ExtRA++: A Conceptual Architecture for a Deep Learning System for Aspect-Based Sentiment Analysis in User Reviews

G. Kanev*, I. Valova
Department of Computer Systems and Technologies, University of Ruse "Angel Kanchev", Ruse, Bulgaria

***Abstract*—Aspect-Based Sentiment Analysis (ABSA) aims to identify opinion targets within textual reviews and determine the sentiment polarity associated with each target. Although transformer-based models have significantly improved contextual understanding in sentiment analysis, they remain limited in explicitly modeling structured knowledge and token-level dependencies. This study presents ExtRA++ (Enhanced Extractive Review Analysis), a conceptual deep learning architecture for fine-grained aspect-based sentiment analysis in user-generated reviews. The proposed framework integrates four complementary components: BERT-based contextual semantic modeling, adaptive external knowledge integration through Wikidata embeddings, graph-based structural reasoning using Graph Attention Networks (GATs), and sequence-consistent aspect extraction through Conditional Random Fields (CRFs) combined with aspect-aware sentiment classification. Unlike transformer-only approaches, ExtRA++ is designed as a modular systems-level architecture that combines contextual semantics, factual grounding, structural token interactions, and structured decoding within a unified framework.**

Keywords—Aspect-based sentiment analysis; aspect extraction; deep learning; graph attention networks; knowledge graph integration; BERT; conditional random fields; natural language processing; opinion mining; transfer learning

I. INTRODUCTION

The rapid growth of user-generated content has intensified the need for advanced methods capable of extracting fine-grained opinions from textual reviews [1]. Beyond coarse document-level sentiment classification, modern applications increasingly require aspect-based sentiment analysis (ABSA)[2], where opinions are identified with respect to specific product or service attributes and characteristics. This level of granularity is essential for actionable insights in domains such as e-commerce, hospitality, and customer experience analytics.

Recent advances in transformer-based language models, particularly BERT and its variants, have significantly improved the contextual understanding of textual data. Contextual embeddings generated by transformer architectures allow models to capture long-range semantic dependencies and substantially outperform static embedding approaches such as Word2Vec or GloVe in many natural language processing tasks. Consequently, transformer-based architectures have become dominant in contemporary ABSA research.

Despite these advances, existing transformer-based ABSA approaches exhibit several limitations. First, contextual language models rely primarily on statistical co-occurrence patterns learned during pretraining and do not explicitly integrate structured external knowledge [3]. This can limit the model's ability to reason about real-world entities and their semantic relations, especially in domain-specific reviews involving products, brands, or technical terminology. Second, although transformer self-attention captures global contextual information, it does not explicitly model structural relationships between tokens. In practice, sentiment expressions may depend on syntactic or local semantic structures that are difficult to represent solely through attention-based contextualization. Third, many sequence labeling approaches perform token-level classification independently, which may result in structurally inconsistent BIO predictions during aspect extraction.

Several research directions have attempted to address these limitations independently. Knowledge-enhanced methods incorporate information from external knowledge bases or domain ontologies to enrich semantic representations [4]. Graph-based approaches employ Graph Neural Networks (GNNs) or Graph Attention Networks (GATs) to model syntactic or semantic token dependencies. Meanwhile, structured prediction approaches such as Conditional Random Fields (CRFs) improve sequence consistency in aspect extraction tasks. However, most existing methods focus on individual improvements rather than providing an integrated framework that jointly combines contextual semantic understanding, external factual grounding, structural token interactions, and sequence-aware decoding within a unified architecture.

To address this gap, this study presents ExtRA++ (Enhanced Extractive Review Analysis), a modular deep learning architecture for aspect-based sentiment analysis in user reviews. The proposed system integrates four complementary processing stages: 1) contextual semantic encoding through BERT, 2) adaptive integration of external knowledge from Wikidata using entity embeddings and a learnable gating mechanism, 3) structural reasoning through a Graph Attention Network for modeling token-level dependencies, and 4) sequence-consistent aspect extraction using a Conditional Random Field decoder, followed by aspect-aware sentiment classification.

---

*Corresponding author.

Unlike transformer-only approaches or isolated graph-enhanced architectures, ExtRA++ is designed as a systems-level integrated framework in which complementary components address different limitations of aspect-level sentiment analysis. Contextual representations provide semantic understanding, external knowledge supplies factual grounding, graph attention captures structural relationships, and CRF decoding enforces consistency in BIO sequence prediction. This design enables richer and more robust aspect representations while maintaining a modular architecture suitable for extension and domain adaptation.

II. ARCHITECTURE DESCRIPTION

The proposed ExtRA++ framework is designed as a modular architecture that jointly performs aspect extraction and sentiment classification within a unified pipeline. The architecture integrates contextual semantic modeling, external knowledge enrichment, structural token reasoning, and sequence-aware decoding to address several limitations of existing transformer-based ABSA systems.

The architectural decisions are motivated by three key observations from the sentiment analysis literature review. First, pre-trained language models provide strong contextual representations but lack domain-specific and factual knowledge crucial for aspect identification [5]. Second, knowledge graph embeddings can supplement textual understanding with structured world knowledge, but naive integration often introduces noise [6]. Third, sequential models inadequately capture long-range syntactic dependencies that frequently connect aspects with their sentiment-bearing context [7].

As illustrated in Fig. 1, the proposed architecture consists of four sequential processing stages:

1) Contextual representation layer – contextual token embeddings are generated through a pre-trained BERT encoder.

2) External knowledge integration layer – semantic enrichment is performed through adaptive integration of Wikidata entity embeddings using a learnable gating mechanism.

3) Structural reasoning layer – Graph Attention Networks (GATs) model token-level structural dependencies.

4) Prediction layer – aspect terms are extracted through Conditional Random Fields (CRF), while sentiment polarity is predicted through an aspect-aware sentiment classification head.

Formally, given an input review represented as:

$$X = \{w_1, w_2, \ldots, w_n\} \tag{1}$$

The architecture predicts a set of aspect-level outputs:

$$O = \{(A_i, S_i, C_i)\}_{i=1}^{m} \tag{2}$$

Where:

- A_i denotes the extracted aspect term;
- S_i denotes the sentiment polarity;
- C_i represents a confidence score;
- m is the number of extracted aspects.

The overall processing pipeline can be summarized as:

$$H^{(enriched)} = f_{KG}(H^{(BERT)}, Q) \tag{3}$$

$$H^{(GAT)} = f_{GAT}(H^{(enriched)}, E) \tag{4}$$

$$H^{(combined)} = [H^{(enriched)} \parallel H^{(GAT)}] \tag{5}$$

$$Y^{(BIO)} = f_{CRF}(H^{(combined)}) \tag{6}$$

$$S = f_{sent}(H^{(combined)}, Y^{(BIO)}) \tag{7}$$

where,

Qdenotes Wikidata entity identifiers;

Erepresents graph edges between tokens;

$Y^{(BIO)}$is the predicted BIO tag sequence.

A key characteristic of ExtRA++ is its systems-level integration strategy, in which complementary components address different limitations of aspect-level sentiment analysis. BERT provides contextual semantic understanding, Wikidata introduces factual grounding, GAT captures local structural relationships, and CRF enforces valid BIO decoding. Rather than functioning as independent modules, these components operate within a shared representation space and jointly contribute to aspect-aware sentiment understanding.

B. Input Representation and Tokenization

The input to the proposed architecture consists of user-generated textual reviews represented as token sequences. Since neural language models operate on discrete token representations rather than raw text, the input sequence undergoes WordPiece tokenization [8], following the standard preprocessing strategy of BERT. WordPiece decomposes rare or previously unseen words into subword units, reducing vocabulary sparsity and improving generalization across domains.

Following tokenization, special BERT tokens are added to mark sequence boundaries:

$$X' = \{[CLS], t_1, t_2, \ldots, t_L, [SEP]\} \tag{8}$$

where:

- [CLS] serves as a global contextual representation of the input sequence;
- [SEP] denotes the end of the sequence.

Each token is subsequently mapped to its vocabulary identifier:

$$I = \{i_1, i_2, \ldots, i_L\} \tag{9}$$

along with an attention mask:

$$M = \{m_1, m_2, \ldots, m_L\} \tag{10}$$

used to distinguish valid tokens from padding positions.

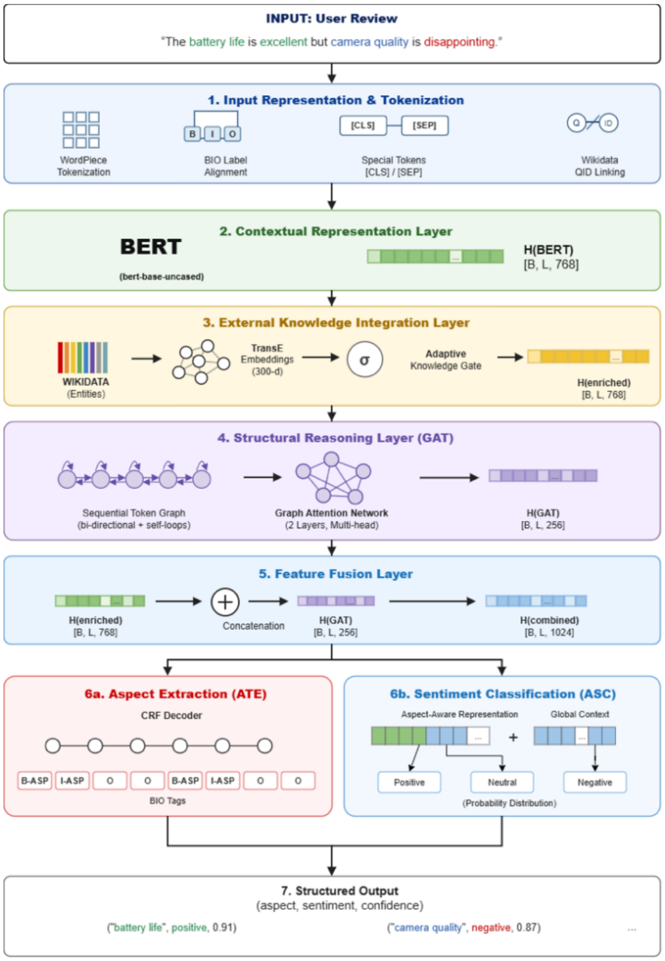

Fig. 1. Conceptual architecture extra++.

To maintain computational efficiency and ensure compatibility with the transformer encoder, all input sequences are padded or truncated to a maximum sequence length:

$$L_{max} = 128 \tag{11}$$

This value is selected empirically as a compromise between contextual coverage and computational efficiency, since the majority of user reviews in commonly used ABSA datasets remain substantially shorter than this threshold.

For aspect extraction, the model adopts the BIO annotation scheme, in which each token receives one of three labels:

$$y_i \in \{O, B-ASP, I-ASP\} \tag{12}$$

where:

- O denotes a non-aspect token;
- B-ASP indicates the beginning of an aspect span;
- I-ASP marks continuation tokens.

C. Contextual Representation Layer

Following tokenization, the structured input representation is processed through a contextual representation layer based on a pre-trained BERT (Bidirectional Encoder Representations from Transformers) model [9]. The objective of this stage is to transform discrete token identifiers into contextualized semantic representations that capture both local and long-range dependencies within the review text.

In the proposed architecture, the bert-base-uncased configuration is employed due to its favorable balance between representational power and computational efficiency. The model consists of 12 transformer layers, 12 self-attention heads, and a hidden dimensionality of:

$$d_h = 768 \tag{13}$$

resulting in approximately 110 million trainable parameters.

Each token is represented by a contextual embedding:

$$h_i^{BERT} \in \mathbb{R}^{768} \tag{14}$$

which incorporates semantic information from the entire input sequence.

Unlike static embedding methods such as Word2Vec [10] or GloVe[11], contextual embeddings generated by BERT vary depending on the surrounding context. Consequently, the same lexical unit may receive different vector representations depending on its semantic role within a sentence. This property is particularly important in Aspect-Based Sentiment Analysis, where identical terms may express different meanings depending on context and associated opinion targets.

The representational capability of BERT originates from the self-attention mechanism, which enables each token to dynamically attend to all other tokens in the sequence. For each attention head, token representations are projected into query, key, and value spaces:

$$Q = HW_Q \tag{15}$$

$$K = HW_K \tag{16}$$

$$V = HW_V \tag{17}$$

where:

- Q represents query vectors;
- K represents key vectors;
- V represents value vectors;
- W_Q, W_K, and W_V are trainable projection matrices.

Attention scores are computed using scaled dot-product attention:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{18}$$

This mechanism enables contextual interactions between aspect terms and sentiment expressions even when they are separated by long token distances. For example, in the sentence *"Despite the high price, the camera quality exceeded expectations."*, the sentiment expression *"exceeded expectations"* may still be associated with *"camera quality"* despite intervening contextual information.

To adapt the pre-trained language model to the ABSA task while reducing the risk of overfitting, a selective fine-tuning strategy is employed. Rather than updating all transformer parameters uniformly, lower BERT layers are partially frozen, while higher layers remain trainable:

- $L_1, \dots, L_8$ are frozen;
- $L_9, \dots, L_{12}$ are fine-tuned during training.

This design choice is motivated by observations that lower transformer layers predominantly capture general linguistic and syntactic patterns, while higher layers encode task-specific semantic information more relevant to aspect-level sentiment analysis. Partial freezing additionally reduces computational cost and mitigates catastrophic forgetting of pretrained knowledge.

To improve generalization and reduce overfitting, dropout regularization is applied to the contextual embeddings:

$$H' = Dropout(H^{BERT}, p) \tag{19}$$

where, $p = 0.1$. This value follows the standard BERT configuration and provides a balance between representational stability and regularization.

D. External Knowledge Integration Layer

Although transformer-based models provide strong contextual representations, they primarily rely on statistical language patterns acquired during pretraining and do not explicitly incorporate structured factual knowledge. In aspect-based sentiment analysis, understanding references to real-world entities, product components, or domain-specific terminology may require semantic grounding beyond contextual text representations alone.

To address this limitation, ExtRA++ incorporates an external knowledge integration layer that enriches contextual embeddings using semantic information from Wikidata. The purpose of this stage is to complement contextual language understanding with explicit entity-level knowledge while preserving the flexibility of transformer representations.

1) Entity linking and knowledge retrieval: An entity linking component maps textual mentions to their corresponding Wikidata identifiers (QIDs). This process involves:

- Mention detection – Identifying candidate spans in the text that may refer to entities (typically noun phrases).
- Candidate generation – Retrieving potential entity matches from the knowledge base using string matching and alias lookup.
- Entity disambiguation – Selecting the correct entity from candidates based on contextual coherence (e.g., "Apple" → Q312 [company] vs. Q89 [fruit] depending on context).

An example of entity linking of the review "*The iPhone camera takes stunning photos but battery drains fast*" is shown in Table I.

TABLE I. EXAMPLE OF ENTITY MAPPINGS

Mention	*WikiData QID*	*Entity Type*	*Key Properties*
iPhone	Q2766	Smartphone	manufacturer: Apple Inc., introduced: 2007
camera	Q15328	Optical instrument	subclass of: imaging device
battery	Q267298	Electrochemical cell	has use: power source
photos	Q125191	Photograph	created by: camera

Tokens without semantic matches retain null entity assignments and are processed using zero-valued knowledge representations.

Wikidata is selected due to its broad domain coverage, multilingual structure, explicit semantic relations, and continuous maintenance, making it particularly suitable for opinion mining tasks involving consumer products and technical entities.

2) Knowledge embedding representation: To represent external semantic information in vector form, the proposed architecture employs TransE knowledge graph embeddings.

TransE models relations through vector translations:

$$h + r \approx t \tag{20}$$

where:

- h is the head entity;
- r denotes the relation;
- t is the target entity.

3) Adaptive knowledge integration: Directly injecting external knowledge into language representations may introduce irrelevant information or semantic noise. Not all entities contribute equally to sentiment interpretation, and some knowledge may be weakly related to the contextual meaning of the review.

To regulate the influence of external information, ExtRA++ employs an adaptive gating mechanism that dynamically controls the contribution of knowledge embeddings for each token. The gate coefficient is computed as:

$$g_i = \sigma(W_g[h_i^{BERT} || k_i] + b_g) \tag{21}$$

where:

- W_g and b_g are trainable parameters;
- $\sigma(\cdot)$ denotes the sigmoid activation function.

This mechanism allows the model to selectively incorporate external knowledge:

- When semantic grounding is beneficial, external knowledge strongly contributes to representation learning.

$$g_i \to 1$$

- When irrelevant – the architecture relies primarily on contextual language information.

$$g_i \rightarrow 0$$

This selective integration mechanism was preferred over direct feature concatenation because it reduces semantic noise and enables context-aware use of external knowledge.

The resulting enriched representations:

$$H^{(enriched)} \in \mathbb{R}^{B \times L \times 768} \tag{22}$$

E. Structured Reasoning Layer

Although transformer encoders effectively model global contextual dependencies through self-attention, they do not explicitly enforce structural relationships between neighboring tokens or syntactic patterns within a sentence. In ABSA, sentiment expressions frequently depend on localized structural relations that may not be fully captured through contextual attention alone.

Consider the review: "Despite the high price, the camera quality exceeded my expectations." A purely sequential model must learn that "camera quality" (aspect) is associated with "exceeded expectations" (positive sentiment) despite the intervening clause introducing a potentially confusing negative element ("high price"). Graph-based representations can model the direct dependency between "quality" and "exceeded" while appropriately downweighting the misleading proximity to "high price."

To improve the modeling of token-level interactions, ExtRA++ incorporates a Graph Attention Network (GAT) layer after contextual encoding and knowledge integration [12]. The primary role of this component is to introduce an explicit structural reasoning mechanism that captures local dependencies between tokens and complements the global semantic understanding provided by BERT.

The enriched token representations generated by the previous layer are transformed into a graph structure, where each token corresponds to a node and neighboring tokens are connected through bidirectional edges. Unlike dependency-parsing approaches, which rely on external syntactic parsers, the proposed architecture employs a sequential token graph in which tokens are connected to their immediate neighbors. This design was selected intentionally for two reasons. First, dependency parsers may introduce noise or unstable syntactic structures when processing informal user-generated reviews. Second, sequential graphs provide a lightweight and domain-independent alternative while preserving architectural modularity for future extensions.

Within this graph representation, the GAT layer learns how strongly neighboring tokens should influence each other. Rather than treating all neighboring words equally, the model dynamically assigns attention weights that reflect the contextual importance of individual token interactions. As a result, sentiment-bearing expressions that are more relevant to a specific aspect receive stronger influence during representation learning.

To improve representational robustness, the architecture employs multi-head graph attention, allowing the model to learn different structural interaction patterns simultaneously. This enables the network to capture multiple complementary perspectives of token relationships, including local contextual proximity and semantic interactions.

The output of the graph reasoning stage is transformed into a compact structural representation that summarizes token-level dependencies while preserving contextual information from previous layers. In the proposed implementation, graph representations are compressed into a lower-dimensional feature space before being passed to the prediction stage, reducing computational complexity while maintaining discriminative structural information.

The inclusion of a GAT layer addresses an important limitation of transformer-only ABSA systems. While BERT provides strong contextual semantic representations, graph attention introduces an additional mechanism for explicitly modeling local structural interactions between aspects and opinion expressions. Consequently, ExtRA++ combines global contextual understanding with localized structural reasoning, improving the interpretability and robustness of aspect-level sentiment modeling.

F. Feature Fusion Layer and Prediction Layer

After contextual semantic modeling and structural graph reasoning, the proposed architecture produces two complementary representations of the input text. The first representation captures global contextual semantics, enriched through BERT and external knowledge integration, while the second representation emphasizes local structural dependencies obtained through graph attention.

Because these representations encode different but complementary information, ExtRA++ integrates them into a shared feature space before prediction. Rather than forcing an early decision about which representation is more important, the architecture preserves both information streams and allows the prediction layers to learn how to combine them during training.

To achieve this, the semantic and graph-based representations are merged through feature concatenation, forming a unified representation for each token. This strategy was selected instead of direct summation or gated fusion because it preserves the full informational content of both components without imposing prior assumptions about their relative importance.

Concatenation preserves the full information content from both sources, delegating the optimal weighting to subsequent layers. The projection layers in the prediction heads learn task-specific combinations of BERT and GAT features, effectively implementing a learned linear fusion.

The resulting fused representation serves as a shared input to two prediction branches operating in a multi-task learning setting:

- Aspect Term Extraction (ATE) – responsible for identifying aspect spans within the review.

- Aspect Sentiment Classification (ASC) – responsible for determining the sentiment polarity associated with each extracted aspect.

1) Aspect extraction head: Aspect extraction is formulated as sequence labeling using the BIO (Beginning-Inside-Outside) tagging scheme, a standard approach for span identification in NLP, Table II.

TABLE II. AN EXAMPLE OF SEQUENCE LABELING

Tag	*Numeric Code*	*Description*	*Example Tokens*
O	0	Outside any aspect span	"the", "takes", "stunning", "but"
B-ASP	1	Beginning of an aspect span	"camera" (in "camera quality")
I-ASP	2	Continuation of an aspect span	"quality" (in "camera quality")

2) Aspect-aware sentiment classification head: Once aspect spans are identified, the second prediction branch determines the sentiment polarity associated with each aspect. Instead of relying solely on the aspect phrase itself, the sentiment classifier considers both:

- the local aspect representation, describing the extracted opinion target;
- the global sentence context, summarizing the overall meaning of the review.

This design choice is important because sentiment often depends on contextual interpretation. The same aspect may receive different polarity depending on surrounding expressions, negation, or contrastive statements.

Overall, the prediction layer transforms enriched semantic and structural token representations into interpretable aspect-level outputs. By combining CRF-based structured decoding with aspect-aware sentiment classification, ExtRA++ jointly addresses both aspect extraction and sentiment prediction within a unified framework.

G. Output Layer

The final stage of ExtRA++ transforms intermediate prediction outputs into a structured and interpretable aspect-level representation. While the prediction layer generates BIO tags for aspect extraction and probability distributions for sentiment classification, the output layer organizes these predictions into semantically meaningful aspect–sentiment pairs.

Following CRF decoding, the predicted BIO sequence is converted into explicit aspect spans. Consecutive tokens labeled as B-ASP and I-ASP are grouped to form complete aspect terms, while tokens labeled O are ignored.

Once aspect spans are identified, each extracted aspect is associated with a sentiment polarity predicted by the sentiment classification branch. The model assigns one of three sentiment categories:

- positive;
- neutral;
- negative.

In addition to categorical sentiment labels, ExtRA++ provides a confidence score derived from the predicted probability distribution. This value indicates the model's confidence in a given sentiment prediction and may be used to filter uncertain outputs or support downstream decision-making.

Consequently, the output of the proposed system is represented as a collection of structured aspect-level tuples:

$$(aspect, sentiment, confidence)$$

For example, the previous review may produce:

$$[(\text{"}battery\ life\text{"}, positive, 0.91),$$

$$(\text{"}camera\ quality\text{"}, negative, 0.87)]$$

This structured representation enables significantly more fine-grained sentiment interpretation than traditional document-level sentiment analysis. Instead of assigning a single polarity label to an entire review, ExtRA++ identifies individual opinion targets and their associated sentiment, making the architecture particularly suitable for applications such as customer feedback analytics, product monitoring, recommender systems, and explainable opinion mining.

The modular output design also facilitates integration with downstream analytical systems, allowing extracted aspect–sentiment information to be incorporated into dashboards, business intelligence pipelines, or knowledge-driven recommendation frameworks.

III. CONCEPTUAL ADVANTAGES AND FUTURE VALIDATION

The proposed ExtRA++ architecture is designed to address several limitations of existing Aspect-Based Sentiment Analysis approaches through the integration of complementary processing mechanisms. Rather than relying exclusively on contextual language modeling, the framework combines transformer-based semantic representations, external knowledge grounding, structural token reasoning, and sequence-consistent decoding within a unified architecture.

From a conceptual perspective, the proposed design introduces several advantages. First, contextual embeddings generated by BERT provide robust semantic understanding and long-range contextual modeling. Second, the integration of Wikidata entity representations enables explicit incorporation of factual knowledge, potentially improving semantic grounding for domain-specific entities and technical concepts. Third, the Graph Attention Network layer introduces localized structural reasoning, complementing transformer attention by explicitly modeling token-level interactions. Finally, Conditional Random Fields improve consistency in BIO sequence prediction during aspect extraction.

An important characteristic of ExtRA++ is its modular design. Individual architectural components may be independently replaced or extended without requiring redesign of the overall framework. For example, alternative transformer encoders such as RoBERTa or DeBERTa may replace BERT, dependency-based graph construction strategies may substitute sequential token graphs, and alternative knowledge bases may be incorporated depending on the target domain. This flexibility supports future adaptation to specialized application areas and multilingual environments.

The present work focuses on the conceptual formulation of the proposed architecture and its systems-level integration strategy. Accordingly, the objective of this study is not to provide empirical benchmarking, but rather to establish the theoretical and architectural foundation of the proposed framework.

Future work will include comprehensive experimental validation using benchmark ABSA datasets, including SemEval collections, comparative evaluation against state-of-the-art baselines, ablation studies assessing the contribution of individual architectural components, and statistical analysis of model robustness and generalization performance.

IV. CONCLUSION

This study presents ExtRA++ (Enhanced Extractive Review Analysis), a conceptual deep learning architecture for Aspect-Based Sentiment Analysis (ABSA) in user-generated textual reviews. The proposed framework was designed to address several limitations of existing transformer-based sentiment analysis systems by integrating complementary processing mechanisms within a unified architecture.

The conceptual design combines contextual semantic modeling through BERT, external semantic enrichment using Wikidata knowledge embeddings, graph-based structural reasoning through Graph Attention Networks, and sequence-consistent decoding via Conditional Random Fields. By integrating these components into a shared representational framework, ExtRA++ aims to support more robust aspect-level sentiment understanding while maintaining modularity and extensibility.

Unlike approaches that focus on isolated architectural improvements, the proposed system emphasizes a systems-level integration strategy, in which contextual semantics, factual grounding, structural token dependencies, and structured sequence prediction complement one another. This design is intended to provide a flexible foundation for fine-grained opinion mining in complex review environments.

An additional contribution of the proposed framework lies in its modular organization, which facilitates future adaptation and extension. Individual components may be independently modified or replaced depending on task requirements, domain characteristics, or advances in neural language modeling and graph representation learning.

The present work focuses on the conceptual formulation and architectural design of ExtRA++. Consequently, no empirical benchmarking or experimental validation is reported in this study. Instead, the objective is to establish the theoretical and systems-level foundation of the proposed architecture.

Future research will focus on experimental evaluation and comparative analysis using benchmark ABSA datasets, including SemEval collections. Planned investigations will include comparisons against state-of-the-art transformer-based and graph-enhanced baselines, ablation studies examining the contribution of individual architectural components, and evaluation of generalization capabilities across multiple domains.

REFERENCES

[1] M. Hu and B. Liu, “Mining and summarizing customer reviews,” in Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 2004, pp. 168–177.

[2] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, “A survey on aspect-based sentiment analysis: Tasks, methods, and challenges,” IEEE Trans. Knowl. Data Eng., vol. 35, no. 11, pp. 11019–11038, 2022.

[3] F. Petroni et al., “Language models as knowledge bases?,” in Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), 2019, pp. 2463–2473.

[4] Z. Zhang, X. Han, Z. Liu, X. Jiang, M. Sun, and Q. Liu, “ERNIE: Enhanced language representation with informative entities,” arXiv Prepr. arXiv1905.07129, 2019.

[5] H. Xu, B. Liu, L. Shu, and P. S. Yu, “BERT post-training for review reading comprehension and aspect-based sentiment analysis,” arXiv Prepr. arXiv1904.02232, 2019.

[6] C. Zhang, Q. Li, and D. Song, “Aspect-based sentiment classification with aspect-specific graph convolutional networks,” arXiv Prepr. arXiv1909.03477, 2019.

[7] K. Sun, R. Zhang, S. Mensah, Y. Mao, and X. Liu, “Aspect-level sentiment analysis via convolution over dependency tree,” in Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), 2019, pp. 5679–5688.

[8] M. Schuster and K. Nakajima, “Japanese and korean voice search,” in 2012 IEEE international conference on acoustics, speech and signal processing (ICASSP), 2012, pp. 5149–5152.

[9] N. M. Gardazi, A. Daud, M. K. Malik, A. Bukhari, T. Alsahfi, and B. Alshemaimri, “BERT applications in natural language processing: a review,” Artif. Intell. Rev., vol. 58, no. 6, pp. 1–49, 2025.

[10] K. W. Church, “Word2Vec,” Nat. Lang. Eng., vol. 23, no. 1, pp. 155–162, 2017.

[11] J. Pennington, R. Socher, and C. D. Manning, “Glove: Global vectors for word representation,” in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543.

[12] P. Velickovic et al., “Graph attention networks,” Stat, vol. 1050, no. 20, pp. 10–48550, 2017.