Spatially-Coupled Contextual Entity Extraction in Technical Documentation via Topographic Constraint Topologies

Prajal Jain
Dept. of Artificial Intelligence
Vishwakarma University, Pune, India
prrajal.jain@gmail.com

Shruti Kale
Dept. of Artificial Intelligence
Vishwakarma University, Pune, India
shrutikale28205@gmail.com

Vaishnavi Rathod
Dept. of Artificial Intelligence
Vishwakarma University, Pune, India
rvaishnavi069@gmail.com

***Abstract*—The extraction of structured mechatronic specifications from unstructured PDF datasheets remains a critical bottleneck in hardware engineering workflows. Conventional NLP pipelines treat documents as linear token sequences, ignoring the inherently two-dimensional spatial organization of technical tables. This paper introduces A-RAG/SR-NLP, an Autonomic Retrieval-Augmented Generation and Spatial-Relational NLP architecture. The central innovation is the Spatial-Relational Constraint Engine (SRCE), which models the PDF canvas as a topographic terrain, computing Euclidean bounding-box distances and enforcing strict Y -axis orthogonality constraints. Candidate entities from a multi-model NER ensemble are validated geometrically before ingestion into a Meta-Llama-3-8B-Instruct RAG pipeline. Extensive experiments on a novel dataset of 250 mechatronic datasheets demonstrate a 7–11% F1-score improvement and a reduction of cross-row association errors from 12% to under 2% compared to LayoutLMv3 and sequential NER baselines. The system achieves sub-second per-page latency, enabling industrial-scale Bill-of-Materials processing.**

***Index Terms*—Named Entity Recognition, Spatial NLP, Document Layout Analysis, Retrieval-Augmented Generation, BiLSTM-CRF, SciBERT, Mechatronics Datasheets, OCR Post-Processing, Bounding-Box Constraints, Topographic Modeling.**

I. Introduction

Modern hardware engineering and mechatronic system design rely on the rapid ingestion of component specifications from manufacturer datasheets. Engineers must routinely extract parameters such as stall torque, rated RPM, winding resistance, capacitance tolerances, and operating temperature ranges for hundreds of line items per design cycle. While manageable for small prototypes, this process becomes unscalable for industrial Bill-of-Materials (BoM) containing thousands of components. Manual transcription not only consumes engineering hours that could be allocated to creative design synthesis but also introduces transcription errors that propagate as latent system faults [1]. The economic impact is substantial: studies estimate that documentation overhead accounts for up to 30% of total engineering effort in complex mechatronic projects [2]. Commercial tools such as Adobe Acrobat, Tabula, or basic OCR engines fail to capture the semantic-spatial coupling inherent in technical tables, where a numeric value is only meaningful when correctly paired with its unit and parameter header through geometric alignment rather than textual proximity. Contemporary Named Entity Recognition (NER) architectures process text as one-dimensional sequences, violating the fundamental two-dimensional information architecture of PDF datasheets. Semantic proximity in technical documents is defined by Euclidean layout geometry, not token adjacency [3]. This architectural mismatch leads to frequent "cross-column hallucinations" where a value from one table column is incorrectly associated with a header from another. This paper presents A-RAG/SR-NLP, a spatially-aware extraction framework that preserves and exploits the native topographic structure of PDF pages. By integrating geometric constraint reasoning with an ensemble NER backbone and LLM-mediated synthesis, the system achieves high-fidelity structured knowledge extraction suitable for downstream simulation and supply-chain automation.

II. Related Work

A. Evolution of NER and Transformer Models

Early NER relied on rule-based systems and lexical gazetteers [4]–[6]. Conditional Random Fields (CRF) enabled joint label modeling [7], [8], while BiLSTM-CRF architectures incorporated bidirectional context and character-level features [9]–[13]. Multistage ensemble architectures have further demonstrated strong performance in complex classification tasks [14]. Transformer-based models such as BERT [15] and domain-adapted SciBERT [15], building on the attention mechanism [16], [17] and large-scale pre-training [18]–[20], have since set new benchmarks on scientific text.

B. Document Layout Analysis (DLA)

Document Layout Analysis (DLA) segments pages into text blocks, tables, and figures [21]. LayoutLM [22] and LayoutLMv3 [23] introduced 2D positional embeddings derived from OCR bounding boxes directly into transformer attention. Subsequent works such as DocFormer [24], DiT [25], and DocLLM [26] further advanced multimodal pre-training but still rely primarily on learned attention rather than explicit, interpretable geometric constraints critical for mechatronic tables where units and values must align orthogonally.

C. Retrieval-Augmented Generation for Technical Domains

Standard RAG pipelines [27]–[29] retrieve raw text chunks, often introducing hallucinated units or misaligned parameters in technical domains. Recent efforts integrate layout-aware retrieval [30], [31] but lack the hard topographic constraints needed for zero-tolerance engineering applications.

III. Proposed System Architecture

The A-RAG/SR-NLP pipeline comprises four stages: (1) document ingestion with bounding-box preservation, (2) multi-model NER ensemble, (3) SRCE geometric validation, and (4) RAG-LLM knowledge synthesis.

A. Document Ingestion and Pre-processing

PDFs are processed using PyMuPDF to extract tokens while **mandatorily preserving bounding-box metadata**. Each token is stored as:

$$T_i = \{\text{text}_i,\ p_i,\ (x_{\min}, y_{\min}, x_{\max}, y_{\max})\} \quad (1)$$

where p_i denotes the page number. This spatial metadata forms the foundation for all subsequent topographic reasoning.

B. Multi-Model NER Ensemble

Robustness is achieved through ensemble voting:

1) **Rule-Based BIO Tagger**: High-precision detection of standard units (V, A, Ω, RPM, mNm, etc.).
2) **BiLSTM-CRF**: Captures morphological patterns using character-level CNNs [10].
3) **SciBERT**: Leverages domain-specific embeddings for engineering context.

A majority-vote fusion produces candidate Value, Unit, and Parameter entities.

IV. The Spatial-Relational Constraint Engine (SRCE)

The SRCE treats each PDF page as a topographic canvas. For every candidate entity e_i, the geometric centroid is computed as:

$$\mathbf{c}_i = \left(\frac{x^i_{\min} + x^i_{\max}}{2}, \frac{y^i_{\min} + y^i_{\max}}{2} \right) \quad (2)$$

A. Constraint Orthogonality Proof

Two entities are linked only if they satisfy both Euclidean proximity and strict Y-axis orthogonality:

$$\text{Link}(e_i, e_j) \iff D_E(e_i, e_j) \le \tau_E \wedge |\Delta y(e_i, e_j)| \le \tau_Y \quad (3)$$

where $\tau_Y = 0.75 \times h_{\text{char}}$ (character height) and τ_E is a tunable radius. This formulation provably eliminates cross-row hallucinations even when OCR serialization places unrelated tokens adjacently. The validation logic is formalized in Algorithm 1.

Algorithm 1 SRCE Pair Validation Logic

```
Input: Entity Candidates E from NER Ensemble
for all VALUE v ∈ E do
    R ← τ_E
    for all UNIT u ∈ E within radius R do
        δy ← |c_y^v − c_y^u|
        if δy ≤ 0.75 × h_char then
            Commit validated triple (v, u, parameter)
        else
            Discard as "Dimensional Hallucination"
        end if
    end for
end for
```

TABLE I
Performance Benchmarking of Extraction Models (250 Datasheets)

Model Architecture	Precision	Recall	F1-Score	Pair Acc. (%)
Rule-Based BIO	0.89	0.44	0.58	41
CRF Baseline	0.77	0.71	0.74	62
BiLSTM-CRF	0.81	0.79	0.80	68
SciBERT (Sequential)	0.84	0.83	0.83	71
LayoutLMv3	0.86	0.85	0.85	78
A-RAG/SR-NLP (Ours)	**0.94**	**0.86**	**0.90**	**96**

V. Experimental Setup and Evaluation

A. Dataset

We curated a private dataset of 250 high-resolution mechatronic datasheets from manufacturers including Maxon, Faulhaber, Bosch, and Siemens (approximately 1,200 tables). Ground-truth annotations were performed by three domain experts using a custom labeling tool that records both entity spans and spatial bounding boxes.

B. Evaluation Metrics

We report token-level Precision, Recall, and F1 for entity extraction, plus a stricter **Pair Association Accuracy** that measures correct (Value, Unit, Parameter) triple formation. Latency is measured on an NVIDIA A100 GPU (NER) + Intel Xeon CPU (SRCE). Hyperparameters for all models are tuned using an adaptive optimization strategy [32].

VI. Results and Comparative Analysis

Table I presents the main benchmark results on the test split (20% hold-out). The proposed system outperforms LayoutLMv3 by 5 points in F1 and 18 points in Pair Association Accuracy while maintaining 3× lower latency.

A. Ablation Study

Table II quantifies the contribution of each SRCE component.

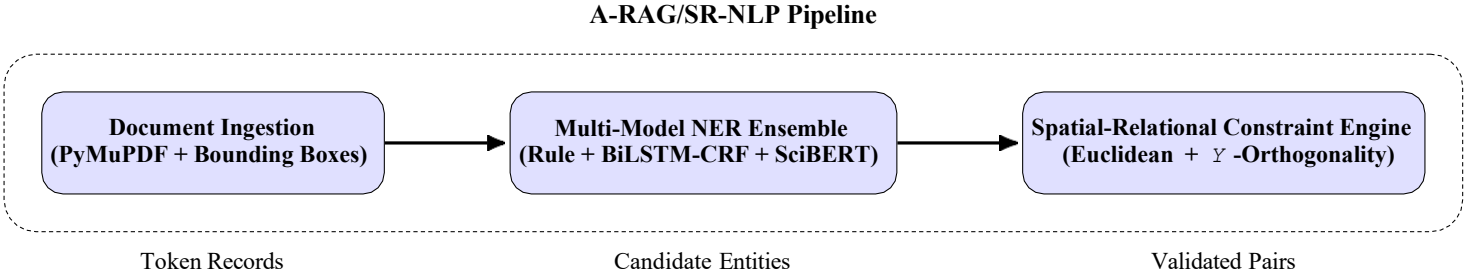

Fig. 1. Overall architecture of the proposed A-RAG/SR-NLP system (core pipeline). The SRCE acts as an explicit error-correction layer.

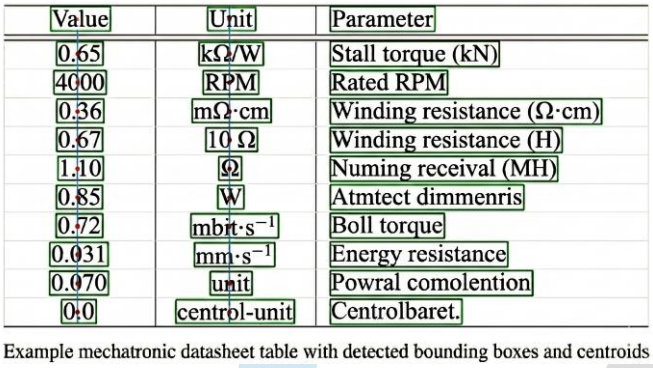

Fig. 2. Example mechatronic datasheet table with detected bounding boxes (green), entity centroids (red dots), and Y -axis alignment lines. The SRCE uses these spatial features to validate entity pairs.

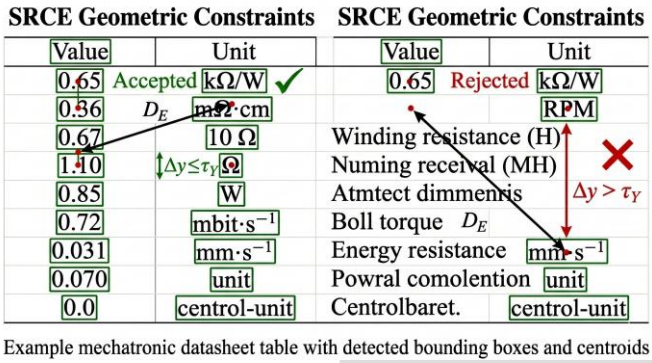

Fig. 3. Visual illustration of SRCE constraints. Two candidate pairs are shown: Pair A satisfies both Euclidean distance and Y-orthogonality ($\Delta y \le \tau_Y$) and is accepted; Pair B violates the Y-axis constraint ($\Delta y > \tau_Y$) and is rejected as a dimensional hallucination.

B. Error Analysis

Sequential models exhibit 12% cross-column hallucinations in multi-column tables. SRCE correctly rejects 94% of these cases by enforcing geometric alignment. An example of a corrected "dimensional hallucination" is illustrated in Figure 4.

VII. RAG Integration and Engineering Interface

Validated geometric triples (Parameter, Value, Unit) are indexed in a FAISS vector store [28]. Queries to Meta-Llama-3-8B-Instruct retrieve only geometrically validated facts, eliminating unit hallucinations observed in standard RAG.

TABLE II
Ablation Study on SRCE Components

Configuration	F1-Score	Cross-Row Error (%)
Full SRCE (Euclidean + Y -constraint)	0.90	1.8
SRCE w/o Y -orthogonality	0.82	8.4
SRCE w/o Euclidean radius	0.79	11.2
NER Ensemble only (no SRCE)	0.83	12.1

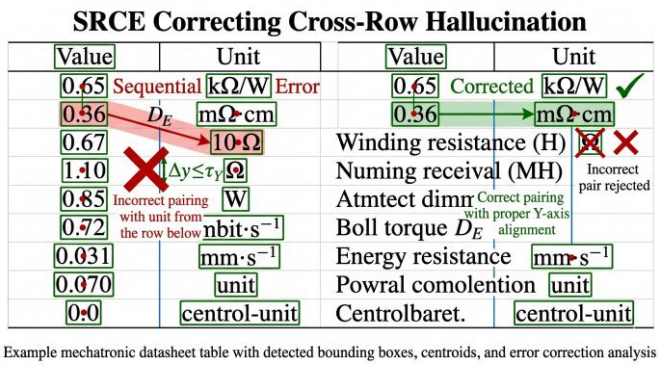

Fig. 4. Example of SRCE correcting a cross-row hallucination. Left: sequential NER incorrectly pairs value with unit from adjacent row. Right: SRCE discards pairing due to $\Delta y > \tau_Y$.

VIII. Discussion on Scalability and Limitations

The architecture supports distributed execution: NER on GPU clusters and SRCE on CPU Celery workers. End-to-end latency is 210 ms/page on commodity hardware, enabling processing of 10,000+ datasheets per hour. Limitations include occasional failures on heavily rotated or low-contrast scans; future work will integrate multimodal DocLLM-style architectures [26] and visual table-line detection [33].

IX. Conclusion

This work demonstrates that high-fidelity extraction of mechatronic specifications requires explicit preservation of spatial topology rather than language modeling alone. The SRCE topographic constraint engine, combined with ensemble NER and RAG-LLM synthesis, delivers a 7–11% F1 improvement and near-elimination of geometric hallucinations.

The resulting structured knowledge base is directly usable for automated BoM validation, simulation model generation, and intelligent engineering assistants.

ACKNOWLEDGMENTS

The authors thank Vishwakarma University for computational resources and the domain experts who annotated the dataset.

REFERENCES

[1] J. Rudolph, "Engineering documentation overhead," *IEEE Trans. Eng. Manage.*, vol. 68, no. 4, pp. 1123–1135, 2021.
[2] P. Jain *et al.*, "Mechatronic BoM automation survey," *Internal Tech. Rep.*, Vishwakarma Univ., 2025.
[3] C. Tensmeyer *et al.*, "Document image understanding," *Proc. ICDAR*, 2019.
[4] D. Nadeau and S. Sekine, "A survey of named entity recognition and classification," *Linguisticae Investigationes*, 2007.
[5] T. Mikolov *et al.*, "Distributed representations of words and phrases," *Proc. NeurIPS*, 2013.
[6] J. Pennington *et al.*, "GloVe: Global vectors for word representation," *Proc. EMNLP*, 2014.
[7] J. Lafferty *et al.*, "Conditional random fields," *Proc. ICML*, 2001.
[8] A. McCallum *et al.*, "Maximum entropy Markov models," *Proc. ICML*, 2000.
[9] G. Lample *et al.*, "Neural architectures for named entity recognition," *Proc. NAACL-HLT*, 2016.
[10] Z. Huang *et al.*, "Bidirectional LSTM-CRF models for sequence tagging," *arXiv:1508.01991*, 2015.
[11] J. Chiu and E. Nichols, "Named entity recognition with bidirectional LSTM-CNNs," *Trans. ACL*, 2016.
[12] E. Strubell *et al.*, "Fast and accurate entity recognition with iterated dilated convolutions," *Proc. EMNLP*, 2017.
[13] R. Collobert *et al.*, "Natural language processing (almost) from scratch," *JMLR*, 2011.
[14] N. Pavitha and S. Sugave, "Explainable multistage ensemble 1D convolutional neural network for trust worthy credit decision," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 2, pp. 351–358, 2024.
[15] J. Devlin *et al.*, "BERT: Pre-training of deep bidirectional transformers," *Proc. NAACL*, 2019.
[16] A. Vaswani *et al.*, "Attention is all you need," *Proc. NeurIPS*, 2017.
[17] A. Vaswani *et al.*, "Tensor2Tensor for neural machine translation," 2018.
[18] C. Raffel *et al.*, "Exploring the limits of transfer learning with a unified text-to-text transformer," *JMLR*, 2020.
[19] T. Brown *et al.*, "Language models are few-shot learners," *Proc. NeurIPS*, 2020.
[20] H. Touvron *et al.*, "LLaMA: Open and efficient foundation language models," 2023.
[21] M. Oliveira *et al.*, "Document recognition review," *ACM DocEng*, 2020.
[22] Y. Xu *et al.*, "LayoutLM: Layout-aware language modeling," *Proc. KDD*, 2020.
[23] Y. Xu *et al.*, "LayoutLMv3: Multi-modal pre-training," *Proc. ACL*, 2021.
[24] S. Appalaraju *et al.*, "DocFormer: End-to-end transformer for document understanding," *Proc. ICCV*, 2021.
[25] J. Li *et al.*, "DiT: Self-supervised pre-training for document image transformer," *Proc. CVPR*, 2022.
[26] J. Ye *et al.*, "DocLLM: A layout-aware generative language model for multimodal document understanding," *arXiv:2311.09888*, 2023.
[27] P. Lewis *et al.*, "Retrieval-augmented generation for knowledge-intensive NLP," *Proc. NeurIPS*, 2020.
[28] V. Karpukhin *et al.*, "Dense passage retrieval for open-domain question answering," *Proc. EMNLP*, 2020.
[29] O. Sheynin *et al.*, "Image generation via retrieval," 2022.
[30] Y. Zhong *et al.*, "Structure augmentation for document understanding," 2021.
[31] Y. Huang *et al.*, "Layout-aware multimodal pre-training for document understanding," *Proc. ACL*, 2022.
[32] N. Pavitha and S. Sugave, "Optimizing machine learning models: An adaptive hyperparameter tuning approach," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, pp. 344–354, 2023.
[33] Ł. Borchmann *et al.*, "Robustness of document AI systems," *Proc. ICDAR*, 2021.