eq. (14), where the distance between the sea lion and the target prey is denoted as $\overrightarrow{Ds}$. $\vec{S}(t)$ shows vector position of sea lion. $\vec{T}(t)$ shows vector position of target prey. t displays the present iteration. $\vec{L}$ represents a random vector.

$$\overrightarrow{Ds} = \left|2\vec{L}.\vec{T}(t) - \vec{S}(t)\right| \quad (14)$$

With each iteration, the sea lion progressively moves closer to the target prey. The arithmetical modelling is displayed using Eq. (15), in which the leading iteration is given by $(t+1)$ and $\vec{H}$ value gets reduced from 2 to 0 linearly over an iteration.

$$\vec{S}(t+1) = \vec{T}(t) - \overrightarrow{Ds}.\vec{H} \quad (15)$$

Vocalization phase: Sea lions have the ability to adjust to stay in both land and water. The speed of sound when sea lions vocalize underwater is approximately four times faster than when they vocalize in the air. These sea lions use a variety of vocalizations to communicate with each other and to hunts of target prey. Furthermore, they are expert at detecting sounds both above and beneath the water's surface. Consequently, when they spot their target prey, sea lions emit calls to gathering other members for the purpose of encircling and beginning an attack on the prey. This process is quantitatively determined using Eq. (16), (17), and (18), in which $\overrightarrow{S_{leader}}$ [26] shows the speed of sea lion leader's sound. The sound speed in water is demonstrated as $\overrightarrow{C_1}$. The sound speed in air is demonstrated as $\overrightarrow{C_2}$.

$$\overrightarrow{S_{leader}} = \left|\left(\overrightarrow{C_1}(1+\overrightarrow{C_2})\right)/\overrightarrow{C_2}\right| \quad (16)$$

$$\overrightarrow{C_1} = \sin\theta \quad (17)$$

$$\overrightarrow{C_2} = \sin\phi \quad (18)$$

Attacking phase: During the exploration phase, the sea lions' encircling and hunting movements are segmented into the subsequent two stages:

a) Dwindling encircling approach [26]: This approach is executed using the value of $\vec{H}$ Eq. (15). Predominantly, $\vec{H}$ value gets decreased from 2 to 0 over the following iteration. This reducing factor helps sea lions to give direction to move on and encircle the target prey. b) Circle updating position: The prime target of sea lions is the bait balls of fishes. They start an attack from edges which is explicated as per Eq. (19), in which $\vec{T}(t) - \vec{S}(t)$ states distance between a Sea lion and target prey. Sea lion is a search agent and the best optimal solution is target prey. $\left|\;\right|$ represents absolute value. q indicates the random number. q value is between -1 to 1.

$$\vec{S}(t+1)=\left|\vec{T}(t)-\vec{S}(t).\cos(2\pi q)\right|+\vec{T}(t) \quad (19)$$

Prey searching: The update of the sea lion's position is derived from the most effective search agent during the exploration stage. The sea lions (search agent's) position adjustment during exploration is determined by evaluating the chosen random sea lion. The SLnO algorithm performs a global search agent update as presented in Eq. (20). The SLnO algorithm determines the global optimum solution using Eq. (21) while is $\vec{H}$ larger than 1.

$$\overrightarrow{Ds}=\left|2\vec{B}.\overrightarrow{S_{rnd}}(t)-\vec{S}(t)\right| \quad (20)$$

$$\vec{S}(t+1)=\overrightarrow{S_{rnd}}(t)-\overrightarrow{Ds}.\vec{H} \quad (21)$$

5.3 Proposed Improved SLnO Algorithm

Position updating toward the target prey is a significant challenge within the Sea Lion algorithm. Sea lion algorithm has some most fascinating features with it, and yet, some issues exist in this that need to be corrected, which is their lowest convergence speed. Hence, in order to enhance the performance of this optimization model, this work made some improvements in the existing SLnO algorithm and is named as ISLnO algorithm. Self-improvement is proven to be promising in traditional optimization algorithms. The improved version is explained as follows: in the conventional model, based on the $\overrightarrow{S_{leader}}$ in Eq. (16), the whole process is evaluated. In the proposed ISLnO concept, instead of $\overrightarrow{S_{leader}}$, a new random value r value is generated. If r value is less than 0.5, the other condition $H<1$ is verified. If the H value is less than 1, Subsequently, adjust the position of the current search agent using the newly calculated equation provided in Eq. (22). If the condition is not met, the evaluations remain consistent with those of the conventional SLnO algorithm. The algorithmic representation of the proposed ISLnO algorithm can be found in Algorithm 1. Table 2 shows Pseudo-code representation of Improved Sea Lion Algorithm.

$$\vec{S}(t+1)=\left|\sigma(t)-\sigma(t-1)\right|+\sigma(t) \quad (22$$

Table 2 Pseudo-code representation of Improved Sea Lion Algorithm

Algorithm 1: Pseudo-code representation of Improved Sea Lion Algorithm			
Population initialize			
Selected $\overrightarrow{S_{rnd}}$			
Calculate the fitness function for each individual search agent			
if $(i<\max iter)$			
	the random value r (initialize)		
	If(r <0.5)		
		If ($H<1$)	
			Adjust current search agent Position based on new evaluated Eq. (22) - update
		else	
			choose a search agent randomly $\overrightarrow{S_{rnd}}$
			Update the position of the current search agent using Eq. (21)
		endif	
	else		
		current search agent position using Eq. (19) - update	
	endif		
	If the search agent do not consist to any $\overrightarrow{S_{leader}}$		
		Go to the first if condition	
	else		
		Evaluate the every search agent fitness function	
		Update $\vec{S}$ according to better solution	
		Return $\vec{S}$, (best solution)	
	endif		
endif			
stop			

6.0 RESULTS AND DISCUSSION

6.1 Experimental Setup

The implementation of this established cross domain ontology construction approach was evaluated using JAVA. Cross domain data can be cross platform data or cross network data. In this, the simulation is performed using datasets times of India and newsgroup. The datasets is taken from Kaggle. The link is given below:

1. https://www.kaggle.com/therohk/india-headlines-news-dataset - times of India
2. https://www.kaggle.com/crawford/20-newsgroups-Newsgroup

Herein, the cross-domain applications data used has to be uploaded to the IoT channel and then should be retrieved back from it. In this work, the Thing Speak channel is used which is an open-source IoT application and API to store and retrieve data from things using the HTTP and MQTT protocol over the Internet or via a Local Area Network. The performance of the implemented model is assessed through a comparison with conventional models, such as EM clustering [29], Semantic similarity [29], and CI-ROA [28]. CI-ROA, which stands for Circling Insisted-Rider Optimization Algorithm is hybridization of whale optimization and rider optimization algorithms. The calculation of both the newly proposed and conventional models involves varying cluster sizes and training percentages to evaluate their performance in terms of recall and precision metrics. Additionally, changes in training percentage and cluster size are used to analyse the time taken for ontology construction and overall process execution.

6.2Performance Analyzed using Training Percentage

Figure 4 shows the performance of the constructed model by changing training percentage. In this cross-domain applications are used such as times of India, newsgroup. The performance analysis is measured by precision and recall. The different training percentage like 60,70, 80, and 90 is used for evaluation. Cluster size is 3 for all training percentages.

The results of the cross domain-based ontology construction model are presented using information retrieval standard evaluation metrics, namely the precision and recall. It is measured based on user query. The precision and recall values range is between 0 and 1. Recall value is 0% means no relevant information are retrieved. Recall value is 100% means all relevant information is retrieved. Precision value is 0% means all irrelevant information are retrieved. Precision value is equal to 100% means all relevant information is retrieved. Also, ontology construction time means how much time it takes to construct ontology is computed. Finally, execution time means how much time model takes to execute user query is evaluated.

Figure 4 exhibit the analysis regarding the performance of the adopted model against traditional terms for cross-domain applications like times of India, newsgroup. This performance analysis is evaluated based on measures like precision and recall. The proposed ontology construction model is reviewed by varying the training percentage. We have considered training percentages of 60, 70, 80, and 90. Cluster size is constant 3. Figure 4 shows the performance analysis regarding the cross platforms of the proposed model by comparing it over the other state of art models. In the view of the newsgroup domain in Figure 4(a), the implemented model regarding precision measure at learning percentage 60 achieves a superior precision, which is 17.46%, 4.53%, and 2% improved than EM clustering [29], Semantic similarity [29], and CI-ROA [28], respectively. While taking the recall measure in Fig. 5(b), the implemented model at training percentage 60 is 7.22%, 3.78% and 2.66% better than the conventional models like EM clustering [29], Semantic similarity [29] and CI-ROA [28], respectively. The same evaluation is carried out for the other cross platforms such as times of India, Flipkart and Amazon. It shows the performance of the implemented model is improved and validated.

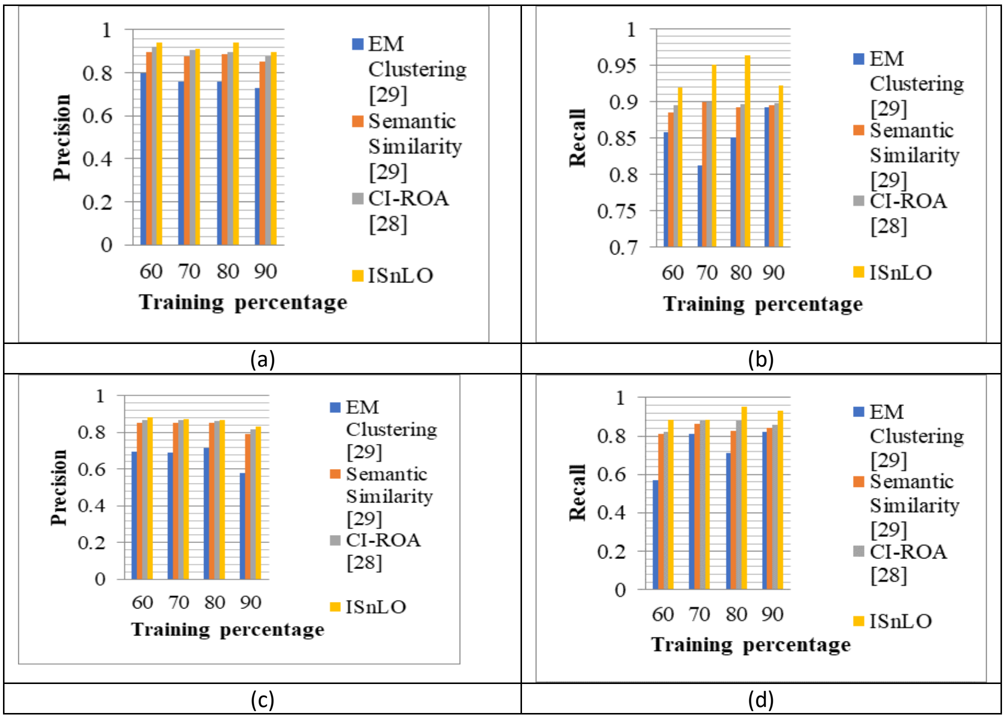


Figure 4 Performance analysis of the ontology construction model over conventional models by changing training(learning)percentage for two cross platforms (a) and (b) Precision and Recall(Newsgroup) (c) and (d) Precision and Recall (Times of India)

6.3 Performance Analyse Using Cluster Size

Different cluster size is used to check and validate implemented model performance. Figure 5 represent the implemented work performance using cross-domain applications that are explained earlier by changing the size of the cluster. In this cluster size is considered as 2, 3, 4, and 5 for performance analysis. The precision and recall are measured by using different cluster size as 2, 3, 4, and 5. Cluster size represents the total number of clusters. The training percentage is fixed and that is 70. In Fig. 5, the performance analysis of the adopted model with conventional terms for precision and recall measure for cross platforms like Newsgroup and Times of India are illustrated. Figure 5(a), the precision measure of ontology construction model using newsgroup data is better than EM clustering [29], Semantic similarity [29], and CI-ROA [28] by 16.07%, 3.16%, and 1.24%, respectively using cluster size 4. The developed model is validated with the best precision and recall performance than other state of art models by performing the same set of process for other cross-platform applications as well

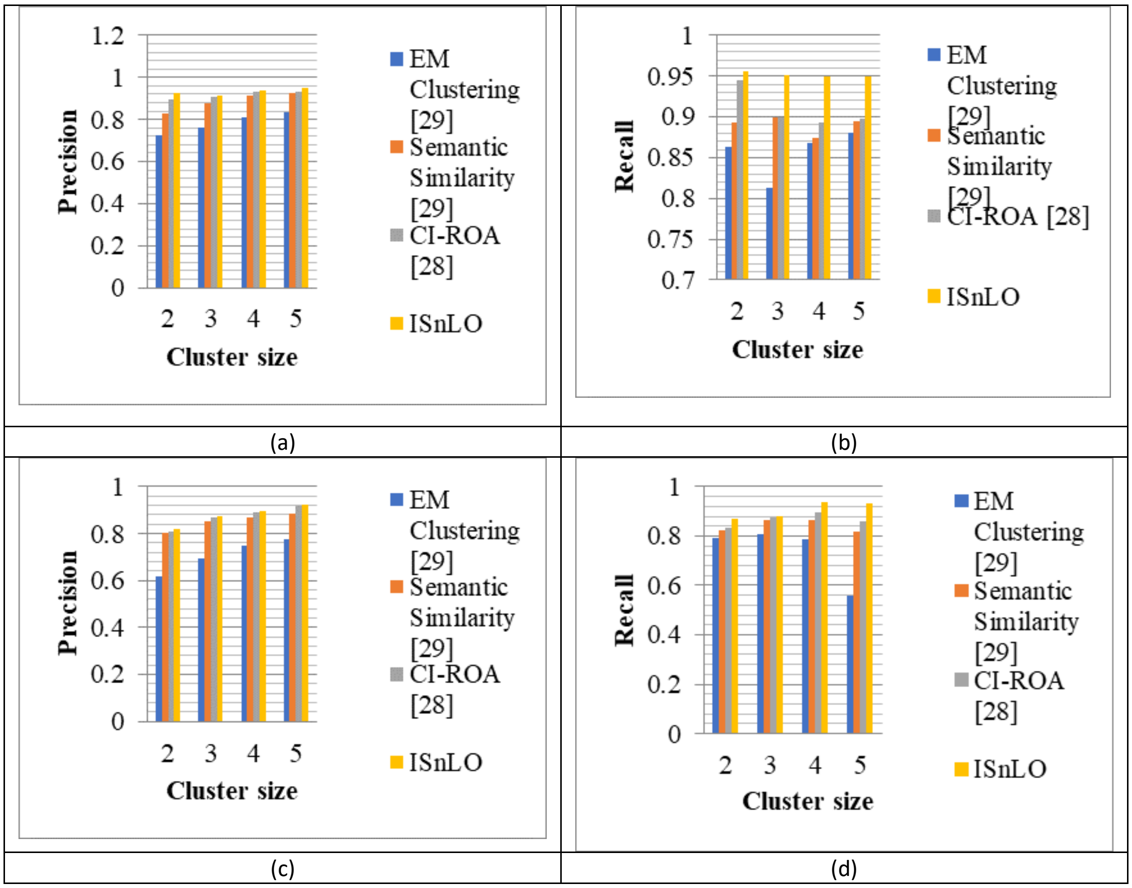


Figure 5 Performance analysis of the implemented ontology construction model over traditional models by changing cluster size for two cross platforms (a) and (b) Precision and Recall (Newsgroup) (c) and (d) Precision and Recall (Times of India)

6.4 Ontology Construction and Execution Time by changing Training Percentage

The performance analysis of the established model in the view of ontology construction and execution time for four cross-domain applications over the traditional models is illustrated in Table 3. The Ontology Construction and Execution Time are evaluated by changing the training percentage with fixed cluster size 3. The average of total 6 independent runs are taken for result analysis. In this, the time is estimated in nano seconds. Under the times of India application, the time taken for ontology construction is achieved for the proposed model at training percentage 60 is achieved with less time than conventional models like EM clustering, Semantic similarity, and CI-ROA by 64.05%, 63.4%, and 61.62%, respectively. Similarly, regarding the process execution, the time taken by the implemented model at training percentage 70 has attained the least time, which is 65.38%, 64.08%, and 62.42% improved than EM clustering [29], Semantic similarity [29] and CI-ROA [28], respectively. This similar analysis is evaluated for the other cross domain applications and the outcomes are analysed. The overall analysis thus confirmed the better performance of the implemented model with the least time for ontology construction and process execution time.

Table 3 Analysis of implemented model and conventional model regarding ontology construction and execution time with different training percentage

Newsgroup					
Ontology construction time	Training %	EM Clustering	Semantic Similarity	CI-ROA	ISLnO
	60	3.95×10^{11}	3.88×10^{11}	3.70×10^{11}	1.42×10^{11}
	70	4.94×10^{11}	4.83×10^{11}	4.56×10^{11}	1.72×10^{11}
	80	$4.57\times10e^{11}$	4.49×10^{11}	4.26×10^{11}	1.74×10^{11}
	90	$4.73\times10e^{11}$	4.62×10^{11}	4.35×10^{11}	2.12×10^{11}
Execution time	Training %	EM Clustering	Semantic Similarity	CI-ROA	ISLnO
	60	3.95×10^{11}	3.86×10^{11}	3.68×10^{11}	1.42×10^{11}
	70	4.94×10^{11}	4.76×10^{11}	4.55×10^{11}	1.71×10^{11}
	80	4.57×10^{11}	4.45×10^{11}	4.26×10^{11}	1.73×10^{11}
	90	4.73×10^{11}	4.59×10^{11}	4.34×10^{11}	2.11×10^{11}
Times of India					
Ontology construction time	Training %	EM Clustering	Semantic Similarity	CI-ROA	ISLnO
	60	1.66×10^{10}	1.60×10^{10}	1.58×10^{10}	9.90×10^{08}
	70	3.74×10^{09}	3.31×10^{09}	3.09×10^{09}	3.09×10^{09}
	80	3.86×10^{09}	3.25×10^{09}	3.05×10^{09}	9.90×10^{08}
	90	4.07×10^{09}	3.61×10^{09}	3.19×10^{09}	1.84×10^{09}
Execution time	Training %	EM Clustering	Semantic Similarity	CI-ROA	ISLnO
	60	1.65×10^{10}	1.60×10^{10}	1.57×10^{10}	9.27×10^{08}
	70	3.72×10^{09}	3.28×10^{09}	2.98×10^{09}	2.89×10^{09}
	80	3.84×10^{09}	3.23×10^{09}	2.92×10^{09}	9.27×10^{08}
	90	4.05×10^{09}	3.39×10^{09}	3.05×10^{09}	1.72×10^{09}

6.5 Ontology Construction and Execution Time by Changing Size Of Cluster

Table 4 elucidates the time analysis of the proposed model for four cross-domain applications. The review of the implemented model is made regarding the ontology construction time and execution time analysis of the whole process, by changing the cluster size with fixed training percentage of 70. The time is taken in nanoseconds. The proposed work with process execution time is achieved with betterment at cluster size 2 than conventional models like EM clustering, Semantic similarity, and CI-ROA by 38.91%, 27.72%, and 10.98%, respectively. Thus, the analysis has achieved superior performance for the established model while comparing it over the conventional works.

Table 4 Analysis of proposed ontology construction model and conventional model regarding ontology construction and execution time with different cluster size

Newsgroup					
Ontology construction time	Cluster size	EM Clustering	Semantic Similarity	CI-ROA	ISLnO
	2	4.04×10^{11}	3.95×10^{11}	3.62×10^{11}	1.31×10^{11}
	3	4.94×10^{11}	4.83×10^{11}	4.56×10^{11}	1.72×10^{11}
	4	6.04×10^{11}	5.92×10^{11}	5.75×10^{11}	2.35×10^{11}
	5	5.75×10^{11}	5.65×10^{11}	5.39×10^{11}	2.48×10^{11}
Execution time	Cluster size	EM Clustering	Semantic Similarity	CI-ROA	ISLnO
	2	4.04×10^{11}	3.81×10^{11}	3.61×10^{11}	1.30×10^{11}
	3	4.94×10^{11}	4.76×10^{11}	4.55×10^{11}	1.71×10^{11}
	4	6.03×10^{11}	5.92×10^{11}	5.74×10^{11}	2.34×10^{11}
	5	5.75×10^{11}	5.64×10^{11}	5.38×10^{11}	2.47×10^{11}
Times of India					
Ontology construction time	Cluster size	EM Clustering	Semantic Similarity	CI-ROA	ISLnO
	2	2.59×10^{09}	2.16×10^{09}	1.84×10^{09}	1.16×10^{09}
	3	3.74×10^{09}	3.31×10^{09}	3.09×10^{09}	3.09×10^{09}
	4	2.73×10^{09}	2.26×10^{09}	1.92×10^{09}	1.08×10^{09}
	5	2.87×10^{09}	2.31×10^{09}	1.93×10^{09}	9.00×10^{08}
Execution time	Cluster size	EM Clustering	Semantic Similarity	CI-ROA	ISLnO
	2	2.57×10^{09}	1.98×10^{09}	1.68×10^{09}	1.09×10^{09}
	3	3.72×10^{09}	3.28×10^{09}	2.98×10^{09}	2.89×10^{09}
	4	2.71×10^{09}	2.07×10^{09}	1.78×10^{09}	9.84×10^{08}
	5	2.85×10^{09}	2.28×10^{09}	1.80×10^{09}	8.28×10^{08}
	2	1.96×10^{09}	6.40×10^{08}	4.92×10^{08}	4.05×10^{08}
	3	2.39×10^{09}	2.02×10^{09}	1.64×10^{09}	1.46×10^{09}
	4	2.15×10^{09}	1.83×10^{09}	1.51×10^{09}	6.97×10^{08}
	5	2.40×10^{09}	2.05×10^{09}	1.68×10^{09}	7.54×10^{08}

7.0 CONCLUSIONS

This research work described optimization-based clustering approach that is particularly designed to improve precise information retrieval using semantic web technology. In this thesis, ontology construction models and semantic web information retrieval models are reviewed in the literature review. As a main part of this framework, it is divided into two phases that is using data filtering and data annotation. Data preprocessing is proposed to preprocess the data taken through IOT the channel. Further, meaningless or less important information is filtered using Jaccard similarity evaluation. The selected data is divided into wanted and unwanted data after applying Jaccard similarity. The unwanted words are called outliers. After this, semantic annotation- term expansion is computed for total selected count words. In this, each word is processed by term expansion to find the semantic of a particular word. There is a group of semantically related words that should be clustered. However, selecting the best centroid for clustering is a difficult process to form a precise cluster. To overcome this, optimization-based clustering to select optimal centroid model is developed. Then, data annotation is performed on selected data. Next, this research presents cross domain-based ontology construction with optimization algorithm to retrieve accurate information. Further, CI-ROA and ISnLo algorithms are proposed to reduce ontology construction time and execution time.

The experiment results show proposed ontology construction models perform better than conventional models. The performance is measured using information retrieval metrics such as precision and recall. The results show that ontology construction time and execution time of proposed models using CI-ROA and ISnLO are reduced. In the view of the newsgroup domain, the implemented model regarding precision measure at training percentage 60 achieves a superior precision, which was 17.46%, 4.53%, and 2% improved than EM clustering, Semantic similarity, and CI-ROA, respectively. The recall measure o the implemented model at training percentage 60 was 7.22%, 3.78%, and 2.66% better than the conventional models like EM clustering, Semantic similarity, and CI-ROA, respectively. The experiment results show proposed ontology construction models perform better than conventional models. The performance is measured using information retrieval metrics such as precision and recall. The results show that ontology construction time and execution time of proposed models using CI-ROA and ISnLO are reduced. In the future, semantic web will bridge the understanding gap between human and machine completely. Semantic web makes data machines understandable.

Web 4.0 is the next version of semantic web (Web 3.0). There is no specific definition of web 4.0 yet. It will be adding value in education 4.0 and industry revolution 4.0. In the future, everything will be connected to the web using different technology such as artificial intelligence, block chain, machine learning, data science, big data, cyber security and robotics. Everything will be processed automatically and intelligently through web 4.0 like the human brain.

Acknowledgement

The author would like to express gratitude to all people who have directly or indirectly contributed towards the successful completion of this research.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper

References

[1] Zhou, Jianhui, et al, 2021. "Building Real-Time Ontology Based on Adaptive Filter for Multi-Domain Knowledge Organization," *IEEE Access.* 9: 66486-66497.

[2] del Mar Roldán-García, María, et al, 2021"Ontology-driven approach for KPI meta-modelling, selection and reasoning," *International Journal of Information Management* 58: 102018.

[3] Li, Man, Xiaoyong Du, and Shan Wang, 2005 "A semi-automatic ontology acquisition method for the semantic web," Advances in Web-Age Information Management: *6th International Conference, WAIM 2005, Hangzhou, China, October 11–13,* 2005. Proceedings 6. Springer Berlin Heidelberg.

[4] Chi, Nai-Wen, Yu-Huei Jin, and Shang-Hsien Hsieh. 2019. "Developing base domain ontology from a reference collection to aid information retrieval," Automation in Construction vol. 100: 180-189

[5] Xing, Zimeng, et al, "Application of ontology in the web information retrieval," *Journal on Big Data*. 1(2):79.

[6] Benkhaled, Sihem, et al, 2022. "An ontology–based contextual approach for cross-domain applications in internet of things," *Informatica.* 46(5)

[7] Kumar, CS Saravana, and R. Santhosh, 2020. "Effective information retrieval and feature minimization technique for semantic web data," *Computers & Electrical Engineering.* 81: 106518.

[8] Palmer, Claire, et al, 2018. "An ontology supported risk assessment approach for the intelligent configuration of supply networks," *Journal of Intelligent Manufacturing*.29: 1005-1030.

[9] Zhuang, Lisa, Kim Schouten, and Flavius Frasincar, 2020. "SOBA: Semi-automated ontology builder for aspect-based sentiment analysis," *Journal of Web Semantics.* 60: 100544.

[10] Wiśniewski, Dawid, et al, 2019. "Analysis of ontology competency questions and their formalizations in SPARQL-OWL," *Journal of Web Semantics*. 59: 100534.

[11] Liu, Gang, and Hanwen Zhang, 2020. "An ontology constructing technology oriented on massive social security policy documents," *Cognitive Systems Research.* 60: 97-105.

[12] Alobaid, Ahmad, et al, 2019. "Automating ontology engineering support activities with OnToology," Journal of Web Semantics.57: 100472.

[13] Ming, Zhenjun, et al, 2020. "An ontology for representing knowledge of decision interactions in decision-based design," *Computers in Industry*. 114: 103145.

[14] Meyer, David E., et al, 2020. "Enhancing life cycle chemical exposure assessment through ontology modeling," *Science of The Total Environment*. 712: 136263.

[15] Masmoudi, Maroua, et al, 2020. "MEMOn: modular environmental monitoring ontology to link heterogeneous earth observed data," *Environmental Modelling & Software* .124: 104581.

[16] Karimi, Hamed, and Ali Kamandi, 2019. "A learning-based ontology alignment approach using inductive logic programming," *Expert systems with applications* .125: 412-424.

[17] Ming Tao, Kaoru Ota, Mianxiong Don, 2017. "Ontology-based data semantic management and application in IoT- and cloud-enabled smart homes , " *Future Generation Computer Systems.*76: 528-539.

[18] Tao, Ming, Kaoru Ota, and Mianxiong Dong, 2017. "Ontology-based data semantic management and application in IoT-and cloud-enabled smart homes," *Future Generation Computer Systems*. 76: 528-539.

[19] Gyrard, Amelie, and Amit Sheth, 2020. "IAMHAPPY: Towards an IoT knowledge-based cross-domain well-being recommendation system for everyday happiness," *Smart Health*. 15: 100083.

[20] Zheng, Zeqi, et al, 2020. "Modeling and analysis of a stock-based collaborative filtering algorithm for the Chinese stock market," *Expert Systems with Applications* .162: 113006.

[21] Liu, Jixiong, Weike Pan, and Zhong Ming, 2020. "CoFiGAN: Collaborative filtering by generative and discriminative training for one-class recommendation," *Knowledge-Based Systems* .191: 105255.

[22] Casino, Fran, Constantinos Patsakis, and Agusti Solanas, 2019. "Privacy-preserving collaborative filtering: A new approach based on variable-group-size microaggregation," *Electronic Commerce Research and Applications*. 38: 100895.

[23] Manotumruksa, Jarana, Craig Macdonald, and Iadh Ounis, 2020. "A contextual recurrent collaborative filtering framework for modelling sequences of venue checkins," *Information Processing & Management* 57(6): 102092.

[24] Iwanaga, Jiro, et al, 2019. "Improving collaborative filtering recommendations by estimating user preferences from clickstream data," *Electronic Commerce Research and Applications*. 37: 100877.

[25] Duma, Mlungisi, and Bhekisipho Twala, 2019. "Sparseness reduction in collaborative filtering using a nearest neighbour artificial immune system with genetic algorithms," *Expert Systems with Applications.* 132: 110-125.

[26] Masadeh, Raja, Basel A. Mahafzah, and Ahmad Sharieh, 2019."Sea lion optimization algorithm," *International Journal of Advanced Computer Science and Applications.* 10(5): 388-395.

[27] Ferilli, Stefano, Floriana Esposito, and Domenico Grieco, 2014. "Automatic learning of linguistic resources for stopword removal and stemming from text," *Procedia Computer Science.* 38: 116-123.

[28] Kakad, Shital, and Sudhir Dhage, 2021. "Cross domain-based ontology construction via Jaccard Semantic Similarity with hybrid optimization model," *Expert Systems with Applications.* 178: 115046.

[29] Kakad, Shital, and Sudhir Dhage, 2021. "Ontology construction from cross domain customer reviews using expectation maximization and semantic similarity," *2021 International Conference on Emerging Smart Computing and Informatics (ESCI). IEEE,*

[30] Nilashi, Mehrbakhsh, Othman Ibrahim, and Karamollah Bagherifard, 2018. "A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques," *Expert Systems with Applications*. 92: 507-520.

[31] Geng, Qian, et al, 2020. "Cross-domain ontology construction and alignment from online customer product reviews," *Information Sciences.* 531: 47-67.

[32] Shoaip, Nora, et al. 2021. "A comprehensive fuzzy ontology-based decision support system for alzheimer’s disease diagnosis." *IEEE Access.* 9: 31350-31372.

[33] Kakad, Shital, and Sudhir Dhage. 2022. "Knowledge Graph and Semantic Web Model for Cross Domain" *Journal of Theoretical and Applied Information Technology*. 100(16).