Knowledge Base Construction from Unstructured Text

Lamiya Ali, Linda Sara Mathew

Abstract*: Knowledge base assumes a critical job in numerous cutting edge applications. Developing learning base from unstructured content is a testing issue because of its inclination. Subsequently, numerous methodologies propose to change unstructured content to organized content so as to make an Knowledge base. Such methodologies can't yet give sensible outcomes to mapping an extricated predicate to its indistinguishable predicate in another information base. Predicate mapping is a basic system since it can lessen the heterogeneity issue and increment accessibility over the portrayal. A learning base development framework is proposed. In the framework, a mixture mix of a standard based methodology and a closeness based methodology is exhibited for mapping a predicate to its indistinguishable predicate in an information base portrayal. Changing unstructured content into a formal portrayal is a vital objective of the Semantic Web so as to encourage the mix and recovery of data. The development of Knowledge Graphs (KGs) seeks after such a thought, where named elements (genuine things) and their relations are separated from content. The procedure incorporates substance acknowledgment, element goals, element connecting, connection extraction lastly the RDF readiness. For such reason, procedures for favoring the extraction and connecting of named substances with KG people, and also, their relationship with syntactic units that lead to creating increasingly rational certainties are displayed. It likewise gives choices to choosing the extricated data components for making possibly valuable RDF triples for the KG. The incorporation of data extraction units with linguistic structures give a superior comprehension of recommendation based development of KGs.

***Index Terms*: Knowledge base, knowledge graph, semantic web, RDF.**

I. INTRODUCTION

The way toward populating an organized social database from unstructured sources has gotten restored enthusiasm for the database network through prominent new businesses (e.g., Tamr and Trifacta), built up organizations like IBM's Watson [1, 2], and an assortment of research endeavors [1, 3]. In the meantime, networks, for example, normal language preparing and machine learning are tackling comparative issues under the name information base development (KBC) [5, 14]. While different communities place differing emphasis on the extraction, cleaning, and integration phases, all communities seem to be converging toward a common set of techniques that include a mix of data processing, machine learning, and engineers-in-the-loop.

A definitive objective of KBC is to get high caliber organized information from unstructured data. These databases are lavishly organized with several diverse element types in complex connections. Regularly, quality is evaluated utilizing two correlative measures: exactness (how frequently an asserted tuple is right) and review (of the conceivable tuples to remove, what number of are really extricated). These frameworks can ingest enormous quantities of documents– far surpassing the report checks of even all around subsidized human curation endeavors. Modernly, KBC frameworks [1] are developed by talented designers in a months-in length (or more) process– not a one-shot algorithmic undertaking. Seemingly, the most imperative inquiry in such frameworks is the manner by which to best utilize gifted specialists' a great opportunity to quickly enhance information quality. In its full simplification, this inquiry traverses various territories in software engineering, including programming dialects, frameworks, and HCI. We center around a smaller inquiry, with the saying that the more fast the developer travels through the KBC development circle, the more rapidly she acquires great information.

Data expended each day by individuals in an assortment of administrations, for example, general stores, banks, libraries, and web indexes is inside put away in an organized style to be effectively questioned and changed. Changing unstructured content into a formal portrayal is a vital objective of the Semantic Web so as to encourage the incorporation and recovery of data. The development of Knowledge Graphs (KGs) seeks after such a thought, where named substances (true things) and their relations are separated from content. As of late, numerous methodologies for the development of KGs [4] have been proposed by misusing Discourse Analysis, Semantic Frames, or Machine Learning calculations with existing Semantic Web information. Albeit such methodologies are helpful for preparing scientific classifications and associating convictions, they give a few etymological portrayals, which lead to semantic information heterogeneity and in this way, muddling information utilization.

Populating a database with unstructured data is a long-standing issue in industry and research that envelops issues of extraction, cleaning, and mix. Ongoing names utilized for this issue incorporate managing dull information and learning base development (KBC). Perceiving element specifies in a content and connecting them to substances in a learning base are two major errands in content investigation. In the learning extraction pipeline, a Named Entity Recognition (NER) [3] framework is regularly used to perceive notices of named substances in content and after that an Entity Linking (EL) framework is executed to connect perceived notices to elements in an information base like Wikipedia. Since NER

Revised Manuscript Received on December 22, 2018.

Lamiya Ali, Computer Science and Engineering, Mar Athanasius College of Engineering, Kothamangalam, Kerala, India.

Linda Sara Mathew, Computer Science and Engineering, Mar Athanasius College of Engineering, Kothamangalam, Kerala, India.

frameworks center around recognizing named elements, for example, individuals, associations and areas, EL is regularly considered as just connecting named elements and not capable preparing ostensible elements.

Element resolution (ER), otherwise called duplication discovery, record linkage, and so forth is the errand of discovering records alluding to a similar certifiable element in a dataset. It has been broadly contemplated as of late. It additionally assumes a basic job in the two information quality administration and information coordination. Substance connecting is the issue of making joins among records speaking to certifiable elements that are connected in certain ways. As an imperative uncommon case, it incorporates element goals, which is the issue of distinguishing or connecting "copy" elements. Substance connecting has been perceived as an essential computational issue that has been explored by a few diverse research networks. Relation Extraction (RE) [4] is one of the imperative undertakings in regular language handling, empowering data extraction and learning disclosure from content. It goes for sorting out applicable sections of unstructured content in connection triples that speak to the connection between two contentions through a connection.

II. RELATED WORKS

A. Toward an Architecture for Never-Ending Language Learning

NELL is a never-ending system that learns to read the web. To extract triples in NELL, NLP limitations are utilized. The issue of building an endless language student [5] that is, a keen PC operator that runs always and that every day must (1) concentrate, or read, data from the web to populate a developing organized information base, and (2) figure out how to play out this assignment superior to on the earlier day is considered. The issue in this framework is that it doesn't consider the substance mapping. There might be comparable substances for the extricated element. Be that as it may, the NELL overlooks such likeness mapping which causes vagueness in the information base framed.Submit your manuscript electronically for review.

B. Lodifier: Generating linked data from unstructured text

The robotized extraction of data from content and its change into a formal portrayal is a critical objective in both Semantic Web inquire about and computational phonetics [6]. The extricated data can be utilized for an assortment of errands, for example, philosophy age, question noting and data recovery. LODifier is a methodology that consolidates profound semantic investigation with named element acknowledgment, word sense disambiguation and controlled Semantic Web vocabularies so as to remove named substances and relations between them from content and to change over them into a RDF portrayal which is connected to DBpedia and WordNet.

C. Joint entity recognition and relation extraction as a multi-head selection problem

Best in class models for joint element acknowledgment and connection extraction firmly depend on outer regular language handling (NLP) [7] instruments, for example, POS (grammatical form) taggers and reliance parsers. Subsequently, the execution of such joint models relies upon the nature of the highlights got from these NLP instruments [9]. Be that as it may, these highlights are not constantly exact for different dialects and settings. A joint neural model which performs substance acknowledgment and connection extraction all the while is utilized, without the need of any physically removed highlights or the utilization of any outside apparatus. In particular, we show the element acknowledgment errand utilizing a CRF (Conditional Random Fields) layer and the connection extraction assignment as a multi-head choice issue (i.e., possibly recognize numerous relations for every element).

III. PROPOSED SYSTEM

The Proposed System develop an information base from unstructured content. Learning base is organized portrayal of information which is anything but difficult to process contrasted with unstructured text.The learning base [8] is spoken to as a chart for example learning chart. Learning base development should be possible utilizing the accompanying advances, Entity acknowledgment, Entity goals, Entity connecting, Relation extraction, RDF planning.

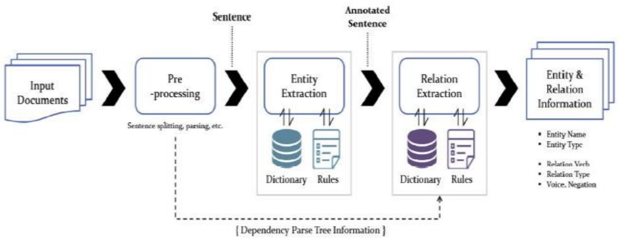

Fig. 1. Entity and Relation extraction overview

A. Entity recognition

The EREL algorithm integrates Entity Recognition, Co-reference Resolution (CR) and Disambiguation. The calculation perceives element makes reference to as the longest name dependent on the name word reference developed from the Wikipedia information. The CR is incorporated into the calculation to enhance the execution in handling short-structure or truncated names. The calculation utilizes another methodology in disambiguation substances utilizing new highlights as element level setting data and case delicate information about the notice in disambiguation. The EREL calculation [9] has four primary advances:

Step 1 : Document Structure Analyzing looks at the archive structure and parts the record into passages.

Step 2 : Entity Mention Scanning skims over a section content to distinguish element makes reference to that can connection to substances in Wikipedia.

Step 3 : Co-reference Resolution finds co-referencing elements that

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

can allude to a previously mentioned substance.

Step 4 : Disambiguation positions linkable elements of a notice and chooses the most suitable element as the connected substance of that notice. The last arrangement of 'notice element' joins is return as the aftereffect of the calculation.

```
The EREL algorithm:
Input: A text document
Output: A set of (mention, entity)
Main content:
  Step 1 - Document Structure Analyzing:
    Determine document header parts and the starting point of the main content
    Split the document text into paragraphs
  Step 2 - Entity Mention Scanning:
    For each paragraph:
      Split the paragraph text into words and assign a Part-Of-Speech tag for each word
      For each length l from max_length down to 1: Scanning from beginning to the end
        of the paragraph text:
          Form the Temporary Mention Candidate Set (TMCS) as the combination of l
            continuous uncovered words
          Query the existing of each mention candidate of TMCS on the Knowledge
            Base (KB)
          If a candidate exists in KB,
          Mark "covered" for all words of the mention candidate
          Add the candidate to Mention Candidate Set (MCS)
          If length l equals to 1 and no mention candidate of TMCS exists in the KB
            Add the original mention candidate to the Unrecognized Mention Candidate
              Set (UMCS)

  Step 3 - Co-reference Resolution (CR):
    CR for abbreviated terms
      For each entity e_i in MCS or UMCS:
        If there is entity e_j in MCS where j < i, e_i and e_j not in the header part, e_i is the
          abbreviated name of e_j: e_i is co-referencing entity of e_j
        If there is entity e_j where e_i is in the header part, e_j is not in the header part of the
          document, e_i is the abbreviated name of e_j: e_i is co-referencing entity of e_j
    CR for short-name entities
      For each entity e_i in MCS or UMCS:
        If there is entity e_j where j < i, e_i and ej not in the header part, POS-tag of e_i is
          NNP or NNPS, the mention of e_i is a part of the mention of e_j: e_i is co-
          referencing entity of e_j
        If there is entity e_j where e_i is in the header part, e_j is not in the header part of the
          document, POS-tag of e_i is NNP or NNPS, the mention of e_i is a part of the
          mention of e_j: e_i is co-referencing entity of e_j
  Step 4 – Entity Disambiguation:
    For each entity candidate e_i in both MCS and the header part:
      Filter the set of linkable entities by the entity type
    For each entity candidate e_i in MCS:
      If e_i is not a co-referencing entity:
        Calculate the linked measure of each linkable entity
        Select the entity with the highest linked measure as the linked entity of e_i
```

B. Entity Resolution

Entity resolution (ER) is finished via cautiously investigating a few information quality standards. Corpus-based strategies Corpus-based semantic closeness techniques depend on word affiliations gained from huge content accumulations following the distributional speculation. Two words are thought to be increasingly comparative if their encompassing settings are progressively comparable or they seem together more every now and again. The calculation of corpus-put together techniques are based with respect to insights of word conveyances or word co-events. As indicated by various computational models, there are tally based techniques, for example Pointwise Mutual Information or Normalized Google Distance, and prescient strategies, for example Word2Vec [10].

Check based strategies tally word co-events and build a word-word framework, in which those co-event insights are specifically connected with probabilistic models, grid factorization measurement decrease. Prescient based techniques specifically learn thick vectors through foreseeing a word from its encompassing setting. We utilize the prescient based word inserting instrument Word2Vecto learn thick vector portrayal of words, since it has been accounted for to have great execution in numerous applications and our proposed Category2Vec show depends on it. The Continuous Bag of Words (CBOW) display is all the more computationally proficient and reasonable for bigger corpus than the skip-gram demonstrate. In this manner, the CBOW display is utilized to prepare word vectors in a neural system design which comprises of an info layer, a projection layer, and a yield layer to foresee a word given its encompassing words with a specific setting window estimate. Formally, given an arrangement of preparing words { w 1 , w 2 , . . . w T } , each word vector is prepared to boost the normal log likelihood. Having the prepared word vectors, word likeness are registered utilizing standard cosine similitude.

Although the preparation procedure depends on a neural system based administered expectation demonstrate, the genuine preparing results are the vector portrayal of words rather than the neural system forecast model. As a result of such thought, the preparation of word implanting is unsupervised and can be connected in different printed corpus without marked dataset [13], which makes Word2Vec material to most KGs containing literary portrayals. Moreover, because of the basic neural system engineering and the utilization of various leveled softmax, Word2Vec can address expansive corpus and the preparation is effective. Be that as it may, since the preparation of word vectors just use word arrangements, a wide assortment of word relations are considered as similarly related agreeing their co-events, which makes the similitude between prepared word vectors coarse and helpless to address synonymous words and progressive relations precisely.

Algorithm 1 The SCSNED approach for disambiguation.

```
procedure DISAMBIGUATE(context, candidates, K)
    C ← words(context)
    E ← candidates
    ef ← frequency(word ∈ candidates)
    score ← 0, entity ← ∅
    for all e ∈ E do
        F(e) ← words(e)
        for all w ∈ C do
            for all f ∈ F(e) do
                S ← sim_word(w, f) * (1 + log(|E| / (1+ef(f))))
            end for
        end for
        value ← sum(top(S, K))
        if value > score then
            entity ← e
            score ← value
        end if
    end for
    return entity
end procedure
```

C. Entity linking

Entity linking (EL) uses co-events data to enhance element depictions which are additionally used to compute nearby similarity among notices and elements to decide results. In any case, element intelligence is additionally esteemed to have a critical influence in EL, which is yet right now ignored. Substance connecting is the undertaking of connecting the literary notice in an archive to the referent element in the current information base [11]. This progression is vital for some undertakings, for example, content understanding, semantic pursuit and so forth. This assignment is trying because of name uncertainty and word polysemy. Here the substance notice in the archive is connection to the DBpedia information base element. The semantic likeness between a report and a competitor substance is estimated by looking at their customized Entity Rank vectors, and after that score the hopeful elements for each notice joined with other nearby highlights. Element Linking (EL) [12] is done as adjusting a printed notice to the referent substance in an information base (e.g., Freebase). Most past investigations on EL primarily center around planning different component portrayals for the notices and elements.

A Deep Semantic Match Model (DSMM) is utilized for EL by utilizing learning chart and unmistakable content. In particular, the DSMM applys bidirectional Long Short Term Memory Network (BiLSTM) with multi-granularities to coordinate notices with competitor substances from two angles: surface structure coordinate by a character-level BiLSTM (singe LSTM) [15] and semantic match dependent on the "auxiliary" setting of elements and the printed setting of notices by a word-level BiLSTM (word-LSTM).

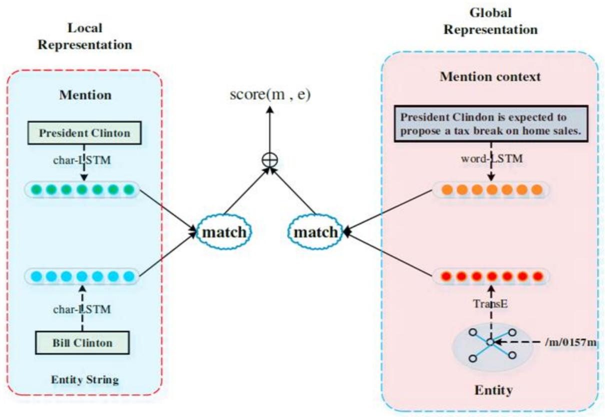

Fig. 2. Entity Linking using BiLSTM

Given a notice M = {c1, c2, . . . , cT } comprise of T characters, an implanting framework Wchar is utilized to change each character into its disseminated portrayal. Information is given to a BiLSTM and utilize the last shrouded condition of the BiLSTM as neighborhood portrayal of the notice yield. A similar method to get Loce of a substance. Thereafter, we process surface structure coordinate as ml = cosine(Locm, Loce). Semantic Match Part catches the worldwide portrayal of notice Glom and substance Gloe by utilizing the literary setting of notice and the "basic" setting of element. Specifically, we utilize a word-LSTM [8], like the singe LSTM aside from the info and a consideration layer, to deal with the printed setting of notice. We change each expression of the setting sentence of the notice into a blend of the word installing and position implanting as the contribution by looking into word inserting grid Wwordand position implanting framework Wp, at that point the yield of BiLSTM layer H = [h1, h2, . . . , hT] is nourished into the consideration layer and the yield Glom is characterized as a nonlinear change of the weighted aggregate of H. With respect to the Gloe, since there needs setting for elements in Freebase, we abuse the structure limitations between substances to catch the semantic portrayal. Specifically, we utilize TransE which prepares the embeddings of elements and relations by authorizing E(s)+E(r) = E(o) for each watched triple (s, r, o) ∈ K . We utilize the prepared element embeddings as an instatement of Gloe in semantic match. Essentially, the semantic match is figured as mg = cosine(Glom,Gloe). Preparing Objective The surface structure and semantic match are both clear and essential sign to EL. So we figure the general match score of a couple (m, e) utilizing the total of ml and mg as score(m, e) = ml + mg. At the preparation arrange, the element notices may have numerous applicant substances in Freebase. To viably prepare the model, we use the pivot misfortune with negative examples as the preparation objective. At the point when for deduction, we figure the match score of each pair and select the best one as the last outcome.

D. Relation extraction

The relations among multiple elements are for the most part alluded to as Complex or Higher Order or n-ary relations. Another point of view to take a gander at n-ary RE issue as Semantic Roles Labelling(SRL). The SRL [13] errand is to distinguish predicate and its contentions in a given sentence naturally. Inside the setting of Information Extraction (IE) [9], connection extraction is situated towards recognizing an assortment of connection phrases and their contentions in subjective sentences. A provision based system for data extraction in literary archives is utilized. Our system centers around two essential difficulties in data extraction: 1) Open Information Extraction and (OIE), and 2) Relation Extraction (RE). In the plenty of research that emphasis on the utilization of syntactic and reliance parsing for the motivations behind identifying relations, there has been expanding proof of muddled and uninformative extractions. The removed relations may even be mistaken now and again and neglect to give a significant understanding. The English proviso structure and condition types is utilized to produce suggestions that can be esteemed as extractable relations. In addition, refinements to the linguistic structure of syntactic and reliance parsing lessens the quantity of mixed up and uninformative extractions from statements.

Typical work around there concentrates triples as (arg1, rel, arg2), speaking to fundamental suggestions or affirmations from content. In this unique situation, suggestions are characterized as cognizant and non-over-indicated bits of essential data. This system handles two errands in RE: (T1) extricating open relations and (T2) removing indicated relations. It center around the English syntax provision structure . The Oxford word reference characterizes a condition as "A unit of syntactic association next beneath the sentence in rank and in conventional punctuation said to comprise of a subject and predicate". The strategy refines the tree structure created from syntactic and reliance parsing [7]. A tale linguistic structure reorganization on the result of the syntactic parser to include fundamental connection hubs and expelling commotion hubs so as to infer a lot of reasonable constituents for creating recommendations that can deliver right extractable relations. For the primary errand (T1), it separates open relations where the framework makes an information driven disregard its condition designs without requiring foundation learning and physically named preparing information. In this appreciation, different kinds of relations are thought about without the need to confine the pursuit to pre-determined semantic relations. For the second assignment (T2), regarding every proviso, the comparing statement type will be resolved according to the linguistic capacity of its sound constituent. The developing examples for the decided condition type will be utilized to remove determined relations.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

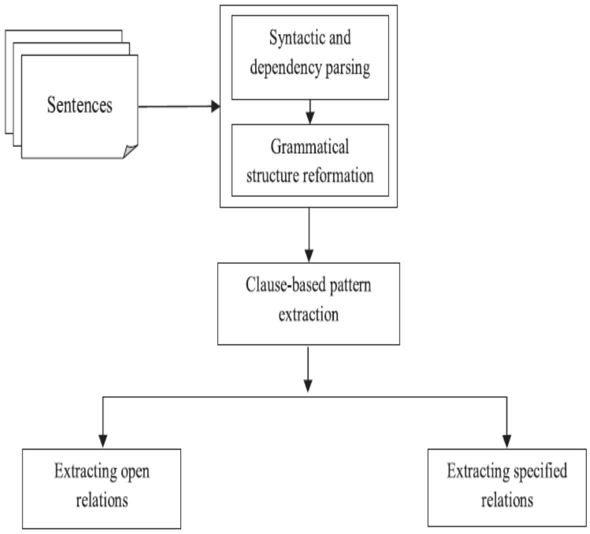

Fig. 3. Relation extraction

Subsequently, a self-preparing calculation dependent on boot-tying that utilizes the examples distinguished in the initial step to naturally infer the required seeds. Setting signs from the educated seeds are found out and utilize the pieces of information to distinguish the classification of a specific connection. This technique disposes of the requirement for a physically arranged seed set at the beginning and rather selects to consequently separate the required seeds from high certainty designs removed in statement based extraction [14]. Through the iterative extension of the first seed set, bootstrapping takes into account an expanding number of seeds to be distinguished that can at last lead to higher certainty connection extraction designs.

Steps included are :

Step 1: Grammatical structure transformation

Step 2: Clause-based example extraction incorporates Determining conditions and proviso types,Extracting open relations

Step 3: Self-preparing calculation

After every principle part are removed (NP-elements, semantic relations, SRL comments), the last advance is to assemble all information for making RDF triples [15] in a persevering organization. RDF gives a model dependent on paired relations yet in some cases a connection should be demonstrated including a few assets and portrayals.

IV. RESULT

So as to exhaustively assess the execution of our strategy in building learning base from unstructured content, we embraced distinctive techniques to get the organized content in our examinations. Knowledge base construction (KBC) is the way toward populating an information base, i.e., a social database together with deduction rules, with data extricated from reports what's more, organized sources. KBC obscures the qualification between two customary database issues, data extraction and data joining. Throughout the previous quite a while, our gathering has been building information bases with logical associates. Utilizing our methodology, we have assembled information bases that have tantamount and now and then preferred quality over those developed by human volunteers.

The info given is a content archive. The record experience pre-preparing and after that the elements are separated from the report utilizing EREL calculation. ie. the – Entity Recognition and Entity Linking calculation. Next substance goals is done to ensure that elements removed is disambiguous. At that point the extricated elements are connected with existing learning base like DBpedia. At that point the connection between elements are removed. For this first syntactic and reliance parsing is done, at that point condition based example extraction. Utilizing the yield of above we can separate open relations. Presently we have elements and relations extricated from the info archive. Presently a triplet is shaped utilizing the above information in the structure (substance, connection, element) which could be spoken to as a tree, that is learning diagram. Such information diagrams are due to their structure, learning diagrams catch actualities identified with individuals, forms, applications, information and things, and the connections among them. At the point when information charts are pondered along these lines, it turns out to be clear why a learning diagram is so essential for AI. The last yield of this framework is a learning chart which is an organized portrayal of information.

V. CONCLUSION

The greater part of the data devoured by human clients on the Web has an unstructured nature, which makes it hard to be handled by applications except if complex errands are performed. To address this issue, the Semantic Web gives an approach to structure all the data through information models, gauges, vocabularies, and apparatuses. In spite of the fact that this is by all accounts an answer to enhancing data utilization, speaking to data on a formal structure is a complex and tedious procedure on the grounds that unstructured information don't have highlights what's more, depictions to help a formal portrayal. A methodology for building Knowledge Charts on the SemanticWeb through an assignment instituted as Relation Extraction and Linking. It depends on Information Extraction (IE) assignments for getting named substances and relations to at that point interface them utilizing information and norms of the SemanticWeb. Also, coordinated data from such IE undertakings together with syntactic units of data for keeping soundness at the portrayal organize. For such purposes, two vital segments, the yield of EEL instruments (i.e., named elements with IRI identifiers) can be related with semantic data given by Noun Phrases (NP) so as to keep sound proclamations. The instinct is that NPs speak to linguistic units of data and along these lines, can be utilized to sort out elements where the first thought is safeguarded. The blend of results from conventional IE instruments and IE instruments consolidating semantic web information has been useful for acquiring semantic relations what's more, named elements separately.

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*

REFERENCES

1. Tanwar, M. , Duggal, R. , Khatri, S. K. (2015). Unravelling unstructured data: A wealth of information in big data, In Reliability, infocom technologies and opti- mization (pp.16). IEEE .
2. Daconta, M. C. , Obrst, L. J. , Smith, K. T. (2003). The semantic web: A guide to the future of xml, web services, and knowledge management,Wiley Publishing, Inc .
3. Daconta, M. C. , Obrst, L. J. , Smith, K. T. (2003). The semantic web: A guide to the future of xml, web services, and knowledge management,Wiley Publishing, Inc .
4. Corcoglioniti, F. , Rospocher, M. , Aprosio, A. P. (2016). Frame-based ontology population with PIKES., IEEE Transactions on Knowledge and Data Engineering, 28 (12), 32613275 .
5. Dutta, A. , Meilicke, C. , Stuckenschmidt, H. (2014). Semantifying triples from open information extraction systems, Frontiers in Artificial Intelligence and Applications, 264 , 111120.
6. Gangemi, A. , Presutti, V. , Recupero, D. R, Nuzzolese, A. G. , Draicchio, F. , Mongiov‘ , M. (2017). Semantic web machine reading with FRED, Semantic Web, 8 (6), 873893 .
7. Daconta, M. C. , Obrst, L. J. , Smith, K. T. (2003). The semantic web: A guide to the future of xml, web services, and knowledge management, Wiley Publishing, Inc.
8. Lehmann, J. , Isele, R. , Jakob, M. , Jentzsch, A. , Kontokostas, D. , Mendes, P. N. , et al. (2015). DBpedia A large-scale, multilingual knowledge base extracted from wikipedia, Semantic Web, 6 (2), 167195.
9. Mahdisoltani, F. , Biega, J. , Suchanek, F. M. (2014). Yago3: A knowledge base from multilingual wikipedias, Biennial conference on innovative data systems research . CIDR .
10. Vrandecic, D., Krotzsch, M. (2014). Wikidata: A free collaborative knowledgebase, Communications of the ACM, 57 (10), 7885. doi: 10.1145/2629489.51.
11. Hoffart, J. , Yosef, M. A. , Bordino, I. , Furstenau, H. , Pinkal, M. , Spaniol, M. , et al. (2011). Robust disambiguation of named entities in text. In Empirical methods in natural language processing, (pp. 782792). Association for Computational Linguistics.
12. Moro, A. , Raganato, A. , Navigli, R. (2014). Entity linking meets word sense disambiguation: A unified approach, Transactions of the Association for Computational Linguistics, 2 , 231244 .
13. Grosz, B. J. , Joshi, A. K. , Weinstein, S. (1983). Providing a unified account of definite noun phrases in discourse, In Annual meeting on association for computational linguistics (pp. 4450). ACL .
14. Heath, T. , Bizer, C. (2009). Evolving the web into a global data space., Morgan and Claypool publishers .
15. Ilievski, F. , Rizzo, G. , van Erp, M. , Plu, J. , Troncy, R. (2016). Contextenhanced adaptive entity linking. International conference on language resources and evaluation, nternational conference on language resources and evaluation . European Language Resources Association (ELRA) .

AUTHORS PROFILE

Lamiya Ali currently pursuing Master of Technology in Computer Science and Engineering from Mar Athanasius College of Engineering Kothamangalam affiliated to APJ Abdul Kalam Technological University.

Linda Sara Mathew, Assistant professor at Mar Athanasius College of Engineering Kothamangalam affiliated to APJ Abdul Kalam Technological University.