

R - programowanie w języku R

Poziom zaawansowany

Cele szkolenia

Po zakończeniu szkolenia jego uczestnik nabędzie wiedzę na temat zaawansowanych aspektów programowania w języku R. Posiędzie umiejętność efektywnej pracy z typami danych często spotykanymi w procesie analizy (tabele danych, szeregi czasowe, wartości temporalne). Będzie potrafił rozbudować swój kod R o własne funkcje oraz klasy. Wiedza na temat optymalizacji działania kodu oraz jego profilowania okaże się bardzo przydatna podczas wdrażania i uruchamiania własnych rozwiązań.

Umiejętności

Dzięki szkoleniu uczestnik będzie:

- efektywnie wykorzystywać dane temporalne (daty, czas, szeregi czasowe),
- przetwarzać większe zbiory danych wykorzystując bibliotekę `data.table`,
- rozwijać własne rozwiązania w języku R wykorzystując programowanie obiektowe oraz funkcyjne,
- sprawnie wychwytywać błędy oraz wąskie gardła w kodzie, optymalizować wykonanie kodu poprzez wektoryzację i/lub zrównoleglenie.

Profil uczestników

Szkolenie przeznaczone jest dla osób wykorzystujących język R do przetwarzania, analizowania oraz wizualizacji danych.

Przygotowanie uczestników

Od uczestników szkolenia wymagana jest znajomość podstaw programowania w języku R (tworzenie i uruchamianie kodu w środowisku RStudio, praca z podstawowymi typami i strukturami danych, podstawy wizualizacji danych z wykorzystaniem biblioteki standardowej lub `ggplot2`, odczyt i zapis danych w plikach płaskich).

Zagadnienia, których znajomość jest wymagana od uczestnika szkolenia objęte są kursem *R – analiza danych, uczenie maszynowe – poziom podstawowy* (nie jest konieczna znajomość stosowania algorytmów uczenia maszynowego).

Szczegółowy program szkolenia

1. Wykorzystanie wartości dat i czasu
Typy temporalne w języku R

Pakiet lubridate

Działania na wartościach daty i czasu

Szeregi czasowe

Wizualizacja szeregów czasowych

2. Wyrażenia regularne

3. Zaawansowane metody przekształcania danych

Funkcje apply

Obsługa brakujących wartości

Normalizacja, centrowanie i skalowanie

Przekształcenia dla rozkładów skośnych

Próbkowanie danych

Agregacja i reshaping (funkcje gather, unite, separate, spread)

4. Pakiet data.table

Różnice pomiędzy data.frame a data.table

Przekształcenia danych

Agregacja danych

5. Definiowanie własnych funkcji

Definicja i wywoływanie funkcji

Parametry funkcji

Zakres widoczności zmiennych w i poza funkcją, środowiska

Funkcje w R jako obiekty

6. Programowanie obiektowe

Klasy S3, S4

Klasy referencyjne

7. Programowanie funkcyjne

Funkcje map, reduce

Funkcje search, filter, compose

8. Obsługa błędów, debugowanie i profilowanie kodu

9. Optymalizacja działania kodu w języku R

Wektoryzacja kodu

Zarządzanie pamięcią

10. Zrównoleglanie działania

pakiet parallel

pakiet snow

Metoda realizacji szkolenia

Szkolenie realizowane jest w formie naprzemiennie następujących po sobie mini wykładów oraz ćwiczeń praktycznych. Szkolenie łączy w sobie fachową wiedzę merytoryczną z praktycznymi przykładami jej wykorzystania w środowisku pracy.

Liczba dni, liczba godzin szkoleniowych

3dni, 24 godziny szkoleniowe

Ścieżka rozwoju po szkoleniu

Po zakończeniu szkolenia rekomendowane jest skorzystanie ze szkolenia *Analiza danych w języku R – wizualizacja i raportowanie - poziom zaawansowany*.