A unified LangChain, Groq Cloud and Llama-Based Approach to Document Information Retrieval in Multilingual Documents

Bijimol T.K*[1], Govind Krishnan[2], Gibin George[3]

[1*]Associate Professor, Department of Computer Applications, Amal Jyothi College of Engineering, Autonomous, Kanjirapally, Kerala, India tkbijimol@amaljyothi.ac.in

[2]PG Scholar, Department of Computer Applications, Amal Jyothi College of Engineering, Autonomous, Kanjirapally, Kerala, India govindkrishnan2025@mca.ajce.in

[3]Assistant Professor, Department of Computer Science, Santhigiri College of Computer Sciences, Vazhithala, Kerala, India george.gibin@gmail.com

HOW TO CITE:

Bijimol T.K, Govind Krishnan, Gibin George (2026), A unified LangChain Groq Cloud and Llama-Based Approach to Document Information Retrieval in Multilingual Documents.)

International Journal of Special Education, 41(19s), 1163-1173

ABSTRACT:

Document information extraction has emerged as a critical component of Natural Language Processing (NLP) as organizations increasingly digitize their operations. This paper provides an in-depth overview of the existing methods applied to retrieve structured information from various document types, including OCR to extract both handwritten and printed text, concept and semantic structure extraction, and Named Entity Recognition (NER). Although there has been an improvement in these areas, most of the currently available solutions have difficulties in handling complex, multilingual documents and usually cannot ensure the preservation of contextual integrity between different formats. The paper introduces a novel system that integrates LangChain to divide the document, Groq Cloud to find the desired data in a short time, and the multilingual model of Llama to improve the quality of extracted information. This experimental finding demonstrates that this method is better than the traditional models particularly where legal and regulatory frameworks are involved and they need a close interpretation. The paper concludes with the discussion of the directions of the research in the future, as more flexible and context-dependent frameworks are needed in document information extraction.

Keywords: Document Information Extraction, Natural Language Processing, LangChain Segmentation, Groq Cloud, Llama Multilingual Model.

1. Introduction

One of the applications of Natural Language Processing has been document information extraction. With organisations going paperless, digitisation of documents presents significant research potentials in image processing, NLP, ontology and semantic analysis (Silva & Silva, 2021). Most of the current research is however, task specific, and little has been done to develop generalised solutions to a variety of documents. The variation in document processing is based on the varied ways of creating documents and the abundance of content area, such as business, legal, technical, medical and academic documents (Perot et al., 2024). While an ideal solution would be universally applicable across different document types, languages, and structures, achieving such generalisation remains a significant challenge due to the intrinsic diversity in document

characteristics (Silva & Silva, 2021). The following sections present the research work.

1.1 Problem Statement

Document Information Extraction is a component of NLP that helps to move to a paperless organization. The inherent heterogeneity of document structures across business, legal, and medical fields makes research difficult. Existing solutions often fail to handle multilingual content or preserve contextual integrity. In addition, data is organized into PDFs, tables and document files are unstructured. The lack of structured representations requires sophisticated NLP techniques.

1.2 Literature Review of Extraction Techniques

The paper categorizes and reviews various established and modern methodologies such as OCR & Text Recognition, Structural Analysis, Named Entity Recognition, Advanced Architectures and Zero-Shot Learning.

1.3 The Proposed System

The proposed system is a novel three-stage architecture designed for document processing in legal and regulatory contexts:

1. Document Segmentation (LangChain): Breaks long, complex documents into logical, thematic segments to focus processing on relevant content.

2. Contextual Retrieval (Groq Cloud): Uses high-speed parallel infrastructure to retrieve relevant document chunks and provide inference.

3. Response Generation (Llama 3.1): Employs the Llama model to provide contextually accurate, multilingual responses to user queries.

1.4 Results

The proposed system was tested using questions such as scribe eligibility in Indian education system, Matrices such as ROUGE and Cosine Similarity are used to evaluate the performance and semantic relationship of the proposed model.

1.5. Unique Benefits of the Proposed System

The integration of LangChain and Llama gives many potential benefits for document Information Extraction.

1.6 Conclusion and Future Work

The paper concludes that integrating LangChain, Groq, and Llama creates a versatile, language-neutral platform. The future contributions include Improving multimodal integration, enhancing entity localization, and developing more diverse real-world benchmarks.

2. Review of Document Information Extraction Techniques

2.1 OCR for Handwritten Documents

OCR plays a critical role in converting handwritten and printed documents into digital format. Special language independent, rule level approaches are described for the differentiation of handwritten and printed text in data entry forms [3]. These approaches exploit structural and statistical properties; the baseline and characteristics of the texts are used to distinguish their source.

New OCR techniques have emerged with advances in deep learning. Efficient approaches of offline text recognition on the whole page with the help of deep neural networks are suggested [4]. They utilise object detection neural network, multi scale CNN and bidirectional LSTM to accomplish text localization and recognition. The systems typically detect text regions and partition text lines using data sets such as IAM. Pre-trained models, such as ResNet34, can be used for feature extraction, followed by text recognition using CNN-biLSTM architectures with integrated language modelling.

2.2 Concept Relation Extraction

Getting information on the relations between the involved concepts is very important for describing the structure of documents and their content. Methods have been proposed for segmenting document streams based on structural and factual descriptions, defining four document levels: documentary archive, technical documents, case notes, and fundamentals [5]. These approaches present new stream segmentation methods that

can utilize a logbook of prior pages for descriptors.

The use of 2D Conditional Random Fields for logical labelling has been investigated to understand the page structure in machine-printed documents [6]. These techniques create neighbourhood graphs and pairwise clique templates, integrating both local and contextual features taken from PDF files. Support Vector Machine (SVM) classifiers are then applied as a baseline that does not dependent on contextual information.

In historical document digitisation, the methods have been created to identify complex formatting patterns, like chapters and sections [7]. These plans connect the extracted information into a table of contents giving more insightful structure in terms of aggregation using techniques based on set operators and properties of table of content entries.

2.3 Semantic Structure Extraction

It is essential that the semantic structure extraction is used to understand the meaning and structure of the content of documents. Multimodal fully convolutional networks have been introduced to extract semantic structure via pixel-wise segmentation [8]. These approaches are typically executed in two phases: page segmentation and concept-relation analysis.

Techniques for dimensionality reduction in web text document classification, utilising WordNet ontology, have also been presented [9]. These methods employ Vector Space Models for mining tasks and apply Term Frequency-Inverse Document Frequency (TF-IDF) for task weighting. Ontology-based approaches have been validated using Principal Component Analysis (PCA).

2.4 Named Entity Recognition (NER) in Documents

Named Entity Recognition (NER) is a core task in Document Information Extraction. Extensive evaluations of various tools for document entity recognition have been conducted, assessing lexical, semantic, and heuristic features [10]. These studies have identified top-performing tools for specific document processing tasks.

NLP pipelines for NER in tweets have been developed, incorporating part-of-speech tagging and chunking [11]. These approaches apply shallow syntax in tweets using annotated data, generating features for named entity segmentation.

Research on the efficient enablement of semi-structured textual information has also led to proposals of frameworks for named entity detection in web content linked to text data records [12]. These methods build upon the natural occurrence of such structure by transformation processes and allow for detection in the collective, eliminating the need for per-record annotation in the form of training labels.

2.5 Advancements in Language Models

Recent improvements in language model have greatly advanced document-based language modelling. Transformer architecture has revolutionized NLP tasks including text summarisation and translation [13]. Because transformers make parallelisation, allowing for shorter training times than Recurrent Neural Networks. BERT (Bidirectional Encoder Representations from Transformers) of most recent have taken the field forward [14]. BERT uses bidirectional contextual representations, incorporating information from both directions of the input sequence. It has been applied to tasks such as question answering and classification.

ELMo (Embeddings from Language Models) is the Deep contextualised word representations that are provided through BiLSTM model from the language models [15]. ELMo generates output vectors in real-time based on the input text and has shown impressive performance across various tasks, such as question answering, coreference resolution, and sentiment analysis.

2.6 Document-Based Question Answering Systems

Whereas Document Information Extraction is a general idea, then question answering systems are its practical implementation. To increase the

accuracy of documents' selection miscellaneous strategies have been developed, such as Web And Semantic Knowledge Driven (WAD) [16]. These methods assess user inquiries, perform entity resolution, and use query bipartition strategies and, more importantly, ontological information to rate candidate answers.

New approaches for Graph-based Question Answering systems specific to reading comprehension tests have been proposed [17]. These methodologies synthesize applications of morphological analysis and reference checking, synonym check to improve the chances of assessing the right answer sentence from a passage.

2.7 Multimodal Approaches to Document Understanding

Recent work in document understanding has paid significantly more attention to structures like text, layout and image. They work to preserve much of the structure and characteristics of documents and, thus, construct stronger and more general information extraction systems.

The recent development of LayoutLM: multimodal pre-training of text and layout for document image understanding [18] has been put forward. This approach adds 2D position embeddings and image embeddings into BERT-style architecture, and achieves superior performance across a range of document understanding tasks.

Extending from this, LayoutLMv2 proposes the new idea of spatial-aware self-attention and a new image-text matching task [19]. This model obtains the state-of-the-art recognition performance on numerous documents understanding applications.

In LayoutLMv3, a novel pre-training task of text and image masking is added to learn cross-modal associations [20]. This combined text and image masking approach helps generalize the model for the interdependencies of textual and visual content in documents.

2.8 Graph-based Approaches for Document Understanding

There are also graph based approaches which have recently shown evidence that they are an excellent way of modelling the interdependence of elements to a document. Such approaches can represent the textual information and location of documents on a plane, simultaneously.

ROPE (Reading Order Equivariant Positional Encoding) employs a new positional encoding to encompass the reading sequence of the rhetorical items [21]. This method enhances the ability to extract information by capturing the sequential texture of document content while preserving spatial information.

This is in addition to FormNet, a structural encoding approach, that provides a better model for form Document Information Extraction as compared to sequential modelling [22]. It uses a graph neural network to capture long-range dependencies and structural information in forms, leading to improved performance on complex document layouts. FormNetV2 uses form document information extraction from raw data through multimodal graph contrastive learning [23]. Later, advancing the idea of combining textual and visual information, we achieve even better document representations and work on high benchmarks.

2.9 Zero-shot and Few-shot Learning for Document Understanding

As the field of document understanding evolves, there is an increasing need to construct new models that are trainable on new document types and its schemas without any further trainings at all. This has prompted machine learning studies in zero-shot and few-shot learning for Document Information Extraction.

A class of query-form based zero-shot form entity query framework called QueryForm has been introduced [24]. This method enables the extraction of information from forms without having to use the training data for tasks and using natural language queries to condition the extraction task.

The layout-aware generative language model for multimodal document understanding by the name DocLLM has been presented [25]. By using the LLM approach, it is possible to perform zero-shot information extraction from the documents to make reasoning and use general knowledge and reason with them.

2.10 Benchmarks and Evaluation

Regarding the topic in the field of document understanding, the description of benchmarks and the methods of evaluation have been introduced to examine the effectiveness of the strategies on different types of documents and tasks.

CORD, the Consolidated Receipt Dataset, has been introduced as a post-OCR parsing benchmark for receipts consolidated receipt dataset [26]. This set of documents serves well as a test collection to assess how well information extraction systems can perform on realistic documents whose formats and content differ in many ways.

VRDU or Visually-Rich Document Understanding can be regarded as one of the most complex proposed worldwide as a benchmark to evaluate document understanding systems [27]. It involves activities like form understanding and invoice processing with the degrees of freedom for accessing the data to measure the performance of models in conditions with limited amount of data and unlimited amount of data available.

2.11 Large Language Models for Document Information Extraction

New possibilities of Document Information Extraction have emerged due to the emergence of large language models. These models trained by millions of textual data are quite effective in understanding and even generating text that is similar to human-written text in different fields.

LMDX (Language Model-based Document Information Extraction and Localisation) has been proposed as the methodology for the further use of LLMs for obtaining the desired information from the documents that contain a lot of visual information [28]. LMDX resolves important issues that arise when applying LLMs to document processing, including encodings of layout information and re-contextualisation of the extracted entities within the document.

This approach enables the extraction of singular, repeated, and hierarchical entities with or without training data, with also grounding assurance as well as the ability to identify the location of the entities. In this paper, we observe the efficacy of LMDX on VRDU and CORD benchmarks and prove the effectiveness of LLMs for generating accurate and concise document parsers.

3. Proposed System

Based on the described system, an improved solution for document info extraction is proposed, which integrates both the LangChain's option of segmenting docs and the improved multilingual model of the Llama. This integration also assists in offering high precision to the multilingual queries that are complex and is what makes it stand out to the traditional models. As we can observe, the segmentation approach of LangChain allows the system to focus on only the various segments that add significant value to the parameters to be discussed, thereby enhancing the accuracy of Llama, a strategy that is valuable in the fields such as regulatory, legal where accuracy and context preservation is of utmost importance.

3.1 System Functionality

In the tests that were conducted, the system was found to have a high degree of contextual clarity in responding to particular academic regulatory questions within the framework of the Indian education system e.g. when responding to questions such as:" Can B Tech students be scribes to differently abled students with M.Tech education when given official document of a university in India?" The division completed by LangChain was precise enough to identify the topical parts of the documents, allowing Llama to put the query in context. The system gave a very specific answer that B.Tech students cannot be employed as scribes to the M.Tech students, which is a subtle difference that is consistent with the initial purpose of the document. This kind of specificity is an advantage over baseline models, like the DistilBERT model, in which the outcomes

of the models are, more frequently than not, more generic and less context-aware.

3.2 System Architecture and Workflow

Based on the architecture of the proposed document extraction system, it can embrace the capabilities of LangChain to divide the document, Groq Cloud platform to execute processes in parallel, and the advanced multilingual cognition achieved by the Llama 3.1. This design addresses some principal issues with the current models employed to extract documents, namely, offering a multiple tiered system that is expected to classify and interpret the information contained in the significant regulation documents effectively and with a high degree of contextual relevance. The system work flow is divided into three broad operational phases – Document Segmentation, Contextual Retrieval and Response Generation which addresses key issues in document management, as outlined in Figure 1.

a) Document Segmentation (LangChain)

Document segmentation capacity of LangChain is the first and fundamental component of the system framework. This segmentation process is developed particularly for such long and hard copy documents like legal documents or guides and rules and regulations. LangChain breaks input documents into logical segments of related content using document structure and contextually relevant breakpoints. This segmentation step defines required areas of the document and can reduce the processing of unnecessary information. The segmented text is then pre-processed using a chunking process in which the text is split into smaller, manageable chunks. Each chunk is created with the intention of preserving context through a small overlap with neighbouring chunks. The rationale for such segmentation and chunking methodology is to feed the subsequent processing stages with input data that is logically well structured and semantically well aligned so as to encourage better response generation. Due to this, the segmentation and chunking process is intended to improve the efficiency of processing large and complex documents.

b) Contextual Retrieval (Groq Cloud API)

After segmenting, retrieval processing of each chunk is done by Groq Cloud on high-speed and parallel infrastructure. The API of Groq Cloud will serve as a middleware that manages the retrieval of the document chunks efficiently and returns the parts that have been segmented into smaller pieces of information to alignment to the context to the information can be appropriately extracted. The Groq Cloud API provides high-speed inference infrastructure that can process incoming requests efficiently. This high-speed processing can reduce inference latency and support quick responses to user queries. During this phase, the set of segments of the text, which are already separated, are processed by the targeted Groq Cloud structure into the matched vectors, using NLP vectors embedding that is compatible with Llama 3.1. These representations preserve semantic information from the chunks and support subsequent processing. Scalability is also achieved in the Groq Cloud architecture because it allows the system to support large scale document retrieval duties over a distributed system. Groq Cloud enables a smooth interoperability process by optimising the process of data flow between LangChain and Llama 3.1, which effectively allows bridging the gap between segmentation and language processing.

c) Response Generation (Llama 3.1)

The last layer in the system would be Llama 3.1 which is used to provide understanding of language and to generate responses. It makes semantically correct responses by interpreting the meaning of the context during processing pre-processed embeddings of Groq Cloud using deep learning. The model is aimed to maintain the informational content as well as the regulatory or legal context of the source text. Embeddings Transformer-based embeddings and Multi-Head Attention Llama 3.1 is capable of cross-lingual understanding, allowing it to give consistent and unbiased answers across languages. The model processes query and uses its attention layers so as to produce outputs that do not lose the meaning of the original document. Also, the formatting of

responses makes them grammatically, syntactically and contextually correct.

The work offers a new way of addressing the issues of extracting the information of the document by combining document segmentation based on LangChain, high-speed retrieval on Groq Cloud and multilingual responses using Llama 3.1. Beyond providing an easy way to conduct the efficient retrieval and response generation to support the context-specific assessment, this multi-level organization is also involved in the creation of an overall and versatile platform that is designed to support multiple languages and that can be tailored to high-stakes fields such as legal and compliant work.

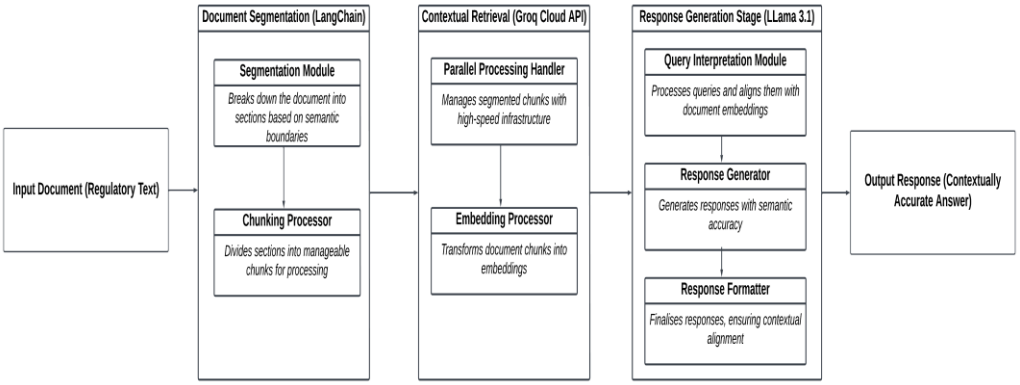

Figure 1. Block diagram of the proposed document extraction system, detailing the Document Segmentation, Contextual Retrieval, and Response Generation stages.

4. Evaluation Metrics

To evaluate the effectiveness of the system, two major evaluation measures were utilized: ROUGE scores and Cosine Similarity. All measures were used to quantify accuracy and relevance in such a manner that it reflects both the linguistic and semantic correspondence:

ROUGE Score: ROUGE, an acronym that means Recall-Oriented Understudy for Gisting Evaluation, is a common metric that can be used to measure the performance of a system writing text based on a sequence of reference answers [29]. ROUGE can give information on the completeness and relevance of the responses by contrasting their recall and F1 scores. As seen in Figure 2, the proposed system reports a higher ROUGE score than the baseline models especially in case of multiple languages. This result shows stronger lexical overlap with the reference responses, which is a key element of regulatory and legal texts.

Table 1. Comparison of ROUGE score between proposed system and DistilBERT.

Metric	Proposed System	DistilBERT
ROUGE Score	0.253	0.105
Performance Level	High	Low

Cosine Similarity: Cosine Similarity is a measure that is used to determine the degree to which the meaning of the responses of the system is similar to the desired responses, but does not use word-to-word matches [30]. It gives the feeling of semantic proximity by finding the cosine of the angle between the vectors of each text. This method can be used especially in the evaluation of semantic accuracy especially in multilingual tasks. Figure 3 shows scores of Cosine Similarity which shows that the proposed system has

stronger meaning correspondence among languages, which indicates stronger semantic similarity between the evaluated responses.

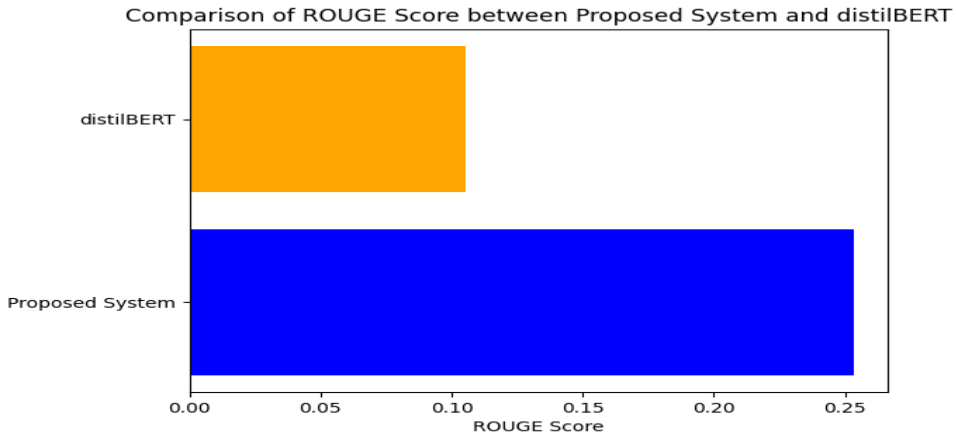

Figure 2: The Proposed System and Traditional Model ROUGE Score Comparison which reveals that the proposed system has a higher accuracy in content summarisation than the traditional models.

Table 2. Comparison of Cosine Similarity score between proposed system and DistilBERT

Metric	Proposed System	DistilBERT
Cosine Similarity Score	0.322	0.071

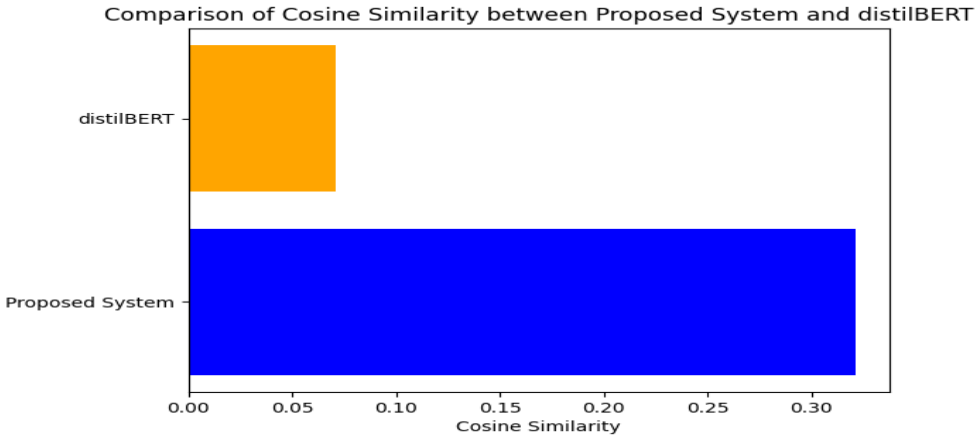

Figure 3. Comparison of Cosine similarity of Proposed System and Traditional Models demonstrating the improved semantic alignment of the proposed system that demonstrates its capacity to isolate and maintain contextual finer details.

5. Unique Benefits of the Proposed System

The integration of LangChain and Llama gives number of specific benefits, for improving Document Information Extraction.

Multilingual Query Flexibility: The system is intended to support queries in different languages without requiring an explicit translation step, subject to the capabilities of the selected embedding and language models. This is beneficial in regulatory settings where users may submit queries in different languages.

Context-Aware Summarisation: The document segmentation approach of LangChain enhances the capacity of the system to remember the context and categorises documents into meaningfully related parts. This can be achieved so that the responses of the system remain exact and in tandem with the original meaning of the document and chances of misinterpretation are minimized as it can easily happen in traditional models.

Relevance in High-Stakes Domains: The combination of segmentation and multilingual comprehension is distinctive and, therefore, such a system is especially successful in critical areas of such domains as legal, academic, and regulatory compliance. The system may be useful for information extraction in these fields where accuracy and context are essential, and the sophisticated language understanding capability of the system would prove to be an effective instrument in extracting accurate information.

6. Conclusion and Future Directions

In this research paper, we have explored the different approaches to Document Information Extraction, and we have touched on the approaches possible and include OCR and semantic structure and the named entity recognition as well as the latest advances in the language models. Though much has been achieved in all these areas, it has not yet become very easy to create more flexible systems which are able to handle a large number of document types, formats, and languages.

When analysing the existing methods of document extraction, it becomes apparent that there are significant shortcomings especially when document extraction is of multilingual documents and when the document is context-intensive. In spite of the fact that most of the existing techniques work reasonably well with general queries, they do not have the richness and level of detail needed on complex legal or regulatory documents.

To fill these holes, the suggested system will integrate the segmentation functionality of LangChain with the multilingual processing offered by Llama to offer a system that produces relevant and accurate replies. The system is optimized by combining the document segmentation of LangChain with the language comprehension of the advanced Llama, which increases its performance through the improvement of ROUGE and Cosine Similarity metrics. This combination renders the system particularly adapted to the use in regulatory compliance, legal research or any other field that requires information retrieval that is language sensitive and needs accuracy to the last word.

Among the main areas that can be promising in future research are:

1. Improving Multimodal Integration: Future work is required to incorporate more types of information: text, layout, and images, into document understanding models to form a more complete process.

2. Moving forward on Zero- Shot and Few-Shot Learning: The next step in enhancing the generalisability of Document Information Extraction in a wide range of contexts is to develop more efficient and versatile zero-shot and few-shot learning methods.

3. Models Exploring New Architectures: Design model architectures can become more innovative in order to reflect the hierarchical and spatial structure of documents in a better manner.

4. Enhancing Entity Localisation and Grounding: Further development of entity localization and grounding methods could make

extracted information more graspable and appropriate to users.

5. Development of Diverse Benchmarks: It will be beneficial to develop more comprehensive benchmarks, that is, spanning across a variety of domains and languages, in order to test document understanding systems in a more realistic manner that would be relevant to real-world contexts.

Further advancement of research may lead to more advanced systems that will be able to read and comprehend documents with human-like accuracy and understanding.

8. Acknowledgements

The authors express their gratitude to Amal Jyothi College of Engineering and Santhigiri College of Computer Sciences for institutional support and computational infrastructure provided during this research.

9. Author Declaration

The authors confirm that this work is original and has not been published elsewhere in whole or in part. The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

7. References

Chung, J., & Delteil, T. (2019). A computationally efficient pipeline approach to full page offline handwritten text recognition. International Conference on Document Analysis and Recognition Workshops (ICDARW), 35–40.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT, 4171–4186.

Elhadad, M. K., Badran, K. M., & Salama, G. I. (2017). A novel approach for ontology-based dimensionality reduction for web text document classification. IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), 373–378.

Huang, Y., Lv, T., Cui, L., Lu, Y., & Wei, F. (2022). LayoutLMv3: Pre-training for document AI with unified text and image masking. Proceedings of the 30th ACM International Conference on Multimedia, 1273–1282.

Jayalakshmi, S., & Sheshasaayee, A. (2017). Automated question answering system using ontology and semantic role. International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 528–532.

Karpinski, R., & Belaïd, A. (2016). Combination of structural and factual descriptors for document stream segmentation. 12th IAPR Workshop on Document Analysis Systems (DAS), 221–226.

Lahitani, A. R., Permanasari, A. E., & Setiawan, N. A. (2016). Cosine similarity to determine similarity measure: Study case in online essay assessment. 4th International Conference on Cyber and IT Service Management, 1–6.

Lee, C. Y., Li, C. L., Dozat, T., Perot, V., Su, G., Wang, R., Fujii, Y., & Pfister, T. (2022). FormNet: Structural encoding beyond sequential modeling in form document information extraction. Proceedings of ACL 2022, 3735–3754.

Lee, C. Y., Li, C. L., Wang, C., Wang, R., Fujii, Y., Qin, S., Popat, A., & Pfister, T. (2021). ROPE: Reading order equivariant positional encoding for graph-based document information extraction. Proceedings of ACL 2021, 6213–6223.

Lee, C. Y., Li, C. L., Zhang, H., Dozat, T., Perot, V., Su, G., Zhang, X., Sohn, K., Glushnev, N., Wang, R., Ainslie, J., Long, S., Qin, S., Fujii, Y., Hua, N., & Pfister, T. (2023). FormNetV2: Multimodal graph contrastive learning for form document information extraction. Proceedings of ACL 2023, 9011–9026.

Lin, C. Y. (2004). ROUGE: A package for automatic evaluation of summaries. Text Summarization Branches Out: Proceedings of the ACL Workshop, 74–81.

Marrero, M., Urbano, J., Sánchez-Cuadrado, S., Morato, J., & Gómez-Berbís, J. M. (2013). Named Entity Recognition: Fallacies, challenges and opportunities. Computer Standards & Interfaces, 35(5), 482–489.

Nguyen, T. T. H., Doucet, A., & Coustaty, M. (2017). Enhancing table of contents extraction by system aggregation. 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), 242–247.

Park, S., Shin, S., Lee, B., Lee, J., Surh, J., Seo, M., & Lee, H. (2019). CORD: A consolidated receipt dataset for post-OCR parsing. Workshop on Document Intelligence at NeurIPS 2019.

Perot, V., Kang, K., Luisier, F., Su, G., Sun, X., Boppana, R. S., Wang, Z., Mu, J., Zhang, H., Lee, C. Y., & Hua, N. (2024). LMDX: Language model-based document information extraction and localization. arXiv preprint arXiv:2402.12345.

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. Proceedings of NAACL-HLT, 2227–2237.

Ritter, A., Clark, S., Mausam, & Etzioni, O. (2011). Named entity recognition in tweets: An experimental study. Proceedings of EMNLP, 1524–1534.

Saba, T., Almazyad, A. S., & Rehman, A. (2015). Language independent rule based classification of printed & handwritten text. IEEE EAIS, 1–4.

Silva, & Silva (2021). A review on document information extraction approaches. Proceedings of the Student Research Workshop Associated with RANLP-2021, 174–179.

Tao, X., Tang, Z., Xu, C., & Wang, Y. (2014). Logical labeling of fixed layout PDF documents using multiple contexts. 11th IAPR International Workshop on Document Analysis Systems, 360–364.

Toral, A., & Munoz, R. (2006). A proposal to automatically build and maintain gazetteers for Named Entity Recognition by using Wikipedia. Workshop on New Text Wikis and Blogs, 11–18.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 6000–6010.

Veena, G., Athulya, S., Shaji, S., & Gupta, D. (2017). A graph-based relation extraction method for question answering system. ICACCI, 944–949.

Wang, D., Raman, N., Sibue, M., Ma, Z., Babkin, P., Kaur, S., Pei, Y., Nourbakhsh, A., & Liu, X. (2023b). DocLLM: A layout-aware generative language model for multimodal document understanding. arXiv preprint arXiv:2312.11486.

Wang, Z., Zhang, Z., Devlin, J., Lee, C. Y., Su, G., Zhang, H., Dy, J., Perot, V., & Pfister, T. (2023a). QueryForm: A simple zero-shot form entity query framework. Findings of ACL 2023, 4146–4159.

Wang, Z., Zhou, Y., Wei, W., Lee, C. Y., & Tata, S. (2023c). VRDU: A benchmark for visually-rich document understanding. Proceedings of the 29th ACM SIGKDD Conference, 5184–5193.

Xu, Y., Li, M., Cui, L., Huang, S., Wei, F., & Zhou, M. (2020). LayoutLM: Pre-training of text and layout for document image understanding. Proceedings of ACM SIGKDD, 1192–1200.

Xu, Y., Xu, Y., Lv, T., Cui, L., Wei, F., Wang, G., Lu, Y., Florencio, D., Zhang, C., Che, W., Zhang, M., & Zhou, L. (2021). LayoutLMv2: Multi-modal pre-training for visually-rich document understanding. Proceedings of ACL 2021, 2579–2591.

Yang, X., Yumer, E., Asente, P., Kraley, M., & Kifer, D. (2017). Learning to extract semantic structure from documents using multimodal fully convolutional neural networks. IEEE CVPR, 4342–4351.