

Łukasz Kobyliński

Instytut Podstaw Informatyki Polskiej Akademii Nauk

ORCID: 0000-0003-2462-0020

lukasz.kobyliński@ipipan.waw.pl

Maciej Ogrodniczuk

Instytut Podstaw Informatyki Polskiej Akademii Nauk

ORCID: 0000-0002-3467-9424

maciej.ogrodniczuk@ipipan.waw.pl

Rozdział 3

Inżynieria lingwistyczna

Streszczenie

Inżynieria lingwistyczna, zwana także przetwarzaniem języka naturalnego (ang. *natural language processing*, NLP), zajmuje się komputerową analizą i generowaniem tekstu, mowy oraz sygnałów niewerbalnych. Stosowana jest w wyszukiwarkach internetowych, tłumaczeniach maszynowych, chatbotach czy asystentach głosowych. Początki tej dziedziny sięgają połowy XX w., gdy Alan Turing zaproponował test mierzący inteligencję maszyn poprzez język. W kolejnych dekadach rozwijały się systemy regułowe, a następnie algorytmy uczenia maszynowego, aż po współczesne modele głębokiego uczenia, takie jak BERT, GPT i T5. Te nowoczesne rozwiązania umożliwiają automatyczne tłumaczenie, analizę sentymentu, generowanie tekstów i inne zaawansowane zadania. Przetwarzanie języka oparte jest na różnych metodach reprezentacji tekstu, takich jak modele statystyczne (np. TF-IDF), wektorowe (word2vec) i kontekstowe (BERT). Najnowsze modele, w rodzaju GPT-4, są wielomodalne i mogą przetwarzać zarówno tekst, jak i obraz czy dźwięk. Przyszłość inżynierii lingwistycznej to dalsza integracja ze sztuczną inteligencją (ang. *artificial intelligence*, AI), rozwój wielojęzycznych modeli oraz tworzenie bardziej zaawansowanych systemów automatycznej analizy językowej. Również w Polsce trwają prace nad modelami

Nota bibliograficzna:

Kobyliński, Łukasz, Ogrodniczuk, Maciej (2025): Inżynieria lingwistyczna. [W:] Topczewska, U., Jopek-Bosiacka, A., Leńko-Szymańska, A., Szupica-Pyrzanowska, M. (red.): *Lingwistyka stosowana. Kompedium*, t. II: *Język w badaniach interdyscyplinarnych*, pod red. M. Szupicy-Pyrzanowskiej. Warszawa: Wydawnictwo Naukowe Scholar, 79–93. DOI: 10.7366/978836809139703.

dla języka polskiego, takimi jak Bielik i PLLuM, a rozwój tej dziedziny wymaga stałej aktualizacji korpusów językowych i współpracy międzynarodowej.

Słowa kluczowe: inżynieria lingwistyczna, przetwarzanie języka naturalnego, uczenie maszynowe, duże modele językowe, sztuczna inteligencja

Linguistic engineering

Abstract

Linguistic engineering, also known as natural language processing (NLP), focuses on the computational analysis and generation of text, speech, and non-verbal signals. It is widely used in search engines, machine translation, chatbots, and voice assistants. The field dates back to the mid-20th century, when Alan Turing introduced a test to measure machine intelligence through language. Over the decades, it evolved from rule-based systems to machine learning algorithms, culminating in deep learning models such as BERT, GPT, and T5. These modern approaches enable tasks such as automatic translation, sentiment analysis, and text generation. Text processing relies on various representation methods, including statistical models (e.g. TF-IDF), vector-based models (word2vec), and contextual embeddings (BERT). The latest models, such as GPT-4, are multimodal, enabling the processing of not only text but also images and audio. The future of linguistic engineering involves further AI integration, the development of multilingual models, and the creation of more sophisticated automatic language analysis systems. In Poland, efforts are underway to develop language models such as Bielik and PLLuM. The field's growth requires continuous updates to linguistic corpora as well as international collaboration to ensure the effective processing of different languages and dialects.

Keywords: linguistic engineering, natural language processing, machine learning, large language models, artificial intelligence