

| | |
|------------------------------------|--|
| imię (imiona) i nazwisko stażysty: | Natalia Wanda Potocka |
| adres e-mail: | potockan@student.mini.pw.edu.pl |
| nr telefonu: | 509385984 |
| PESEL: | 91081211867 |
| numer albumu: | 237476 |
| wydział: | Matematyki i Nauk Informatycznych |
| kierunek: | Matematyka (SMAD) |
| semestr: | III |
| stopień studiów (I czy II): | II |

PROGRAM STAŻU

MIEJSCE ODBYWANIA STAŻU:

Nazwa firmy: Instytut Badań Systemowych Polskiej Akademii Nauk

Adres firmy: ul. Newelska 6, 01-447 Warszawa

(jeśli staż nie odbywa się w siedzibie firmy proszę podać adres, gdzie faktycznie odbędzie się staż)

OPIEKUN STAŻU (z ramienia firmy):

Imię i nazwisko: dr Marek Gągolewski.

Telefon/adres e-mail: +48 503 988 054, gagolews@ibspan.waw.pl

TERMIN REALIZACJI STAŻU¹: 01.02.2015-30.04.2015

MINIMALNY WYMAGANY WYMIAR STAŻU WYNOSI 15 GODZIN TYGODNIOWO

CEL STAŻU: Umożliwienie zdobycia praktycznej wiedzy przez studentkę/a Politechniki Warszawskiej, w obszarze związanym z jej/jego ścieżką kształcenia.

ZAKRES ZADAŃ STAŻYSTY (opis merytoryczny planowanych prac z podaniem zadań jakie student będzie wykonywał):

Celem stażu jest implementacja nowego algorytmu automatycznej kategoryzacji tematycznej dużych zbiorów danych tekstowych na przykładzie polskiej Wikipedii oraz dokumentacji funkcji w różnych pakietach języka R. Nowa metoda bazuje na odległościach edycyjnych między napisami (pseudometryki).

1. Przegląd literatury na tematy związane z analizą skupień (clustering) danych tekstowych
2. Pobranie kompletnego obrazu polskiej Wikipedii – tworzenie skryptów w R i Python. Eksploracyjna analiza zbioru danych oraz jego czyszczenie
3. Klasyfikacja tematyczna tekstów z polskiej Wikipedii na podstawie algorytmu k-średnich oraz przy pomocy klasyfikacji hierarchicznej z uwzględnieniem zaproponowanego podejścia
4. Kompletowanie danych do analizy tekstów pomocy pakietów R
5. Analiza tekstów pomocy pakietów R

¹ W ramach Projektu finansowane są wyłącznie trzymiesięczne staże długoterminowe.

Wyrażam wolę przyjęcia na staż
studentki/a Politechniki Warszawskiej

.....
*Pełnomocnik Dziekana ds. Praktyk/
odpowiedni Prodzikan na Wydziale*
PODPIS I PIECZĘĆ

.....
Osoba decyzyjna w firmie
PODPIS I PIECZĘĆ