

Zaawansowana Analiza Danych

Specjalność uzupełniająca Katedry Inżynierii Oprogramowania



KATEDRA
INŻYNIERII
OPROGRAMOWANIA



Kogo kształcimy?



Prowadzone przedmioty

Przedmioty specjalnościowe		w	l	p	ECTS	semestr	Egzamin
1	Platformy przetwarzania typu Big Data	15	30	15	4	1	+
2	Zaawansowane techniki eksploracji danych		30		2	1	
3	Zaawansowane technologie nierelacyjnych baz danych	15	30	30	5	2	



KATEDRA
INŻYNIERII
OPROGRAMOWANIA



Laboratorium przetwarzania danych typu Big Data

4 serwery tworzące klaster o następujących parametrach

1. CPU 128 rdzeni/ 256 wątków
2. HDD 22,7 TB
3. RAM 0,6 TB

Klaster ufundowany przez firmę INTEL



KATEDRA
INŻYNIERII
OPROGRAMOWANIA



Platformy przetwarzania typu Big Data

dr inż. Adam Przybytek
Program przedmiotu

- ▶ W ramach przedmiotu omówione zostaną 3 platformy programistyczne dostarczające wysokopoziomowe API do równoległego przetwarzania ogromnych zbiorów danych na klastrach komputerowych.
- ▶ Apache Hadoop (MapReduce)
 - ▶ przetwarzanie wsadowe
- ▶ Apache Storm
 - ▶ przetwarzanie danych strumieniowych w czasie rzeczywistym
- ▶ Apache Spark
 - ▶ in-memory framework do przetwarzania iteracyjnego i interaktywnego
 - ▶ integruje przetwarzanie wsadowe, przetwarzanie w czasie rzeczywistym, SQL, uczenie maszynowe oraz złożone analizy



KATEDRA
INŻYNIERII
OPROGRAMOWANIA



Platformy przetwarzania typu Big Data

Praktyka

Przykładowy kod Python zliczający wyrazy przy wykorzystaniu API Spark

```
textFile = sc.textFile("/data/gutenberg/ulysses.txt")
words = textFile.flatMap(lambda line: line.split(" "))
counts = words.map(lambda word: (word, 1))
               .reduceByKey(lambda a, b: a+b)
counts.saveAsTextFile("/wordcount_output")
```



KATEDRA
INŻYNIERII
OPROGRAMOWANIA



Zaawansowane techniki eksploracji danych

dr inż. Paweł Kaptłański

Data Science



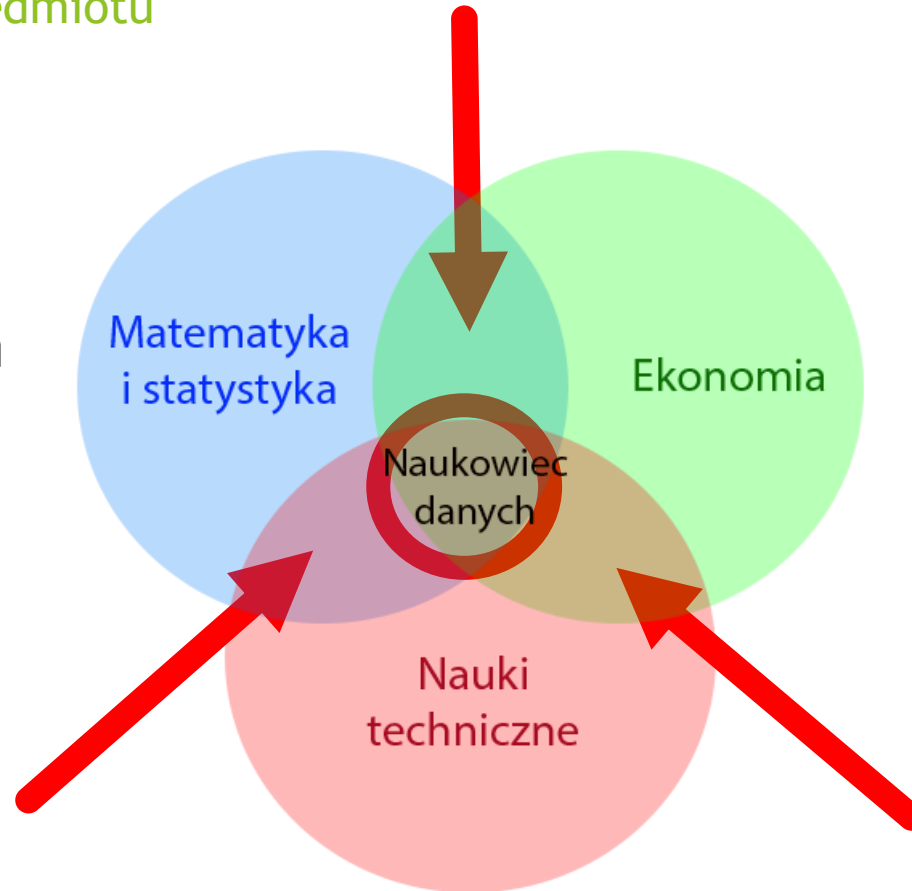
KATEDRA
INŻYNIERII
OPROGRAMOWANIA



Zaawansowane techniki eksploracji danych

Cel przedmiotu

Wprowadzenie studentów w tematykę (1) **eksploracji** oraz (2) **wizualizacji** dużych zbiorów danych z (3) użyciem **skalowanego klastra obliczeniowego** przy wykorzystaniu (4) nowoczesnych **języków funkcyjnych** oraz pakietów **statystycznych**.



Utworzony na podstawie

<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>



KATEDRA
INŻYNIERII
OPROGRAMOWANIA



Zaawansowane techniki eksploracji danych

Program przedmiotu

Przedmiot będzie odbywał się w trybie warsztatów na platformie TAP udostępnionej przez firmę Intel. Pokażemy następujące metody, używając metod statystycznych programowanych w językach funkcyjnych:

1. Przygotowywanie danych (wbrew pozorom jest to bardzo ważne)
2. Model rekomendacyjny (rekomendacja muzyki) oraz ocena jakości rekomendacji
3. Model drzewa decyzyjnego (prognozowanie zalesienia), dobór hiperparametrów modelu oraz wizualizacja danych
4. Wykrywanie anomalii (cyberataki)



Zaawansowane technologie nierelacyjnych baz danych

dr inż. Teresa Zawadzka
Program przedmiotu

Przedmiot prowadzony w formie warsztatów i projektów.



Hurtownie dużych danych



Bazy grafowe - algorytmy grafowe w rozproszeniu



Zapraszamy



KATEDRA
INŻYNIERII
OPROGRAMOWANIA

