

Program seminarium “CLARIN-PL-Biz – technologie językowe dla nauki i biznesu”

16 marca 2021 r.

14:15-14:35 **Maciej Piasecki** Wprowadzenie

- CLARIN ERIC i CLARIN-PL – europejska infrastruktura badawcza technologii językowych dla nauk humanistycznych i społecznych
- Projekt CLARIN-PL-Biz – rozszerzenie wsparcia na obszar sztucznej inteligencji oraz zastosowań w biznesie

14:35-14:50 **Tomasz Naskręt, Tomasz Walkowiak, Mateusz Tykierko, Mateusz Gniewkowski** System CLARIN BIZ do przechowywania i przetwarzania bardzo dużych danych językowych

- Przyszły superkomputer do analizy danych językowych w zastosowaniach sztucznej inteligencji
- Planowana architektura w zarysie, podstawowe funkcje do udostępniania zasobów oraz usługi obliczeniowe ukierunkowane na język i sztuczną inteligencję

14:50-15:05 **Marcin Oleksy, Piotr Pęzik, Agnieszka Dziob** Zasoby językowe – źródła wiedzy i wzorce treningowe dla uczenia maszynowego

- Korpusy językowe, między innymi:
 - Korpus PWr – źródło danych treningowych dla sztucznej inteligencji,
 - Korpusy mowy polskiej i dialogów,
 - Korpus Idiolektów,
 - Korpus Wieszczów (projekt wspierany przez CLARIN-PL)
- Zasoby leksykalne: Słowność powiązana ze światowymi zasobami semantycznymi i zasobami wiedzy

15:05-15:20 **Arkadiusz Janz, Alina Wróblewska, Wiktor Walentynowicz** Podstawowe narzędzia analizy językowej

- Neuronalne modele językowe do opisu słownictwa, składni i semantyki - modele wektorów osadzeń (ang. word embeddings i contextual text embeddings)
- Narzędzia wstępnego przetwarzania: korekta błędów, normalizacja, analiza form gramatycznych wyrazów i struktur składniowych
- Narzędzia do wydobywania i agregacji informacji z tekstu (rozpoznawanie znaczeń słów w tekście i łączenie tekstu z sieciami semantycznymi, rozpoznawanie jednostek identyfikacyjnych, relacji czasowych i przestrzennych)

15:20-15:35 **Alicja Wieczorkowska** Serwis przetwarzania mowy

- Usługi mowy: narzędzia do rozpoznania, normalizacji i analizy nagrań
- Narzędzia do tworzenia korpusów mowy
- Korpusy nagrań

15:35-15:50 **Tomasz Walkowiak, Arkadiusz Janz** Analiza semantyczna tekstów

- Klasyfikacja i filtrowanie informacji, podobieństwo i grupowanie, modele tematyczne
- Wydobywanie słów kluczowych, terminologii i sieci semantycznych z tekstów

15:50-16:05 **Maciej Piasecki** Systemy dialogowe: podstawowy, szkieletowy, modułowy system dla języka polskiego

16:05-16:20 **Jan Kocoń** Analiza polaryzacji wydźwięku i emocji w tekście

- Rozpoznawanie wydźwięku w tekstach polskich i obcojęzycznych
- Wydobywanie aspektów opinii wraz z oceną emocywną
- Personalizacja modeli wydźwięku
- Interfejs słownik-mózg

16:20-16:35 **Maciej Piasecki** Wybrane planowane aplikacje CLARIN-PL-Biz

- Specjalizowane systemy dziedzinowe
- Moduł indeksowania semantycznego kolekcji tekstów
- Modułowy system odpowiadania na pytania jako rozszerzenie wyszukiwarki semantycznej
- Mechanizmy agregacji i inteligentnego filtrowania informacji