



Katedra Systemów Decyzyjnych i Robotyki
Politechnika Gdańska, Wydział ETI

Autonomiczne Systemy Ekspertyzy i Eksploracji Danych

Instrukcja do zajęć laboratoryjnych
Spark - wprowadzenie

Opracował:
Jakub Wszolek
kubawszolek@gmail.com
ver. 3.0

1. Cel ćwiczenia

Celem ćwiczenia jest zapoznanie się z podstawowymi algorytmami przydatnymi przy budowie systemów ekspertyzy i eksploracji danych.

2. Oprogramowanie

W realizacji ćwiczenia wykorzystanie zostanie środowisko Spark (<https://spark.apache.org/>).

3. Zadanie 1

Zadanie polega na zapoznaniu się z kodem napisanym w języka Java. Kod dostępny jest w następującej lokalizacji: <https://github.com/jwszolek/ASEiED>

Do uruchomienia aplikacji lokalnie wymagane jest posiadanie zainstalowanego środowiska Maven (<https://maven.apache.org/>).

Zrealizowany kod dokonuje złączenia (LeftJoin) na dwóch testowych zbiorach (users i transactions). Struktura zbiorów opisana jest poniżej.

1. **Users information** (id, email, language, location)
2. **Transaction information** (transaction-id, product-id, user-id, purchase-amount, item-description)

Zadanie polega na modyfikacji kodu, tak aby wyświetlić informację na temat ilości produktów zakupionych przez użytkowników w danych lokalizacjach.

Wynik działania:

- User_id, count(product_id), location