Rok akademicki 2012/2013

Politechnika Warszawska

Wydział Elektroniki i Technik Informacyjnych

Instytut Informatyki



praca dyplomowa magisterska

inż. Bartłomiej Jańczak

Tytuł pracy dyplomowej

Opiekun pracy

prof. nzw. dr hab. Marzena Kryszkiewicz

Ocena:

Podpis Przewodniczącego

Komisji Egzaminu Dyplomowego

Kierunek: Informatyka

Specjalność: Inżynieria Systemów Informatycznych

Data urodzenia: 1988.03.06

Data rozpoczęcia studiów: 2007.10.01

Życiorys

Podpis studenta

EGZAMIN DYPLOMOWY

Złożył egzamin dyplomowy w dniu 20\_\_ r

z wynikiem

Ogólny wynik studiów:

Dodatkowe wnioski i uwagi Komisji:

STRESZCZENIE

Streszczenie pracy w języku polskim.

Słowa kluczowe:

THESIS TITLE IN ENGLISH

Summary in English.

Keywords:

# 

# 1. Wprowadzenie

Współczesne systemy komputerowe agregują i generują ogromną ilość danych, która rośnie szybciej niż przewidywano jeszcze kilka lat temu. Ich gromadzenie i przechowywanie na nośnikach pamięci masowej nie stanowi problemów dla współczesnych systemów, natomiast działanie na takiej ilości danych, pomimo stale wzrastającej mocy obliczeniowej komputerów, wciąż jest wyzwaniem dla dzisiejszej informatyki. Zbiory danych same w sobie nie stanowią wielkiej wartości, jednakże rozsądnie wykorzystane mogą stać się cennym źródłem szczególnej wiedzy. Jej odkrywaniem znajduje się dziedzina informatyki zwana eksploracją danych, której ideą jest wykorzystanie komputera do znajdowania ukrytych dla człowieka prawidłowości w danych zgromadzonych w repozytoriach. Eksploracja danych jest czwartym etapem procesu odkrywania wiedzy, na który również składają się operacje selekcji, czyszczenia i transformacji danych a także analiza i interpretacja wyników.

Jednymi z najpopularniejszych metod eksploracji danych są grupowanie, klasyfikacja oraz odkrywanie asocjacji i sekwencji. Każda z metod ujawnia różnego rodzaju korelacje pomiędzy danymi, z czego wynika ich odmienne zastosowanie. W tej pracy skoncentrowałem się na zagadnieniu grupowania danych, które określane jest jako wyznaczanie zbiorów obiektów podobnych przy zachowaniu właściwości maksymalizacji podobieństwa obiektów należących do tych samych grup i minimalizacji podobieństwa obiektów należących do innych grup.

## Przegląd literatury

Grupowanie danych jest popularną metodą o wielu zastosowaniach, dlatego nie trudno o jej opis w literaturze. W przypadku algorytmów, na których skupiłem się w niniejszej pracy wyjątkowo przydatne okazały się artykuły naukowe.