

Original Article

A Systematic Review of Relation Extraction for various Indian and Non-Indian Languages

Namrata S. Sarode¹ Ajay S. Patil²

¹Research Scholar, School of Computer Sciences, KBC North Maharashtra University, Jalgaon

²Professor, School of Computer Sciences, KBC North Maharashtra University, Jalgaon

Manuscript ID:
IJERSD-2026-020406

ISSN: 3067-2325

Volume 2

Issue 4

Pp. 30-34

August 2026

Submitted: 08 July 2026

Revised: 19 July 2026

Accepted: 12 Aug. 2026

Published: 31 Aug. 2026

Correspondence Address:

Namrata S. Sarode
Research Scholar, School of
Computer Sciences, KBC North
Maharashtra University, Jalgaon
Email:
namratasarode22@gmail.com

Quick Response Code:



Web: <https://rlgjaar.com>

DOI: 10.5281/zenodo.23331389

DOI Link:
<https://doi.org/10.5281/zenodo.23331389>



Attribution-NonCommercial-
ShareAlike 4.0 International



Abstract

This article comprehensively reviews text relation extraction for various Indian and non-Indian languages. The study involves a comparative investigation of diverse models developed using machine learning (ML) & deep learning (DL) approach for relation extraction. A detailed literature review has been conducted for diverse Indian and non-Indian languages. The comparative study focuses on the effectiveness of diverse models across these languages. This paper highlights the strengths and limitations of various approaches, providing insights into the best practices for relation extraction in diverse linguistic contexts.

Keywords: Natural Language Processing, Machine Learning, Deep Learning, Relation Extraction, Indian languages, non-Indian languages.

Introduction

In this technological era, the Internet is the main source of Information. Lots of information is uploaded on the web every second. However, this information is available in heterogeneous forms and present in unstructured manner. Retrieving the relevant information from this massive unstructured text is a challenging task that gains the focus of Language researchers. Information Extraction (IE) is the branch of Artificial Intelligence that deals with this challenge. IE transforms the unstructured text into a structured form that can be easily used by the machine learning algorithms. Relation Extraction is subfield of Information Extraction that focuses on identifying and categorizing semantic relationships between entities in unstructured data. Relationship extraction recognizing entities (e.g. people, places, organizations) by identifying the connections or interactions between them. Relationship extraction is a process of transform unstructured text to structured text. For example, from the following sentence: "सचिन तेंडुलकर मुंबई मध्ये राहतात."

Entities: सचिन तेंडुलकर → PERSON, मुंबई → LOCATION, Relation: (सचिन तेंडुलकर, राहण्याचे_ठिकाण, मुंबई)

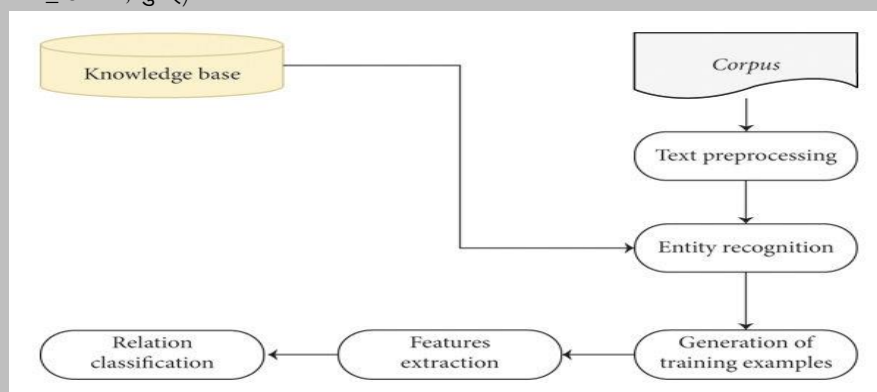


Fig. 1: General architecture of relation extraction system

Creative Commons

This is an open access journal, and articles are distributed under the terms of the [Creative Commons Attribution NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/), The Creative Commons Attribution license allows re-distribution and re-use of a licensed work on the condition that the creator is appropriately credited

How to cite this article:

Sarode, N. S., & Patil, A. S. (2026). A Systematic Review of Relation Extraction for various Indian and Non-Indian Languages. *International Journal of Engineering Research for Sustainable Development*, 2(4), 30–34. <https://doi.org/10.5281/zenodo.23331389>

There are numerous applications of Relation Extraction systems. Some few of them are Question Answering systems (QA), construction of knowledge bases, text summarisation, etc. Construction of knowledge bases is a tedious and time-consuming task that demands domain expertise. Automatic relationship extraction helps to reduce the time. QA systems use RE systems because relations can provide the clue about the answers to most of the questions [1,2]

Related work

There are various techniques used for development of Relation Extraction systems as shown below:

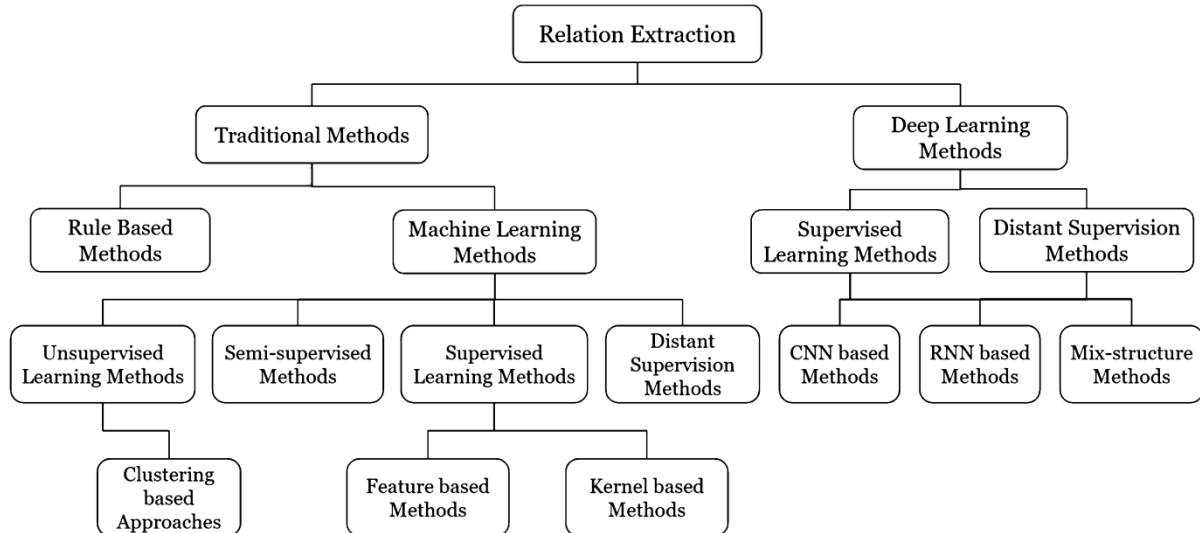


Fig.2 Technique of Relation Extraction

Traditional method:

RE methods can be categorised into two main types: Traditional methods and Deep Learning based methods. Traditional methods used either rule-based techniques or machine learning techniques to extract predefined set of relations from a corpus. Traditional methods can be further categorised in Rule-based methods and Machine Learning-based methods. Machine learning-based methods can be further classified into four types: (1) Unsupervised methods, (2) Supervised methods, (3) Semi-supervised methods and (4) Distant Supervision methods [3].

Rule based methods:

These types of methods define a set of rules of extraction patterns for the predefined set of relations. Later on, these extraction patterns are matched with the text. If a pattern is matched, then a relation corresponding to the pattern is found in the text. These methods are domain specific, where the structure of the document is fixed and target relations are predefined [4].

Unsupervised methods:

Unsupervised methods do not require any labelled data. Most unsupervised RE methods use a clustering-based approach [5,6,7]. Following figure presents the steps involved in this approach.

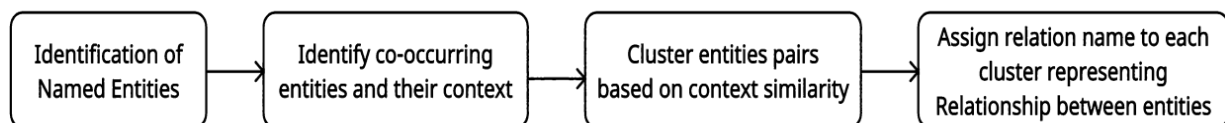


Fig. 3: Clustering based approach for Unsupervised Learning

Supervised methods:

Supervised methods require a large amount of training data, which is annotated with a set of entities and relations. It uses training data to train a classifier, which will extract relations from the test data. There are two types of supervised methods: Feature-based methods and Kernel-based methods. In feature-based approaches a set of features is generated for each relation in the training data, and a classifier is trained to extract a new relation instance. Unlike, kernel-based methods do not use preprocessing steps to design features [8].

Semi-supervised method:

Semi-supervised learning of information extraction` methods, especially for extracting instances of noun categories and relations. Semi-supervised Relation Extraction methods play an important role in extracting relationship from unstructured text, which can leverage both labelled and unlabelled data to improve extraction accuracy. Generating data for supervised methods of RE requires cost, effort, and time. Supervised methods generate labelled data automatically using bootstrapping algorithms. There are two advantages to this approach: (1) reduce the effort required to generate labelled data. (2) Make use of freely available unlabelled data [9].

Distant supervision methods:

These types of methods are also called weakly supervised methods or knowledge-based methods. Distant supervision method in which training data is automatically generated by aligning text with a Knowledge Base (KB). This eliminated the problem of hand-labelling of training data. Distant learning is based on the assumption, that if two entities with a relationship appear in KB, then all phrases that mention these two entities can express that relation [10].

Deep learning method:

In these type of methods, Deep Neural Networks (DNNs) have gained popularity over the past decade for a wide range of applications, that includes computer vision (CV), natural language processing (NLP), and speech recognition. DNN architectures are based on Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long short-term Memory (LSTM), Gated Recurring Unit (GRU), Bidirectional Encoder Representations from Transformers (BERT) are mainly used for different NLP tasks [11,12].

CNN based methods:

Convolution Neural Networks (CNNs) have achieved state-of-the-art performance in image processing tasks. CNN can be employed in many NLP applications to extract word and sentence level features. It builds end-to-end neural network architecture with three main building blocks: a input layer, a convolution layer, and a classic neural network layer [13,14].

RNN based method:

The RNN based model tries to model sequential data and learns long-distance patterns using bi-directional RNN. CNNs are very good at learning local patterns, but RNNs are lacking the capacity to capture long-distance dependency and time sequence information [15].

Relation Extraction for Indian Language

Hindi: In 2001 Dave & Bhattacharyya used a unique approach for Knowledge Extraction of Hindi text. The authors report the applicability of the proposed approach for generating high accuracy of the sentences [16].

Tamil: Jagan et.al. [2019] used a bootstrapping mechanism for extracting the semantic relations between two different concepts of a Tamil text. The authors conclude that rule-based approach is generic, and it can extract all sentence based semantic UNL relations that significantly increases the performance of the generic semantic relation extraction system [17].

Kannada: Ramesh [2025] used topic driven text extraction for Kannada document summarization Using LDA. The author introduces an extractive, topic-driven method for summarizing Kannada news articles from multiple documents [18].

Marathi: Dalvi et.al. [2016] extracted ontology form Marathi language agriculture text using NLP techniques. The author reports 76.98% maximum f-measure using rule-Based and CRF based models for ontology extraction [19].

Relation Extraction for non-Indian Language

Chinese: Yinggang Jie et.al. used an entity-based approach for extracting relationship between Chinese characters. The author researched on remotely supervised methods and kernel functions in relationship extraction strategies for Chinese language. The F1 value, recall, and precision of Chinese text relationship extraction can all be effectively improved in Chinese paper [20,21].

Japanese: Youmi Ma et.al. repots a Japanese Document-Level Relation Extraction Dataset that is assisted by Cross-Lingual Transfer. The author developed the dataset by transferring an English dataset to Japanese language. These models reduce approximately 50% of the human edit steps as compared with the previously reported approaches [22].

English: J. Angel Diaz-Garcia and Julio Amador Diaz Lopez, a survey on cutting-edge relation extraction techniques based on language models. In that studied the evolution and current state of RE techniques by analysing 137 papers presented at the Association for Computational Linguistics (ACL) conferences over the past four years, focusing on models that leverage language models [23,24].

English and Turkish: Aidana Aidynkyzy et.al used Relation Extraction Capabilities of LLMs on Clinical Text: A Bilingual Evaluation for English and Turkish. Evaluations for English performed better than their Turkish counterparts across all evaluated LLMs. In English F1 score was 0.906 and Turkish F1 score was 0.888. LLMs showed higher accuracy in English, while they showed slightly lower accuracy in Turkish [25].

Comparative Studies across Languages

Table 1: Relation Extraction Studies across Languages, Techniques, and Results

Author / Year	Language(s)	Technique / Model	Dataset	Key Results	Reference No.
Zelenko et al. (2002)	English	Tree Kernel Methods	ACE (Automatic Content Extraction)	Improved precision over feature-based models	[26]
Bunescu & Mooney (2005)	English	Dependency Path Kernels	ACE	Better recall using shortest dependency paths	[27]
Nguyen & Grishman (2015)	English	CNN with Position Embeddings	ACE 2005	F1 $\approx$ 69%	[28]
Adel et al. (2016)	English, Arabic	CNN + Cross-lingual Embeddings	ACE	Competitive cross-lingual transfer performance	[29]
Lin et al. (2016)	English	Distant Supervision +	NYT (New York Times news	Precision improvement of 3–	[30]

		Attention	article)	5%	
Ni & Florian (2019)	English, Chinese	mBERT Fine-tuning	TACRED (Text Analysis Conference Relation Extraction Dataset)	Strong zero-shot performance on Chinese	[31]
Soares et al. (2019)	Multilingual	BERT-based Matching the Blanks	FewRel (Few-Shot Relation Extraction Dataset)	F1 > 88% (English), strong multilingual transfer	[32]
77Conneau et al. (2019)	100+ languages	XLNet-R Transformer	mWiki (Multilingual Wiki Dataset)	State-of-the-art cross-lingual F1 scores	[33]

From above table, it is observed that, RE systems related to many languages like English, Chinese, Arabic etc. were developed. Very few RE dataset for Indian language were present. Hence it is need of time to develop RE systems for Indian languages.

Evaluation Measures:

Evaluation metrics like accuracy, Precision, Recall and F-Measure are mainly used for performance evaluation of RE systems.

$$\text{Precision } P = \frac{\text{Number of correctly extracted entity relations}}{\text{Total number of extracted entity relations}}$$

$$\text{Recall } R = \frac{\text{Number of correctly extracted entity relations}}{\text{Actual number of extracted entity relations}}$$

$$f - \text{measure} = \frac{2 * P * R}{P + R}$$

Conclusion

It is observed that the work related to relation extraction for non-Indian languages is carried out. Many different approaches like: rule based, machine learning, supervised, unsupervised, semi-supervised, distant supervised, deep learning, CNN based and RNN based method has used for development of relation extraction systems for these languages. However, there is nascent work related to relation extraction systems for Indian languages due to many challenges like free word order, Morphological richness, Lack of large annotated datasets, and Spelling variations. In future, automated relation extraction systems for Indian languages will be developed.

Acknowledgment

The authors express their sincere gratitude to the School of Computer Sciences, KBC North Maharashtra University, Jalgaon, for providing the academic environment and research facilities necessary for carrying out this study. The authors are especially thankful to the faculty members, research scholars, and colleagues whose valuable guidance, insightful discussions, and continuous encouragement contributed significantly to the completion of this systematic review. The authors also acknowledge the researchers and authors whose published works formed the foundation of this review and provided valuable insights into relation extraction techniques across Indian and non-Indian languages. Finally, the authors are grateful to all those who directly or indirectly supported this research endeavor.

Financial support and sponsorship

Nil.

Conflicts of interest

The authors declare that there are no conflicts of interest regarding the publication of this paper

References:

1. Pareed, A.A., & Idicula, S.M. (2019). A Relation Extraction System for Indian Languages. Advances in Science, Technology and Engineering Systems Journal. A survey of relation extraction
2. Boudjellal, N., Zhang, H., Khan, A., & Ahmad, A. (2020). Biomedical Relation Extraction Using Distant Supervision. Sci. Program., 2020, 8893749:1-8893749:9.
3. Pawar, S.S., Palshikar, G.K., & Bhattacharyya, P. (2017). Relation Extraction: A Survey. ArXiv, abs/1712.05191.
4. Nebhi, K. (2013). A Rule-Based Relation Extraction System using DBpedia and Syntactic Parsing. NLP-DBPEDIA@ISWC.
5. Chen, J., Ji, D., Tan, C.L., & Niu, Z. (2005). Unsupervised Feature Selection for Relation Extraction. International Joint Conference on Natural Language Processing.
6. Quan, C., Wang, M., & Ren, F. (2014). An Unsupervised Text Mining Method for Relation Extraction from Biomedical Literature. PLoS ONE, 9.
7. Yan, Y., Okazaki, N., Matsuo, Y., Yang, Z., & Ishizuka, M. (2009). Unsupervised Relation Extraction by Mining Wikipedia Texts Using Information from the Web. Annual Meeting of the Association for Computational Linguistics.

8. Kotsiantis, S.B. (2007). Supervised Machine Learning: A Review of Classification Techniques. Informatica.
9. Carlson, A., Betteridge, J., Hruschka Junior, E.R., & Mitchell, T.M. (2009). Coupling Semi-Supervised Learning of Categories and Relations. HLT-NAACL 2009.
10. Min, B., Grishman, R., Wan, L., Wang, C., & Gondek, D. (2013). Distant Supervision for Relation Extraction with an Incomplete Knowledge Base. North American Chapter of the Association for Computational Linguistics.
11. Thenmozhi, D., B. S., & Aravindan, C. (2018). SSN_NLP@IECSIL-FIRE-2018: Deep Learning Approach to Named Entity Recognition and Relation Extraction for Conversational Systems in Indian Languages. Fire.
12. Deepa, C., Reghuraj, P.C., & Ramanujan, A. (2019). Relation Extraction Using Deep Learning Methods – A Survey.
13. Lin, Y., Shen, S., Liu, Z., Luan, H., & Sun, M. (2016). Neural Relation Extraction with Selective Attention over Instances. Annual Meeting of the Association for Computational Linguistics.
14. Pareed, A.A., & Idicula, S.M. (2019). A Relation Extraction System for Indian Languages. Advances in Science, Technology and Engineering Systems Journal.
15. Hashimoto, K., Miwa, M., Tsuruoka, Y., & Chikayama, T. (2013). Simple Customization of Recursive Neural Networks for Semantic Relation Classification. Conference on Empirical Methods in Natural Language Processing.
16. Dave, S., & Bhattacharyya, P. (2001). Knowledge Extraction from Hindi Text. IETE Technical Review, 18, 323 - 331.
17. Jagan, B., Noia, T.D., Sciascio, E.D., Donini, F.M., Ruta, M., Bharti, S.K., Pradhan, R., & Babu, K.S. (2019). Bootstrapping of Semantic Relation Extraction for a Morphologically Rich Language: Semi-Supervised Learning of Semantic Relations. Int. J. Semantic Web Inf. Syst., 15, 119-149.
18. Ramesh, D.D. (2025). Topic Driven Text Extraction for Kannada Document Summarization Using LDA. Journal of Information Systems Engineering and Management.
19. Dalvi, P., Mandave, V., Gothkhindi, M., Patil, A., Kadam, S., & Pawar, S.S. (2016). Ontology Extraction for Agriculture Domain in Marathi Language Using NLP Techniques.
20. Jie, Y., Huang, J., & Wu, Y. (2023). An Entity based Approach for Extracting Relationship between Chinese Characters. Academic Journal of Science and Technology.
21. Wang, X., Cai, J., Jiang, Y., Xie, P., Tu, K., & Lu, W. (2022). Named Entity and Relation Extraction with Multi-Modal Retrieval. Conference on Empirical Methods in Natural Language Processing.
22. Ma, Y., Wang, A., & Okazaki, N. (2024). Building a Japanese Document-Level Relation Extraction Dataset Assisted by Cross-Lingual Transfer. International Conference on Language Resources and Evaluation.
23. Diaz-Garcia, J.A., & López, J.A. (2024). A survey on cutting-edge relation extraction techniques based on language models. ArXiv, abs/2411.18157.
24. Jain, M. (2024). Knowledge Enabled Relation Extraction. Companion Proceedings of the ACM Web Conference 2024.
25. Aidynkyzy, A., Dikenelli, O., Alatli, O., & Bora, Ş. (2026). Relation Extraction Capabilities of LLMs on Clinical Text: A Bilingual Evaluation for English and Turkish. ArXiv, abs/2601.09367.
26. Zelenko, D., Aone, C., & Richardella, A.R. (2002). Kernel Methods for Relation Extraction. Journal of machine learning research.
27. Bunescu, R.C., & Mooney, R.J. (2005). A Shortest Path Dependency Kernel for Relation Extraction. Human Language Technology - The Baltic Perspectiv.
28. Nguyen, T.H., & Grishman, R. (2015). Relation Extraction: Perspective from Convolutional Neural Networks. VS@HLT-NAACL.
29. Adel, H., Roth, B., & Schütze, H. (2016). Comparing Convolutional Neural Networks to Traditional Models for Slot Filling. ArXiv, abs/1603.05157.
30. Lin, Y., Shen, S., Liu, Z., Luan, H., & Sun, M. (2016). Neural Relation Extraction with Selective Attention over Instances. Annual Meeting of the Association for Computational Linguistics.
31. Ni, J., & Florian, R. (2019). Neural Cross-Lingual Relation Extraction Based on Bilingual Word Embedding Mapping. ArXiv, abs/1911.00069.
32. Baldini Soares, L., FitzGerald, N., Ling, J., & Kwiatkowski, T. (2019). Matching the Blanks: Distributional Similarity for Relation Learning. ArXiv, abs/1906.03158.
33. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F. (., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2019). Unsupervised Cross-lingual Representation Learning at Scale. ArXiv, abs/1911.02116.