ISSN 2181-

THEME: PART-OF-SPEECH TAGGING AND ITS APPLICATIONS

Senior teacher of JSPU **Abdullajonova Hakima**

Inomjonova Farangiz

Abdullaxo'jayeva Fero'zabonu

Usmonova Nilufar

Jizzakh State Pedagogical University

Foreign Languages

Annotation. Part-of-speech (POS) tagging is a foundational technique in natural language processing that involves assigning grammatical categories such as nouns, verbs, adjectives, and adverbs to every word in a text. This process enables machines to understand language structure, support linguistic analysis, and enhance various AI applications. POS tagging is essential for information extraction, machine translation, text mining, sentiment analysis, and speech recognition. Modern tagging methods combine rule-based, statistical, and neural-network approaches to increase accuracy. As a result, POS tagging continues to play a crucial role in the development of intelligent language technologies.

Key Words. Part-of-speech tagging, natural language processing, machine learning, corpus linguistics, syntactic analysis, neural networks, text mining, rule-based systems, probabilistic models, language technology

Annotatsiya. So‘z turkumini aniqlash (Part-of-Speech tagging) tabiiy tilni qayta ishlashda muhim texnika bo‘lib, matndagi har bir so‘zga ot, fe’l, sifat yoki ravish kabi grammatik toifalarni belgilashni o‘z ichiga oladi. Bu jarayon til tuzilishini kompyuterga tushunishga yordam beradi, lingvistik tahlilni qo‘llab-quvvatlaydi va turli sun’iy intellekt ilovalarini rivojlantiradi. POS belgilash axborot ajratish, mashina tarjimasi, matnni tahlil qilish, kayfiyatni aniqlash va nutqni tanishda muhim ahamiyatga ega. Zamonaviy usullar aniq-likni oshirish uchun qoida, statistik va neyron yondashuvlarni birlashtiradi.

Kalit so‘zlar. So‘z turkumini belgilash, tabiiy tilni qayta ishlash, mashinaviy o‘qitish, korpus lingvistikasi, sintaktik tahlil, neyron tarmoqlar, matnni qazib olish, qoida asosidagi tizimlar, ehtimollik modellari, til texnologiyasi

Аннотация. Часть-речевое тегирование (Part-of-Speech tagging) — это важная технология обработки естественного языка, заключающаяся в присвоении каждому слову в тексте грамматической категории: существительного, глагола, прилагательного или наречия. Данный процесс помогает компьютерам понимать языковую структуру, поддерживает лингвистический анализ и усиливает различные приложения искусственного интеллекта. POS-тегирование широко применяется в извлечении информации, машинном переводе, текстовом анализе, определении тональности и распознавании речи.

Ключевые слова. тегирование частей речи, обработка естественного языка, машинное обучение, корпусная лингвистика, синтаксический анализ, нейронные сети, интеллектуальный анализ текста, rule-based системы, вероятностные модели, языковые технологии

Part-of-speech (POS) tagging is one of the most fundamental tasks in natural language processing (NLP). It involves automatically assigning each word in a sentence to a grammatical category, such as noun, verb, adjective, adverb, pronoun, or preposition. Although humans perform this task intuitively, enabling machines to do the same is a complex challenge because language is inherently ambiguous and context-dependent. Nevertheless, POS tagging remains crucial in helping computers understand and process human language effectively.

The primary purpose of POS tagging is to clarify the function of each word based on both its definition and its context. For example, the word “book” can be either a noun or a verb, depending on how it is used in the sentence. A POS tagger uses linguistic rules, statistical models, or neural networks to determine which category fits best in each context. This disambiguation is essential for enabling machines to grasp syntactic structure and support more advanced language-processing tasks.

Historically, POS tagging began with rule-based systems in the 1960s and 1970s. Linguists manually created sets of grammatical rules that defined how words behave. For instance, if a determiner such as “the” precedes a word, it is likely a noun. While rule-based systems were interpretable and linguistically sound, they required extensive manual labor and failed to adapt well to real-world language variations.

In the 1980s and 1990s, the focus shifted to statistical approaches. These models, particularly Hidden Markov Models (HMMs), learned patterns from large corpora by calculating probabilities of tag sequences. A statistical tagger might assign the most likely sequence of tags to a sentence based on previous observations in training data. This represented a major breakthrough because statistical taggers were significantly more robust than rule-based methods and could handle ambiguous cases more effectively.

With the rise of machine learning in the 2000s and especially deep learning in the 2010s, neural-network-based POS taggers became dominant. Recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and more recently transformer-based models such as BERT have achieved near-human levels of accuracy. These models capture long-range dependencies, contextual meanings, and subtle linguistic patterns that earlier models could not handle. Modern systems use contextual embeddings, meaning the representation of a word changes depending on surrounding words. As a result, POS tagging has become more flexible, adaptive, and accurate than ever before.

The applications of POS tagging span across a wide range of fields. One of the most important applications is syntactic parsing. Before a computer can analyze sentence structure, it must know the grammatical categories of words. POS tagging provides this foundation. In machine translation, tagging helps systems identify grammatical structure, enabling more accurate translation between languages. Information extraction systems use POS tagging to identify key entities and relations in text, such as names, dates, and events.

Another significant application is in sentiment analysis. Identifying parts of speech helps models determine which words carry emotional content, such as adjectives describing opinions. Similarly, POS tagging helps improve text-to-speech and speech-to-text technologies by providing grammatical context that guides pronunciation or transcription.

Despite major advances, POS tagging still faces challenges. Ambiguity remains a key issue: many words have multiple possible categories, and even state-of-the-art models can misinterpret rare or unusual constructions. Additionally, languages with rich morphology require more complex tagging strategies.

Conclusion:

Part-of-speech tagging remains a vital component of natural language processing, providing the structural foundation required for deeper textual analysis. Through rule-based, statistical, and modern neural techniques, POS tagging has evolved into a highly accurate and efficient tool. Its applications in translation, information extraction, speech technology, and research illustrate its broad impact. Although challenges persist, continuing advancements in machine learning ensure that POS tagging will remain central to future innovations in language technology.

References. Books

1. Jurafsky, D., & Martin, J. H. Speech and Language Processing. 3rd ed., 2023.

2. Manning, C., & Schütze, H. Foundations of Statistical Natural Language Processing. 1999.

3. Bird, S., Klein, E., & Loper, E. Natural Language Processing with Python. 2009.

4. Allen, J. Natural Language Understanding. 2nd ed., 1995.

5. Goldberg, Y. Neural Network Methods for Natural Language Processing. 2017.

6. Eisenstein, J. Introduction to Natural Language Processing. 2019.

7. Cambria, E., & White, B. Jumping NLP Curves: A Review of Natural Language Processing Research. 2014.

Websites

ISSN 2181-

1. https://nlp.stanford.edu
2. https://aclweb.org
3. https://towardsdatascience.com/nlp