IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

CROSS DOMAIN INFORMATION EXTRACTION BY TRANSFER LEARNING.

[1]Kaustubh Chaudhari, [2]Prof.Manisha Mali,

[1]Master's Student, [2]Dr, Professor,

[1]Computer Department,

[1]Vishwakarma Institute of Information Technology, Pune, India

Abstract: The extremely high pace at which data increases in domains such as health and sports makes it increasingly apparent that there is a need for systems that can quickly extract and process domain-specific information. Often such traditional approaches would have different models for each of the domains, which brings even more complex and resource-consuming setup processes. This work introduces a joint pipeline towards extracting information across domains in such a situation where transfer learning is efficient. The understanding that makes it even more compelling is that it uses fine-tuning on multi-domain tagged datasets like healthcare and sports to eliminate the need for separate domain-specific models by pre-trained transformer models available like BERT. Architecture provides domain classification as well as named entity recognition (NER) and relationship extraction. A new prioritization mechanism will evaluate the relevance of the individual lines read in the input text by the number of extracted entities and their associations. Lines having greater relevance will be sequenced for the generation of queries and Extraction of information; thus, efficient and meaningful information Extraction is realized. This system will effectively incorporate the query formulation and Extraction components for accurate and relevant data Extraction in healthcare and sports applications. Experimental results show that the unified approach produces a simplified architecture, reducing resource consumption while improving scalability for flexible and efficient use in multi-domain applications of natural language processing.

Index Terms **- Cross-domain information extraction, transfer learning, BERT, named entity recognition, multi-domain NLP, information Extraction, Relationship Extraction.**

I. INTRODUCTION

The domain-specific expansion which has garnered popularity recently has enhanced the need for systems that can fetch and exploit important information for almost all fields like the healthcare or sports industry. A recognition of Named Entities (NER), a fundamental task in Natural Language Processing, can be generalized as the process, which identifies the entities, to list just a few diseases, drugs, players, and events, from text. However, existing NER approaches are largely based on the definition of a single-domain application and consequently restrict their scalability and adaptability to multiple domains. Not only is the construction of large labelled data sets and generation of domain-specific models for every application resource-intensive, impractical in dealing with numerous domains, or dynamic environments [8, 9, 5].Cross-domain information extraction poses challenges in terms of linguistic variability, proper vocabulary, and context. For instance, the term "viral" would be interpreted differently in healthcare-for the microorganism causing disease-and information technology-a program that causes harm to computer systems. Traditional systems tackle this by developing and training independent models for every domain but this method proves computationally expensive, and cannot generalize well to new domains or entity types without retraining or fine-tuning [1, 6]. More than that, an important number of existing systems cannot efficiently handle contextual nuances or adapt to emerging domains, which make them important in real-life applications [7].

At this moment, NER systems have advanced due to deep learning models, such as CNN-BiLSTM-CRF architectures and transformer-based architectures that build pre-trained embeddings like GloVe and FastText for better accuracy [8]. Models in transfer learning have shown tremendous benefits in overcoming cross-domain issues caused by learning domain-invariant features that retain contextual nuances, such as BERT and its domain-specific variants [1, 4]. These works are especially significant for information extraction, query generation, recommendation systems, etc., where semantic alignment and feature fusion play an important role [3, 12]. This work proposes a transfer learning-based novel practical solution for cross-domain information extraction to fill the gap that traditional systems leave. The system uses pre-trained transformer architectures like BERT and its domain-specific versions for creating a single model for cross-domain NER. Model fine-tuning on a unified multi-domain training corpus annotated towards specific applications, such as healthcare and sports, captures nuances and relationships in contextual domains [8, 6]. So there is no need for separate NER models in all domains, which reduces system complexity, scalability, and computational costs.

Moreover, beyond NER, relationships may also be integrated into the system regarding extraction, clarification generation, and extraction of said facts without the limitation of the domain. A prioritization mechanism is applied to evaluate lines based on the richness of both relationships and entities so that the most applicable ones could be processed first in order to improve Extraction Efficiency [12]. The system also brings in a set of ontology-based alignment and semantic-based frameworks so that a mechanism for dealing with contextual uncertainties is also defined, based on the principles of prior knowledge extraction and alignment [11, 10]. The entire system addresses feature variance, contextual uncertainty, and scaling issues and stands as a solution to cross-domain entity extraction and Extraction. This is the major advancement that asserts its value for cross-domain information extraction. Practical flexibility is evident for real-world applications in multi-domain NLP; theory is not separated from practice as it wields the efficiency of transfer learning with the effectiveness of domain adaptability. Semantic indexing, cross-domain feature alignment, and scalability in NLP techniques integrate this system as a solution for healthcare, sports, and other domains [3, 4, 6].

II. MOTIVATION

This escalation is heading toward the information discharged in lifestyles comprising health and sports; now systems can even better extract entailment from the text independent of any boundaries of domain. Most of the information extraction systems heavily dependence on a domain model; most of the time, such a model required a lot of labeled data and computing resources. So, it becomes quite challenging for such systems to scale and adapt with unseen contexts and varying linguistic forms, consequently making them highly unsuitable for real-world applications that usually cut across domains. Transfer learning happens to be one among the very significant advancements in NLP through its models like BERT and its several counterparts specific to a domain. All these breakthroughs in NLP address the above-mentioned problem. Unified NER systems will then be able to identify entities across domains without using different models that are resource-hungry by reason of fine-tuning the pre-trained language model on multi-domain data. This would reduce complexity, training time, and consumption of resources, while further improving the adaptability and scalability of the systems.

The key objectives of this project, are to create a cross-domain information extraction framework with entity and relationship prioritization integrated with transfer learning. The proposed system deals with healthcare and sports and aims to fulfill the requirements for enabling multi-domains integration compatible with already existing single-domain practices. The approach provides an efficient and scalable solution for real-world applications as it considers the lines of most meaningful entities and relationships for query generation and information Extraction. It is expected that this becomes a step toward creating lean NAD robust NLP systems extensible to other domains and applications in the future.

III. RELATED WORK

The study of legal texts involves identifying the connections between facts and laws and thus, there is a need to extract key attributes in a structured way for further legal reasoning. The custom approach defined as ‘Legal Data Extraction Analysis’ makes the legal paper evaluation process expedient by providing means of identifying relevant elements such as the judge, his appella, any respondent, and date of judgment. This method employs natural language processing (NLP) and machine learning (ML) for improved information Extraction. Clustering text categorization with a regex-based back up and a flagging system guarantees effective recognition of entities, This method shows how effective customized NER based on NLP can be in everyday legal practice incorporating live legal data, making legal research and analysis less tedious [9].

Named Entity Recognition (NER) is a task which has appreciated a more significant growth with deep neural networks (DNNs) in utilizing word and character level features for performance gains. Recent approaches address using pre trained embeddings to explore the such as GloVe and FastText for improved NER precision. Augmenting word embeddings with regiments of text profiling features comprising of character and word also works in most cases especially in systems that tend to combine CNN-BiLSTM-CRF models. It is worth mentioning that use of GloVe 840B algorithms along with word and character templates produced excellent resultson CoNLL-2003. This strategy illustrates the importance of stacked networks and detailed feature build up in achieving better results than existing best practices in NER [8].

Domain adaptation has become an important way of dealing with the problems accompanying the lack of labeled data in various domains. Conventional machine learning systems are often unable to generalize well when they are used to datasets from target domains that differ from the training source, leading to severe loss of accuracy. This gap between source and target domains has also inspired a number of domains adaptation solutions, feature alignment, adversarial training, domain-invariant representation learning among others. Similarity measures are fundamental to these approaches as they help in the selection of such pivot features that define the relationship across the domains. Earlier, approaches such as entropy based ones have been encouraging yet they do not always deal with the minute differences in the frequencies across the domains. The devised improved cross-entropy measure is intended for more efficient selection of features and better alignment of data from various sources to enhance classification in the target domain [1].

Entity Recognition (NER) model was introduced to classify the drug mentions in dark web text content. This approach employs Drug-NER dataset, which classifies the drug entities to either ‘Street names’ or ‘Chemical names’ facilitating the easy targeting of drug oriented websites. A dark web crawler gathers information related to markets and this information is inputted into an NER model that has been previously trained on about 3500 labelled examples. The model also adds a novel 'DRUG' entity type within the existing spaCy’s NLP framework, given the shortcomings of its prebuilt classification system. This development improves the capability of analyzing and classifying dark web content, and hence boosts the detection and surveillance on drug abuse activities[6].

The ability to manage and synthesize more semantic content from several knowledge repositories is improved by the process of knowledge extraction and its assimilation into a structured ontology, within ontological information Extraction frameworks. The significant processing overhead lessens considerably through the use of a semantic indexing mechanism based on entity Extraction models which assists in the effective arrangement and Extraction of information. The proposed approach is shown to work for domains like public transportation using domain ontologies for minimum loss of semantics. Tests using ontology mapping and actual data show that the framework can provide useful and relevant search results, pointing towards its use in domains where effective and efficient production of semantic search is essential [11].

This section discusses predicted ratings of items in a target domain in order to circumvent the issues of cross domain recommendation in the contexts where user ratings do not overlap. Trust relationships are pegged on the items whose ratings are in the target domain only. These estimated ratings add to existing ratings to form a rating matrix across different domains. The matrix is then used to create item-item association matrix for the purpose of finding item similarities in different domains. Recommendations are yielded through item-based collaborated filtering which enhances prediction accuracy, recall and coverage as demonstrated through experiments carried out on real datasets [2].

To alleviate the restrictions of low accuracy in fault diagnosis in high-noise environments, a Cross Domain Feature Fusion Network (CDFN) is presented which makes use of time domain and frequency domain features at the same time. The approach commences with the packaging of the unprocessed vibration signals in the form of pictures using the Fast Fourier Transform technique. An additional cluster is then added to act as the CDFN primary feature extractor. The function of the merger's block is in processing and integrating time-domain, frequency-domain, and cross-domain data to create new data in different domains. After the addition of convolutional and fusion blocks, CDFN provides the extension of cross domain feature extraction for classification tasks. Tested on CWRU data set, the technique shows also a good diagnosis ability and stability in the presence of noise carrying high resolution in TSNE and effective features separation [3].

A new method for expanding queries increases the efficiency of information Extraction based on the use of the graph-based ranking algorithm to select sentences rather than commonly used sentence-based selection. This technique includes a synectics technique for filtering and construction of improved search queries, increasing relevance and minimizing irrelevant content. The approach in question was evaluated on the DUC09 test collection designed for update summarization and showed very good results when it comes to information Extraction. The findings support the proposition that synthesized query expansion and graph-based ranking are highly beneficial for effective and contextual information Extraction [12].

Information extraction from military sources is crucial in gaining the information superiority, starting with named entity recognition (NER). Military entities, such as troop names & distribution, geographical regions, or weapon systems, are automatically defined in deep learning based supervised methods without feature engineering & limitations of domain corpus segmentation. Looking at the features of military text, these techniques make use of character embedding by using BI-LSTM networks as well as CRF layers. While this method reaches the same level of accomplishment as general-purpose NER systems, it performs worse than the most basic interfaces designed for military usage, showing the difficulties of using neural networks to solve dedicated tasks [5].

Recognizing named entities (NER) is one of the essential tasks in NLP that has numerous applications including but not limited to conversational systems and search systems as well as content-based classification. While NER systems exist for resource-extensive languages, challenges arise in identifying named entities in the languages that are closely related in structure but lies in different resources as is the case with Hindi and Marathi languages. Supervised models baselines for these languages are established, trained, and evaluated using multilingual and monolingual adaptations of BERT that is baseBERT, ALBERT, and RoBERTa. It is illustrated in the evaluation that modeling multilingual data is more advantageous than developing a single language since contextual similarities across languages can be utilized. However, performance improvement is not guaranteed by just mixing datasets randomly, on the contrary, this emphasizes the need for data selection strategies when training the model. The results highlight many lower-resourced languages shared common benchmark structures could be capitalized to improve the NER task [7].

The components of domain ontologies are essential in providing a flexible and extensible means of representing semantics and integrating information in systems. Ontology components, in particular, enable interoperability through concepts such as sharing terminologies, semantic matching, dynamic annotation, and semantic information Extraction in a rather 'loosely coupled' manner. In relation to agricultural policy, an example is outlined in a prototype system (APODOCSIIS). This system performs functionalities such as matching concepts across different ontology components, enriching unstructured and semi-structured data, and semantically merging the policies' content. The architecture enables the management of an ontology, that is, querying and modifying of the structural entities. This proves the capability of ontology-enabled systems in facilitating automatic, semantic integration of content from different domains [10].

In the case of unsupervised cross-domain sentiment classification, the task consists in transferring the knowledge which is present in a labeled source domain, to an unlabeled target domain. In this case word embeddings are primary vehicles – facilitators that help in linking two domains. Here, the enhancement of this process is achieved through determining such words which are emotionally relevant and can act as anchors when the domains are aligned. Such pivots are chosen according to their mutual information values allowing the calculation of the transfer coefficients which are then used as limitations for the target domain update of the word embeddings of the given domain. This approach is built around the skip-gram model with domain adaptation extensions in place so that more meaningful embeddings are obtained. Various experiments have shown that this method is better than any traditional methods for cross domain word embedding based solely on frequency of words, and that it is effective in retaining sentiment orientation across different domains [4].

IV. METHODOLOGY

Proposed framework will enable cross-applicability and efficiency in information extraction from different domains through transfer learning techniques. The framework installation would combine domain classification with name entity recognition, query generation, and information extraction modules to make it adaptable and scalable. Each component would, therefore, functionally integrate to process user input into an accurate, context-aware solution **Workflow Steps**

1. Input Handling
The system starts with natural language input from the user, either as a single sentence or a paragraph that will be treated as raw text. At this stage, the intent is primarily to catch the user's query and prepare it for further actions.

Example Transformation:
Input: "Rohan scored 30 points in the NBA Finals. He is playing good."

2. Preprocessing
Once this is over, it pre-processes the text for downstream models. These activities include cleansing the text, purging irrelevant characters, and lowercasing it. It standardizes the input while keeping the original text intact in context and meaning.

Example Transformation:
Output: "rohan scored 30 points in the nba finals. he is playing good."

3. Paragraph Splitting
This preprocessed input, most often in form of lines or sentences, is then further segmented into smaller parts for finer processing. This would allow feeding each sentence to the domain classification and subsequent modules.

Example Transformation.
Output:
Line 1: "rohan scored 30 points in the nba finals."
Line 2: "he is playing good."

4. Domain Classification
Each line is then classified into any one of the prescribed domain categories, such as Healthcare, Sports, or Healthcare-Sports, using a BERT-based classification model. This ensures that input is read in a way most suited for the context of the question using domain-specific modules.

Example Transformation:
Output:
Line 1: Domain: Sports
Line 2: Domain: Sports

5. Named Entity Recognition (NER)
Every line in the segment has its NER specialization for processing. According to the kind of specific classification for every input line, it is processed by these specialized models such as BioBERT for Healthcare, SportsBERT for Sports, or then hobbled models for overlapping domains. That is what helps the system to process lines of input with required entities according to context..

Example Transformation:
Output:
Line 1: {rohan: Player, nba finals: Event, 30 points: Score}
Line 2: {rohan: Player, playing good: Performance}

6. Relationship Extraction
All the relationships that are derived from these entities sketch a quite precise idea on the context regarding which the processed line might have. Transformer-based Relationship Extraction Model identifies the relationship and establishes a semantic and syntactic link among the different entities which enriches the output extracted data.

Example Transformation:
Output:
Line 1: {rohan → scored_in → nba finals, rohan → achieved → 30 points}
Line 2: {rohan → has_performance → playing good}

7. Prioritization
Lines are sorted on the basis of wealth of extracted entities and relations. Thus, this guarantees that the most essential and informative lines are most quickly processed within the generation phase of the queries.

Example Transformation:
Output:
Line 1: Priority Score: 5 (Higher priority)
Line 2: Priority Score: 4 (Lower priority)

8.Query Generation and Execution
Queries thus generated become built across such prioritized lines ensuring compatibility with the domain-specific databases.
Example Transformation:
Output:
Line 1 Query: db.sports.find({player: "rohan", event: "nba finals", score: "30 points"})
Line 2 Query: db.sports.find({player: "rohan", performance: "playing good"})

9. Information Extraction
The generated queries are executed against the target specific databases to retrieve relevant information. The system should therefore ensure structured meaningful data that fits the entities and relations defined earlier.

Example Transformation:
Output:
Line 1: "Rohan scored 30 points in the NBA Finals for the Warriors."
Line 2: "Rohan is performing exceptionally well in his recent games."

10. Output
In the very last step, the above stated information will be presented in the same manner as retrieved. This means that the system will give the user exactly what he/she asked for-the user will have the data accurate and relevant in terms of context to his query.

Example Transformation:
Final Output:
"Rohan scored 30 points in the NBA Finals for the Warriors."
"Rohan is performing exceptionally well in his recent games."

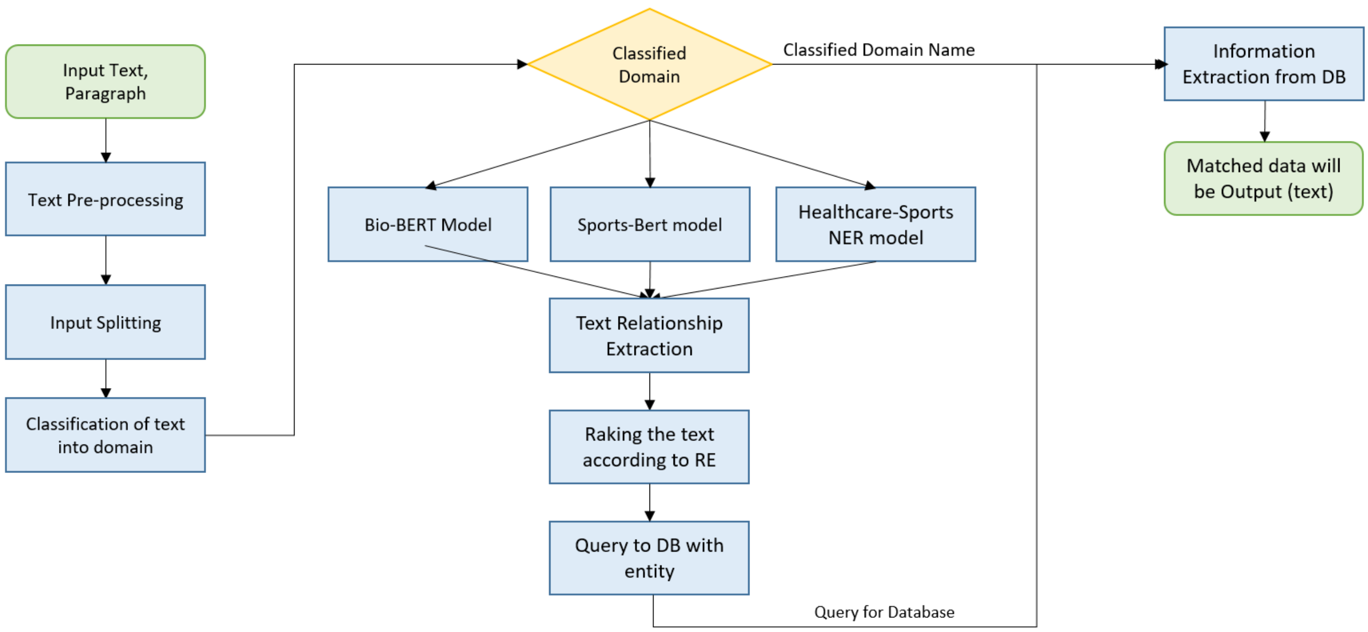

Figure 1 flowchart diagram.

V. DATASETS

T The datasets have been prepared to train models specifically with respect to domain classification, named entity recognition (NER), relationship extraction, and Information Retrieval (IR). These datasets can be effectively used to implement a cross-domain information extraction system that works with healthcare and sports domains.

1. Dataset for Domain Classification
This dataset comprises annotated texts labelled under a few specific domains, such as Healthcare and Sports. It has been set up for fine-tuning the BERT-based domain classifier.

Examples:
Text: "What are the symptoms of COVID-19?" → Domain: Healthcare
Text: "Who scored the most points in the NBA Finals?" → Domain: Sports

2. Dataset for Named Entity Recognition (NER)
The NER dataset comprises sentences that are annotated using entity tags and provide their types to train the classifier for extracting entities specific to a domain..

Examples:
Text: "What are the symptoms of COVID-19?"
Entities: COVID-19: Disease, symptoms: Query

Text: "Rohan scored 30 points in the NBA Finals."
Entities: Rohan: Player, NBA Finals: Event, 30 points: Score

3. Dataset for Relationship Extraction
The dataset comprises sentences annotated with relationships being extracted from entities; the relationships being context; it is used to train a relationship extraction model.

Examples:
Text: "Rohan scored 30 points in the NBA Finals."
Relationships: scored_in, achieved

Text: "LeBron James leads the Lakers to victory."
Relationships: leads

4. Dataset for Information Extraction (IE)
Essentially, the IR dataset comprises all those structured entries relating to all these: entities, type, and details. It supports its generation using this query-and-retrieve-specific information from the domain for knowledge extraction.

Examples:
Entity: COVID-19, Type: Disease, Details: "Symptoms include fever, cough, fatigue."
Entity: Rohan, Type: Player, Details: "Scored 30 points in the NBA Finals."

VI. ARCHITECTURE

Extrapolating from an architecture based on six main components: Input, Domain Classification, NER Module, IR System, Response Generator, and Output, the entire information extraction system geared toward the different domains has been structured as pipeline. Each one has a function that caters to user queries and produces results.

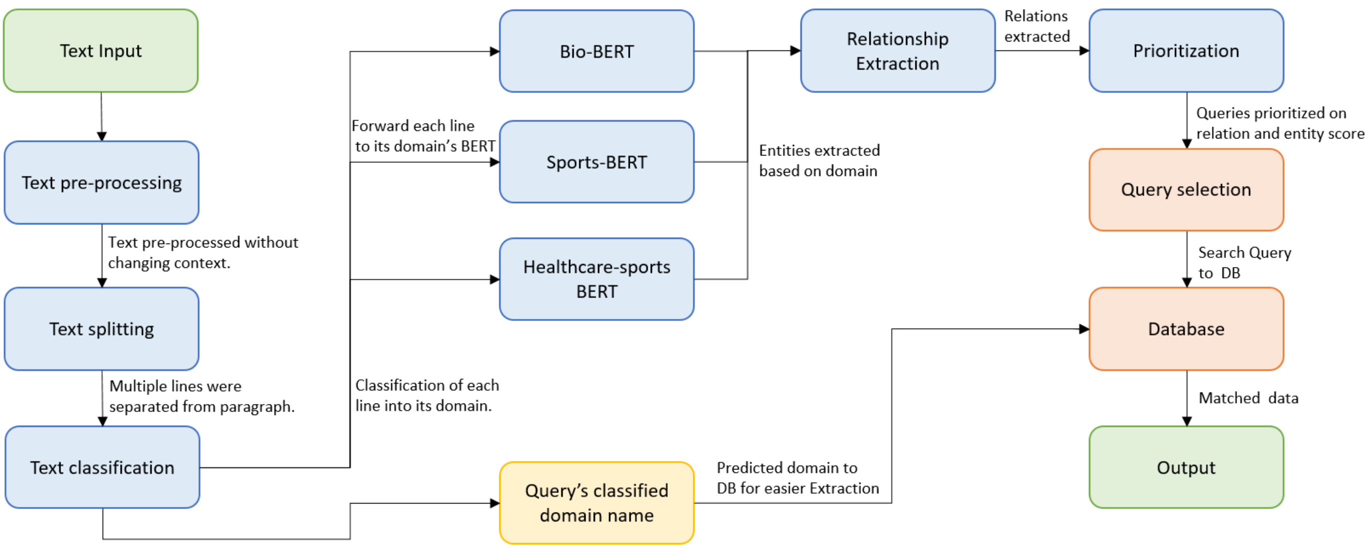

Figure 2:architecture diagram.

1. **Input.**
   Function: Taking input from user.
   Description:
   The system accepts user queries or statements in unstructured natural language; processes through tokenization using BERT tokenizer, text cleaning, and further prepares the text for further processing while preserving the meaning or context.

2. **Input Splitting.**
   Function: Splits input text into individual lines.
   Description:
   This module will split the paragraphs into smaller units, such as sentences or lines, for further processing. The disjointed isolation of each line will allow the system to treat all these lines independently at the next stage, thus making the processing more specific and domain relevant.

3. **Domain Classification.**
   Function: Classifies each line into one of the predefined categories: Healthcare, Sports, or Healthcare-Sports.
   Description:
   The BERT-based classifier assigns each line to a specific domain. This means that such an extension could be done using special models according to the domains: for example, BioBERT for application in healthcare, SportsBERT for athletics, and Healthcare-Sports BERT for overlapped areas. This classification helps the system send lines to the right specific module within their domain.

4. **Named Entity Recognition (NER).**
   Function: Extracts entities from each line using domain-specific NER models.
   Description:
   Thus, the NER module processes classified lines using adapted models for the domains. For such cases, BioBERT is bound to deal with Healthcare, SportsBERT, with Sports, while Healthcare-Sports BERT would have to deal with a content overlap. This particular step focuses on identifying as well as classifying entities with regard to the captured domain.

5. **Relationship Extraction.**
   Function: Identifies relationships between extracted entities.
   Description:

A transformer-based Relationship Extraction Model establishes contextual links between entities. These relationships would provide critical insights for query generation by connecting entities logically and semantically within the input text.

6. **Prioritization.**
   Function: Prioritizes lines based on the weight of extracted entities and relationships.
   Description:
   Every line-or phrase is evaluated very critically according to its richness of entities and relations with other words. The ones displayed first are the most meaningful and related ones, assuring their retrieval first during query generation.

7. **Query Generation and Execution.**
   Function: Generates structured queries and retrieves data from the appropriate database.
   Description:
   This Module converts the biased entities and relations into structured queries meant for specific domain databases. It integrates the use of mongodb for the effective demand execution and relevant data extraction.

8. **Information Extraction.**
   Function: Extracts relevant information from the database.
   Description:
   The whole system retrieves information corresponding to the queries produced. The resulted output, from one end points of structured details that is relevant to entities and relationships, is probably the given input text from which a final output would be produced.

9. **Output.**
   Function: Delivers the final output to the user.
   Description:
   This part retrieves the data for the user into a ready response. The response is sure to be co-related to the original query with information structured and coherent, not fragmentary.

VI. CONCLUSION

In this, multi-domain transfer learning-efficient architecture is presenting cross-domain information extraction from multiple fields like healthcare and sports. Named Entity Recognition (NER) and relationship extraction are processed using BERT-based domain classifiers with domain-specific model constructs such as BioBERT and SportsBERT ensuring optimal performance for processing and extraction to information repo. It also has structured Information Retrieval (IR) as well as response generator, thus making it more usable due to resulting contextually aligned to input. Such a framework could be capable of fulfilling the major needs of cross-domain Natural Language Processing (NLP) in terms of adaptability, scalability, and modularity since it reduces dependency on numerous domain-specific models. Transfer learning through pretrained models makes it resource-efficient while still being performing across domains.

In future improvements addressing more intricate and flexible queries across domains, the system is expected to include a highly sophisticated reasoning mechanism with query generation techniques. Apart from this, the architecture can be further extended to add more adaptive features for individuals with dynamic and conversational interfaces that engender higher user involvement through sophisticated dialogue systems. Thus, it exhibits the modularity and scalability of NLP systems, thereby proving effective at solving problems associated with cross-domain information extraction and laying the foundation for newer applications and research in the domain. Finally, between domain-specific constraints and unified processing, the system establishes a solid foundation for future studies in cross-domain NLP tasks

REFERENCES

[1] [Dipak Patel , Kiran Amin A Cross-Domain Semantic Similarity Measure and Multi-Source Domain Adaptation in Sentiment Analysis , 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS).

[2] Zhenzhen Xu, Fuli Zhang ,Wei Wang ,Haifeng Liu ,Xiangjie Kong ,Exploiting Trust and Usage Context for Cross-Domain Recommendation,11 May 2016 ,2169-3536

[3] Yue Zhao, Guorong Fan, Yuxing Cao, Yong Yang, Wenhua Gao, Zengshou Dong, A cross domain deep learning method for rolling bearing fault diagnosis,2024 43rd Chinese Control Conference (CCC),978-9-8875-8158-1

[4] Junge Liang, Lei Ma, Xin Xiong, Dangguo Shao, Yan Xiang, Xiongbing WangJunge Liang, Lei Ma, Xin Xiong, Dangguo Shao, Yan Xiang, Xiongbing Wang,Mutual Information-Based Word Embedding for Unsupervised Cross-Domain Sentiment Classification,2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)

[5] Xuefeng Wang, Ruopeng Yang, Yiwei Lu, Qingfeng Wu, Military Named Entity Recognition Method Based on Deep Learning, 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS).

[6] Ashwini Dalvi, Vaishnavi Shah, Dhruvin Gandhi, Siddharth Shah, S G Bhirud,Name Entity Recognition (NER) Based Drug Related Page Classification on Dark Web, 2022 International Conference on Trends in Quantum Computing and Emerging Business Technologies (TQCEBT).

[7] Maithili Sabane, Aparna Ranade, Onkar Litake, Parth Patil, Raviraj Joshi, Dipali Kadam,Enhancing Low Resource NER using Assisting Language and Transfer Learning, 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC).
[8] Chirawan Ronran, Seungwoo Lee, Effect of Character and Word Features in Bidirectional LSTM-CRF for NER, 2020 IEEE International Conference on Big Data and Smart Computing (BigComp).
[9] Kishan Kanhaiya, Naveen, Arpit Kumar Sharma, Kamlesh Gautam, Pramod Singh Rathore, AI Enabled- Information Retrival Engine (AI-IRE) in Legal Services: An Expert-Annotated NLP for Legal Judgements, 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS).
[10] Hongmei Zhu , Qijia Tian, Yongquan Liang, Shujuan Ji, Wei Sun,Domain Ontology Component-Based Semantic Information Integration, 2019 First International Workshop on Education Technology and Computer Science.
[11] Amir Zidi , Mourad Abed, A generalized framework for ontology-based information retirval: Application to a public-transportation system, 2013 International Conference on Advanced Logistics and Transport.
[12] Min Peng, Ming Yang, Songtao Sun, Hua Long, Nasir Ghani, Bin Ni, A query expansion based on sentence and vector integration strategy, The 5th International Conference on New Trends in Information Science and Service Science.
[13] B. Liu and L. Zhang, “A survey of opinion mining and sentiment analysis,” in Mining text data, Springer, 2012.
[14] L. Zhang and B. Liu, “Sentiment Analysis and Opinion Mining,” in Encyclopedia of Machine Learning and Data Mining, C. Sammut and G. I. Webb, Eds. Boston, MA: Springer US, 2016.
[15] M. Peng, Q. Zhang, Y. G. Jiang, and X. Huang, “Cross-domain sentiment classification with target domain specific information,” in ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers), 2018.