

# Metody przetwarzania języka naturalnego oraz wyszukiwanie *Projekt*

Prowadzący: Mgr inż. Jakub Klikowski

Semestr letni 2021/2022

## 1 Informacje o projekcie

### 1.1 Cel projektu

Celem projektu jest nabycie praktycznych umiejętności w projektowaniu oraz implementacji wybranych zagadnień przetwarzania języka naturalnego oraz wyszukiwania wraz z eksperymentalną oceną zastosowanych technik na samodzielnie pozyskanych danych pochodzenia rzeczywistego.

### 1.2 Środowisko programistyczne i narzędzia

Implementacja projektu należy wykonać w języku programowania **Python** z wykorzystaniem bibliotek przetwarzania języka naturalnego oraz uczenia maszynowego. Dobór narzędzi uzależniony jest od tematyki projektu i zostanie uzgodniony indywidualnie. Raport końcowy należy przygotowywać z wykorzystaniem edytora  $\text{\LaTeX}$ (np. *Overleaf*<sup>1</sup>) w formacie LNCs<sup>2</sup>. Wraz z postępami prac należy rozwijać wcześniej przygotowany dokument o kolejne etapy.

### 1.3 Etapy projektu zgodnie z kolejnością zajęć

- Tworzenie grup projektowych
- I Przegląd literatury
- II Pozyskiwanie i analiza danych
- III Implementacja oraz opis wybranych technik
- IV Projekt eksperymentów
- V Ewaluacja eksperymentalna
- VI Opracowanie raportu końcowego

---

<sup>1</sup> <https://scikit-learn.org>

<sup>2</sup> <http://www.springer.com/computer/lncs?SGWID=0-164-6-793341-0>

## 2 Warunki zaliczenia projektu

### 2.1 Zasady

1. Zajęcia projektowe przez pierwsze dwa tygodnie będą odbywać się w formie zdalnej za pośrednictwem platformy *MS Teams* w trakcie godzin wyznaczonych zgodnie z planem.
2. Od 14 marca 2022 zajęcia będą odbywać się stacjonarnie w wyznaczonej sali zgodnie z planem.
3. Grupy powinny składać się z dwóch lub trzech osób w zależności od wybranego tematu.
4. Tematy należy wybierać samodzielnie o ich przydzieleniu zdecyduje kolejność zgłoszeń.
5. Sposób zgłaszania tematów zostanie przedstawiony w trakcie zajęć organizacyjnych.
6. Cała grupa ponosi odpowiedzialność za uzyskaną ocenę końcową z projektu i za ilość zdobytych punktów za poszczególne etapy.
7. Za etapy I-V można uzyskać 15 punktów oraz a za etap IV 25 punktów.
8. Ocena końcowa zależna jest od liczby uzyskanych punktów (Tabela poniżej). Maksymalnie można zdobyć 100 punktów.
9. Etapy projektów należy oddawać zgodnie z ustalonym wcześniej harmonogramem.
10. Oddany etap należy omówić podczas spotkania.
11. Opóźnienie oznacza uzyskanie 0 punktów za dany etap.
12. Raport końcowy i kolejne etapy należy umieszczać na platformie *Eportal*.
13. W przypadku stwierdzenia niesamodzielności pracy studentów, prowadzący będzie kierować sprawę, za pośrednictwem dziekana, do komisji dyscyplinarnej ds. studentów<sup>3</sup>.

### 2.2 Tabela ocen

Punkty Ocena	
>90	5.0
90-80	4.5
80-70	4.0
70-60	3.5
60-50	3.0
<50	2.0

<sup>3</sup> Regulamin studiów, §14, punkt 20.

### **3 Lista tematów projektowych**

1. Analiza sentymentu na podstawie treści pozyskanych z mediów społecznościowych - Twitter. (3 osoby)
2. Automatyczne podsumowanie treści pozyskanych z internetowych serwisów informacyjnych. (3 osoby)
3. Rozpoznawanie nazwanych jednostek w treściach odczytanych za pomocą OCR. (3 osoby)
4. Wykrywanie wiadomości typu SPAM wykorzystując metody przetwarzania języka naturalnego. (2 osoby)
5. System oznaczania części mowy bazujący na sztucznych sieciach neuronowych. (2 osoby)