International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT228647

Approach for Efficient Triple Pattern Extraction Information Retrieval in Ontology

Dr. K Chitra[1], R Uma Maheswari[2]

[1]Assistant Professor, Department of Computer Science, Govt. Arts College, Melur, Madurai, Tamilnadu, India

[2]Assistant Professor, Department of CA&IT, Thiagarajar College, Madurai, India

Article Info

Publication Issue :
Volume 8, Issue 6
November-December-2022

Page Number : 451-457

Article History
Accepted: 20 Nov 2022
Published: 07 Dec 2022

ABSTRACT

E-learning is an effective tool for education than what is currently provided by any of the available computer aided tutoring, or learning management systems. Semantic Web is a collection of information that is linked to each other and process by machine in global scale. It can be seen as globally connected database. Here we proposed a combined methodology for better information retrieval by using similarity measurement and visiting frequency and triplet pattern extraction. The primary intention of this paper is to improve the learning capability. triple patterns extraction is essential to convert the incoming learners query to Ontology browsing query (SPARQL).After extracting the triple patterns from the input query is used to build the query for receiving exact information in the ontology content.

Keywords : Ontology, Similarity Measurement, Visiting Frequency, Triple Pattern Extraction, Elearning

I. INTRODUCTION

Learning is portrayed as a change in human association or capacity that returns throughout a timespan and it isn't just referable to procedures of progress. It is additionally used to insinuate choosing up dominating and to acquiring information on what's at this point thought roughly something; the improvement and clarification of a man's own personal explicit experience; or a mentioned, proposed procedure for evaluating mind determined to have issues. The learning framework is ideal to be advanced while the fledglings are prompted around the relevance in their dominating and while they might be provided chance to not fail to remember the condition and component gifted enter on their Execution. One of the informational methods is adaptable dominating wherein showing and dominating expectations are applied, and the educational effects are facilitated with the fledglings as with regards to their own personal explicit requirements. We can isolate the time span e-Learning into stages, the essential of which portrays the transport system as a computerized mode of conveying dominating substances and the second one of which handles the quality's sensibility tips of the dominating substances, keeping any thought from getting the area. E-Learning consolidates an additional a significant area than essentially net understanding, which fuses the use of all helpful computerized media to convey schooling and preparing in an additional a convincing manner. The articulation "e- Learning" is

being applied as a component of the machine that embraces the all around valuable of Supporting an escalated kind of computerized media (Web, intranets, extranets, satellite television for pc show, sound/work tape, clever television, and Compact circle ROM) to make master adjusting extra versatile. It is showing and dominating through the net or a few other predominant media, it makes total utilization of the dominating environmental factors with its huge level of effects and totally new correspondence machine that is advanced through lessening part insights development, to play out some other example of dominating. In the extended run, the total photo of showing and training is thought about to substitute. The arrangement teachers' parts and comparatively the association withinside the focal point of instructors and students in ordinary dominating framework are depended upon to substitute most certainly. E-Learning is a sort of running over what's updated, upheld, or weighed through the utilization of predominant media. E-Learning may likewise contain the use of the most recent or gotten development as well as the improvement of the most recent dominating material. It is most likely made each through neighboring necessities and at detachment.

E-learning systems are mainly used to provide:

- A supplementary learning tool for a traditional class
- Stand-alone distance education
- An e-learning system as more convenient and more efficient
- Reduced cost of education

Ontology can give a typical connection point to recover the data situated in isolated information sources. That can likewise handle the semantic contentions while coordinating various information sources. Thus, many web applications use ontology as a standard instrument for information coordina- tion [1]. Ontology can likewise hold the data about the ideas and relations of the various information sources. To get to the information on ontology, ontology understanding question, for example, SPARQL is required. Be that as it may, End clients can't comprehend and type the configuration of SPARQL

II. RELATED WORK

This segment presents a portion of the current works connected with e-learning based school system for hearing disabled and quiet individuals. Nasr [2] recommended an improved e-learning climate to give many learning facilities to hearing weakened and quiet individuals.

The principal point of this paper was to incorporate the intelligent and social apparatuses together for expanding the framework. In this paper, two unique ideas were examined like student's prerequisite and personalized suggestion procedure. From this review, it was analyzed that the information based proposal framework as the appropriate method for e-students. Li, et al [4] proposed an original Personalized Semantic Recommendation System (PSRS) for e-students.

The proposed framework utilized the Video Structurized Description (VSD) procedure for separating the depiction of initial catchphrases and learning contents. Thusly, the sequen- tial things were added into the philosophy rules in view of Rules Auto-Updating component.

Chrysafiadi and Virvou [5] proposed a clever methodology, specifically, Fuzzy Knowledge State Definer (FuzKSD) in light of Fuzzy Cognitive Maps (FCMs) for electronic schooling. This method utilized an original derivation plan to refresh the client generalizations dynamically with the assistance of fluffy sets. Sanchez Barreiro, et al [6] gave a contextual investigation of programming items for e-learning stages. The essential goal of the proposed speculation, to be specific, Software Product Line (SPL) was to create a viable comparable programming framework.

Verbert, et al [7] constructed a design for allowing the clients to find out about the digital substance cooperatively founded on a digital book perusing client conduct. Here, a proposed recommender framework upheld the clients to find the materials in light of their advantage without an express evaluating. The essential goal of this paper was to recognize the answer for the criticism in the proposal

frameworks inside a climate of digital books. Wang [8] recommended an insect province streamlining with divided goal and meta-control procedures for separating the direction information. From this paper, it was analyzed that the proposed portioned goal preparing approach with the versatile meta-control technique gave the best tracked down ways in better quality.

It can see that inquiry handling is quite possibly of the most significant stages for getting to the information on incorporated ontology. Here, it is expected to extricate the trios from the unstructured information inquiry to create the ontology figuring out inquiry. Thus, many methodologies should be raised for extricating trios from the information sentence. They depend on the parser and the handling season of these method- ologies are more prominent. The accompanying applicable undertakings can be noted.

Delia... in [9] proposed triplets extraction algorithms usability and intuitiveness with gesture based communication interpretation. Shishehchi, et al [3] studied different per- sonalized proposal strategies for students in an e-learning based on the dependency tree generated by the Treebank parser [10] [11], Link

Link parser and Minipar. Treebank parser generates the dependency tree for the sentence that has a syntactic structure. The complexity of the parsing time is O (n^3) where is the length of the input sentence [12], [13]. When applying the triplet extraction algorithm they proposed, it assumes the first noun in noun phrase (NP) subtree as subject, the deepest verb in verb phrase (VP) subtree as predicate and the first noun in propositional phrase (PP) or noun phrase (NP) subtree obtained in the VP subtree as object. This algorithm's time complexity is O (n + (n 1)), so worst case complexity is O (n). Total time complexity is O (n^3 +n); worst case complexity is O (n^3). Also, this algorithm cannot completely extricate all trios from the question sentences since it cannot recognize assuming that input sentence might contain more than one subject, predicate and object.

Salehi and Kmalabadi [14] designed a hybrid attribute based recommender system for endorsing the e-learning materials. The main contribution of this paper was to model the learning material attributes for an efficient recommendation framework. Tulasi, et al [15] presented an ontology-based information re-trieval model for e- learning. The proposed model contains the following modules: ontology management module, resource collection, semantic precision module and retrieval module. Moreover, this paper investigated the drawbacks of traditional keyword-based search engines. Yarandi, et al [16] suggested a personalized adaptive e-learning approach based on a semantic web technology.

This method has an ability to support the personalization based on the following considerations: learner's ability, pref- erences, learning style and knowledge level. In this work, the user profile was updated based on an achieved learner's ability. Qwaider[17]designed a semantic web technology based e- learning system for realizing the sophisticated learning scenarios. The authors illustrated that the semantic web was acting like a backbone of e-learning. Satler, et al [18]suggested a fuzzy ontology-based approach for representing the user profiles in the e-learning environment. The limitations of this were, the pruning process has no significance, and it does not consider the information provided by the user. Hendez and Achour [19]addressed the problem of automatic indexing of online educational resources by using term frequency.

For this purpose, the authors integrated the following processes, such as:

- Recognition;
- Extraction of index terms;
- Score adjustment.

III. PROPOSED METHOD

This section presents the detailed description of the pro- posed web-based e-learning system.Fig. 1 shows the overall flow of the proposed e-learning system for hearing impaired and mute people, which includes the following stages

- Meta-data creation
- Query parsing
- Similarity matching
- Profile matching
- Visiting frequency
- Ranking
- Information retrieval- Triple Pattern Extraction

Initially, the meta-data is created, and the query from the user is collected. Then, it will be parsed into attributes for allowing the users to create a variety of structured queries. Hence, the similarity between the searching concepts is found out by using a triscore based similarity measurement algo- rithm.

After that, the matching profiles and the visiting frequency of the web page are identified. Based on the visiting frequency, the ranking is given to the recommended web pages. Finally, the request learner content is retrieved either in the form of a web page or multimedia. In the following subsections, the step by step working procedure of the e-learning system is.

A. Meta-Data Creation

Meta-data is data used to describe the other data, and it summarizes the basic information about that data. The use of meta-data for web pages is an imperative thing that contains the description of the page contents, keywords, web page descriptions and summary.

The relevancies between the web pages are identified by evaluating the meta-tags. Moreover, the meta-data can be created manually, which tends the data to be more accurate. It allows the users to input any information that is either relevant or needed for describing the file.

B. Similarity Matching

Similarity search is a critical task in many multimedia based e-learning systems. In a multi- dimensional space, the concepts are mapped and, the similarity search is modeled as a nearest neighbor search.

Semantic information retrieval is a type of matching mechanism that is based on the concepts and their relations. Finding similarity between two different concepts is vital for semantic information retrieval.

In this work, the semantic similarity is mainly used to determine the precision of concept matching.

The similarity measure is an important concept in informa- tion retrieval systems. It compares two different objects and determines whether they are related to same or not. Here, triscore based semantic distance measurement technique is used to find the semantic relationship between the concepts. It includes the following similarity measures:

- Paramount similarity
- Angle based similarity
- String based similarity

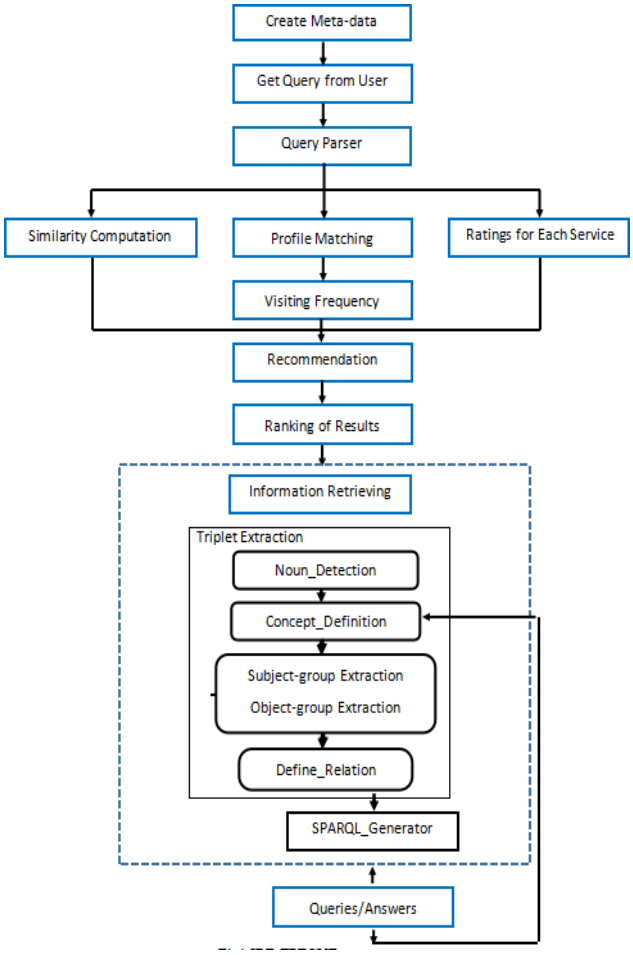

Fig. 1. IRF-TPECVF [Information Retreieval Flow-Triple Pattern Extraction Pattern with Visiting Frequency]

C. Visiting Frequency

In this stage, the visiting frequency is calculated based on the matching profiles. A group of learners visiting a website one or more time is counted as a visiting frequency. The learner who visits the website at once is excluded from this count because it is not probable to generate a rate for them. The visiting frequency is calculated for both user query attributes and services.

The most visited web pages are recommended to other users, who are searching the contents related to that same concept. Recommended learning resources are computed by using the recent navigation history of current learner's activ- ity, as well as by exploiting the similarities and dissimilarities between the learner's preferences and educational cost. The user's implicit query contains a set of learning objects that is recently visited by an active learner. This query is extracted from the navigation history, and it is represented by a vector of relevant terms or by a vector of referred learning objects.

D. Triple Pattern Extraction

Triple patterns extraction is essential to convert the incom- ing learners query to Ontology browsing query (SPARQL). Here we narrate how Triple Pattern extraction is extracted from the incoming query. We implemented an algorithm for triple pattern extraction with three major processes.

They are A. Identifying the concept, B. Subject, C. Object.

A. Identifying the Concept:

Identifying concepts is we have to find the word from a user input sentence or query. Using our customized machine readable dictionary we check if the given word is a noun or not. Identifying concept can be done by the below steps.

Ontology having class and data type properties.

Whenever the word comes from input query, its checked with ontology, if the input word matches with already defined classes in ontology, Then we make that word as a class

If the word is more or less the same matches that already created the class word, then it is defined as subclass.

If the word is matches with properties of already defined class means, that are added as predicates

If nothing matches we call it as constraint

Some meaningless words like special characters and non- noun word are grouped as separate one.

How Net algorithm used to identify similarity between concepts

$$sim\ (s_1, s_2) = \sum_{i=1}^{4} \beta_i \prod_{j=1}^{i} sim_j\ (s_1, s_2)$$

where $\beta i(1 \leq i \leq 4)$ is an adjustable parameter; moreover, $\beta 1 + \beta 2 + \beta 3 + \beta 4 = 1$, $\beta 1 \geq \beta 2 \geq \beta 3 \geq \beta 4$, which reflects the degree contributions to the overall semantic similarity in descending order from sim 1 to sim 2 and sim j(s 1, s 2) is respectively semantic similarity of four parts of original concept atoms. The threshold value for semantic similarity of two concepts is defined as 0.5.

B. Subject:

After extracting the concept and object, the remaining task is to extract the subjects as well as predicates.

After the above all steps only two will be remaining, they are,

i) Class and sub classes without predicates
ii) All predicates in the input query

Further it's proceeded in the algorithm below,

i. Finding out the class names and sub class names which are equivalent to predicate and grouping it.
ii. If the above step is done then it'll be marked as a subject and predicates that can be removed from the concept list.
iii. Those predicates not covered above criteria will be treated with their string type of class or subclass relationship in ontology.

The above steps are constructed as a better Triple Pattern Algorithm. After extracting the triple patterns from the input query is used to build the query for receiving exact information in the ontology content.

Function: EXTRACT-SP (Subject-
group) while (predicate.exist)

Find the class of predicates

<by ontology

cluster the predicates with the same
class Subject class name

If Subject is equal to class-type (word) in the list
Remove predicates from the list

Define the relation based on range value <by ontology

IV. PERFORMANCE ANALYSIS

Effficient Triple pattern Extraction Algorithm result cm- pared with other algorithms. It is thrilling to look the overall performance of proposed triplet extraction set of rules to different triplet extraction algorithms. The overall performance is measured at the processing time of extracting triplets from the enter sentence. Firstly, the symbol 'n' is believed because the wide variety of phrases withinside the question sentence. In this set of rules, time complexity of noun detection from the enter sentence is 'n' instances. The step of idea definition takes 'n' instances for the period of listing that keep the noun shape of phrases received from the noun detection step. Here, it takes n + (n — 1) instances for ontology graph traversal. So, idea definition step takes n^2 instances totally. For item institution reputation step, it additionally takes 'n' instances as it's far had to seek the idea listing described through idea definition step and It takes n + (n — 1) instances for ontology graph traversal. So, item organization popularity step takes n^2 instances totally. Subject organization popularity step takes 'n' instances for looking withinside the ultimate idea listing and it takes n + (n — 1) instances for ontology graph traversal. So, situation organization popularity step additionally takes n^2 instances totally. So, overall time complexity for this set of rules is O (n + $3n^2$), worst case is O (n^2). Compared with different strategies defined in associated work: triplet extraction set of rules primarily based totally on dependency tree generated with the aid of using the Treebank parser with overall time complexity O (n^3), the set of rules primarily based totally at the Minipar with overall time complexity O (n^3) and triplet extraction the use of SVM with overall time complexity O (n^3), the proposed triplet extraction set of rules is greater time-saving than different algorithms.

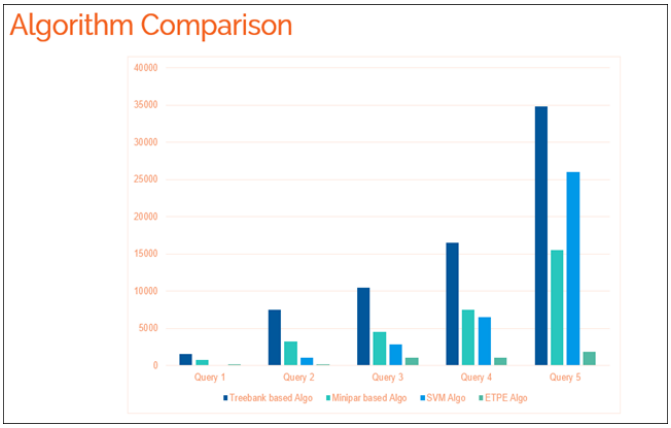

Fig 2. Algorithm Comparison

V. Conclusion Future Work

The presented approach provides efficient content retrieval that fully obtain the all triplets even user inputs unstructured sentence with the help of well-organized ontology and customized machine readable dictionary. The ontology design can establish a better-organized e-learning system in term of the content utilization as well as Creates agent which handle the ontology-based knowledge repository.This approach also has the comparison which other triplet extraction methods based on time of execution by applying different types of query. This can be further extended as helping hands for deaf and dumb people by giving the sign language input. Sign language based information retrieval algorithm will play crucial role in that future work.

VI. REFERENCES

[1]. M. Gagnon, Ontology-Based Integration of Data Sources, 2007

[2]. M. M. Nasr, "An enhanced e-learning environment for Deaf/HOH pupils," in 2010 2nd International Conference on Computer Technology and Development (ICCTD), 2010, pp. 724-727.

[3].]S. Shishehchi, et al., "Review of personalized recommendation techniques for learners in e-learning systems," in 2011 International Conference on Semantic Technology and Information Retrieval (STAIR), 2011, pp. 277-281.

[4]. Y. Li, et al., "A personalized recommendation system in E- Learning environment based on semantic analysis," in 2012 6th International Conference on New Trends in Information Science and Service Science and Data Mining (ISSDM), 2012, pp. 802- 807.

[5]. K. Chrysafiadi and M. Virvou, "Fuzzy Logic for Adaptive Instruction in an E-learning Environment for Computer Programming," IEEE Transactions on Fuzzy Systems, vol. 23, pp. 164-177, 2015.

[6]. P. Sanchez Barreiro, et al., "Building Families of Software Products for e-Learning Platforms: A Case Study," Tecnologias del Aprendizaje, IEEE Revista Iberoamericana de, vol. 9, pp. 64- 71, 2014.

[7]. K. Verbert, et al., "Context-aware recommender systems for learning: a survey and future challenges," IEEE Transactions on Learning Technologies, vol. 5, pp. 318-335, 2012.

[8]. F.-H. Wang, "On extracting recommendation knowledge for personalized web-based learning based on ant colony optimization with segmented-goal and meta-control strategies," Expert Systems with Applications, vol. 39, pp. 6446-6453, 2012.

[9]. D. Rusu, L. Dali, B. Fortuna, M. Grobelink, and D. Mladenic, Triplet Extraction from Sentences, 2008.

[10]. Stanford Parser web page. [Online]. Available: http://nlp.stanford.edu/software/lex-parser.shtml.

[11]. Sharp NLP. [Online]. Available: http;//www.codeplex.com/sharpnlp.

[12]. M. Galley and C. D. Manning, “Quadratic-Time Dependency Parsing for Machine Translation,” in Proc. of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP, pp. 773–781, Singapore, 2-7 August 2009.

[13]. M. A. Covington, “A Fundamental Algorithm for Dependency Parsing,” in Proc. of the 39th Annual ACM Southeast Conference, 2001.

[14]. M. Salehi and I. N. Kmalabadi, "A Hybrid Attribute–based Recommender System for E–learning Material Recommendation," IERI Procedia, vol. 2, pp. 565-570, 2012.

[15]. L. Tulasi, et al., "Study of E-learning Information Retrieval Model based on Ontology," Int J Comput Appl, 2013.

[16]. M. Yarandi, et al., "A personalized adaptive e-learning approach based on semantic web technology," Webology, vol. 10, p. Art. 110, 2013.

[17]. W. Q. Qwaider, "Semantic Web Technologies Applied to E- Learning System," International Journal of Computer Applications (0975–888), vol. 47, 2012.

[18]. M. Ferreira-Satler, et al., "Fuzzy ontologies-based user profiles applied to enhance e-learning activities," Soft Computing, vol. 16, pp. 1129-1141, 2012.

[19]. M. Hendez and H. Achour, "Keywords extraction for automatic indexing of e-learning resources," in 2014 World Symposium on Computer Applications & Research (WSCAR), 2014, pp. 1-5.

Cite this article as :

Dr. K Chitra, R Uma Maheswari, "Approach for Efficient Triple Pattern Extraction Information Retrieval in Ontology", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 6, pp. 451-457, November-December 2022. Available at doi : https://doi.org/10.32628/CSEIT228647

Journal URL : https://ijsrcseit.com/CSEIT228647