

Uczenie maszynowe w bioinformatyce

Wprowadzenie

Tymon Rubel

Zakład Elektroniki Jądrowej i Medycznej
Instytut Radioelektroniki i Technik Multimedialnych PW

e-mail:	tymon.rubel@pw.edu.pl
Konsultacje:	piątek 14:00 - 16:00, pok. 5 lub Teams
Strona WWW:	www.ire.pw.edu.pl/~trubel/umb
Zespół Teams:	103A-IBIBM-MSP-UMB
Kurs LeOn:	103A-IBIBM-MSP-UMB # Uczenie maszynowe w bioinformatyce

Bioinformatyka jest dziedziną skupiającą się na użyciu technik informatycznych do rozwiązywania problemów z zakresu biologii i medycyny. Zajmuje się m. in.:

- gromadzeniem informacji biologicznych w bazach danych;
- wspomaganiem badań nad właściwościami biomolekuł, w tym przede wszystkim kwasów nukleinowych (DNA i RNA) oraz białek;
- porównywaniem genomów i ustalaniem relacji ewolucyjnych między organizmami;
- przetwarzaniem, analizą i interpretacją danych z eksperymentów wykorzystujących nowoczesne wielkoskalowe techniki pomiarowe dostępne biologii molekularnej;
- analizą szlaków metabolicznych i sposobu przekazywania sygnałów w komórkach;
- modelowaniem działania systemów biologicznych;
- wspomaganiem projektowania leków i terapii;
- ... wieloma innymi zagadnieniami.

Wprowadzenie: bioinformatyka

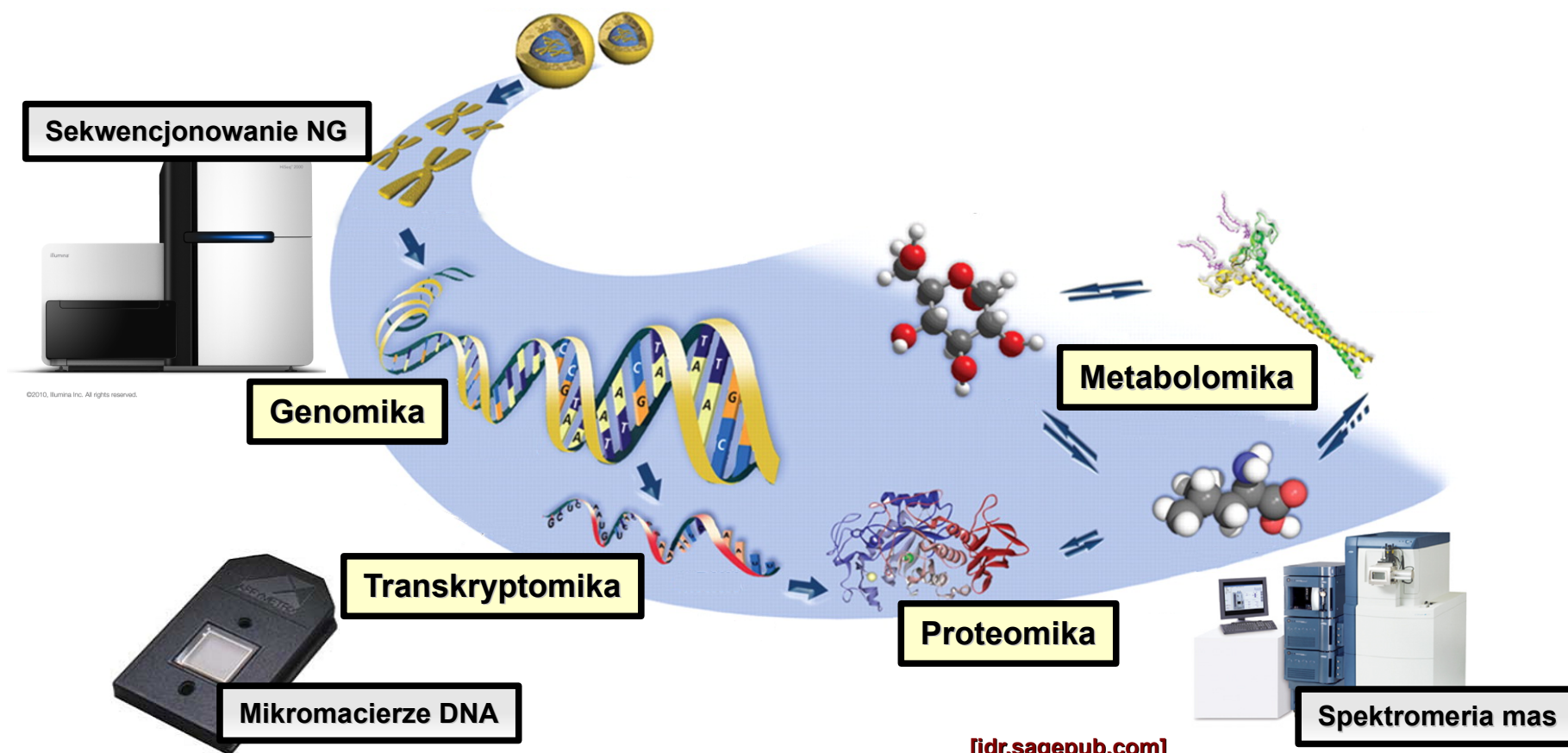
Bioinformatyka jest dziedziną skupiającą się na użyciu technik informatycznych do rozwiązywania problemów z zakresu biologii i medycyny. Zajmuje się m. in.:

- gromadzeniem informacji biologicznych w bazach danych;
- wspomaganiem badań nad właściwościami biomolekuł, w tym przede wszystkim kwasów nukleinowych (DNA i RNA) oraz białek;
- porównywaniem genomów i ustalaniem relacji ewolucyjnych między organizmami;
- przetwarzaniem, analizą i interpretacją danych z eksperymentów wykorzystujących nowoczesne wielkoskalowe techniki pomiarowe dostępne biologii molekularnej;
- analizą szlaków metabolicznych i sposobu przekazywania sygnałów w komórkach;
- modelowaniem działania systemów biologicznych;
- wspomaganiem projektowania leków i terapii;
- ... wieloma innymi zagadnieniami.

Wprowadzenie: wielkoskalowe metody pomiarowe

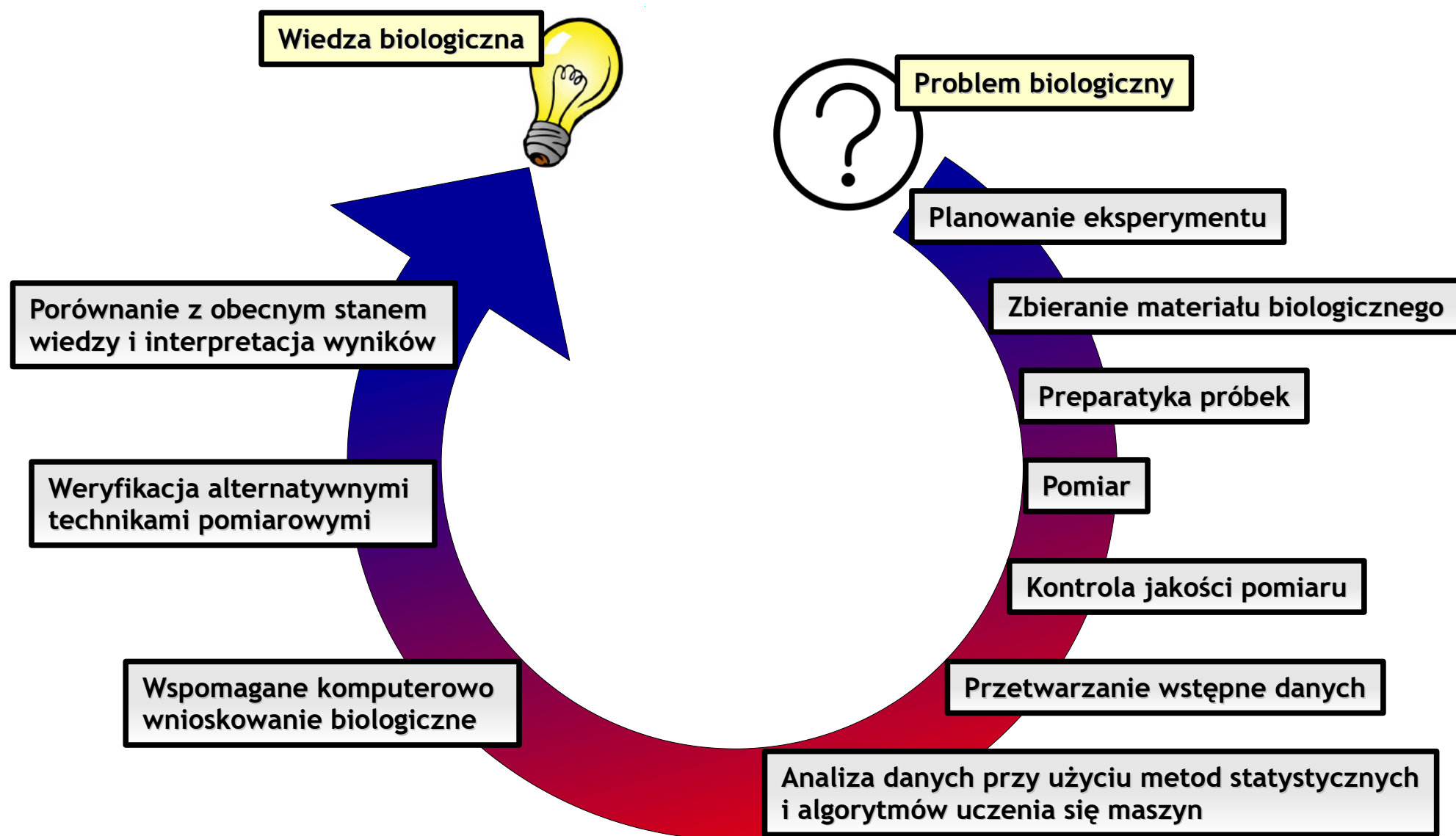
Wielkoskalowe (wysokowydajne, *high-throughput*) metody pomiarowe stały się w ciągu ostatnich dwóch dekad podstawowymi narzędziami biologii molekularnej.

Oferowana przez nie możliwość jednoczesnego badania dużych zbiorów obiektów biologicznych (np. całego genomu organizmu a nie tylko pojedynczych genów) jest głównym motorem rozwoju **genomiki**, **transkryptomiki**, **proteomiki** i **metabolomiki**, dziedzin wiedzy dążących do kompleksowego opisu procesów biochemicznych leżących u podstaw życia.

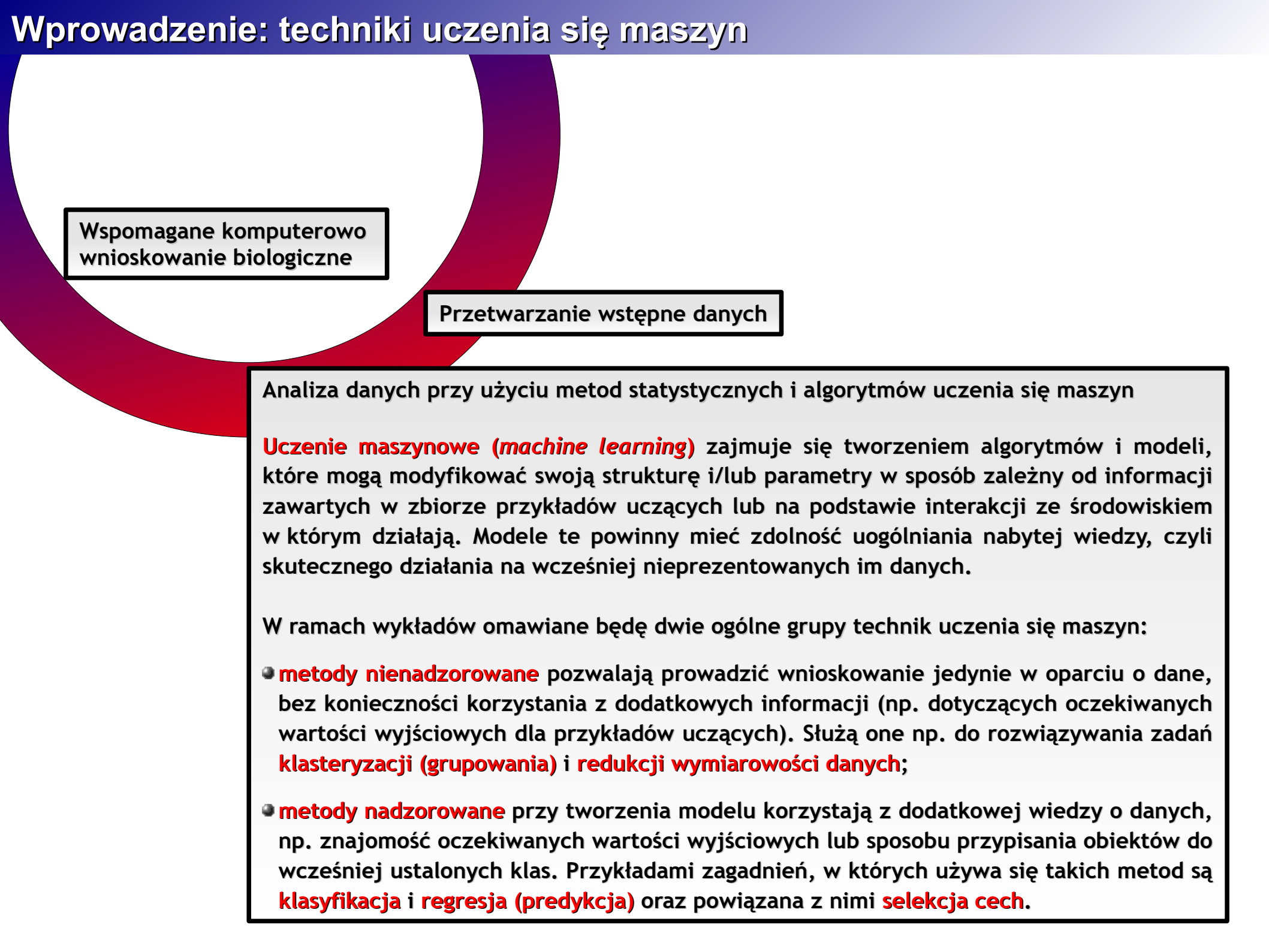


Wprowadzenie: wielkoskalowe metody pomiarowe

Eksperymenty korzystające z technik wielkoskalowych mają wieloetapowy przebieg oraz interdyscyplinarny charakter. W szczególności **wymagają komputerowej analizy uzyskanych danych**, które cechują się dużymi rozmiarami i złożonością.



Wprowadzenie: techniki uczenia się maszyn



Wspomagane komputerowo
wnioskowanie biologiczne

Przetwarzanie wstępne danych

Analiza danych przy użyciu metod statystycznych i algorytmów uczenia się maszyn

Uczenie maszynowe (*machine learning*) zajmuje się tworzeniem algorytmów i modeli, które mogą modyfikować swoją strukturę i/lub parametry w sposób zależny od informacji zawartych w zbiorze przykładów uczących lub na podstawie interakcji ze środowiskiem w którym działają. Modele te powinny mieć zdolność uogólniania nabytej wiedzy, czyli skutecznego działania na wcześniej nieprezentowanych im danych.

W ramach wykładów omawiane będę dwie ogólne grupy technik uczenia się maszyn:

- **metody nienadzorowane** pozwalają prowadzić wnioskowanie jedynie w oparciu o dane, bez konieczności korzystania z dodatkowych informacji (np. dotyczących oczekiwanych wartości wyjściowych dla przykładów uczących). Służą one np. do rozwiązywania zadań **klasteryzacji (grupowania)** i **redukcji wymiarowości danych**;
- **metody nadzorowane** przy tworzenia modelu korzystają z dodatkowej wiedzy o danych, np. znajomość oczekiwanych wartości wyjściowych lub sposobu przypisania obiektów do wcześniej ustalonych klas. Przykładami zagadnień, w których używa się takich metod są **klasyfikacja** i **regresja (predykcja)** oraz powiązana z nimi **selekcja cech**.