Development of a Machine Learning Model for Knowledge Acquisition, Relationship Extraction and Discovery in Domain Ontology Engineering using Jaccord Relationship Extraction and Neural Network

Sivaramakrishnan R. Guruvayur, R. Suchithra

Abstract: ***Creating a fast domain independent ontology through knowledge acquisition is a key problem to be addressed in the domain of knowledge engineering. Updating and validation is impossible without the intervention of domain experts, which is an expensive and tedious process. Thereby, an automatic system to model the ontology has become essential. This manuscript presents a machine learning model based on heterogeneous data from multiple domains including agriculture, health care, food and banking, etc. The proposed model creates a complete domain independent process that helps in populating the ontology automatically by extracting the text from multiple sources by applying natural language processing and various techniques of data extraction. The ontology instances are classified based on the domain. A Jaccord Relationship extraction process and the Neural Network Approval for Automated Theory is used for retrieval of data, automated indexing, mapping and knowledge discovery and rule generation. The results and solutions show the proposed model can automatically and efficiently construct automated Ontology.***

Keywords: ***Automatic Ontology Generation, Jaccord Relationship Extraction, Neural Network, Semantic Web***

I. INTRODUCTION

Data capturing required to create a knowledge base is very complex. A computation model is also essential to draw solution from inference models and knowledge is represented in the form of knowledge. It is difficult to obtain the significant data desired from the vast database available on the Internet. Search engine (SE) acts as a significant position in overcoming this difficulties. SE employs the browser to retrieve information on the site. In general, users enter a number of keywords into the browser, SE execute keyword searches as well as offer appropriate outcome as a result. It is annoying for the average user to recognize the work of AS [16]. It is impossible to recognize the connection between domain-specific terms as well as employ these appropriate terms for improved results [17].

Ontologies are employed in a growing number of applications, especially websites, and have become the chosen template tool. The plan with the maintenance of theory is a large procedure [18-24]. Ecological storage, which has lately become a significant knowledge for theorizing, engages the automatic recognition of ideas in a region as well as the interconnection of ideas [25].

Unfortunately, building and maintaining a theory is a complicated mission. Classical theoretical construction relies on domain expertise, except is expensive, time consuming, and complex [26]. As well as the need of standards, the field also requires a method for acquiring full automation information theory building, which is a time prone as well as expensive process. Although present methods of building a theory can attain partial computerized categorization, there are boundaries such as knowledge engineering requirements and limitations. To address the above issues, this manuscript develops a new methodology based on the process of extracting functional contacts and the neural approach to autocorrelation..

This paper is organized as follows "Section 2 discussed about the various work done on the ontology domain. The first steps for establishing the ontology and updating and improving the ontology are introduced in Section III. Section 4 describes the investigational results of the proposed method. In this section, the proposed method is evaluated with the existing techniques to show the effectiveness of the new technique. Conclusion as well as future work of this study is offered in Section V.

II. RELATED WORK

In ontology training, language techniques are also employed to derive words, ideas, and relationships. Symmetric structure analysis and subtype frames were employed to derive words. Other approaches employed are dependency analysis and the synthesis of synthesized syntactic relationships. In addition, vocabulary can also be employed to derive ideas and relationships. In addition, domain extraction and domain specificity are enhanced by introducing seed terms into the ECG training pipeline.

Revised Manuscript Received on September 23, 2019

***** Correspondence Author

Sivaramakrishnan R Guruvayur*, Department of Computer Science, Jain University, Karnataka, India, Email: srkgr1@gmail.com

R. Suchithra, Department of Computer Science, Jain University, Karnataka, India, Email: suchithra.suriya@gmail.com

Statistical methods are based on the statistics of large corporations as well as do not take into account the technicalities. Most statistical techniques are widely employed probabilities and are often employed in the early stages of microbiology training after pre-processing languages. These techniques are mainly employed for word retrieval, concept retrieval, and online link download.

Statistical techniques include c values. N. E. A.S. X contrast analysis, grouping, joint event analysis, and ARM.

IIC is a machine learning discipline that emerges from a hypothesis based on informationalong with a set of examples employing logical software. In the field of science, ILP is employed in the last stages, where the universalsolution is derived from the schematic idioms. Lima et al. [1] employed IT techniques to dericeontology online. In their job, they employed two sources of evidence: WordNet in addition to an independent language model for the domain. They employed the template to identify candidates for the class. The two evidentiary resources are combined as next-generation knowledge for IL-based automated acquisition.

They extracted 2,100 sentences employing the Bing Search Engine API as well asestimated the work with or without WordNet. They received 96% and 98% of the best improvements, respectively, with and without WordNet. Fortuna et al. [2] developed an inventive method, namely, deriving the term for scientific theory from text documents. Their method has been successfully employed by AI. L. Phil to create a science topic. To experiment with the proposed method, they employed the file as an index to be indexed in the AI publishing database. ILPN2.

Seneviratne and Ranasinghe [3] illustrated the benefit of ILP which is a knowledge method to obtain biological contacts in a multi-agent scheme. Within this multi-agent scheme, an agent employs IT. L. Phil for the policy learning process, while other agencies use those rules to identify new relationships. They employed a bird-related Wikipedia site to assess the proposed method. Lily et al. [4] employed the IP method for interdisciplinary studies, as it provided a wealth of conceptual knowledge in the form of theoretical theories, usually developed in logic (DL).

The authors have looked at the difficulty of merging science along with related informationas well as suggested ILP elements as solutions. Their proposed methodology is based on the withdrawal power and presentation of the CRD + SD information registration. It allowed for a very strong combination of DC and unreasonable negative information registration. They claim that their approach lays the groundwork for the expansion of the study known as pathology training. Lily et al. [15] illustrated a logic-based computational method to persuade AI. M. IT Inc. They demonstrated the benefits of their approach using the proposed methodology in tourism. Their approach is a great contribution to managing the development of autism.

The first paper [5] presents a method called PACTOLE (Patented Articles and Classes) to generate new theories from astronomical texts. The first step is to analyze text collections using the L technique. L and to derive domain objects and their properties using a predefined syntax model Then in the second step, technique A. E. E. E. Applies to pairs (objects, objects) for the purpose and ha created a concept grid where each idea is a set of maximum objects that divides the maximum number of properties. The third step consists of presenting a database of existing objects in the sky through the second panel of the concept, using FCA techniques of ideas.

In the fifth step, the concept hierarchy is presented in the FLE descriptive language in order to perform its reasoning tasks. This method has been implemented in a large number of abstract astronomical journals and has an existing SBSDB object database, and validity scores are high (74.71%), which means that items are classified in a sufficient class .The recall is low because most of the assets involved are inadequate for the treatment . The second paper [6] presents a framework called Theory of Optimization and Theory Evaluation using RSS feeds.The enrichment of an otology is proceeded using OpenNLP API, which is a natural language processing Library, and WordNet [7] resource.

Statistical approaches are implemented to extract links and concepts from RSS feeds using the OpenNLP API. After the capacity enhancement phase, the authors use several indicators to measure how the original science was changed. Ontorich was compared to two scientific enhancement systems, Kaon and Neaon, and compared to two other pathological evaluation systems, OntoQA and Romeo, in terms of some functional criteria, and the results showed that Ontorich was more numerous.The third paper [8] presents a framework based on machine learning strategies for extracting nonlinear relationships that remain a major challenge for the pathology training community. The first proposed framework extracts from a set of Contextual Contextual Constructions (CMSs) from the commentary body and WordNet [7] to use as the first indicator on which to find a good candidate sentence that may be causal.

In the second phase, a new algorithm is applied to show the true existence of the cause and the relationship contained in the sentence, and if so, marks both sides of the relationship (cause-effect). To that end, the sentences are divided into two parts, and the most representative words in each section are searched based on the structure of the hyperbola. The sentences employed were as follows: specificity = 78%, recall = 68%, and function = 73%.Document 4 [9] proposes an automated process for the scientific people of an article. The proposed process is independent of the field of discourse and aims to enrich the new theory with non-literal relationships and examples of anthropological properties. The process has three stages: recognition of applicant, building the ranking and ranking of candidates in science. Identification of Candidates Natural language processing techniques were applied to the scene to identify non-literal and scientific links, noting the recommended authorities.

The "Builder Classification" stage applies information extraction techniques to build classifiers based on language rules from science and questions on the lexical database. This stage consists of a skeleton and a pathology, the inputs and outputs of the classifications employed in the "Classification of Objects" stage to correlate the retrieved objects through the science classes. Using this classifier as a first-person descriptive and pathological classification, the classifications of these examples form a population-based science. Implementation of this process is applicable to the legal domain, yielding 90% accuracy, 89.50% accuracy as Rec and 89.74% as FF. The authors conducted other experiments of their efficiency on the tourism domain and achieved 76.50% accuracy, 77.50% accuracy as well as 76.90% FF.

Finally, the paper [10] presents a model based on pathology-driven models of anonymous communication obtained from Arabic corporations. The theory of pairs of anthony varieties is employed to derive a glossary - a synthetic model in which pairs are encountered. These models are then employed to discover innovative pairs of sets of Arabic corporations. The method was tested on three different Arab entities: the Arabian Armed Forces (KSC) [11], the modern Arabian Agency (CCC) [12], and the United Arab Emirates (KC). C.C. 13). Properly extracted models have been employed to improve the pathology based on the theory for Arabian symbology called Semitic [14]. The developed system consists of a set of KSUCCA templates and an enclosure as inputs. Initially, a given body is preprocessed to remove the punctuation from the text, and, using the acronym corresponding to the model, is extracted and evaluated by a professional evaluator and a new pair of axes is added to the theory.

Systems are evaluated using three measures of reliability, model precision, and system efficiency. The reliability of the model is the ratio of the correct variables extracted using the model to the total extractor using the same model. System specificity is the ratio of total accuracy obtained to total extracts, while system efficiency is a measure of pathologic amplitude. The results obtained showed that, despite the high efficiency of the system (42.3%), the accuracy of the calculated system was approximately 29.45%, on average, of the precision obtained relative to all the houses employed (KSUCCA, CCA and KACSTAC).

In short, we can say that, first of all, the above approaches consider only one type of relationship: punitive or non-taxable. Second, the effectiveness of the above method depends on the target area. Our proposed methodology aims to examine two types of relations, von Dominic or non-behavioral, and to maintain the relevance of new theories through the use of techniques to derive the Dorian relation, regardless of discourse.

III. ONTOLOGY MODEL CREATION

The ontology representation projected in this document is based on extracting Jaccord relations from text documents and using conceptual along with relational ontological models. Innovative of the scientific model is a combination of the use of two different extraction schemes: the automation ratio (AER) and subjective communication automation (AERV) and the result verification by the third scheme, external service descriptor analysis.

Fig. 1. Ontology Designing Process

We employed these three approaches to show the feasibility of our model. Other more sophisticated approaches, such as machine learning (M) and information retrieval (EIS), can also be employed to implement the representation. Though, the use of direct approaches emphasizes that many approaches can be “accounted for” and that the results are attributed to the process of combining and verifying the model. The overall process of ontology is illustrated in Figure 1.

There are four key phases in the process.. First phase of this process is pre-processing the input web source documents. In the preprocessing step, the information normalization, information harmonization and Natural Language Processing (NLP) are done. In the second step, automated attribute selection, automated entity relationship model creation and automatic validation process is used. In the third step, Web context extraction employs a search engine query encoding that aggregates results by descriptors. The fourth step is to classify the descriptorsthat describe the context of the web service. Finally, the evolutionary steps of science expand upon scientific discoveries as they discover new ones and change the relationships between them.External Web Service descriptors serve as mediators if there is a conflict between a new theory and a new idea. Communication is defined as an ongoing process according to the general context of the concept.

A. Preprocessing

This is the first step of this process. In the preprocessing step, the information normalization, and information harmonization are performed for cleaning the information. In the information normalization step, the total web source document is processed and each line are extracted and is stored separately. This process is done on separate web source documents. In information harmonization, all these normalized information are combined and then form a common information documents. And then the stop word removal process is performed to reduce the execution time and extract only the concepts.

After this, speech recognition (POS) process is implemented. POS is also referred to as grammatical marking or marking, the type of word recognition in text (bodies) that corresponds to specific parts of a discourse based on both its definition and context. Its relation to the words is closest to and related in a sentence or paragraph. In this paper opennlp.tools.postag and opennlp.tools.tokenize API are employedto tag each and every word. The example of POS tagging table is shown below Table 3.2.

Tag	Description	Example	Tag	Description	Example
CC	coordin. conjunction	*and, but, or*	SYM	symbol	+,%, &
CD	cardinal number	*one, two*	TO	"to"	*to*
DT	determiner	*a, the*	UH	interjection	*ah, oops*
EX	existential 'there'	*there*	VB	verb base form	*eat*
FW	foreign word	*mea culpa*	VBD	verb past tense	*ate*
IN	preposition/sub-conj	*of, in, by*	VBG	verb gerund	*eating*
JJ	adjective	*yellow*	VBN	verb past participle	*eaten*
JJR	adj., comparative	*bigger*	VBP	verb non-3sg pres	*eat*
JJS	adj., superlative	*wildest*	VBZ	verb 3sg pres	*eats*
LS	list item marker	*1, 2, One*	WDT	wh-determiner	*which, that*
MD	modal	*can, should*	WP	wh-pronoun	*what, who*
NN	noun, sing. or mass	*llama*	WP$	possessive wh-	*whose*
NNS	noun, plural	*llamas*	WRB	wh-adverb	*how, where*
NNP	proper noun, sing.	*IBM*	$	dollar sign	$
NNPS	proper noun, plural	*Carolinas*	#	pound sign	#
PDT	predeterminer	*all, both*	"	left quote	' or "
POS	possessive ending	*'s*	"	right quote	' or "
PRP	personal pronoun	*I, you, he*	(	left parenthesis	[, (, {, <
PRP$	possessive pronoun	*your, one's*	)	right parenthesis	],), }, >
RB	adverb	*quickly, never*	,	comma	,
RBR	adverb, comparative	*faster*	.	sentence-final punc	. ! ?
RBS	adverb, superlative	*fastest*	:	mid-sentence punc	: ; ... – -
RP	particle	*up, off*			

B. Attribute Selection and Entity Relationship Model Generation

After processing step is completed, the next step is to select the attribute from the pre-processed information. This attribute selection is employed to decrease the dimensionality of the information. With the aim of reduce the informationdimension,the ontology construction speed is improved. In attribute selection,the word vector is reduced to remove the unnecessary sequence of words in to statement. After reducing the dimension of the information, the next step is to construct the Entity Relationship Model. In this relationship model generation, Segregate the Subject/Object is done, and the Predicates (Relationship) in every given context is found.

C. Ontology Construction

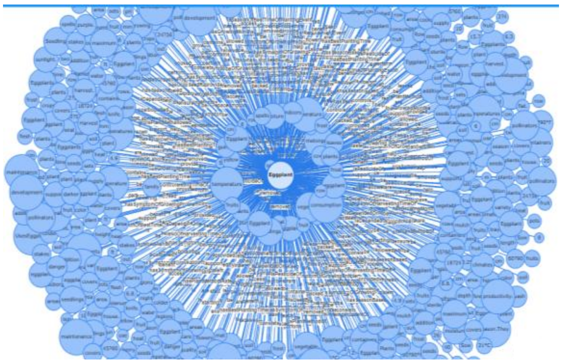
In this step, the Ontology Model is constructed using Jaccard relation estimator. Take a look at two terms, K and L, showing input of the same length (width) at the threshold. A Jacquard measure can be applied directly to assess the similarity between two terms:

$$J(K,L)=(|K\cap L|)/(|K\cup L|) \qquad (1)$$

Jaccard Relation Estimator Algorithm:

Step 1: collect the terms for the given dataset and generate the term list
Step 2: token each term with 'token_term'
tokenized = sent_tokenize(txt)
for i in tokenized:
wordsList = nltk.word_tokenize(i)
Step 3: provide the POS tagging for each term after classification of terms
tagged = nltk.pos_tag(wordsList)
Step 4: perform the dimensional reduction by establishing the relationship among the terms and remove unwanted terms.
Generate group C1= set of token[i]
Get group C1
Generate F = (D,A,C)
Step 5: indulge the process of mapping to obtain the structured information in the structure of subject, predicates and object.
Generate subject, object, and predicates from F
Step 6: Using the Jaccard relation estimator, establish the similarity across the subject, object and predicates that were obtained from the given dataset
Get token [q], token [q+1]
Perform j(token [q], token{q+1]
Display J
Step 7: from the established similarity, construct the sub domain ontology for weather, soil and pest
Step 8: integrate all the obtained sub domain and generate the general ontology for agricultural domain

In this paper, the ontology is constructed from various domains such as Agriculture, health care, food and Books etc. The constructed ontology employed in the proposed approach is shown in below figure. The domain properties of these constructed ontology are given in below table.



Properties for sub domain ontology construction

ABOUT-properties		
belongsTo	hasQuantity	isLandPreparationEventOf
consists	hasQuantityForControlMethodEvent	isLandPreparationInforOf
contains	hasRelatedCausalAgent	isLocationOfAverageYieldEvent
controls	hasRelatedCauseForGrowingProblem	isLocationOfBestTimeOfPlantingEvent
dependsOn	hasRelatedControlMethod	isLocationOfControlMethodEvent
hasApplicationMethodForControlMethodEvent	hasTimeOfApplicationForControlMethodEvent	isLocationOfFertilizerEvent
hasApplicationMethodForFertilizerEvent	hasTimeOfApplicationForFertilizerEvent	isLocationOfLandPreparationEvent
hasAverageYield	hasUnitForAverageYieldEvent	isLocationOfPlantMethodEvent

hasAverageYieldEvent	hasUnitForControlMethodEvent	isMaturedTimeEventOf
hasBestPlantingTime	hasUnitForLaborRequirementEvent	isMaturedTimeOf
hasBestTimeOfPlanting	hasUnitForMaturedTimeEvent	isNoOfHarvestingTimeOf
hasBestTimeOfPlantingEvent	hasUnitForSeedRateEvent	isPlantMethodEventOf
hasCausalAgent	hasVariety	isPlantMethodOf
hasContactsForTechInfor	isApplicationMethodOfControlMethodEvent	isQuantityOf
hasControlMethod	isApplicationMethodOfFertilizerEvent	isQuantityOfControlMethodEvent
hasControlMethodEvent	isAverageYieldOf	isRelatedCausalAgentOf
hasControlMethodEventForCrop	isBasedOn	isRelatedControlMethodOf
hasCropAverageYield	isBelonged	hasSeedBedPreparationInfor
hasCropMaturedTime	isBestPlantingTimeOf	hasSeedRate
hasCropPlantMethod	isBestTimeOfPlantingEventOf	hasSeedRateEvent
hasFarmPreparationInfor	isBestTimeOfPlaningOf	hasWaterSourceForAverageYieldEvent
hasHarvestingDurationTime	isCausalAgentOf	hasWaterSourceForFertilizerEvent
hasHarvestMethod	isConsistedOf	hasWaterSourceForLangPreparationEvent
hasLaborForce	isContactsForTechInforOf	isSeedBedPreparationInforOf
hasLaborRequirement	isControlledBy	isSeedPreparationInforOf
hasLaborRequirementEvent	isControlMethodEventOf	isSeedRateEventOf
hasLandPreparationEvent	isControlMethodOf	isSeedRateOf
hasLandPreparationInfor	isCropAverageYieldOf	isSymptomOf
hasLocationForAverageYieldEvent	isCropDiseaseResistanceOf	isTimeOfApplicationOfControlMethodEvent
hasLocationForBestTimeOfPlantingEvent	isCropMaturedTimeOf	isTimeOfApplicationOfFertilizerEvent
hasLocationForControlMethodEvent	isCropPlantMethodOf	isUnitOfAverageYieldEvent
hasLocationForFertilizerEvent	isCropSeedRateOf	isUnitOfControlMethodEvent
hasLocationForLandPreparationEvent	isCropSymptomOf	isUnitOfLaborRequirementEvent
hasLocationForMaturedTimeEvent	isDependedOn	isUnitOfMaturedTimeEvent
hasLocationForPlantMethodEvent	isFarmPreparationInforOf	isUnitOfSeedRateEvent
hasMaturedTime	isHarvestingDurationTimeOf	isUsedBy
hasMaturedTimeEvent	isHarvestMethodOf	isVarietyOf
hasNoOfHarvestingTimes	isLaborForceOf	isWaterSourceOfAverageYieldEvent
hasPlantMethod	isLaborRequirementEventOf	isWaterSourceOfFertilizerEvent
hasPlantMethodEvent	isLaborRequirementOf	isWaterSourceOfLandPreparationEvent
		employs
Weather- properties		
hasSeasonBased	hasSeasonForPlantMethodEvent	isSeasonOfBestTimeOfPlantingEvent
hasSeasonForAverageYieldEvent	isSeasonOfAverageYieldEvent	isSeasonOfPlantMethodEvent

Soil -Properties		
grows	hasMinSoilPhForFertilizerEvent	isFertilizerQuantityOf
growsIn	hasRelatedFertilizerQuantity	isFertilizerSpInforOf
hasCropSeedRate	hasRelatedGrowingProblem	isFertilizerUnitOf
hasFertilizer	hasRelatedGrowingProblemForCause	isGrowingMonthsOf

hasFertilizerEvent	hasSoilFactor	isGrowingProblemEventOf
hasFertilizerQuantity	hasSoilTypeForControlMethodEvent	isGrowingProblemOf
hasFertilizerSpInfor	hasSpInforForControlMethodEvent	isMaxSoilPhOfFertilizerEvent
hasFertilizerUnit	hasSpInforForPlantMethodEvent	isMinSoilPhOfFertilizerEvent
hasGrowingMonths	isFertilizerEventOf	isRelatedFertilizerQuantityOf
hasGrowingProblem	isFertilizerOf	isRelatedGrowingProblemOf
hasGrowingProblemEvent	isSoilTypeOfControlMethodEvent	isSoilFactorOf
hasGrowingProblemOfSymptom	isSpInforOfControlMethodEvent	
hasMaxSoilPhForFertilizerEvent	isSpInforOfPlantMethodEvent	
Pest- properties		
affects	hasPesticide	isAffectedBy
causes	hasPesticideForControlMethod	isCausedBy
hasCauseOfSymptom	hasPesticideQuantity	isDiseaseResistanceEventOf
hasCropDiseaseResistance	hasResistanceDisease	isDiseaseResistanceRateOf
hasCropSymptom	hasSymptom	isPesticideOf
hasDiseaseResistanceEvent	hasSymptomOfCause	isPesticideOfControlMethod
hasDiseaseResistanceRate	hasSymptomOfGrowingProblem	isPesticideQuantityOf

After constructing algorithm, the next step is to labeling the constructed ontology based on Neural Network. The algorithm of Neural Network is shown below. ANN, also called neural network, is a mathematical model based on biological neural networks. Artificial neural networks based on human brain observations. The human brain is a complex network of neurons. Similarly, artificial neural networks are interconnected sets of three simple entities: input, hiding, and output devices. The attribute is passed as an input to the next form of the first layer. In clinical diagnosis, risk factors are treated as implants into artificial neural network.

There are generally three learning situations for neural networks. 1) Controlled learning 2) Unsupervised training 3) Perception training is the basic unit of the artificial neural network used for classification, where the sample is linear. The main neuron model used in perception is the Mark Kelffield model. Perception requires an input value vector and 1 result if the result is greater than the preset threshold or -1 otherwise. Convergence of the proof of the algorithm is a well-known unifying theory for understanding.

The resulting node is used to represent the model output, the node in the neural network architecture is commonly known as the neuron. Each input node is connected to the output node by weight connection. This is used to mimic the strength of synchronization between neurons. The simple perceptual algorithm is shown below.

D. Algorithm of ANN

Let D = {{Si, Yi} / i = 1, 2, 3 ---- n}} as an example of training.

Initiate the weight vector with any value W (o).
Repeat.
For each training model (C, Y) d.
Calculate Yi Yi Forecast (C)
For every weight we do.
Update the weight that (k + 1) = Wj (k) + (y i - yi ^ (k)) anyway.
End of.
Finish.
Until the criteria are stopped.

E. Ontology Evolution

The development of ontology consists of four phases:

1) to create new ideas,
2) Define the relationship
3) identification of contact types and.
4) Restart the configuration process for the next WSDL file.

Creating a new idea is based on improving a defined idea. Reviving an idea in the previous step does not guarantee that it must be integrated with current pathology. Instead, emerging ideas must be analyzed in relation to current pathology. To assess the relationship between concepts, we first use the K-Means clustering algorithm to group concepts. The K-Means algorithm is shown below.

K-Means algorithm:

Input: k (number of clusters),
D (data set)
Results: A set of k clusters.
Method:

Retrieval Number: C6362098319/2019©BEIESP
DOI:10.35940/ijrte.C6362.098319

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Choose the random k element of D that is the first center of the cluster;
Repeat:
(Re) Divide each unit into clusters whose units are as close as possible based on the average of the units in the cluster.
Update group assets - Estimate the average value of a subject for each group.
While there is no change;
We then use the KBayes and KTree algorithms to assess the efficiency of the built-in Ontology. The KBayes algorithm is shown below.
KBayes Algorithm:
Begin
Initialization
nc->Number of classes
na->Number of attributes
N->Number of samples
for each class C_i do
Assess prior probability $P(C_i) = \frac{\Sigma C_i}{\Sigma N}, i\epsilon\{1, nc\}$
for each class C_i do
for each attribute A_j do
Assess the conditional probability of the tuple K i.e.

$$P\left(\frac{K}{C_i}\right) = P\left(\frac{A_1}{C_i}\right) * P\left(\frac{A_2}{C_i}\right) * \dots . * P\left(\frac{A_{na}}{C_i}\right)$$

for each class C_i do
Assess the posterior probability of the tuple K i.e P (C_i)

$$* P\left(\frac{K}{C_i}\right)$$

Prediction

$$If\left(\left(P(C_p) * P\left(\frac{K}{C_p}\right)\right) > \left(P(C_q) * P\left(\frac{K}{C_q}\right)\right)\right)$$

$$Prediction \longrightarrow C_p$$

Else

$$Prediction \longrightarrow C_q$$

$$Where\ p, q\ \in \{1, \dots, nc\}\ and\ p \neq q$$

End
The algorithm of KTree is shown below.
KTree Algorithm:
Input: Data record, training data set, T attribute are available to calculate the next branch.
Exit: Original KT Solution.
Method:
1. Create N nodes.
2. If all records in T have the same target class.
3. Return N as the leaf node with the target class.
4. If the attribute is empty
5. Returns N as the leaf node with the maximum record class.
6. Get the best attributes (T attributes available).
7.attributes_available = attributes_available - best_attribute.
8. Split the record based on the attribute (best attribute, T)
9. For each Ti divided by T of best_attribute.
10. Connect a new node returned by decision-making (split element, usable attributes)
11. End for.
Step 12: Final function.

IV. RESULT AND ANALYSIS

A. Information Set Employed

In this document ontology is constructed from different fields. The training tool is a set of more than 68,000 articles collected by the University of Agriculture, Cancer Institute, Pizza and the Central Library. From the information collected, the document employs 47,600 labeled articles, 13,600 of them as test sets, and 6,800 of them as test sets.

B. Efficiency Parameters

To assess the efficiency of the proposed ontology constructing process, several efficiency metrics are available. This paperemploys the Detection Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure and Error Rate to analyses the efficiency.

Detection Accuracy

Detection Accuracy is the measurement system, which measure the degree of closeness of measurement between the original labeledtexts and the correctly labeledtexts

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \qquad (4.1)$$

where, TP – True Positive
FN – False Negative
TN – True Negative
FP – False Positive

Error Rate

Error Rate is the measurement system, which measure no of falsely recognised characters form the given input character images.

$$\text{Error Rate} = \frac{\text{No of Images of Falsely labeled texts}}{\text{Total No of texts}} \qquad (4.2)$$

Precision Rate

The precision is the fraction of retrieved instances that are relevant to the find.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \qquad (4.3)$$

Where, TP – True Positive
FP – False Positive

Recall Rate

The recall is the fraction of relevant instances that are retrieved according to the input image.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad (3.3.4)$$

where, TP – True
FN – False Negative

Sensitivity

Sensitivity also called the true positive rate or the recall rate in some field's measures the proportion of actual positives.

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})}$$

where, TP – True Positive (equivalent with hit)
FN – False Negative (equivalent with miss)

Specificity

Specificity measures the proportion of negatives which are correctly identified such as the percentage.

$$\text{Specificity} = \frac{\text{TN}}{(\text{FP} + \text{TN})}$$

where, TN – True Negative (equivalent with correct rejection)
FP – False Positive (equivalent with false alarm)

F-Measure

F-measure is the ratio of product of precision and recall to the sum of precision and recall. The f-measure can be calculated as,

$$F_m = (1+\alpha) * \frac{\text{Precision} * \text{Recall}}{\alpha * (\text{Precision} * \text{Recall})}$$

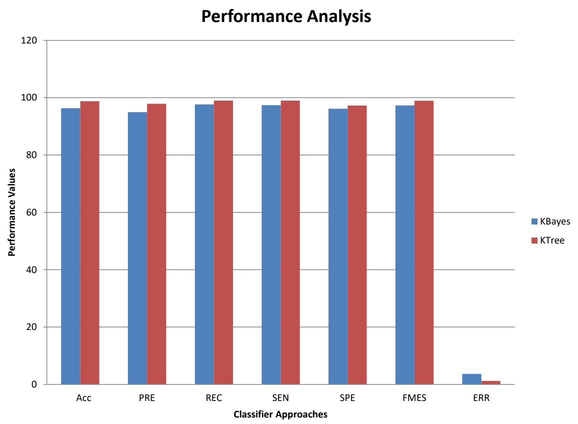
Experiment No #1 : Efficiency Analysis of OntologyGeneration

In this experiment, we will assess the contribution of each classifier approaches which are employed in the work. To assess the efficiency of this feature retrieval scheme, the Detection Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure and Error Rate measures are employed. It is shown in equation 4,5,6 and 7 correspondingly. Ideally, a excellent feature retrieval scheme is accepted to have a high Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure value. Table 1 lists the efficiency analysis Ontology Generation.

Table 1: Analysis of Detection Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure and Error Rate

Classifier Metrics	Acc	PRE	REC	SEN	SPE	FMES	ERR
KBayes	96.333	94.973	97.633	97.393	96.132	97.322	3.667
KTree	98.765	97.864	98.965	98.993	97.256	98.9213	1.235

As observed from Table 1, the Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure of the KTree in range 97-98, which is superior than KBayes method. So the KTreeclassifier is considered to be the best for automated ontology creation. Fig.8 depicted the Detection Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure and Error Rate measures of classifier approaches.



As observed from above figure, the Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure of the KTree in range 97-98, which is superior than KBayes method. So the KTreeclassifier are best for ontology creation.

V. CONCLUSION

In this paper, we established an algorithm for ontology and classification of existing articles. On this basis, we propose a neural network training method to classify texts and then use text to formulate theories. There are four key phases in the process.First phases of this process is Preprocessing the input web source documents. In the preprocessing step, the information normalization, information harmonization and Natural Language Processing (NLP) are done. In the second step, automated attribute selection, automated entity relationship model creation and automatic validation process are done. In the third phase, Web context extraction employs a search engine query encoding that aggregates results by descriptors. In phase four, the classification of a set of descriptions sets the context of a web service is carried out. The effectiveness of the proposed pathology is analyzed in various fields. And the results show that it is capable of automatically and efficiently creating multidisciplinary (autonomic) using our proposed method.

REFERENCES

1. Lima,R., Espinasse,B., Oliveira,H. et al. (2013) An inductive logic programming-based approach for ontology population from the web. In: International Conference on Database and Expert Systems Applications, Springer, Prague, Czech Republic, 319–326.
2. Fortuna,B., Lavrac,N. and Velardi,P. (2008) Advancing topic ˇ ontology learning through term extraction. In: Pacific Rim International Conference on Artificial Intelligence, Springer, Hanoi, Vietnam, 626–635.
3. Seneviratne,M. and Ranasinghe,D. (2011) Inductive logic programming in an agent system for ontological relation extraction. Int. J. Mach. Learn. Comput., 1, 344.
4. Lisi,F.A. and Esposito,F. (2008) Foundations of ontorelational learning. In: International Conference on Inductive Logic Programming, Springer, Prague, Czech Republic, 158–175.
5. R. Bendaoud, Y. Toussaint and A . Napoli, " Pactole: A methodology and a system for semiautomatically enriching an ontology from a collection of texts," Lecture Notes in Computer Science, vol. 5113, pp. 203-216, 2008.
6. G. Barbur,., B. Blaga, and A. Groza, " OntoRich; A support tool for semi-automatic ontology enrichment and evaluation," In IEEE International Conference on Intelligent Computer Communication and Processing, 2011, pp. 129-132,
7. Princeton University, "WordNet : A lexical Information Base for English" , Wordnet Princeton University , 2010. [online]. Available: http://wordnet.princeton.edu/wordnet, [Accessed : June, 10 th , 2016]
8. A. S. Al Hashimy and N. Kulathuramaiyer, "Ontology enrichment with causation relations, " In IEEE Conference on Systems, Process & Control (TCSPC 2013), 2013, pp. 186-192.
9. C. Faria, I, Serra, and R. Girardi, "A domain-independent process for automatic ontology population from text," Science of Computer Programming, vol. 95, pp. 26-43, 2014.
10. M. Al-Yahya, S. Al-Malak, and L. Aldhubayi , "Ontological lexicon enrichment: The BADEA system for semi- automated extraction of antonymy relations from Arabic language corpora," Malaysian Journal of Computer Science. vol. 29, no. 1, 2016.
11. Classical Arabic corpus (KSUCCA), [online]. Available: http://www. ksucorpus.ksu.edu.sa. [Accessed: july, 13, 2017] .
12. Contemporary Arabic corpus (CAC) . [Online]. Available: http://www.comp.leeds. ac.uk/eric/ latifa/ research.htm . [Accessed: july, 13, 2017] .

13. Mixed Arabic corpus (KACSTAC) . [online]. Available : http://www.kacstac.org.sa/pages / Default.aspx. [Accessed: july 13, 2017] .
14. A. Al-Zahrani,., Al-Dalbahie, M., Al-Shaman, M., Al-Otaiby, N., and W. Al-Sultan,. SemTree: analyzing Arabic language text for semantic relations. PhD Thesis, IT Department, KSU, Saudi Arabia,, 2012.
15. Lisi,F.A. and Straccia,U. (2013) A logic-based computational method for the automated induction of fuzzy ontology axioms. FundamentaInformaticae, 124, 503–519.
16. S. S. Laddha, A. R. Laddha and P. M. Jawandhiya, "New paradigm to keyword search: A survey," 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), Noida, 2015, pp. 920 - 923.doi: 10.1109/ICGCIoT.2015.7380594.
17. Shilpa S. Laddha, Pradip M. Jawandhiya, "An Exploratory Study of Keyword based Search Results", Indian J.Sci.Res. 14 (2): 39-45, 2017.
18. Shilpa S. Laddha, Pradip M. Jawandhiya, "Semantic Search Engine", Indian Journal of Science and Technology, Vol 10(23), DOI: 10.17485/ijst/2017/v10i23/115568, June 2017.
19. Prof. RavinderVinayek, Archana Bhatia Nee Malhotra: "Competitiveness Of Indian Tourism In Global Scenario: Academicia" Volume 3, Issue 1 (January, 2013) ISSN 2249-7,pages 137-138.
20. EeroHyvönen, Avril Styrman, and SamppaSaarela, " Ontology – Based Image Retrieval".
21. Dharmish Shah , JheelSomaiya , Sindhu Nair, "Fuzzy Semantic Search Engine", International Journal of Computer Applications (0975 – 8887) Volume 107 – No 15, December 2014.
22. KishanDharavath, Sri KhetwatSaritha," Semantic Web: A Topic specific search", 2012 Ninth International Conference on Information Technology-New Generations pages 145-148.
23. Bense, Hermann &Gernhardt, Benjamin & Hoppe, Thomas &Hemmje, Matthias &Humm, Bernhard &Schade, Ulrich &Schäfermeier, Ralph &Paschke, Adrian & Schmidt, Michael &Haase, Peter & Siegel, Melanie & Vogel, Tobias &Wenning, Rigo. (2016), "Emerging Trends in Corporate Semantic Web. InformatikSpektrum," 39, 474-480.
24. Noy, N.F., Klein, M.: Ontology Evolution: Not the Same as Schema Evolution. Knowledge and Information Systems 6(4) (2004) 428–440.
25. Ehrig, M., Staab, S., Sure, Y.: Bootstrapping Ontology Alignment Approaches with APFEL. In: Proc. of 4th Intl. Semantic Web Conference (ISWC'05), Galway, Ireland (2005).
26. Navigli, R., Velardi, P., Gangemi, A., "Ontology Learning and Its Application to Automated Terminology

AUTHORS PROFILE

Sivaramakrishnan R Guruvayur has overall progressive experience of 23+ years spans across multiple Hands-on as well as Leadership roles across IT in the areas of AI, Machine Learning, Banking Product Development & Implementation, Strategic IT planning, Oracle Core Banking Platform (Flexcube- UBS), Product Engineering & Implementation, Product Professional Services, Ontology Engineering, Chabot, Digital Banking, IT Governance, Enterprise Risk & Compliance Management, Financial Services Analytical Application Suite (OFSAA), EBS, CRM, EPM and Financial Services Data Warehousing, , Management Consulting, Research development & Innovation, Large Tier 1 account & Relationship Management. & eading CMMI Appraisals He has played multiple roles in across his career with a Tier 1 Global IT giant such as Oracle, and his current role as Chief Data Scientist.

He has completed M.S in Software Engineering from BITS, Pilani and Advanced Management Program from IIM, Bangalore.

Paper Presentations

- Machine Learning Methods in Ontology Engineering: A Literature Review - ICRMR 2109, Goa, India
- Design of a Machine Learning Model for Automatic Generation of Domain-Specific Ontologies, 2nd International Conference on Recent Multidisciplinary Research, ICRMR 2018, Thailand
- A Detailed Study on Machine Learning Techniques for Data Mining –International Conference on Trends in Electronics & Informatics, (ICEI 2017)
- SAAS+ - Go-To-Market strategy for an integrated OFSS Product & Services offerings using SAAS Model - IIMB
- Component based Product Development – Software Product Conference, Tokyo, 2000
- Integrated Control & Compliance Framework for BFSIs – Workshop for Central Banks, World Bank, USA 2006
- Integrated View of Governance, Risk and Compliance (GRC), Processes and IT - Bank Systems & Technology USA, June 2009
- Design & Development of Corporate Metrics Program – What Works for You? - SEPG Conference 2002, Florida

R. Suchithra is the director of MCA department in Jain Deemed to be University, Bangalore. She has 16 years of experience in teaching and research. She has published more than 24 papers in reputed journals and her area of specialization is on cloud computing, machine learning and data science. She is guiding research scholars in the PhD programme.