A Comparative Study of ChatGPT-based and Hybrid Parser-based Sentence Parsing Methods for Semantic Graph-based Induction

Walelign Tewabe[1], László Kovács[2]
Institute of Technology, University of Miskolc, Hungary[1,2]
Institute of Technology, Debre Markos University, Ethiopia[1]

***Abstract*—Sentence parsing is a fundamental step in the conversion of a text document into semantic graphs. In this research, novel phrase parsing techniques for semantic graph-based induction are presented, namely the ChatGPT-based and Hybrid Parser-based approaches. The performance of these two approaches in the context of inducing semantic networks from textual data is assessed through a comprehensive analysis in this study. The primary purpose is to enhance the construction of semantic graphs, specifically focusing on capturing detailed event descriptions and relationships within text. The research finds that the Hybrid Parser-Based approach exhibits a slight advantage in accuracy (acc_hybrid = 0.87) compared to ChatGPT (acc_GPT = 0.85) in sentence parsing tasks. Furthermore, the efficiency analysis reveals that ChatGPT's response quality varies with different prompt sizes, while the Hybrid Parser-Based method consistently maintains an "excellent" response quality rating.**

Keywords*—Adverb prediction; ChatGPT; hybrid parser-based; natural language processing; sentence parsing; semantic graph induction

I. Introduction

Semantic Graph Induction is a computational approach in Natural Language Processing (NLP) and artificial intelligence that aims to extract and represent structured knowledge and semantic relationships from unstructured textual data. Semantic Graph visually represents the semantic structure of a document extracted from sentences [1]. Semantic graphs play a multifaceted role in various applications, spanning information retrieval, knowledge representation, question answering, text summarization, document clustering, and classification.

Furthermore, in the finance industry, semantic graphs have emerged as a crucial tool for managing financial knowledge securely [11], enabling applications like transaction surveillance, financial crime detection and prevention, and non-compliant user detection [12]. In the entertainment industry, particularly social media, knowledge graphs power social graphs that help platforms like Facebook connect users within the context of their relationships, while also enhancing recommender systems to offer personalized content recommendations based on user interests [13]. Moreover, semantic graphs play a vital role in cybersecurity by mapping historical cyber attacks and predicting potential future breaches, thus bolstering cyber defense strategies [14].

This study explores the creation of semantic graphs, which are visual representations of knowledge and the interconnections between concepts. Specific tools within the domain of NLP parsing are working for constructing these semantic graphs. However, there are limitations in their ability to present detailed event descriptions, particularly concerning time and place. Recognizing the limitations present in current NLP parsing tools, the primary objective of this research is to enhance the existing approach. To address these limitations, this paper introduces a solution that involves identifying all functional components, including Subject, Predicate, Direct Object, Indirect Object, and Conjunction. Simultaneously, the method explores the prediction of adverb types, encompassing Time, Place, Manner, Degree, and Frequency, thus enriching the depth of linguistic analysis.

To gain a deeper understanding of knowledge, concepts, and the complex web of relationships between them, this research extends beyond traditional limitations by incorporating a more comprehensive set of components. Specifically, the study introduces novel ChatGPT-based and Hybrid Parser-based Semantic Graph Construction and conducts a comparative analysis. This analysis assesses the details of these two approaches, dissecting their respective strengths, weaknesses, and applications.

In this regard, ChatGPT is one of the state-of-the-art Large Language Models (LLMs) [15], that has emerged as a transformative force in the field of NLP. It plays a pivotal role in the construction of semantic graphs by leveraging their natural language understanding capabilities. These models are trained on extensive text corpora and can extract and encode intricate relationships between concepts and entities within textual data. ChatGPT's previous experiences with these tasks are informed by its extensive pre-training on a diverse range of internet text [16]. This pre-training allows it to understand and generate human-like text and perform tasks related to semantic graph construction with high accuracy. By leveraging this understanding, ChatGPT can contribute significantly to the creation and enrichment of semantic graphs across various domains, from healthcare [10] and finance to information retrieval and content recommendation [17]. It has demonstrated remarkable skill in a wide array of language understanding tasks, including question-answering, language generation, and text summarization [18]. However, the question arises: can ChatGPT be effectively harnessed to tackle the difficulties of semantic graph-based induction? On the other hand, Hybrid Parser-based methods integrate multiple NLP components, combining rule-based and machine-learning techniques, to extract and represent semantic relationships from text. The marriage of these disparate approaches promises enhanced robustness and adaptability. This study sets out to investigate which of these approaches outshines in the domain of semantic graph construction, and whether a hybrid approach provides a balanced solution. The contributions of this work are Semantic Graph Construction Enhancement, Testing a novel ChatGPT-based parsing for functional sentence parsing, and Comparative Analysis of Methodologies.

The paper is structured as follows: In the second section, a semantic graph construction model is presented, and a detailed procedure for building the presented model is provided. We discuss the latest NLP background technology and results. Additionally, we explore different knowledge base resources and their applications. The third section describes the proposed Hybrid Parser-based method, explaining all process steps. In the fourth section, we describe the ChatGPT-based method, encompassing the environment, dataset size, benchmark, and evaluation methods. Next, we present the experimental results, analyze the evaluation findings from multiple perspectives, and demonstrate the potential applications of our approach. In the sixth section, we conduct an efficiency analysis and engage in a discussion. Finally, in the concluding section, we summarize our findings and offer suggestions for future research directions.

II. Literature Review

A. Basics of Semantic Graph

A semantic graph is a graph model where nodes represent concepts and edges (or arcs) represent relationships between those concepts [19]. This model type is often used in artificial intelligence applications for representing knowledge.

B. Definition 2.1

A graph $G = (V, E)$ is defined by a set of nodes V and a set of edges E between these nodes. Let $E \subseteq V \times V$ represent directed edges or arcs [20]. Each directed edge $(u, v) \in E$ signifies a connection from a start (tail) vertex u to an end (head) vertex v, where u and v are elements of the node set V. The graph's structure is characterized by these directed connections, providing a representation of relationships between nodes. Each node is associated with a label $Label(v)$.

Building semantic graphs is essential for many practical uses and ongoing research [8], [21], [22]. As we have more and more data available, creating these meaningful graphs becomes increasingly important for learning from different sources. Scientists keep looking for new ways to make this field better, and they use it in things like understanding language, organizing knowledge, and using artificial intelligence. They make structured graphs and networks to show how words, ideas, and things are connected. These graphs help in finding information, answering questions, and suggesting things you might like. So, making these graphs is a big part of helping computers and people work together better. When texts are represented graphically, it allows the preservation of additional information like the text's inner structures, semantic relationships, and term order. However, events like these are not effectively captured using current NLP parsing and semantic graph construction. As an illustration, Fig. 1 provides a visual insight into a fundamental event knowledge graph centered around the concept of "eating" [9].

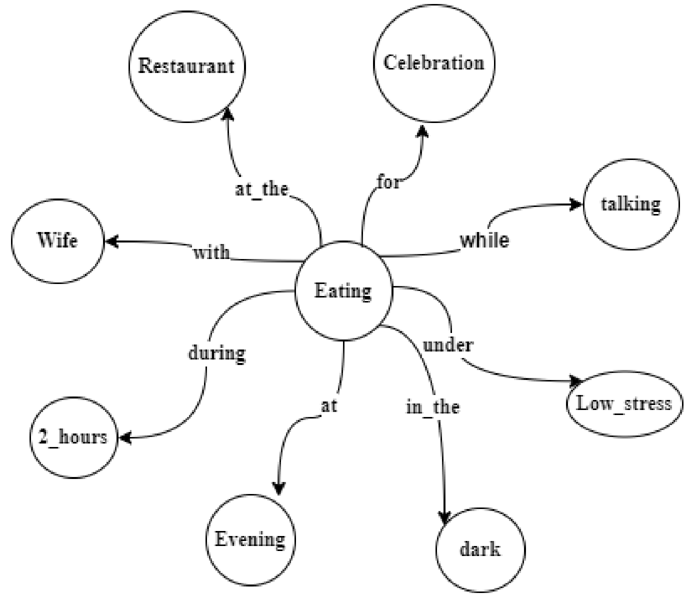

Fig. 1. A visualization of the basic event knowledge graph for eating [9].

Understanding natural language is a big challenge, and that's where semantic graphs come into play. Enhancing our grasp of natural language relies heavily on the development of semantic graphs, a field that's been increasingly in the spotlight. Researchers are actively exploring the creation of these graphs and how they can represent knowledge, diving into structured data, relationships, and more detailed elements, which align with prior work on Semantic Role Labeling (SRL) and adverb sense disambiguation. These efforts aim to provide a more comprehensive understanding of semantic parsing, event descriptions, and the complexities involved, as outlined in related works [23].

Knowledge graphs have also got substantial attention in recent years, serving as vital tools for organizing and connecting vast amounts of information from diverse sources, including text corpora, databases, and the web [24]. Some well-known knowledge graphs, such as DBpedia, Freebase, and Wikidata, have been crucial in this effort. We're also using some smart techniques like word embeddings and word vector representations to make semantic graphs even better [25].

Resource Description Framework (RDF) and ontologies are the foundation for constructing structured, machine-readable semantic graphs, playing a pivotal role in knowledge representation and the advancement of the semantic web. RDF, with its subject-predicate-object triples and Uniform Resource Identifiers (URIs), ensures global consistency and interoperability. Ontologies, including OWL and RDFS, enrich RDF's capabilities by defining the vocabulary and structure for resources and relationships within specific domains, making it easier to understand and work with the information [36]. Together, RDF and ontologies are super important for making and using semantic graphs across different fields.

At the same time, the Semantic Web initiative is pushing for structured data to be shared and linked on the web. They're using things like Linked Data, RDF, and SPARQL Protocol and RDF Query Language (SPARQL) queries to create big semantic graphs that cover a lot of the web[37]. But there are challenges too. We need better ways to handle big sets of data, put together text and visual data, and make sure the knowledge graphs we create are complete and correct. Researchers are used new techniques, like word embeddings and entity embeddings, to help to understand the fine details of how words and things are related [7]. As we have more and more data, making meaningful semantic graphs becomes super important for getting useful information from different places.

In general, the fields of RDF, ontologies, and the ideas behind the Semantic Web initiative where semantic graph play an important role that understand and manage information. The semantic graphs serve as a crucial foundation for knowledge representation and data integration, facilitating the consistence management of structured data on the web. However, this field is evolving, with ongoing efforts focused on improving graph construction techniques, addressing data handling challenges, and harnessing the power of embedding techniques to capture richer semantic relationships. As the landscape of available data continues to expand, the construction of semantic graphs becomes essential for unlocking valuable insights and enabling data-driven applications across various domains.

III. ChatGPT-based Sentence Parsing

A significant aspect of language models is the LLM, recognized for its capacity to achieve a wide-ranging understanding of language and proficiently generate text. LLMs acquire this capability through an extensive training process where they learn from vast amounts of data, effectively processing billions of parameters. This training demands substantial computational resources [28]. These language models primarily employ artificial neural networks, predominantly relying on transformer architectures, and undergo (pre-)training utilizing self-supervised and semi-supervised learning approaches [29].

Functioning as autoregressive language models, LLMs operate by taking an input text and iteratively predicting subsequent tokens or words [30]. Until the year 2020, the primary approach to adapt these models for specific tasks was fine-tuning. However, with the emergence of larger models like GPT-3, they can now be engineered with prompts to achieve similar outcomes [31]. LLMs are believed to acquire an inherent understanding of syntax, semantics, and the "ontology" within human language corpora [32].

Prominent examples of LLMs include OpenAI's GPT models like Generative Pre-trained Transformer (GPT)-3.5 and GPT-4, Google's Pathways Language Model (PaLM) employed in Bard, Meta's Language Model for Language Modeling (LLaMa), as well as BigScience Large Open-science Open-access Multilingual Language Model (BLOOM), Ernie 3.0 Titan, and Anthropic's Claude 2. In this study, due to the model's capabilities, researchers utilized the ChatGPT 3.5 OpenAI API for the sentence parsing.

A. Basics of GPT-based Models

Chat GPT (Generative Pre-trained Transformer) models are designed to understand and generate human-like text by processing vast amounts of data during training [34]. They operate by predicting the next word in a sequence of words and have been instrumental in various NLP tasks. Understanding these fundamental concepts is essential for harnessing the power of GPT-based models in language-related applications. The accuracy of the ChatGPT 3.5 model heavily relies on the quality and representativeness of the labeled dataset used for fine-tuning [35]. The pre-trained ChatGPT model is fine-tuned on a labeled dataset of adverbs to improve its categorization accuracy.

B. The Architecture of ChatGPT

ChatGPT is based on the transformer architecture, that allows for parallel processing, which makes it well-suited for processing sequences of data such as text. ChatGPT uses the PyTorch library, an open-source machine learning library, for implementation. ChatGPT is made up of a series of layers, each of which performs a specific task.

C. Prompt Engineering Techniques

Prompt engineering is a crucial technique employed to guide the behavior of large-scale language models like ChatGPT [34]. By strategically constructing input prompts, researchers and developers aim to obtain more accurate and relevant responses from these models [33]. Several prompts engineering strategies, including prompt rewriting, contextual incorporation, explicit instructions, and templates, have been proposed to address control and responsiveness challenges, aligning the model's outputs with user targets and expectations. The careful design of prompts plays a pivotal role in influencing the quality and relevance of ChatGPT's responses, making it a valuable skill for those working with AI systems. For instance, in a real-world context, prompt engineering bears the potential to enhance the efficiency, accuracy, and effectiveness of healthcare delivery by guiding AI models to provide valuable insights and solutions. However, it's crucial to acknowledge the limitations and risks associated with AI, such as the model's inability to access real-time data or offer personalized medical advice. This necessitates verification by qualified professionals and raises concerns about privacy and data security. Despite these challenges, the significance of prompt engineering has seen exponential growth since the inception of ChatGPT, with ongoing research endeavors aimed at refining and expanding this critical skill, particularly within the medical field. In this specific study, researchers have developed and employed high-quality training sets as templates for prompts to augment the accuracy of responses.

D. Methodology

The methodology for this study involves the following steps.

1) Construction of a Labeled Dataset: A high-quality labeled dataset is carefully collected to fine-tune ChatGPT for sentence parsing by including the adverb type prediction. This dataset includes Subject, Predicate, Direct Object, Indirect Object, Conjunction, and adverb types such as Time, Place, Manner, Degree, and Frequency. The dataset is essential for training ChatGPT to categorize adverbs accurately and for sentence parsing.

```
# Prompt for generating a semantic graph description
prompt_template= """
Sentence 1: The coffee shop is always busy in the morning.
Parsing Answer 1: {'Predicate': is, 'Subject': The coffee shop,
      'Direct Object': [] , 'Indirect Object': [], 'Time': in the
      morning, 'Place': [], 'Manner': always busy, 'Frequency': [],
      'Degree': []}
Sentence 2: The train arrived at the station on time.
      Parsing Answer 2: {'Predicate': arrived, 'Subject': The train,
      'Direct Object': [] , 'Indirect Object': [], 'Time': on time,
      'Place': at the station, 'Manner': [], 'Frequency': [], 'Degree':
      []}
Sentence 3: Ethiopia defeated Italy at the Battle of Adwa.
      Parsing Answer 2: {'Predicate': defeated, 'Subject': Ethiopia,
      'Direct Object': Italy, 'Indirect Object': [], 'Time': [],
      'Place': at the Battle of Adwa, 'Manner': [], 'Frequency': [],
      'Degree': []} """

custom_prompt=""" Generate the Predicate, Subject, Direct Object,
Indirect Object, Time, Place, Manner, Frequency, Degree, Conjunction,
Clause parts of the following sentence:- Sentence: 'The child reads
the book carefully and attentively  at the library everyday. """

prompt = prompt_template + custom_prompt
```

Fig. 2. Sample prompt template [9].

2) Fine-tuning ChatGPT: Fine-tuning is a phase where the pre-trained model is further trained on the specific task it will be used for. The objective of this phase is to adapt the model to the specific task and fine-tune the parameters so that the model can produce outputs that are in line with the expected results. The pre-trained ChatGPT 3.5 model is fine-tuned using the labeled dataset of functional sentence structure. One of the most important things in the fine-tuning phase is the selection of the appropriate prompts. The prompt is the text given to the model to start generating the output. Providing the correct prompt is essential because it sets the context for the model and guides it to generate the expected output. It is also important to use the appropriate parameters during fine-tuning, such as the temperature, which affects the unpredictability of the output generated by the model. As shown in Fig. 2 the researcher developed and used representative prompt templates from the collected dataset in this regard. This fine-tuning process helps the model to learn and recognize the functional structure of a sentence including the adverb types based on contextual information.

3) Response Generation: With the ability to predict the functional structure of the sentence, ChatGPT can generate coherent and contextually relevant responses. These responses are informed by the adverb-type predictions, making them more precise and contextually appropriate. Throughout the methodology, emphasis is placed on the quality and representativeness of the labeled dataset, as this significantly influences the accuracy of adverb categorization and response generation. This ChatGPT-based methodology combines the power of pre-trained language models with fine-tuning on a domain-specific dataset to enhance adverb type prediction and response generation. It is a dynamic approach that leverages ChatGPT's natural language understanding and generation capabilities, making it a valuable tool for various NLP applications.

IV. Hybrid Parser-based Method

The creation of a Hybrid Parser-based sentence parsing framework is a noteworthy breakthrough in the field of NLP.

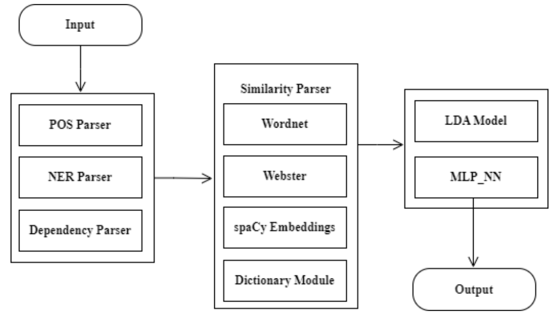

Fig. 3. The structural framework of the proposed hybrid parser based sentence parsing.

This innovative approach combines rule-based and machine-learning methods to extract meaning from text [38], addressing the limitations of current NLP parsing techniques. By incorporating both rule-based and machine-learning components, this framework becomes capable of handling a wider range of linguistic structures and domains, ensuring robust performance. Its primary objective is to enhance the accuracy of semantic parsing by capturing context-specific elements in language, ultimately improving the comprehension of the underlying meaning in the text. The framework strikes a careful balance between accuracy and efficiency, allowing for the precise construction of a semantic graph from textual content. The architecture of this framework encompasses text preprocessing, rule-based and machine learning-based sentence parsing, adverb-type prediction, and semantic graph construction.

One distinguishing feature of this framework is its dedicated component for predicting adverb types within the text. This feature plays a pivotal role in accurately extracting the essence of a sentence. The integration of outputs from both rule-based and machine learning-based parsing yields a comprehensive semantic graph representing the structured knowledge present in the text. This Hybrid parser-based approach harnesses the strengths of rule-based systems, which excel at handling linguistic patterns and prior knowledge, and machine learning models, which adapt to context and data-driven insights. As a result, the framework enhances natural language understanding and information extraction, offering a promising solution to the challenges presented by traditional parsing methods.

Fig. 3 provides an overview of the structural framework of the Hybrid Parser-based approach. It illustrates the key components, including text preprocessing, rule-based and machine learning-based parsing, adverb-type prediction, and semantic graph construction, highlighting their interconnections.

A. Methodology

The researchers utilized a free cloud-based platform called Google Collaboratory for running and writing Python code. For text analysis and parsing, we used essential parsing tools such as spaCy and NLTK. To improve the analysis and understanding of language, we integrated external resources, including dictionaries like Webster and ontologies such as WordNet.

Furthermore, to train the adverb prediction model, the dataset that contained definitions and synsets derived from a list of adverbs and prepositions is carefully collected, playing a fundamental role in model training.

To enhance the precision of adverb prediction, the researchers incorporated the machine learning technique known as Latent Dirichlet Allocation (LDA), with specific application of the MLP (MultiLayer Perceptron) model. The researchers utilize the power of LDA to construct topic-based feature vectors for words, with a particular focus on adverbs. LDA is commonly used in NLP to discover hidden topics within a corpus of text. The process of generating these feature vectors comprised several key steps: first, LDA modeling was applied, wherein words were associated with specific topics to discover the underlying semantic patterns. Then, the *LDA_vector* method is introduced and designed to take a word as input and determine its LDA representation, representing the word as a vector of topic probabilities based on its contextual associations.

Additionally, the *Webster_LDA_vector* method is defined to extend this capability to adverbs not found in Wordnet but present in word embeddings, thereby broadening the scope of the LDA approach. Ultimately, the LDA-derived vectors obtained from these methods were integrated into the feature vectors for adverbs, providing a structured means to measure their similarity or categorization in the context of the discovered semantic topics. This feature-based analysis allowed for comprehensive comparisons with other word similarity measures, including spaCy and Wordnet-based metrics, enhancing our understanding of adverb similarities and categories.

In addition to the methodological approach, the researchers utilized the power of word embeddings. Word embeddings are a way to represent words as dense vectors in a continuous vector space, allowing us to capture relationships between words and how they fit into sentences. Within the scope of this study, the utilization of word embeddings offers several advantages. First, they help us measure how similar words are to each other, which is particularly useful for understanding adverbs in the context of other words. Second, when we encounter words that aren't in the dictionary (Wordnet) we're using in the code, word embeddings provide a smart solution by giving us vector representations for a wide range of words. Third, they enable us to understand the meaning of words within their context, making it easier to figure out what adverbs mean based on the words they're associated with. Fourth, when we're creating graphs that show how words relate to each other, word embeddings enhance these vectors with more information. This enrichment helps us better understand the roles of adverbs and other words in sentences. Lastly, the integration of word embeddings results in more accurate and detailed graphs, representing words and their connections in sentences, ultimately enhancing our overall understanding.

Now, with the understanding of how word embeddings enhance our analysis of word relationships, let's delve into the process of determining the functional type of a given sentence sequence. This process involves analyzing the structure and components of sentences to categorize them into different functional units. To do this, we consider a set of accepted functional unit types, which include Predicate, Subject, Direct Object, Indirect Object, Time, Place, Manner, Frequency, Degree, and Conjunction. This parsing process is the initial step in our study.

Having an input word sentence, $s = w_1, w_2, ..., w_l$. where symbol w denotes a word inside the sentence. The set of accepted functional unit types is given by

$$\begin{aligned} T = \{&\text{Predicate, Subject, Direct Object,Indirect Object,} \\ &\text{Time,Place, Manner,Frequency, Degree, Conjunction}\} \end{aligned} \tag{1}$$

To determine the functional type of a given sentence sequence, the following parsing processes are first:

1) The internal dictionary contains the list of frequent adverb words, like the phrase as soon as, in this case, the dictionary contains also the related functional type.
2) Label of the dependency parsing: l_e This property is generated with the spacy parser as the label of the dependency edge from the generated dependency tree.
3) Wordnet-based Lin similarity (l_l): A score denoting how similar two word senses (s_1, s_2) are, based on the Information Content (IC) of the Least Common Subsumer (s_c) most specific ancestor node) and that of the two input synsets: $l_l(s_1, s_2) = \frac{2 \cdot IC(s_c)}{IC(s_1)+IC(s_2)}$
4) Wordnet-based path similarity (l_p): The path between the two synsets in the concept tree of the Wordnet.
5) Wordnet LDA similarity (l_d): We take the definition sections from the Wordnet database and calculate the topic similarities using the LDA method.
6) Webster LDA similarity (l_w): The definitions in Webster dictionary are used to calculate the topic similarities using the LDA method.
7) Spacy similarity (l_s): The similarity is based on the grammatical properties generated in the spacy NLP library.

The proposed framework also includes a dictionary which contains some selected words with the related unit types labels:

$D = \{(w, T(w)\}$

We divide this dictionary into two parts:

$D = D_B \bigcup D_L$

where D_B is the set of baseline words, we use to determine the similarity positions of new query words. For a given query word w_q, the following local feature vectors are calculated:

$\{l_e(w_q, w), l_l(w_q, w), l_p(w_q, w), l_d(w_q, w), l_w(w_q, w), l_s(w_q, w) | w \epsilon D_B\}$

Using these similarity measures, the generated similarity vectors are merged into a global feature vector

$l(w_q)$

These global feature vectors are used to predict the corresponding unit type label of w_q. For the prediction, an MLP neural network module (NN) is involved, where

$NN(l(w_q))$

outputs the predicted unit label.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 1100)              49500

 dropout (Dropout)           (None, 1100)              0

 dense_1 (Dense)             (None, 440)               484440

 dense_2 (Dense)             (None, 132)               58212

 dense_3 (Dense)             (None, 6)                 798

=================================================================
Total params: 592950 (2.26 MB)
Trainable params: 592950 (2.26 MB)
Non-trainable params: 0 (0.00 Byte)
```

Fig. 4. MLP architecture.

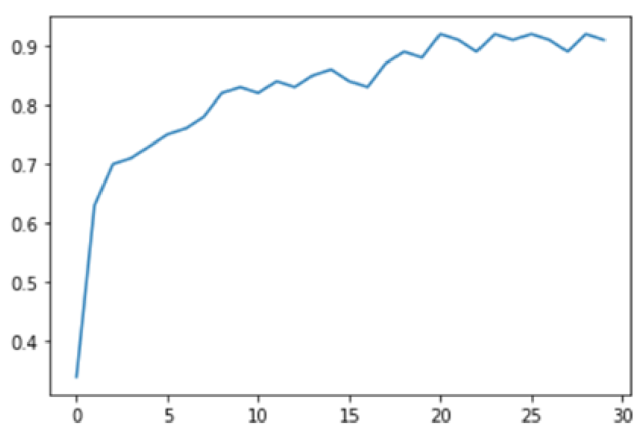

Fig. 5. Validation accuracy curve in the training process.

For the training of the MLP unit, the D_L dataset is used as training and test dataset.

The MLP neural network unit under consideration comprises five layers, with one dedicated to model regularization (as depicted in Fig. 4). The trained MLP unit demonstrated a commendable average accuracy of 92% on the tested datasets.

Fig. 5 displays the validation accuracy curve during the training process of the proposed framework. The curve illustrates how the accuracy of the model evolves as it undergoes training iterations. It provides valuable insights into the model's performance and its ability to generalize to unseen data, showcasing the progress made during the training phase.

V. Semantic Graph Induction

The process of automatically building a semantic graph from unstructured data, like textual documents or datasets, is known as semantic graph induction [26], [27]. It involves taking information from unstructured data, such as entities, concepts, and their relationships, and putting it in an organized manner. This procedure frequently depends on NLP and machine learning approaches to discover and link entities, infer relationships, and build the graph.

The term graphs refer to a common data format as well as a universal language for describing complicated systems. A common data structure and language for characterizing complex systems is called a graph. In its most basic form, a graph is just a set of objects or nodes, and the interactions (or edges) that exist between pairs of these nodes. For instance, we can utilize edges to signify the friendship between two individuals and utilize nodes to symbolize each person, effectively encoding a social network. This is illustrated in Fig. 6, featuring the renowned Zachary Karate Club Network.

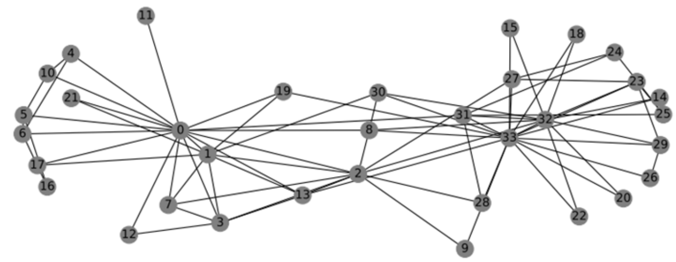

Fig. 6. Application fields of knowledge graphs example. The famous zachary karate club network represents the friendship relationships between members of the karate club studied by Wayne W. Zachary from 1970 to 1972.

An edge that connects two individuals if they socialize outside of the club. During Zachary's study, the club split into two factions centered around nodes 0 and 33 and Zachary was able to correctly predict which nodes would fall into each faction based on the graph structure [20]. Graphs do more than just provide an elegant theoretical framework, however. They offer a mathematical foundation that we can build upon to analyze, understand, and learn from real-world complex systems [20], [7].

Constructing large-scale semantic graphs from vast and diverse datasets is a significant challenge. Researchers are continually developing more efficient algorithms and technologies to handle big data [2]. Recently, word embeddings and entity embeddings have become effective in capturing semantic relationships, and the advancements in embedding techniques continue to improve graph construction [3]. Ensuring the completeness and accuracy of knowledge graphs is an ongoing challenge [4], [5], with methods for knowledge base completion and alignment being actively explored [6].

VI. Experimental Results

A. Methodology

The dataset utilized for fine-tuning ChatGPT is meticulously curated from a wide array of linguistic sources, including academic texts in history and biology, and factual data about world events. This selection, aimed at capturing a rich variety of sentence structures, ensures exposure to complex and diverse linguistic patterns. Each sentence within this dataset is carefully labeled by linguistics experts to identify its functional components such as subjects, predicates, direct and indirect objects, as well as various types of adverbs like those indicating time, place, frequency, and manner. This detailed labeling is crucial for the accurate training of the model.

The size of the dataset was determined considering the resource-intensive nature of manual collection and analysis.

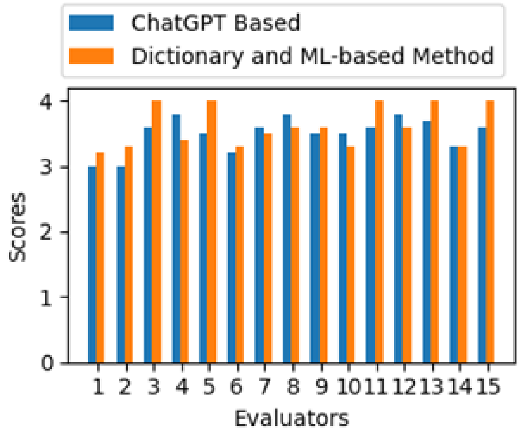

Fig. 7. The overall evaluation result of linguistic experts.

We assembled 160 sentences for the training dataset and 40 sentences for testing purposes. The semantic graph model employed in this study categorizes words and phrases from these sentences into their respective functional structures. This approach facilitates a comprehensive and nuanced understanding of sentence parsing, essential for the model's training and evaluation. The deliberate and diverse selection of sources ensures a well-rounded dataset, contributing to the effectiveness of the model in recognizing and interpreting a broad spectrum of linguistic elements.

B. Results and Analysis

One significant limitation within this area of sentence parsing research is the absence of an automated performance evaluation system, which remains unimplemented. To assess the accuracy of the parsing, the researchers engaged the expertise of linguistic professionals, educators, and students. The survey encompassed five distinct rating categories: "Poor", "Below Average", "Average", "Above Average", and "Excellent". The researchers used similar test datasets for both approachs and make a comparative result analysis.

In the evaluation process, we used the ChatGPT efficiency for prompts of different lengths and complexity. The models evaluated in this study include the OpenAI API and ChatGPT 3.5 Web Interface, as well as a Hybrid Parser-based Method.

Fig. 7 provides an overview of the comprehensive evaluation results of 15 linguistic experts for both methods. The evaluation scores range from a minimum of 1.5 to a maximum of 4, showcasing the experts' assessments of the performance of these methods.

Efficiency, as reflected in the average quality rating of responses generated by these models, is a key measure. We explored prompt set sizes ranging from 5 to 40. Surprisingly, both the ChatGPT 3.5 Web Interface and the Hybrid Parser-based Sentence Parsing model consistently maintained an "excellent" response quality rating, irrespective of the prompt set size. This indicates their enduring efficiency across a spectrum of prompt set sizes. This table provides valuable insights into how different prompt set sizes impact ChatGPT model efficiency, revealing noteworthy disparities in performance between the OpenAI API and other models.

TABLE I. EFFICIENCY OF CHATGPT IN DEPENDENCY OF PROMPT SIZE

Model	Prompt Size	Average Rating
OpenAI API	5	Poor
	15	Below Average
	25	Average
	35	Above Average
	40	Excellent
ChatGPT 3.5 Web Interface	–	Average
Hybrid Parser-based Sentence Parsing	–	Excellent

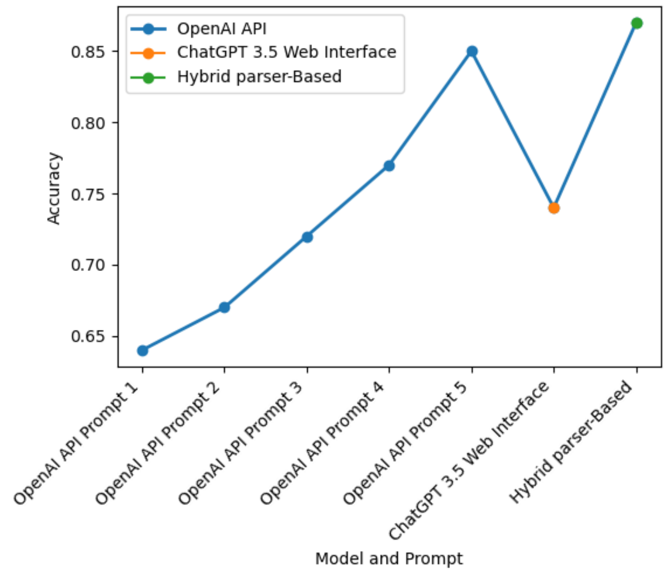

Fig. 8. ChatGPT OpenAI and hybrid parser-based sentence parsing accuracy.

Fig. 8 visually illustrates the influence of prompt set size on ChatGPT's sentence parsing performance, quantified by accuracy. Accuracy is determined by the ratio of correctly assigned sentences to the total assigned sentences. The OpenAI API employs five distinct prompts, each with varying numbers of sentence parsing templates: prompt one with 5 templates, prompt two with 15 templates, prompt three with 25 templates, prompt four with 35 templates, and prompt five with 40 templates. As seen in Table I, the accuracy of OpenAI models sees improvement as the number of templates within the prompts increases. In a separate experiment conducted with the ChatGPT 3.5 Web Interface, an accuracy score of 0.74 was achieved.

Table II presents accuracy values, indicating that the Hybrid Parser-based sentence parsing method exhibit a slight advantage over the ChatGPT-based model (acc_GPT = 0.85, acc_hybrid = 0.87). This evaluation scenario provides valuable insights into the performance and effectiveness of both approaches in sentence parsing.

This experiment underscores that while ChatGPT 3.5 is a recent and versatile language model capable of generating diverse and interesting results, it has limitations, particularly in domains like sentence parsing. The observed accuracy values

TABLE II. EFFICIENCY OF CHATGPT AND HYBRID PARSER-BASED SENTENCE PARSING METHOD

Model	Prompt size	Accuracy
OpenAI API	5	0.64
	15	0.67
	25	0.72
	35	0.77
	40	**0.85**
ChatGPT 3.5 Web Interface		0.74
Hybrid Parser-based Sentence Parsing		**0.87**

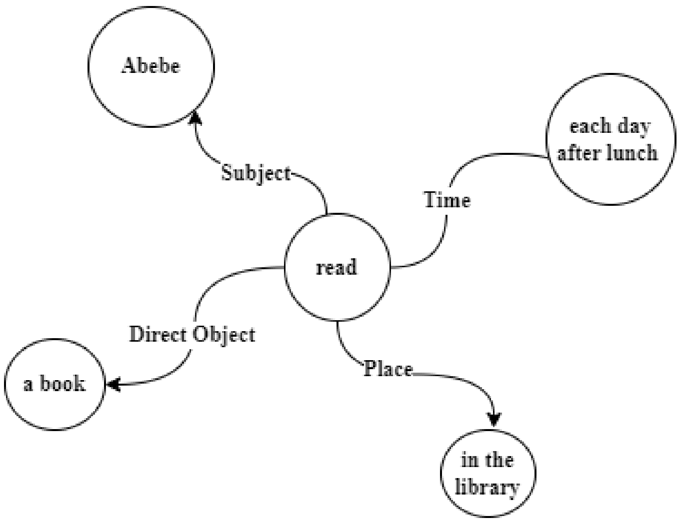

Fig. 9. Semantic graph generated by the proposed model for the sentence "Abebe reads a book deeply in the library each day after lunch".

strongly advocate for the effectiveness of the proposed Hybrid Parser based sentence parsing. This suggests that the proposed model may find broader applicability in sentence parsing tasks (see Fig. 9).

VII. DISCUSSION

This paper presents a novel approach to sentence parsing using ChatGPT, demonstrating significant potential in understanding and manipulating complex linguistic structures. We believe that the integration of LLMs like ChatGPT in sentence parsing tasks can revolutionize how we approach language understanding in AI. The model's ability to notice small details in language, from syntax to semantics, is particularly promising for applications in automated text summarization, sentiment analysis, and even in developing more advanced conversational AI.

However, we also recognize challenges, particularly in terms of computational demands and potential biases inherent in the training data. The scalability of such models in real-world applications remains a concern, especially considering the resource-intensive nature of their training and operation. It's crucial for future research to address these challenges, ensuring the responsible and efficient use of these powerful tools in various NLP applications.

VIII. CONCLUSIONS AND FUTURE WORK

In conclusion, the process of semantic graph construction stands as a cornerstone in the field of knowledge representation and artificial intelligence, giving structured meaning upon the vast landscape of textual data. It draws its strength from an array of foundational technologies, encompassing NLP, dependency parsing, word embeddings, LDA, and the integration of ontologies and knowledge graphs. These technological underpinnings empower the creation of semantic graphs, spanning from entity recognition to intricate topic modeling. The infusion of ChatGPT's NLP capabilities further enriches this process, rendering it a dynamic and adaptable tool for semantic graph construction.

Our deliberate experimentation and meticulous evaluation have illuminated the comparative performance, applicability, and constraints of ChatGPT-based and Hybrid Parser based sentence parsing methods within the context of semantic graph construction. These findings not only contribute to the expanding reservoir of knowledge within the field of NLP but also offer invaluable insights to researchers, developers, and practitioners venturing into real-world applications. These applications include information retrieval, knowledge graph development, and automated question-answering systems, among others.

It's worth noting that the accuracy values indicate a slightly better performance of the hybrid parser-based sentence parsing method compared to the ChatGPT-based model $acc\_GPT = 0.85$, $acc\_hybrid = 0.87$. In this evaluation scenario, our test results provide comprehensive insights into the strengths and limitations of ChatGPT 3.5, particularly in the domain of English sentence parsing and language understanding tasks. This knowledge is instrumental in further enhancing the capabilities of ChatGPT for these specific tasks.

As we investigate into the future, ongoing efforts will focus on refining ChatGPT for improved performance in English sentence parsing, thus bridging the gap between language models and semantic graph construction. The integration of additional linguistic resources, enhanced fine-tuning techniques, and prompt engineering strategies will be explored to further empower ChatGPT in its role as a dynamic tool for language understanding and knowledge representation.

ACKNOWLEDGMENT

The research reported here was carried out as part of the EFOP-3.6.1-16-2016-00011 "Younger and Renewing University – Innovative Knowledge City – Institutional development of the University of Miskolc aiming at intelligent specialization" project implemented in the framework of the Széchenyi 2020 program. The realization of this project is supported by the European Union, co-financed by the European Social Fund.

REFERENCES

[1] D. Rusu, B. Fortuna, M. Grobelnik, and D. Mladenić, "Semantic graphs derived from triplets application in document summarization," *Informatica*, vol. 33, no. 3, 2009.

[2] W. Hu, M. Fey, H. Ren, M. Nakata, Y. Dong, and J. Leskovec, "Ogb-lsc: A large-scale challenge for machine learning on graphs," *arXiv preprint arXiv:2103.09430*, 2021.

[3] P. Goyal and E. Ferrara, "Graph embedding techniques, applications, and performance: A survey," *Knowledge-Based Systems*, vol. 151, pp. 78–94, 2018. doi: 10.1016/j.knosys.2018.03.022.

[4] S. Ji, S. Pan, E. Cambria, P. Marttinen, and S. P. Yu, “A survey on knowledge graphs: Representation, acquisition, and applications,” *IEEE transactions on neural networks and learning systems*, vol. 33, no. 2, pp. 494–514, 2021. doi: 10.1109/TNNLS.2021.3070843.

[5] B. Abu-Salih, “Domain-specific knowledge graphs: A survey,” *Journal of Network and Computer Applications*, vol. 185, p. 103076, 2021. doi: 10.1016/j.jnca.2021.103076.

[6] H. Singh, P. Jain, S. Chakrabarti, et al., “Multilingual knowledge graph completion with joint relation and entity alignment,” *arXiv preprint arXiv:2104.08804*, 2021.

[7] X. Zou, “A survey on application of knowledge graph,” in *Journal of Physics: Conference Series*, vol. 1487, no. 1, 2020. doi: 10.1088/1742-6596/1487/1/012016.

[8] L. Shi, S. Li, X. Yang, J. Qi, G. Pan, B. Zhou, et al., “Semantic health knowledge graph: semantic integration of heterogeneous medical knowledge and services,” *BioMed research international*, vol. 2017, 2017. doi: 10.1155/2017/2858423.

[9] H. Oh and R. Jain, “Detecting events of daily living using multimodal data,” *arXiv preprint arXiv:1905.09402*, 2019.

[10] M. Rotmensch, Y. Halpern, A. Tlimat, S. Horng, and D. Sontag, “Learning a health knowledge graph from electronic medical records,” *Scientific reports*, vol. 7, no. 1, p. 5994, 2017. doi: 10.1038/s41598-017-05778-z.

[11] J. Liu, Z. Lu, and W. Du, “Combining enterprise knowledge graph and news sentiment analysis for stock price prediction,” 2019.

[12] S. A. Elnagdy, M. Qiu, and K. Gai, “Cyber incident classifications using ontology-based knowledge representation for cybersecurity insurance in financial industry,” in *2016 IEEE 3rd International Conference on Cyber Security and Cloud Computing (CSCloud)*, 2016. doi: 10.1109/CSCloud.2016.45.

[13] F. Zablith, “Constructing social media links to formal learning: A knowledge Graph Approach,” *Educational technology research and development*, vol. 70, no. 2, pp. 559–584, 2022. doi: 10.1007/s11423-022-10091-2.

[14] Y. Jia, Y. Qi, H. Shang, R. Jiang, and A. Li, “A practical approach to constructing a knowledge graph for cybersecurity,” *Engineering*, vol. 4, no. 1, pp. 53–60, 2018. doi: 10.1016/j.eng.2018.01.004.

[15] Y. Liu, T. Han, S. Ma, J. Zhang, Y. Yang, J. Tian, et al., ”Summary of ChatGPT-Related Research and Perspective Towards the Future of Large Language Models,” *Meta-Radiology*, p. 100017, 2023. doi: 10.1016/j.metrad.2023.100017.

[16] M. Abdullah, A. Madain, and Y. Jararweh, ”ChatGPT: Fundamentals, applications and social impacts,” in *2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2022. doi: 10.1109/SNAMS58071.2022.10062688.

[17] G. Vassiliou, N. Papadakis, and H. Kondylakis, ”SummaryGPT: Leveraging ChatGPT for summarizing knowledge graphs” in *European Semantic Web Conference*, 2023. doi: 10.1007/978-3-031-43458-7_31.

[18] P. Pichappan, M. Krishnamurthy, and P. Vijayakumar, ”Analysis of ChatGPT as a Question-Answering Tool,”

[19] H. B. de Barros Pereira, M. Grilo, I. de Sousa Fadigas, C. T. de Souza Junior, M. do Vale Cunha, R. S. F. Dantas Barreto, J. C. Andrade, and T. Henrique, ”Systematic review of the ’semantic network’ definitions,” *Expert Systems with Applications*, pp. 118455, 2022. DOI: 10.1016/j.eswa.2022.118455.

[20] W. L. Hamilton, *Graph representation learning*. McGill University: Morgan & Claypool Publishers, 2020.

[21] Y. Chen, X. Ge, S. Yang, L. Hu, J. Li, and J. Zhang, ”A Survey on Multimodal Knowledge Graphs: Construction, Completion and Applications,” *Mathematics*, vol. 11, no. 8, pp. 1815, 2023. DOI: 10.3390/math11081815.

[22] C. Peng, F. Xia, M. Naseriparsa, and F. Osborne, ”Knowledge graphs: Opportunities and challenges,” *Artificial Intelligence Review*, pp. 1–32, 2023. DOI: 10.1007/s10462-023-10465-9.

[23] A. Kamath and R. Das, ”A survey on semantic parsing,” *arXiv preprint arXiv:1812.00978*, 2018.

[24] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. Van Kleef, S. Auer, et al., ”Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia,” *Semantic web*, vol. 6, no. 2, pp. 167–195, 2015. DOI: 10.3233/SW-140134.

[25] Š. Čebirić, F. Goasdoué, H. Kondylakis, D. Kotzinos, I. Manolescu, G. Troullinou, and M. Zneika, ”Summarizing semantic graphs: a survey,” *The VLDB journal*, vol. 28, pp. 295–327, 2019. DOI: 10.1007/s00778-018-0528-3.

[26] P. Pham, L. T. T. Nguyen, W. Pedrycz, and B. Vo, ”Deep learning, graph-based text representation and classification: a survey, perspectives and challenges,” *Artificial Intelligence Review*, vol. 56, no. 6, pp. 4893–4927, 2023. DOI: 10.1007/s10462-022-10265-7.

[27] L. Kovacs, ”Mapping Between Semantic Graphs and Sentences in Grammar Induction System,” *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, vol. 1, pp. 100–112, 2010.

[28] H. Naveed, A. U. Khan, S. Qiu, M. Saqib, S. Anwar, M. Usman, N. Barnes, and A. Mian, ”A comprehensive overview of large language models,” *arXiv preprint arXiv:2307.06435*, 2023.

[29] E. Kotei and R. Thirunavukarasu, ”A Systematic Review of Transformer-Based Pre-Trained Language Models through Self-Supervised Learning,” *Information*, vol. 14, no. 3, pp. 187, 2023. DOI: 10.3390/info14030187.

[30] S. R. Bowman, ”Eight things to know about large language models,” *arXiv preprint arXiv:2304.00612*, 2023.

[31] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., ”Language models are few-shot learners,” in *Advances in neural information processing systems*, vol. 33, 2020, pp. 1877–1901.

[32] C. D. Manning, ”Human language understanding & reasoning,” *Daedalus*, vol. 151, no. 2, pp. 127–138, 2022. DOI: 10.1162/daed_a_01905.

[33] B. Meskó, ”Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial,” *Journal of Medical Internet Research*, vol. 25, pp. e50638, 2023. DOI: 10.2196/50638.

[34] S. Vemprala, R. Bonatti, A. Bucker, and A. Kapoor, ”ChatGPT for Robotics: Design Principles and Model Abilities,” *Microsoft Technical Report*, February 2023. [Online]. Available: https://www.microsoft.com/en-us/research/ publication/

[35] P. P. Ray, ”ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope,” *Internet of Things and Cyber-Physical Systems*, vol. 3, pp. 121-154, 2023. ISSN: 2667-3452. DOI: 10.1016/j.iotcps.2023.04.003, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S266734522300024X.

[36] G. Wohlgenannt, *Learning Ontology Relations by Combining Corpus-Based Techniques and Reasoning on Data from Semantic Web Sources*, Frankfurt am Main: Peter Lang International Academic Publishers, 2018.

[37] A. Harth, K. Hose, R. Schenkel, *Linked Data Management*, Boca Raton, FL, USA: CRC Press, 2014.

[38] F.-X. Desmarais, M. Gagnon, A. Zouaq, ”Comparing a Rule-Based and a Machine Learning Approach for Semantic Analysis,” in *Proceedings of the 6th International Conference on Advances in Semantic Processing*, Barcelona, Spain, 2012, pp. 103–108.