BTP Poster - Named Entity Recognition for Cooking Instructions

Winter Semester 2025 - Research Track

Anant Kaushal (2022067), Anikait Agrawal (2022072), Vatsal Gupta (2022564)
Advisor : Dr. Ganesh Bhimanna Bagler

Introduction

The goal of the research is to develop a domain-adapted Named Entity Recognition (NER) system that can accurately extract structured information from unstructured culinary text, enabling intelligent recipe understanding and downstream food-tech applications. The work can be categorised into,

- **Models used:** To build an accurate domain-specific NER system for culinary text, multiple transformer-based models were fine-tuned and evaluated. These include **DeBERTa**, **RoBERTa**, **DistilRoBERTa**, and **spaCy's transformer pipeline**. All models were trained using the **BIO tagging scheme**, which enables precise span boundary recognition—essential for handling multi-word entities such as ingredient names.
- **Data Augmentation and Analysis:** To improve model robustness and generalization, **data augmentation** techniques were employed, such as **entity replacement** and random oversampling, tailored to ingredient-centric contexts. A thorough **data analysis** was also conducted to examine entity distribution and label frequency, helping identify imbalances and improve annotation consistency. These strategies led to noticeable improvements in performance across all models, particularly in handling noisy and varied recipe text.

The main objective was to design a robust NER model tailored to the culinary domain, capable of identifying key entities such as **ingredients**, **quantities**, **units**, **sizes**, and **dry/fresh state** within informal cooking instructions.

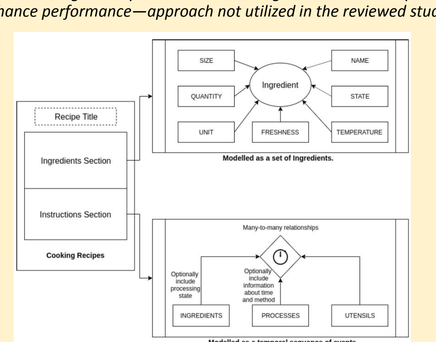
Literary Review

Our exploration of recent studies and research papers are as follows:

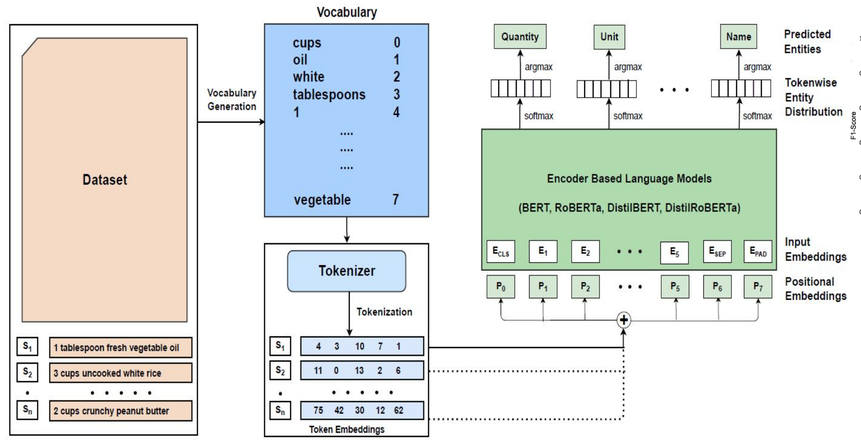
- **Paper -1 : A NER-based Approach on Model Recipes**
  This paper presents a hybrid Named Entity Recognition (NER) approach tailored for culinary recipe data, where language is informal and entities like ingredients, quantities, and tools are embedded in unstructured instructions. The proposed system combines a BiLSTM-CRF model with linguistic features such as POS tags, and chunking information to enhance recognition accuracy. Tested on a domain-specific recipe dataset, the model achieved F1-scores ranging from 82% to 88% across different entity types, showing that domain-aware feature engineering significantly boosts performance, especially in noisy text environments.

- **Paper - 2 : A Deep Learning-based Approach for Entity Extraction in Recipe Datasets**
  This work explores an end-to-end deep learning approach using transformer models like BERT and RoBERTa for entity recognition in cooking recipes. By fine-tuning these models on annotated culinary corpora, the system eliminates the need for handcrafted features while leveraging contextual embeddings to understand complex language patterns. The model demonstrated strong generalization on unseen recipes and achieved an impressive F1-score of 95.52%, outperforming hybrid baselines in both precision and recall, particularly in less frequent compound entities and noise.

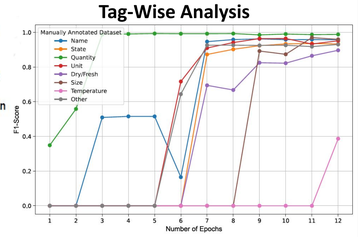
  Prior works focused on hybrid and transformer-based NER methods for recipes, we observed that zero-shot prompting with LLMs was ineffective. Instead, we adopted DeBERTa for improved contextual understanding and explored extra data augmentation techniques to enhance performance—approach not utilized in the reviewed studies.



Architecture and Pipeline



Tag-Wise Analysis



BIO-Tagging

Token	BIO Tag
2	B-QTY
green	B-NAME
onions	I-NAME
chopped	B-PREPROC
finely	I-PREPROC

Methodology

DATA AUGMENTATION

Initial exploration of the Phrases dataset revealed severe class imbalances: common units (e.g., cup, teaspoon) and frequent ingredients (garlic, onion) dominated, while many form and state labels appeared fewer than five times . To address this, various augmentation strategies were evaluated:

- **SMOTE**: Discarded due to its focus on numeric features, unsuitable for categorical text.
- **Lexical Token Window Replacement (LTWR) and Synonym Replacement (SR)**: Introduced diversity but sometimes broke annotation spans.
- **Shuffle within Segments (SiS)**: Maintained label boundaries while varying token order.

Augmented Dataset:

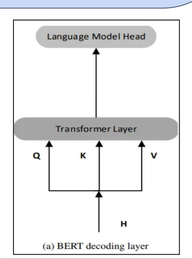
- Ultimately, **Random Oversampling** proved most effective. Underrepresented classes—Unit, Ingredient Name, Form, State—were duplicated up to mean-frequency thresholds (≈195 for Unit, ≈9 for Name, ≈11 for Form, ≈14 for State), with no deletions of original samples. This ensured each label appeared uniformly during training, preventing the model from defaulting to the majority "O" tag .
- **BIO tagging (Begin–Inside–Outside)** was applied to every token to capture precise entity boundaries and reduce fragmentation—essential in cooking texts where multi-word entities are common. For example, "2 green onions chopped finely" is tagged as B-QTY, B-NAME, I-NAME, B-PREPROC, I-PREPROC . This labeling not only improved span consistency but also enabled advanced span-based evaluation and post-processing validation
- **Random oversampling** and **BIO labeling** together bolstered model generalization, particularly for low-frequency entities, yielding consistent F1-score improvements across all architectures, and best suited for Named Entity Recognition purposes.

MODELS

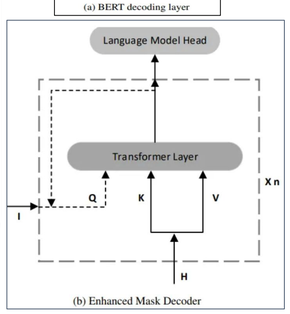
Fine-tuning was performed on three annotated Excel-derived datasets (Phrases, Augmented, and BIO_Tagged), ensuring consistency of span boundaries by masking special tokens and optimizing with AdamW under cross-entropy loss. We implemented four models and architectures of them are as follows:

1. **RoBERTa-base**: A 12-layer Transformer model with 768 hidden units and dynamic masking. Its WordPiece tokenizer cleanly splits multi-word ingredients, and fine-tuning yields high precision and recall on span boundaries .
2. **spaCy**: Two pipelines were created. First, a lightweight CNN-based tagger trained for ten epochs on character-span annotations for labels like QTY, UNIT, ING, etc. Second, a transformer-backed NER using Hugging Face's AutoModelForTokenClassification instructions are tokenized with RoBERTa's tokenizer, aligned to BIO labels, and fine-tuned under cross-entropy. This hybrid approach balances millisecond-scale inference with high accuracy, particularly on rare labels .
3. **DistilBERT**: A six-layer distilled BERT variant retaining ≈95 % of performance. Data is tokenized with DistilBertTokenizerFast, and fine-tuned via the Trainer API under the BIO scheme. DistilBERT matched RoBERTa's results with substantially lower latency for both training and testing.
4. **DeBERTa**: Enhances BERT by disentangling content and position embeddings and introducing an improved mask decoder. After tokenizing with DebertaTokenizerFast, tokens are aligned to BIO labels and fine-tuned under AdamW. DeBERTa's architecture excels at disambiguating multi-word phrases (e.g., "extra-virgin olive oil") and achieved the highest precision on complex ingredient spans.

- To conclude, **RoBERTa** uses the classic BERT decoding layer: simple masked LM with self-attention on a single hidden state. **DeBERTa** enhances decoding by disentangling the source of queries (input token) and keys/values (context), as shown in the Enhanced Mask Decoder diagram. In NER tasks, this leads to better contextualization and sharper token boundary understanding, improving performance



(a) BERT decoding layer



(b) Enhanced Mask Decoder

Results

- Performance Comparison of NER Models on Different Datasets:

Modelling Technique	Phrases_Sort			Augmented			BIO_Without_LLM		
	F1 (%)	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)	P (%)	R (%)
DeBERTa	**95.48**	**95.32**	**95.44**	**97.65**	**97.67**	**97.68**	**92.55**	**92.58**	**92.55**
spaCy-tranformer	92.30	91.40	93.90	97.39	97.41	97.37	91.68	91.46	91.50
spaCy-inbuilt pipelined	94.44	94.42	94.44	95.14	98.92	92.30	89.50	92.36	88.24
DistilRoBERTa	94.59	94.64	94.59	96.80	96.78	96.77	91.70	91.68	91.68
RoBERTa	94.56	94.60	94.65	97.26	97.24	97.25	91.60	91.64	91.64

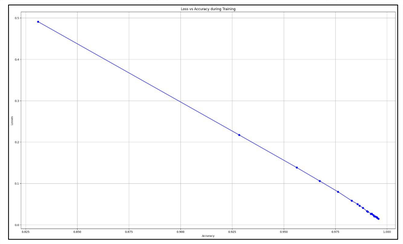


Figure: The Loss vs Accuracy curve for DeBERTa model

Key Observations:

- Achieved a new best accuracy of **97.65%** on augmented data using the **DeBERTa model**, improving over the previous state-of-the-art (95.6%) by **2%**.
- **DeBERTa** consistently outperformed all models across datasets, owing to its disentangled attention mechanism.
- **Augmented data** led to noticeable performance improvements for all models, validating the benefit of data diversity.
- **BIO-labeling** enabled more accurate span boundary detection by marking beginnings and continuations of entity spans—crucial for multi-word ingredients.
- The current ingredient **NER pipeline** provides a strong foundation for future work on identifying actions, tools, durations, and temperatures in **cooking instructions**.

Conclusion and Future Work

Our experiments demonstrate that **domain-specific fine-tuning**, combined with **careful data curation**, enables the development of highly accurate NER systems for ingredient extraction. Among the models evaluated, **DeBERTa** consistently delivered the best performance across all datasets, showcasing the strength of its **disentangled attention mechanism** in capturing fine-grained distinctions within culinary phrases.

DistilRoBERTa, a lighter architecture, and even **RoBERTa** offered a strong **trade-off between speed and accuracy**, making it a viable option for low-resource environments. Additionally, **spaCy's transformer pipeline** served as a reliable and efficient baseline. Despite a slight impact on raw accuracy, **BIO-style labeling** proved effective for **precise span boundary detection**, which is critical for downstream applications such as **recipe parsing** and **nutritional estimation**.

Overall, our findings emphasize that NER performance depends not only on model size but also on **data quality**, **label representation**, and **domain alignment**.

Future Work: We plan to extend our NER pipeline to cover **cooking instructions**, which often contain **complex temporal and procedural expressions**. This expansion will build upon our current insights and leverage techniques from prior studies, including **Goel et al. (2024)**, to enable entity recognition from **full recipe texts**, not just ingredient lists.

References

[1] Diwan, N., Batra, D., and Bagler, G. *A named entity based approach to model recipes.* In *2020 IEEE 36th International Conference on Data Engineering Workshops (ICDEW)* (2020), IEEE, pp. 88–93.

[2] Goel, M., Agarwal, A., Agrawal, S., Kapuriya, J., Konam, A. V., Gupta, R., Rastogi, S., Bagler, G., et al. *Deep learning based named entity recognition models for recipes. arXiv preprint* arXiv:2402.17447 (2024).