

Piotr Chmiel, 200608

Kamil Machnicki, 200752

Łukasz Matysiak, 200646

Michał Polański, 200852

Maciej Stelmaszuk, 200654

Jakub Zgraja, 200609

# Deep learning – tagger

(POS, lematyzacja)

## DOKUMENTACJA PROJEKTOWA

Kurs „Zastosowania informatyki w gospodarce”  
Rok akad. 2015/2016, kierunek INF, studia II stopnia

PROWADZĄCY:

Dr inż. Tomasz Walkowiak

WYDZIAŁ ELEKTRONIKI  
POLITECHNIKA WROCŁAWSKA

## **Spis treści**

<b>1</b>	<b>Cel i zakres dokumentu</b>	<b>3</b>
<b>2</b>	<b>Cel projektu</b>	<b>3</b>
<b>3</b>	<b>Koszty</b>	<b>4</b>
<b>4</b>	<b>Terminy i harmonogram projektu</b>	<b>4</b>

# 1 Cel i zakres dokumentu

Niniejszy dokument jest kompletnym projektem informatycznym, obejmującym wszelkie aspekty dotyczące planu wytworzenia oprogramowania, podjętego w ramach kursu „Zastosowania informatyki w gospodarce”. Jego celem jest sprecyzowanie postaci projektu m.in. poprzez:

- określenie wymagań funkcjonalnych i нефункциональных,
- oszacowanie kosztów projektu i czasu potrzebnego na jego realizację,
- identyfikację i analizę punktów krytycznych,
- określenie ograniczeń projektowych.

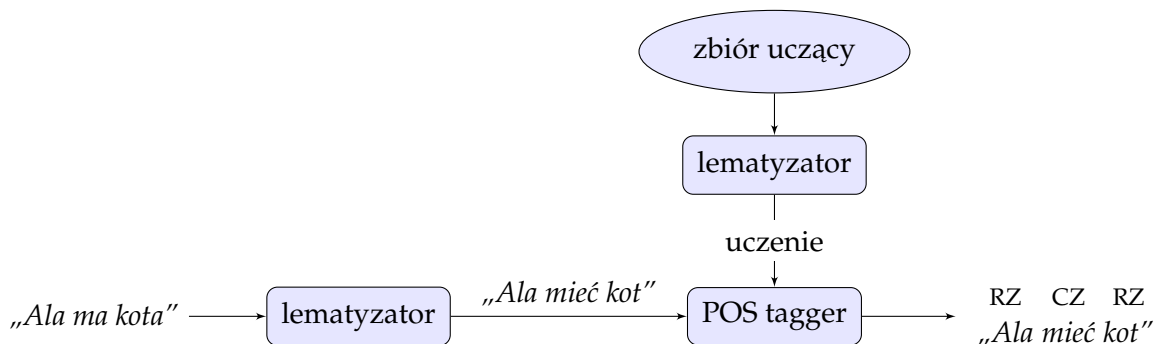
Ponadto dotyczy on sformalizowania samego procesu realizacji projektu. W dokumencie będzie można znaleźć informacje na temat:

- planu zarządzania jakością i ryzykiem,
- struktury organizacyjnej zespołu projektowego – w tym informacje dotyczące komunikacji wewnętrznej i z klientem/prowadzącym,
- dokładnego harmonogramu wykonywanych prac.

## 2 Cel projektu

**Temat projektu:** *Deeplearning – tagger (POS, lematyzacja).*

Zadaniem realizowanego przedsięwzięcia jest stworzenie programu głębokiego uczenia (ang. *deep learning*) z zakresu przetwarzania języka naturalnego. Jego celem jest przypisywanie każdemu wyrazowi w tekście wejściowym odpowiadającej mu części mowy (ang. *part-of-speech, POS*). Wykonanie zadania tagowania zostanie poprzedzone procesem lematyzacji tekstu.



Rysunek 1: Uproszczony schemat działania programu.

Realizacja powinna posiadać formę aplikacji desktopowej lub skryptu. Cel projektu zostanie osiągnięty, jeśli przygotowane oprogramowanie (wyposażone w odpowiedni zbiór uczący) będzie w stanie przetwarzać tekst w języku polskim w czasie i dokładności, które zostaną sprecyzowane przez klienta/prowadzącego.

### 3 Koszty

#### Szacunkowy czas wykonania

Biorąc pod uwagę zakres tematyczny projektu szacuje się, że do jego wykonania potrzebnych jest 400 godzin roboczych. Około 1/8 czasu poświęcone będzie na zebranie i analizę materiałów dotyczących projektu, ogólne poznanie możliwości i technologii. Największa część czasu zużyta zostanie na implementację projektu (stworzenie lematyzatora oraz taggera). Szacuje się, że będzie to 6/8 czasu. Pozostała część przeznaczona jest na kontakty z klientem oraz testy zaimplementowanego rozwiązania.

#### Szacunkowy koszt projektu

Mając na uwadze poziom skomplikowania zadania projektowego stawka za godzinę pracy wynosi 30 zł. Przy szacowaniu kosztów projektu nie uwzględniamy dodatkowych wydatków, które trzeba będzie ponieść w przypadku, gdy np. zajdzie potrzeba wykupienia domeny WWW. Nie są uwzględnione także koszty wdrożenia projektu w środowisku produkcyjnym. W związku z powyższym szacunkowy koszt wykonania projektu jest równy 12 000 zł netto.

Usługi, o których mowa w poprzednich punktach opodatkowane są 23% stawką podatku VAT (podatek od towarów i usług). Szacowana cena brutto wynosi **14 760 zł**.

Tabela 1: Kosztorys.

Nazwa	Jedn.	Ilość	Cena jedn.	Wart. netto	Stawka	Podatek	Wart. brutto
Robocizna	r-g	400	30	12 000	23 %	2 760 zł	14 760 zł
Razem:							14 760 zł

### 4 Terminy i harmonogram projektu

Granicznym terminem realizacji projektu jest **7 czerwca 2016 r.** W tym dniu nastąpi prezentacja rezultatów projektu wraz z przekazaniem jego pełnej dokumentacji.

Tabela 2: Harmonogram projektu.

Data	Opis
23.02.2016	Wybór tematu projektu.
1.03.2016	Określenie zakresu tematycznego projektu.
8.03.2016	Przekazanie specyfikacji projektu (wstępna funkcjonalność, określenie kamieni milowych).
9.03.2016 – 06.06.2016	Konsultacja wyników pracy po osiągnięciu kolejnych kamieni milowych projektu.
<b>7.06.2016</b>	Prezentacja rezultatów projektu i przekazanie dokumentacji projektowej.

Tabela 3: Kamienie milowe.

<b>Data</b>	<b>Opis</b>
29.03.2016	Instalacja środowiska
19.04.2016	Stworzenie lematyzatora
10.05.2016	Stworzenie taggera
31.05.2016	Stworzenie instrukcji użytkowania
07.06.2016	Prezentacja projektu