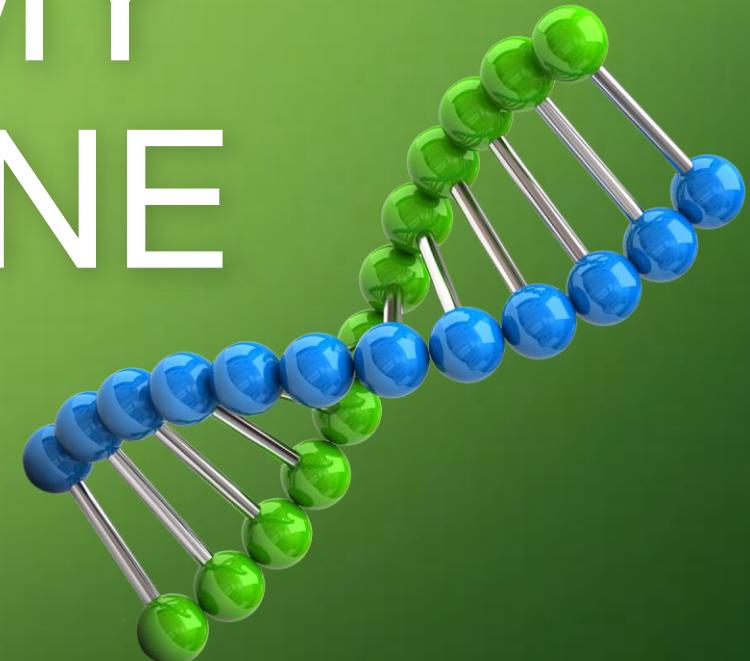


ALGORYTMY GENETYCZNE



ALGORYTMY GENETYCZNE

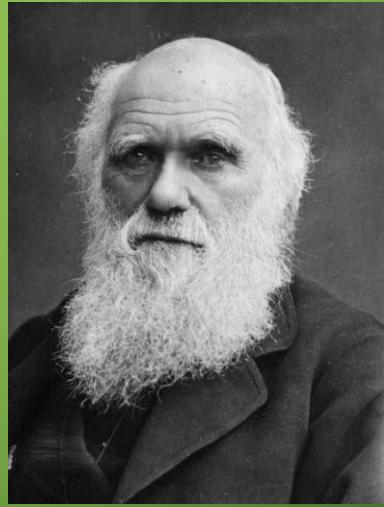
- Co to są algorytmy genetyczne?
- Podstawowe pojęcia algorytmów genetycznych
- Proste algorytmy genetyczne
- Kodowanie osobników i operacje genetyczne.



SYNTETYCZNA TEORIA EWOLUCJI



Gregor Mendel (1822-1884)



Charles Darwin (1809-1882)

Genetyka

Zasady
dziedziczenia

Ewolucja

Zasady doboru
naturalnego

NEODARWINIZM

DNA

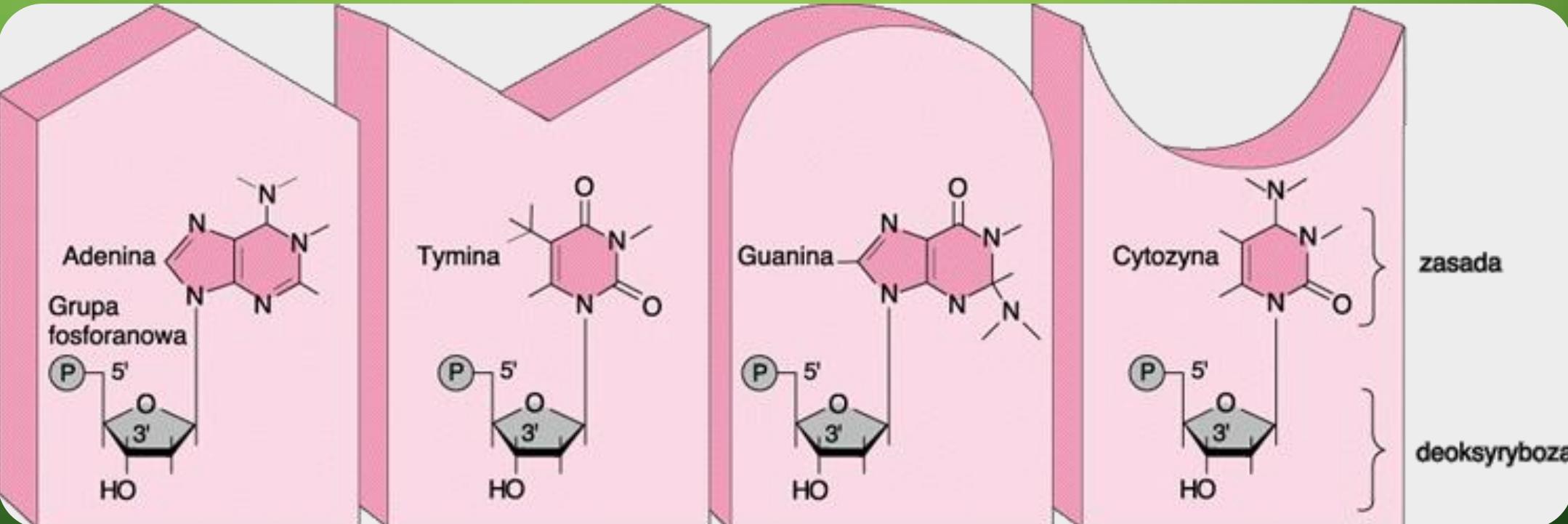
Syntetyczna teoria ewolucji

HISTORIA

- Pojęcie „genu” wprowadzono w 1910 roku na określenie abstrakcyjnej jednostki dziedziczenia, odpowiedzialnej za jakąś cechę danego gatunku.
- W 1869r. J. F. Miescher wyizolował z jąder komórkowych kwas deoksyrybonukleinowy - DNA.
- W 1953 roku na podstawie dostępnych fizycznych i chemicznych informacji J. Watson i F. Crick ustalili, że DNA ma strukturę podwójnej helisy.
- DNA jest polimerem składającym się z czterech cząsteczek zwanych nukleotydami: adenina (A), tymina (T), guanina (G), cytozyna (C).
- Informacja genetyczna zapisywana jest w DNA za pomocą kolejności ułożenia cząstek.



PODSTAWOWE JEDNOSTKI STRUKTURALNE DNA

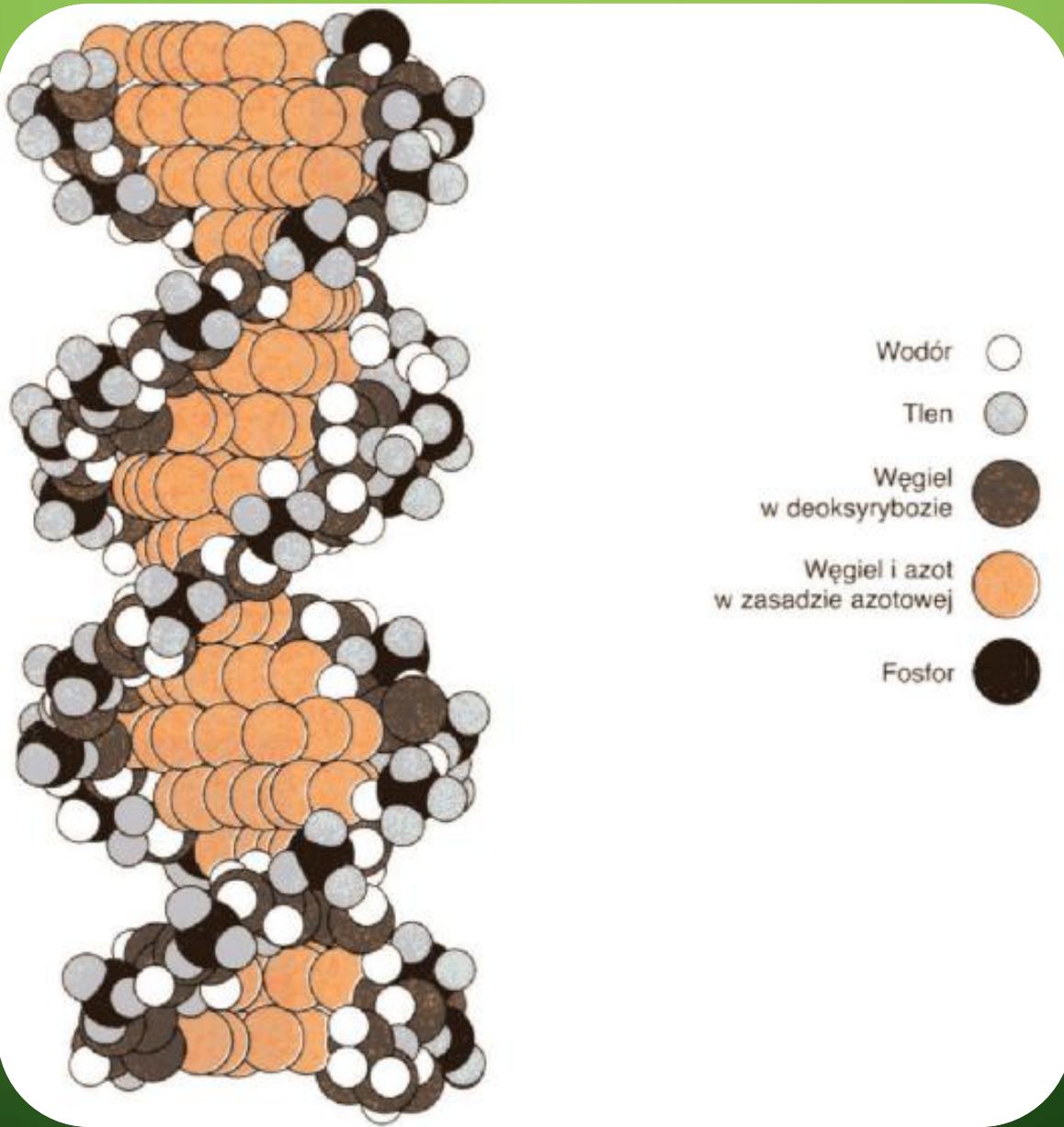


zasada

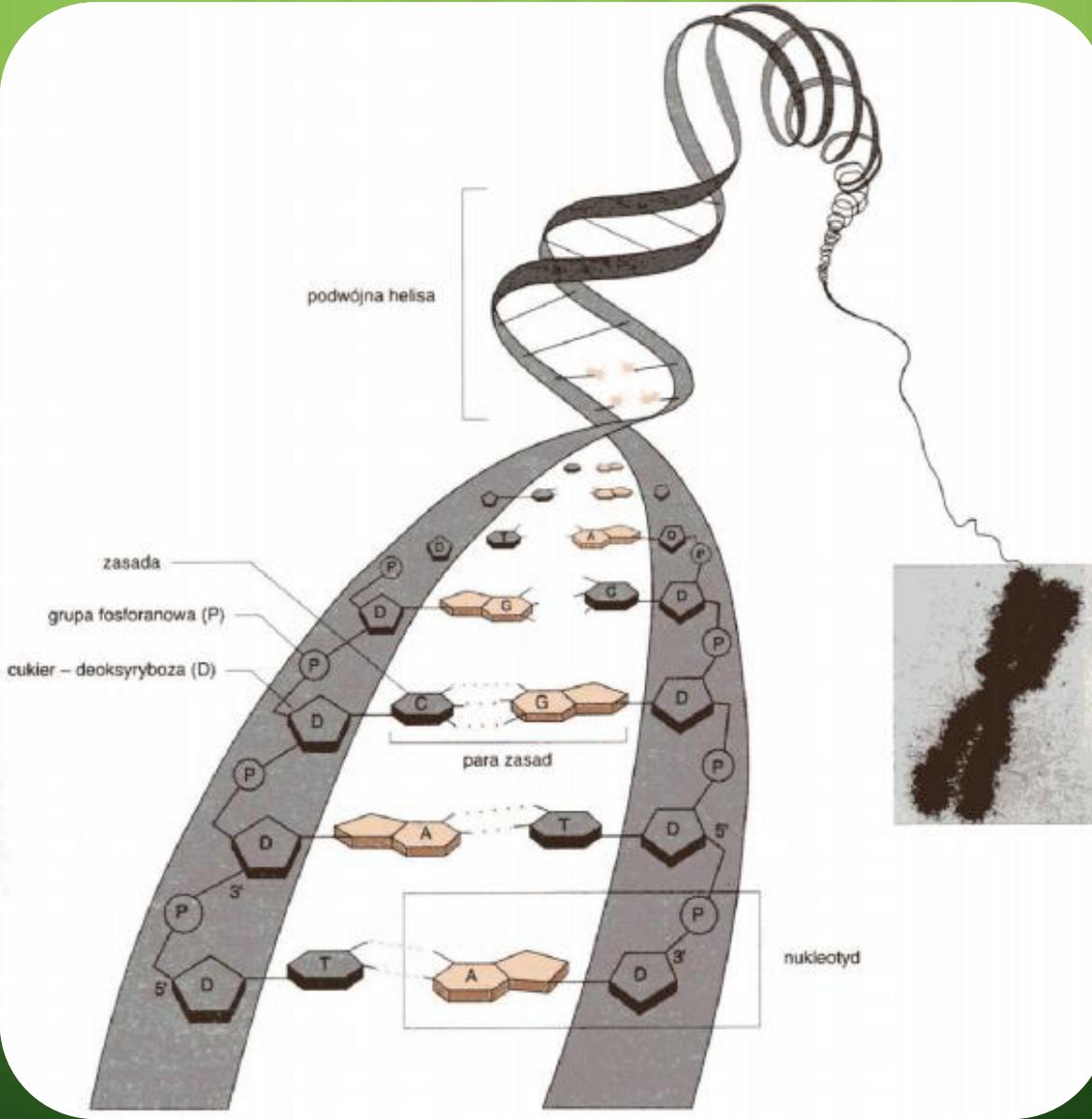
deoksyryboza

deoksalilrosa

MODEL PODWÓJNEJ HELISY DNA



UŁOŻENIE I POWIAZANIE NUKLEOTYDÓW W HELISIE DNA

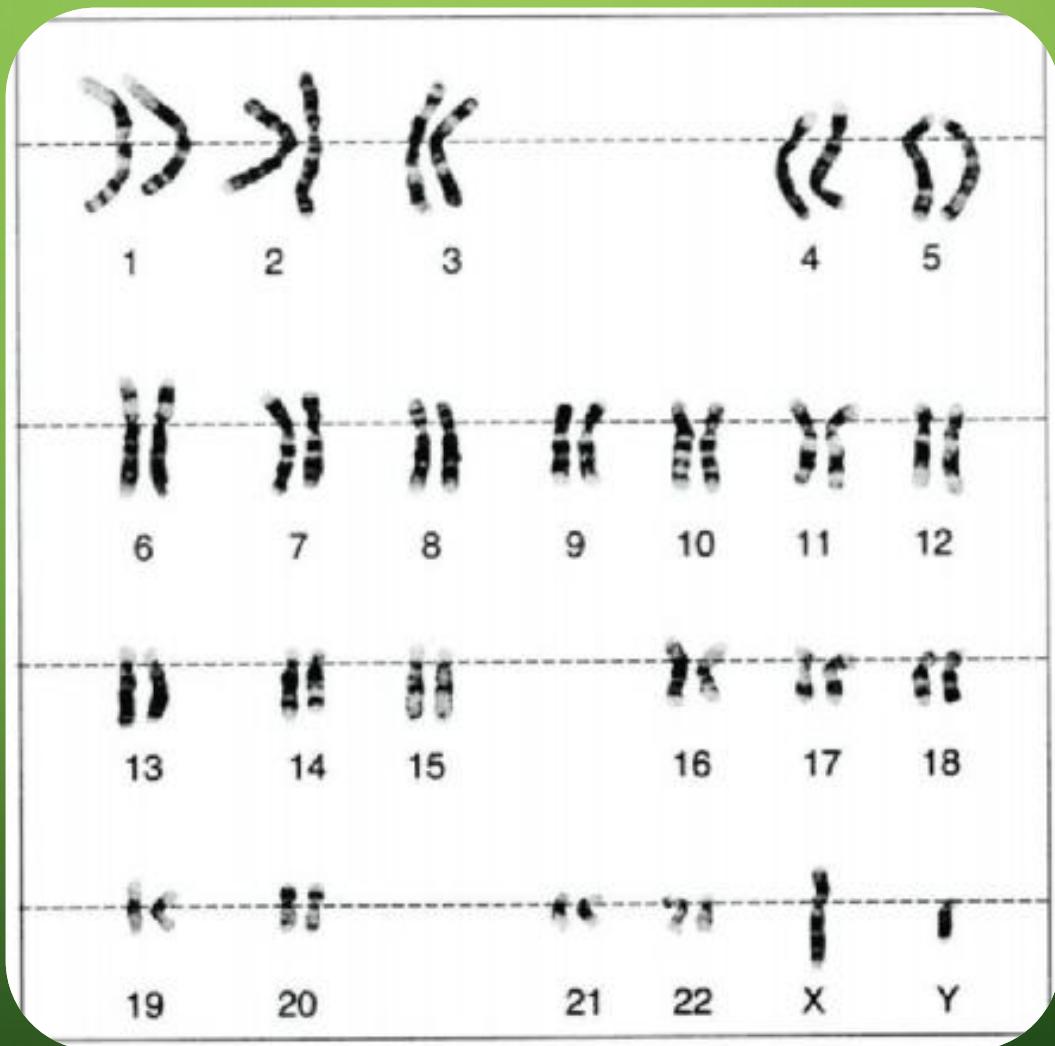


LUDZKI CHROMOSOM

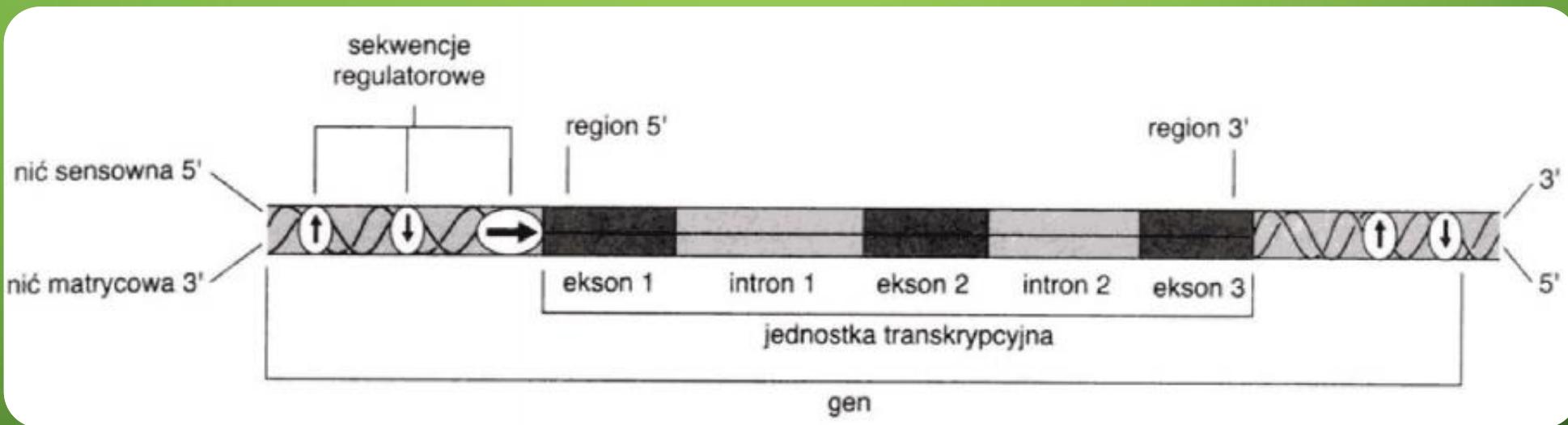


*Powiększenie
30 000 razy*

CHROMOSOMY JEDNEGO GATUNKU



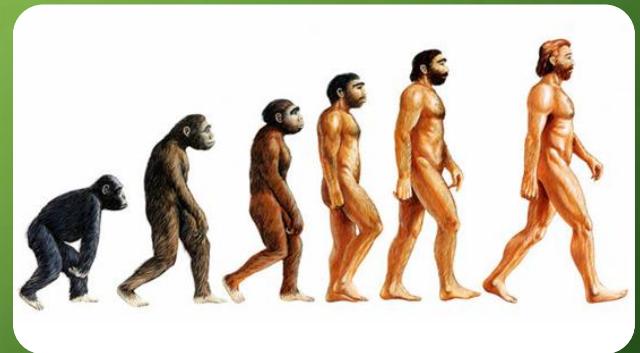
ELEMENTY DNA WCHODZĄCE W SKŁAD GENU



Typowy gen składa się z krótkich odcinków kodujących, czyli eksonów, poprzegradzanych odcinkami niekodującymi, zwanych intronami.

TEORIA EWOLUCJI

- Teoria ewolucji jest podstawą teoretyczną współczesnej biologii, dostarcza ona odpowiedzi na pytania: jak powstały przeróżne formy **organizmów**, jak powstały różne rodzaje organizmów.
- Wyjaśnienia tych zagadnień podał **Karol Darwin** formułując teorię ewolucji, według niej głównym procesem kształtującym właściwości organizmów jest **dobór naturalny**, ważnym zagadnieniem teorii jest teza, że wszystkie współcześnie żyjące organizmy połączone są więzami wspólnego pochodzenia.



TEORIA EWOLUCJI

- Dobór naturalny wynika z właściwości organizmów żywych: ich śmiertelności, zdolności rozmnażania się, dziedziczenia cech, jest także wynikiem ograniczonych zasobów środowiska.
- Rośliny, zwierzęta i mikroorganizmy różnią się między sobą, a różnice te są zdeterminowane genetycznie, jeśli różnice te nie tylko są dziedziczne, lecz także wpływają na przeżywanie i rozmnażanie się tych organizmów, to z pokolenia na pokolenie zmieniać się będzie frekwencja różnych form w populacji

TEORIA EWOLUCJI

- Przyjęcie darwinowskiej teorii ewolucji wiąże się z przyjęciem doboru naturalnego jako najistotniejszego mechanizmu kształtującego właściwości gatunków.
- Kumulatywne działanie doboru może być bardzo efektywne, dając w długim czasie, jakim dysponowała ewolucja, struktury bardzo skomplikowane i mało prawdopodobne.

GENEZA ALGORYTMÓW GENETYCZNYCH

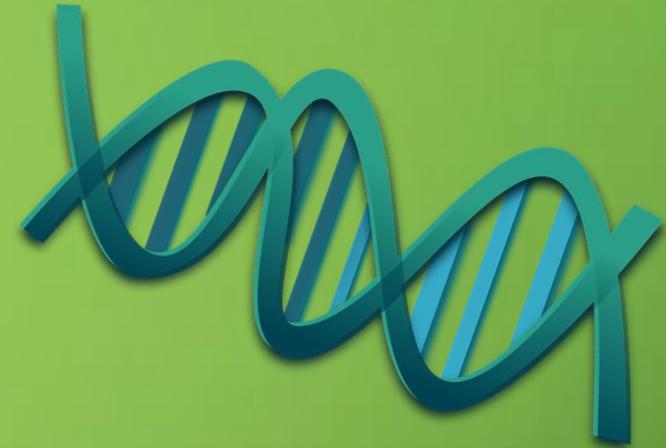
John Holland, Uniwersytet Michigan



- Opisał i wyjaśnił istotę procesów adaptacyjnych występujących w przyrodzie
- Stworzył oprogramowanie, odtwarzające podstawowe mechanizmy rządzace systemami biologicznymi

**J.Holland 1975, Adaptation in Natural and Artificial Systems,
The University of Michigan Press.**

ALGORYTMY GENETYCZNE



- Algorytmy genetyczne są to algorytmy poszukiwania, które w rozwiązywaniu zadań stosują zasady ewolucji i dziedziczności, posługując się populacją potencjalnych rozwiązań, zawierającą pewien proces selekcji, oparty na dopasowaniu osobników, i pewne operatory genetyczne. Każde rozwiązanie ocenia się na podstawie pewnej miary jego dopasowania - funkcja przystosowania (celu). Nową populację, w kolejnej iteracji, tworzy się przez selekcje osobników najlepiej dopasowanych.

ALGORYTMY GENETYCZNE

Elementarny algorytm genetyczny jest skonstruowany z trzech następujących operacji:

- **reprodukcia** - jest procesem, w którym indywidualne ciągi kodowe zostają powielone w stosunku zależnym od wartości, jakie przybiera dla nich funkcja celu, powstaje pula rodzicielska.
- **krzyżowanie** - proces, który polega na losowym skojarzeniu ciągów z puli rodzicielskiej w pary, a następnie krzyżowaniu tzn. wymianie części informacji zawartej w ciągach rodzicielskich i utworzenie potomstwa.
- **mutacja** - polega na sporadycznej, przypadkowej zmianie wartości elementu ciągu kodowego.

CO TO SĄ ALGORYTMY GENETYCZNE?

- Algorytmy genetyczne odwzorowują naturalne procesy ewolucyjne zachodzące w czasie, których celem jest maksymalne dopasowanie osobników do istniejących warunków życia.
- Algorytm genetyczny stanowi wzorowaną na naturalnej ewolucji metodę rozwiązywania problemów, głównie zagadnień optymalizacyjnych. Algorytmy genetyczne są procedurami przeszukiwania opartymi na mechanizmach doboru naturalnego i dziedziczenia. Korzystają z ewolucyjnej zasady Przeżycia osobników najlepiej przystosowanych

SCHEMAT BLOKOWY ALGORYTMU GENETYCZNEGO

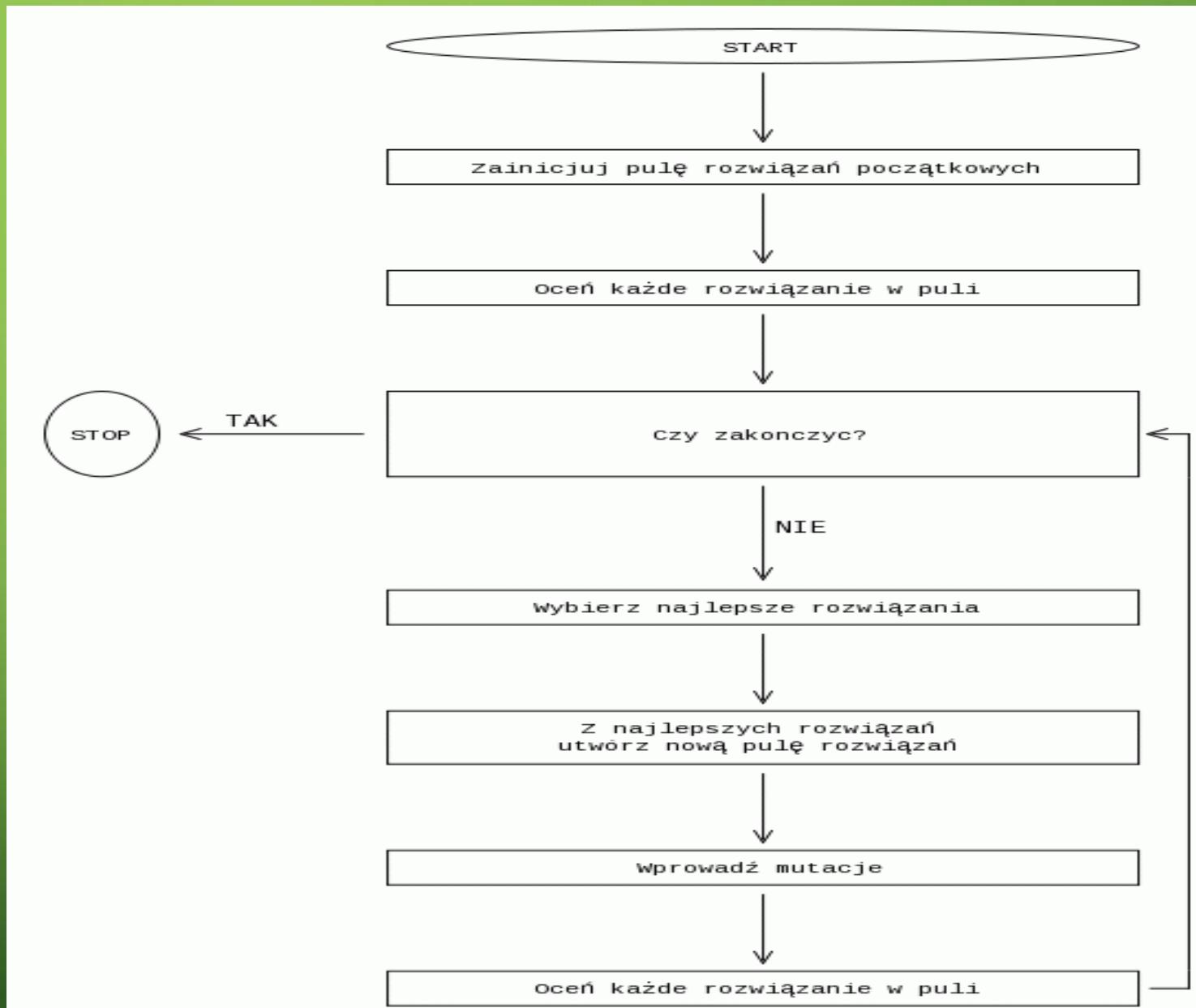
Przyjrzyjmy się teraz kolejnym krokom algorytmu.

Jak z najlepszych rozwiązań tworzyć nowe? Poprzez krzyżowanie. A więc jeżeli nowe rozwiązanie R_{nowe} będzie tworzone z dwóch innych R_1 oraz R_2 to krzyżowanie należy przeprowadzić następująco:

Wylosuj liczbę od 0 do 1, jeżeli wylosowana liczba jest mniejsza od 0.5 to do R_{nowe} weź $param_1$ z R_1 w przeciwnym wypadku weź go z R_2 .

Wylosuj liczbę od 0 do 1, jeżeli wylosowana liczba jest mniejsza od 0.5 to do R_{nowe} $param_2$ z R_1 w przeciwnym wypadku weź go z R_2 .

...
Wylosuj liczbę od 0 do 1, jeżeli wylosowana liczba jest mniejsza od 0.5 to do R_{nowe} $param_n$ z R_1 w przeciwnym wypadku weź go z R_2 .



PODSTAWOWE POJĘCIA ALGORYTMÓW GENETYCZNYCH

- **Populacja**- zbiór osobników o określonej liczebności
- **Osobnikami** populacji w algorytmach genetycznych są zakodowane w postaci chromosomów
- **Chromosomy** — inaczej łańcuchy lub ciągi kodowe — to uporządkowane ciągi genów
- **Gen** — najmniejsza składowa chromosomu, decydująca o dziedziczności jednej lub kilku cech

PODSTAWOWE POJĘCIA ALGORYTMÓW GENETYCZNYCH CD.

- **Genotyp** czyli struktura, to zespół chromosomów danego osobnika
- **Fenotyp** jest zestawem wartości odpowiadających danemu genotypowi
- **Allel** to wartość danego genu, określana też jako wartość cechy lub wariant cechy.
- **Locus** wskazuje miejsce położenia danego genu w łańcuchu, czyli chromosomie
- **Funkcja przystosowania** jest miarą przystosowania (dopasowania) danego osobnika w populacji



PROSTE ALGORYTMY GENETYCZNE

- W roku 1975 John Holland zaproponował algorytm, którego zadaniem było modelowanie procesu ewolucji. Schemat tego algorytmu, zwanego obecnie prostym algorytmem genetycznym (SGA)

SCHEMAT DZIAŁANIA

- Algorytm przetwarza populacje:
- P_t - zwaną populacją bazową.
- O_t - zwaną populacją potomną.
- T_t - zwana populacją tymczasową, w której przechowywane są kopie osobników z P_t .
- W każdej z nich zawarta jest jednakowa liczba osobników.
- W kroku „inicjacji P_0 ” populacja bazowa wypełniana jest losowo wygenerowanymi osobnikami.
- Dla każdego osobnika obliczana jest na etapie oceny wartość funkcji przystosowania.
- Po przygotowaniu populacji bazowej P_0 przystępuje się do głównej pętli programu, w której zdefiniowany jest proces sztucznej ewolucji.

Prosty algorytm genetyczny

begin

$t := 0$

inicjacja P^0

ocena P^0

while (not warunek stopu) do

begin

$T^t := \text{reprodukcja } P^t$

$O^t := \text{krzyżowanie i mutacja } T^t$

ocena O^t

$P^{t+1} = O^t$

$t := t + 1$

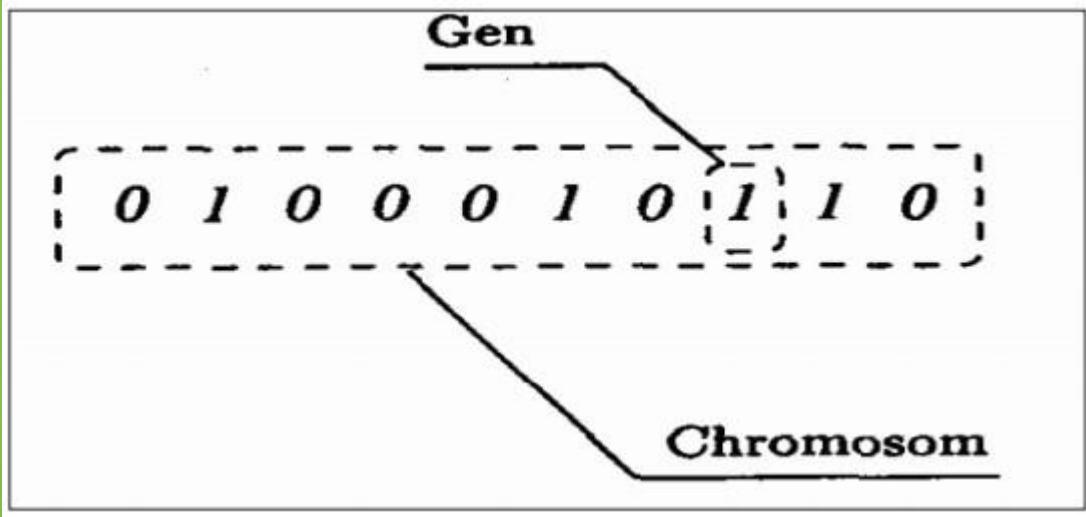
end

end

- Warunek zatrzymania jest spełniony gdy:
- Pojawił się osobnik o odpowiednio dużej - wymaganej wartości funkcji przystosowania.
- Została wykona założona Liczba generacji
- Generacja – jedna iteracja - obieg głównej pętli algorytmu
- Stan populacji w chwili t określa się mianem pokolenia P_t

KODOWANIE OSOBNIKÓW I OPERACJE GENETYCZNE

- Schemat algorytmu zaproponowany przez Holland'a abstrahuje od przyjętego sposobu kodowania osobników i wykonywania operacji krzyżowania i mutacji



- Jednakże algorytm genetyczny najbardziej jest znany z kodowaniem binarnym. W tej wersji chromosom jest więc n-elementowym wektorem genów, z których każdy jest zerem lub jedynką.

INNE METODY KODOWANIA

- **Kod Gray'a** jest jedną z alternatyw dla prostego kodowania binarnego. Ma on tę zaletę, że chromosomy odpowiadające dwóm kolejnym liczbom całkowitym różnią się tylko jednym bitem.
- **Kodowanie logarytmiczne** jest kolejnym przykładem modyfikacji kodowania binarnego. Ten sposób kodowania jest stosowany celem zmniejszenia długości chromosomów w algorytmie genetycznym.
- **Wprowadzenie zapisu zmiennopozycyjnego** można uznać za kolejną modyfikację metod kodowania. Każdy chromosom jest zapisany jako wektor liczb zmiennopozycyjnych. Dokładność takiego podejścia jest zwykle znacznie wyższa niż przy kodowaniu binarnym. Reprezentacja zmiennopozycyjna może objąć całkiem duże lub nawet nieznane dziedziny.

ZASADA MINIMALNEGO ALFABETU

- Jedną z podstawowych zasad przy wyborze kodu jest zasada minimalnego alfabetu. Mówiąc o niej, mówimy, że należy wybrać najmniejszy alfabet, w którym rozpatrywane zadanie wyraża się w sposób naturalny.

METODY SELEKCJI

- W algorytmach genetycznych wyróżniamy etap selekcji, który z bieżącej populacji osobników wybiera te o największej wartości funkcji przystosowania do populacji rodzicielskiej.
- **Selekcja metodą koła ruletki** Jest to metoda prosta i daje w miarę zadowalające wyniki. Ma jednak wady.
 - Jedną z nich jest możliwość stosowania metody ruletki tylko do jednej klasy zadań, tzn. tylko do maksymalizacji (lub tylko do minimalizacji). Takie ograniczenie jest niewątpliwie jej wadą.
 - Bardziej istotną wadą metody ruletki jest zbyt wcześnie eliminowanie z populacji osobników o bardzo małej wartości funkcji przystosowania. Może to prowadzić do zbyt wcześnie zbieżności algorytmu

INNE METODY SELEKCJI

- **Selekcja turniejowa** polega na podzieleniu populacji na podgrupy k-elementowe (k to rozmiar turnieju – zwykle 2 lub 3) i wyborze z każdej podgrupy osobnika o najlepszym przystosowaniu.
- **W selekcji rankingowej** osobniki populacji są ustawiane kolejno w zależności od funkcji przystosowania – tzn. od najlepiej do najgorzej przystosowanego. Każdemu osobnikowi przyporządkowana jest liczba zwana rangą i oznaczająca jego pozycję na liście
- **W strategii elitarnej** nacisk położony jest na zachowanie w kolejnych iteracjach najlepiej przystosowanych osobników. W klasycznym algorytmie genetycznym możliwa jest sytuacja, w której do nowej populacji nie wejdą najlepsze osobniki. Model elitarny ma za zadanie do tego nie dopuścić

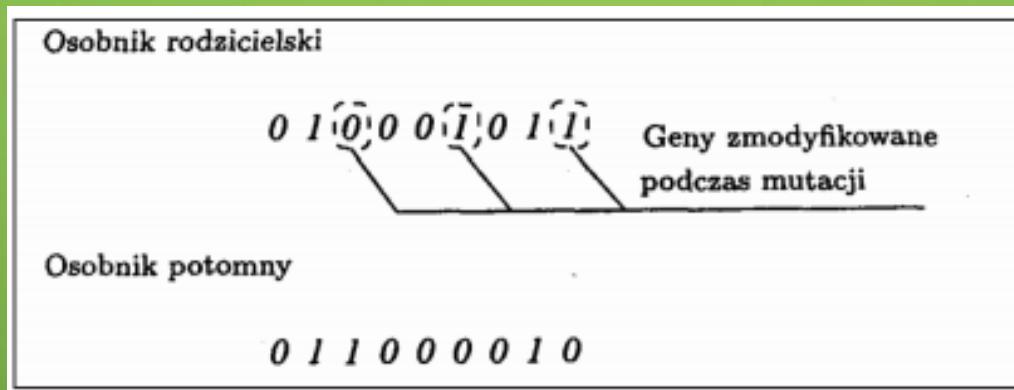
INNE PODZIAŁY METOD SELEKCJI

- statyczne i dynamiczne
- wygaszające i zachowujące
- pokoleniowe i w locie

SKALOWANIE FUNKCJI CELU

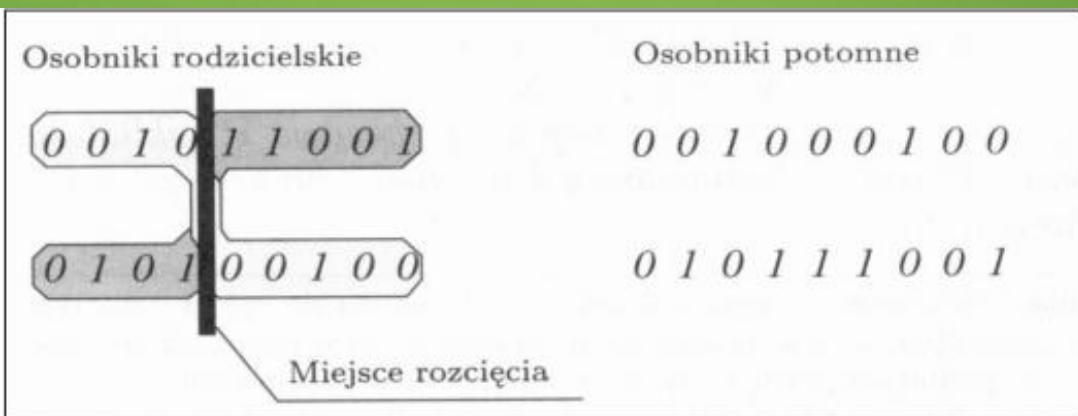
- Stosuje się, aby zapobiec przedwczesnej zbieżności algorytmu genetycznego. Przedwczesna zbieżność algorytmu polega na tym, że najlepsze ale jeszcze nie optymalne chromosomy dominują w populacji.
- **Skalowanie liniowe** (ang. linear scaling) :
 - $f' = af + b$
- **Skalowanie-obcinające typu sigma** (ang. sigma truncation). :
 - $f' = f + (f - c \cdot \sigma)$
- **Skalowanie potęgą:**
 - $f' = f^k$

- **Mutacja osobnika** jest wykonywana dla każdego genu osobno – polega ona na tym, że z prawdopodobieństwem p_m , którego wartość jest parametrem algorytmu, wartość genu zmienia się na przeciwną



- **Krzyżowanie osobników** – biorą w nim udział dwa osobniki rodzicielskie. Operacja wzorowana na procesie przemian w DNA, zachodzących podczas rozmnażania generatywnego

- Chromosomy rodzicielskie rozcinane są na dwa fragmenty. Pierwszy fragment pierwszego chromosomu jest sklejany z drugim fragmentem drugiego chromosomu. Podobnie pierwszy fragment drugiego chromosomu sklejany jest z drugim fragmentem pierwszego. Powstałe w ten sposób łańcuchy tworzą chromosomy osobników potomnych



- Wszystkie chromosomy zawierają jednakową liczbę bitów, dlatego też rozcięcie musi być wykonane w tym samym miejscu dla obu chromosomów
- **Miejsce rozcięcia** wybierane jest w sposób losowy z rozkładem równomiernym
- **Sposób reprodukcji** – populacja tymczasowa T^t tworzona jest w sposób preferujący osobniki lepiej przystosowane

Rodzic

011111000110111110011001010111

Potomek

011111000110101110011001010111

ZAKRES MUTACJI

Rodzic 1

0111110001101 11100011001010111

Rodzic 2

1001000000011 10000001000000011

POZYCJA CIĘCIA

Potomek 1

0111110001101 10000001000000011

Potomek 2

1001000000011 11100011001010111

Rodzic 1

0111110001101|011|11100011010111

Rodzic 2

100100000011101|10000000000011

POZYCJE CIĘCIA

Potomek 1

0111110001101|101|11100011010111

Potomek 2

100100000011011|10000000000011

PRZYKŁAD

Produkcja których lodów
jest najbardziej
opłacalna?

- Niech dany będzie
wytwórcą lodów dążący
do maksymalizacji
swych dochodów.
Wytwórcza może
decydować o smaku
lodów, ich wielkości i
obecności różnych
dodatków



DANE TWORZĄCE CHROMOSOM

- Smaki:
 - Waniliowe
 - Cytrusowe
 - Czekoladowe
 - Karmelowe
- Dodatki
 - Są
 - Nie ma
- Rozmieszczenie:
 - Duże
 - Małe

DANE POTRZEBNE W FUNKCJI PRZYSTOSOWANIA

- Cena:
- Mały 10zł
- Dodatki +5zł
- Duży x2

Popyt:

- Lody waniliowe: 6, z dodatkami: 6
- Lody cytrusowe: 10, z dodatkami: 6
- Lody czekoladowe: 8, z dodatkami: 12
- Lody karmelowe: 10, z dodatkami: 12



ALGORYTM GENETYCZNY

- Lody waniliowe: 00
- Lody cytrusowe: 01
- Lody czekoladowe: 10
- Lody karmelowe: 11
- Nie ma dodatków: 0, są dodatki: 1.
- Lody małe: 0, lody duże: 1.
- Np.: 0011 - lody waniliowe w polewie czekoladowej, duże

- Krzyżowanie będziemy przeprowadzać z prawdopodobieństwem 50%, mutacją: 5%(dotyczy to każdego bitu z osobna).
- Wielkość populacji równa się 4: Np.: 0111, 1000, 1000, 0100

Nr	Osobnik	Wartość przystosowania	% udział	suma udziałów, narastająco
1	0111	90	25%	25%
2	1000	80	23%	48%
3	1000	80	23%	71%
4	0100	100	29%	100%
	Suma:	350	100%	

- Wylosowaliśmy liczby: 0.22, 0.01, 0.15, 0.88. Odpowiadają one osobnikom: 1, 1, 1 i 4
- 0111, 0111, 0111, 0100

- A po mutacji:
- 0111, 0101, 0111, 0101

IMPLEMENTACJA ALGORYTMU GENETYCZNEGO W C++

```
#include <iostream>
using namespace std;

int main(void) {
    int tmpInput = 0;

    cout << "Hello World!" << endl;

    // Wait for input
    cin >> tmpInput;

    } return 0;
```

- Przeważnie stosuje się **reprodukcję proporcjonalną lub ruletkową**
- Określa się zmienną losową na zbiorze dla elementów populacji P^t , o rozkładzie danym wzorem

$$p_r(X) = \frac{\phi(X)}{\sum Y \in P^t(Y)}$$

, gdzie X oznacza osobnika, a $\phi(X)$ jego wartość przystosowania

- Następnie dokonuje się wielokrotnego próbkowania tej zmiennej losowej, przy czym za każdym razem wylosowany osobnik jest kopiowany do populacji tymczasowej T^t

KONIE
C

