Zawartość

[Złożoność obliczeniowa: 3](#_Toc486115353)

[Szacowanie wydajności: 3](#_Toc486115354)

[Przykładowe zadanie: 4](#_Toc486115355)

[Algorytmy sortownia: 5](#_Toc486115356)

[Sortowanie bąbelkowe: (bubble sort) 5](#_Toc486115357)

[Zasada działania: 5](#_Toc486115358)

[Zalety: 5](#_Toc486115359)

[Wady: 5](#_Toc486115360)

[Implementacja: 5](#_Toc486115361)

[Rozrysowanie: 6](#_Toc486115362)

[Sortowanie przez wstawianie (karciane): (insert sort) 6](#_Toc486115363)

[Zasada działania: 6](#_Toc486115364)

[Zalety: 6](#_Toc486115365)

[Wady: 6](#_Toc486115366)

[Implementacja: 6](#_Toc486115367)

[Rozrysowanie: 7](#_Toc486115368)

[Sortowanie przez wybieranie: (selection sort) 7](#_Toc486115369)

[Zasada działania: 7](#_Toc486115370)

[Zalety: 7](#_Toc486115371)

[Wady: 7](#_Toc486115372)

[Implementacja: 8](#_Toc486115373)

[Sortowanie przez zliczanie: (counting sort) 8](#_Toc486115374)

[Zasada działania: 8](#_Toc486115375)

[Zalety: 8](#_Toc486115376)

[Wady: 8](#_Toc486115377)

[Implementacja: 9](#_Toc486115378)

[Sortowanie przez scalanie: (merge sort) 9](#_Toc486115379)

[Zasada działania: 9](#_Toc486115380)

[Zalety: 9](#_Toc486115381)

[Wady: 9](#_Toc486115382)

[Implementacja: 9](#_Toc486115383)

[Szybkie sortowanie: (quick sort) 11](#_Toc486115384)

[Zasada działania: 11](#_Toc486115385)

[Zalety: 11](#_Toc486115386)

[Wady: 12](#_Toc486115387)

[Implementacja: 12](#_Toc486115388)

[Podsumowanie: 14](#_Toc486115389)

[Tabela złożoności (szeregowanie algorytmów): 14](#_Toc486115390)

[Algorytmy stabilne: 14](#_Toc486115391)

[Algorytmy niestabilne: 14](#_Toc486115392)

[Definicje: 14](#_Toc486115393)

[Sortowanie w miejscu: 14](#_Toc486115394)

[Sortowanie stabilne: 14](#_Toc486115395)

[Złożoności obliczeniowe: 14](#_Toc486115396)

[Stos 15](#_Toc486115397)

[Odwrotna notacja polska 16](#_Toc486115398)

[Kolejka 16](#_Toc486115399)

[Haszowanie 16](#_Toc486115400)

[Haszowanie liniowe i obsługa kolizji 16](#_Toc486115401)

[Haszowanie łańcuchowe 18](#_Toc486115402)

[Haszowanie kwadratowe 18](#_Toc486115403)

[Pojęcia związane z haszowaniem 19](#_Toc486115404)

[Drzewo 20](#_Toc486115405)

[przeszukiwanie 20](#_Toc486115406)

[Podstawowe pojęcia 20](#_Toc486115407)

[Grafy 21](#_Toc486115408)

[Graf skierowany (zorientowane) 21](#_Toc486115409)

[Graf nieskierowaniy (niezorientowane) 21](#_Toc486115410)

[Macierze sąsiedztwa 21](#_Toc486115411)

[Listy sąsiedztwa 21](#_Toc486115412)

[Macierze incydencji 21](#_Toc486115413)

[Przeszukiwanie BFS (w szerz) 21](#_Toc486115414)

[Przeszukiwanie DFS (w głąb) 22](#_Toc486115415)

[Złożoność pamięciowa grafu 22](#_Toc486115416)

[Kompresja danych 22](#_Toc486115417)

[Podstawowe dane 22](#_Toc486115418)

[Algorytm Huffmana 23](#_Toc486115419)

[Średnia długość słowa 24](#_Toc486115420)

[Entropia 24](#_Toc486115421)

[Redundancja 25](#_Toc486115422)

[Kompresja RLE 25](#_Toc486115423)

[LZ77 26](#_Toc486115424)

[Stringologia: 26](#_Toc486115425)

[Długość Hamminga 26](#_Toc486115426)

[Bit parzystości 26](#_Toc486115427)

[Kontrola parzystości 27](#_Toc486115428)

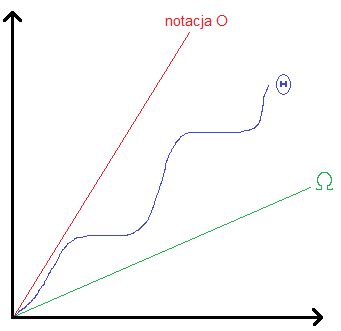
[LCS – Longest common subsequence 28](#_Toc486115429)

[Odległość Levenshteina (edycyjna) 30](#_Toc486115430)

**Ważne!**

# Złożoność obliczeniowa:

## Szacowanie wydajności:



Rys. szacowanie wydajności

Wzór do szacowania [teta]

### Przykładowe zadanie:

Udowodnij używając definicji, że

Liczenie ograniczenia górnego:

Liczenie ograniczenia dolnego:

Szukamy najmniejszego naturalnego n, dla którego równanie jest wiesze od 0

1) metoda: podstawianie po kolei

…

Więc

2) metoda: przyrównanie do zera

Zawsze dodajemy do wyniku 1. Więc

Odp.: Udowodniłem, że funkcja f(n) jest ograniczona funkcją dla

# Algorytmy sortownia:

## Sortowanie bąbelkowe: (bubble sort)

### Zasada działania:

Porównuje sąsiadujące ze sobą elementy, aż cała tablica nie będzie posortowana.

Złożoność obliczeniowa:

Złożoność pamięciowa:

### Zalety:

- sortowanie w miejscu

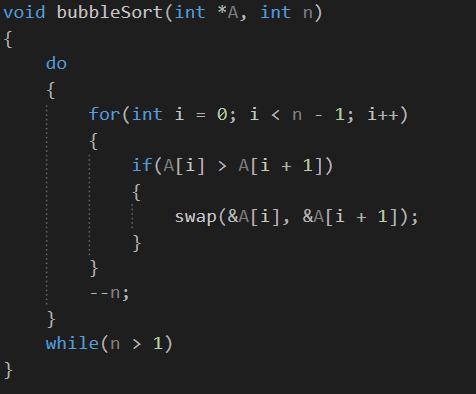
- stabilne sortowanie

- łatwy w implementacji

### Wady:

- wolny

### Implementacja:



Rys. sortowanie bąbelkowe

### Rozrysowanie:

Dla danych 8 4 5 22 30 66 50

0) 8 4 5 22 30 66 50

Przy pierwszym przejściu 8 zamienia się z 4 miejscami, a następnie 8 z 5

1) 4 5 8 22 30 66 50

## Sortowanie przez wstawianie (karciane): (insert sort)

### Zasada działania:

Łap po kolei elementy i przesuwaj je tak długo w lewo aż nie znajdziesz elementu od niej większego (mniejszego).

Złożoność obliczeniowa:

Złożoność pamięciowa:

### Zalety:

- stabilny

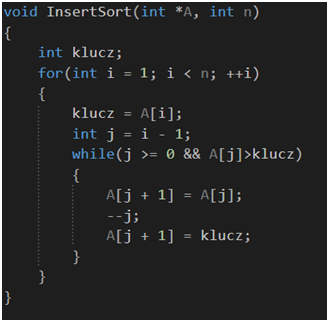
- szybki dla małej ilości danych

- szybki dla wstępnie posortowanych danych

### Wady:

- wolny dla rzeczywistych danych (średniego przypadku)

### Implementacja:



Rys. sortowanie przez wstawianie

### Rozrysowanie:

Dla danych 6 5 3 1 8 7 2 4

0) 6 5 3 1 8 7 2 4

1) 5 6 3 1 8 7 2 4

2) 3 5 6 1 8 7 2 4

3) 1 3 5 6 8 7 2 4

4) 1 3 5 6 8 7 2 4

5) 1 3 5 6 7 8 2 4

6) 1 2 3 5 6 7 8 4

7) 1 2 3 4 5 6 7 8

## Sortowanie przez wybieranie: (selection sort)

### Zasada działania:

Znajdź najmniejszy element i zamień go z tym na początku (ważne zamień a nie przestawiaj), a potem wyszukuj od kolejnego elementu który jest traktowany jako pierwszy. Często mylony z sortowaniem przez wstawianie można sobie skojarzyć, że wybieramy sobie najmniejszy lub największy element żeby go przestawić na początek.

Złożoność obliczeniowa:

Złożoność pamięciowa:

### Zalety:

- sortowanie w miejscu

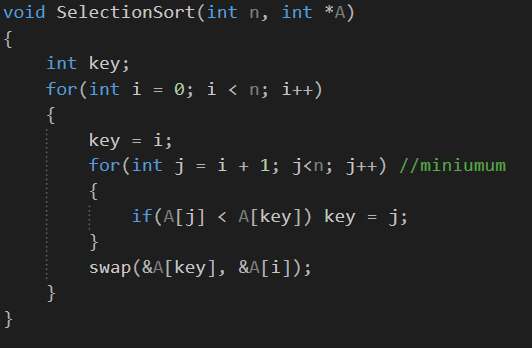
- łatwy w implementacji

### Wady:

- nie jest stabilny

- wolny

### Implementacja:



Rys. sortowanie przez wybieranie

## Sortowanie przez zliczanie: (counting sort)

### Zasada działania:

Zapisuje do dodatkowej tablicy ilość wystąpień poszczególnych wartości. Np. pod indeksem 2 jest ilość wystąpień liczby 2.

Złożoność obliczeniowa:

Złożoność pamięciowa:

k – rozpiętość danych (różnica pomiędzy najmniejszą a największą wartością danych)

### Zalety:

- szybki

- dobry dla danych z dużą ilością powtórzeń

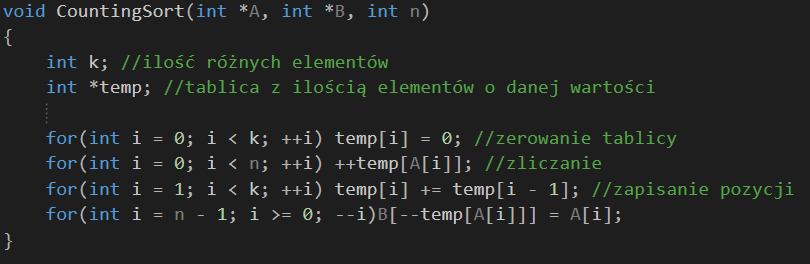
- stabilny

### Wady:

- dodatkowa pamięć (nie jest w miejscu)

- słaby dla zróżnicowanych danych

### Implementacja:



Rys. sortowanie przez zliczanie

UWAGA implementacja nie zawiera alokacji

## Sortowanie przez scalanie: (merge sort)

### Zasada działania:

Dzieli tablice na pół tak długo aż nie zostanie jeden element i potem łączy od dołu sortując poszczególne podtablice używając funkcji do scalania porostowych ciągów.

Złożoność obliczeniowa:

Złożoność pamięciowa:

### Zalety:

- stabilny

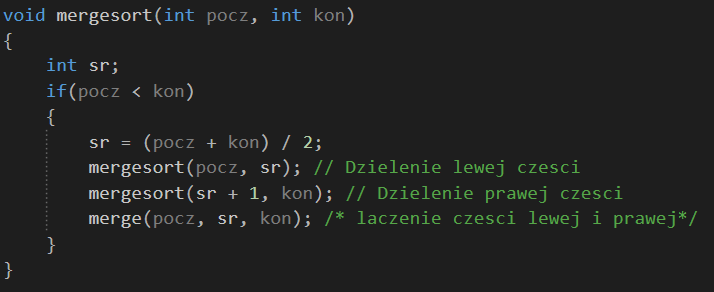
- prosty

- szybki także w praktyce

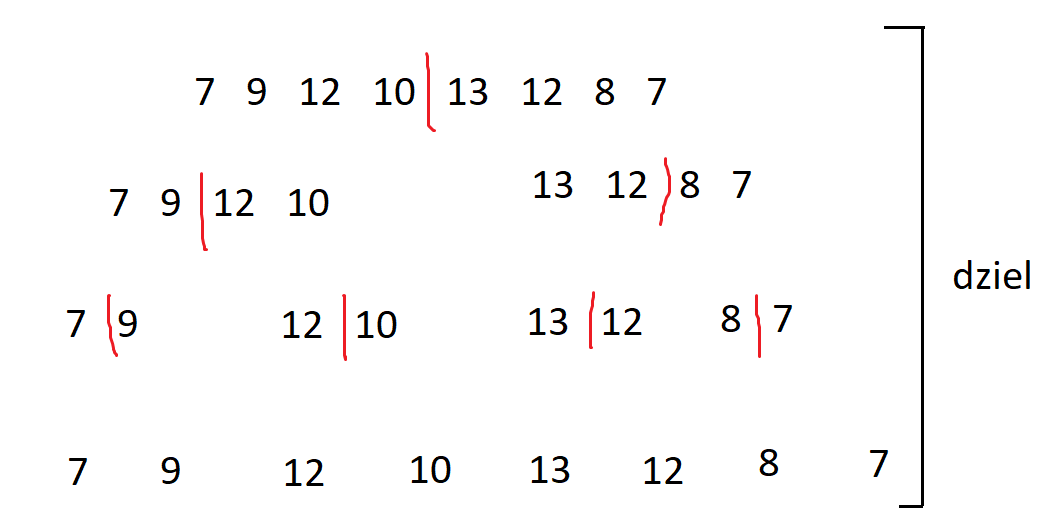
### Wady:

- potrzeba dodatkowej pamięci

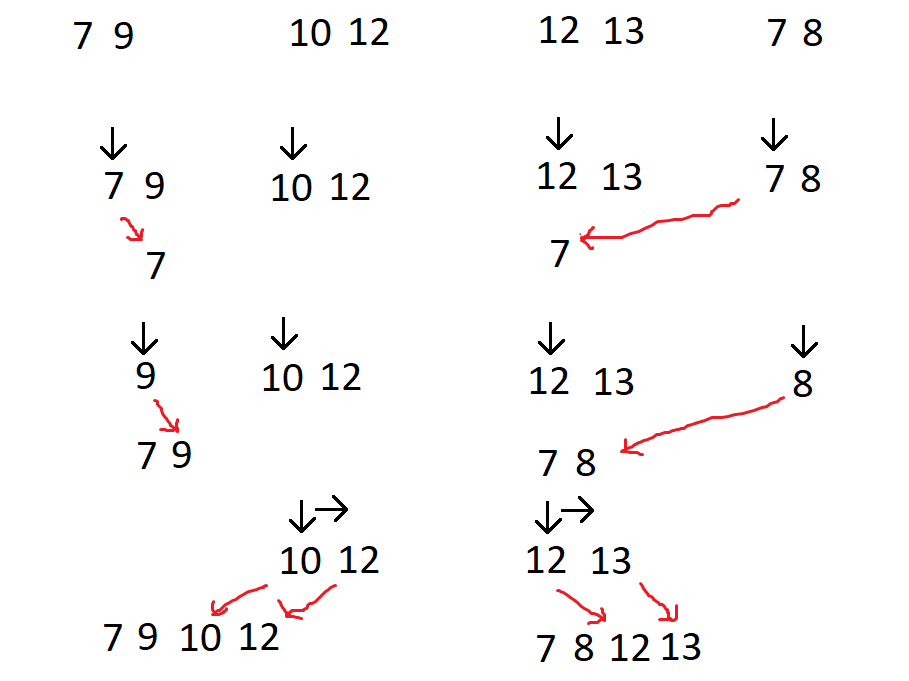
### Implementacja:



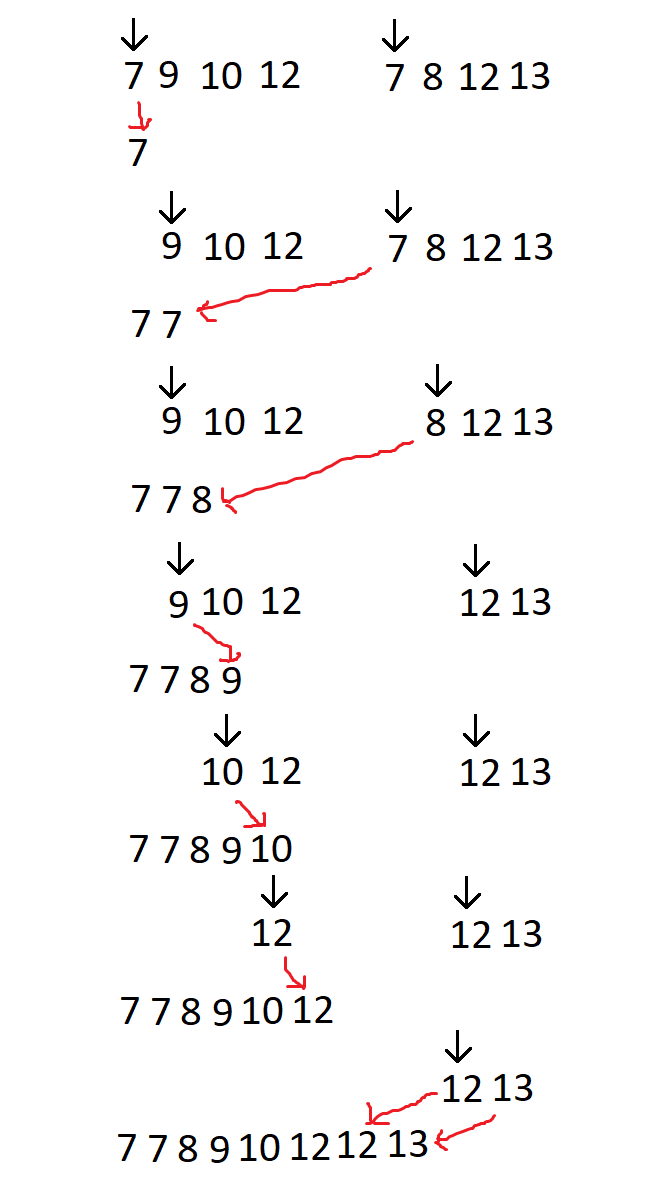
Rys. sortowanie przez łączenie



Rys. faza dzielenia w algorytmie sortowania przez scalanie



Rys. faza łączenia w algorytmie przez scalanie



Rys. faza łączenia w algorytmie przez scalanie ostatnie łączenie

## Szybkie sortowanie: (quick sort)

### Zasada działania:

Wybieramy arbitralnie jakiś punkt i przerzucamy na lewo mniejsze, a na prawo większe bądź równe od danego elementu. Potem wchodzimy do każdej z podzielonych części i powtarzamy aż nie zostanie tylko jeden element.

Złożoność obliczeniowa:

Złożoność pamięciowa:

### Zalety:

- szybki także w praktyce

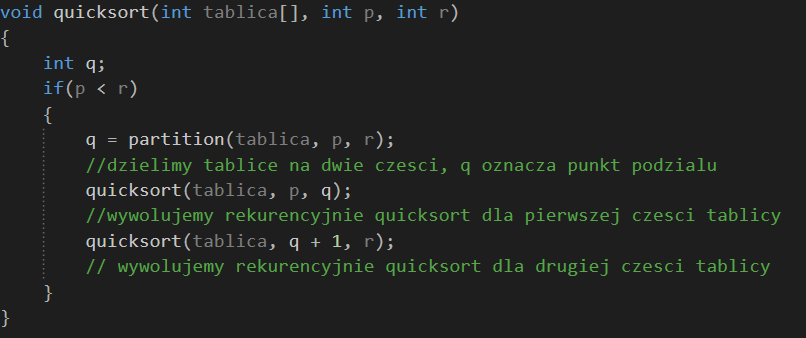
- sortuje w miejscu

### Wady:

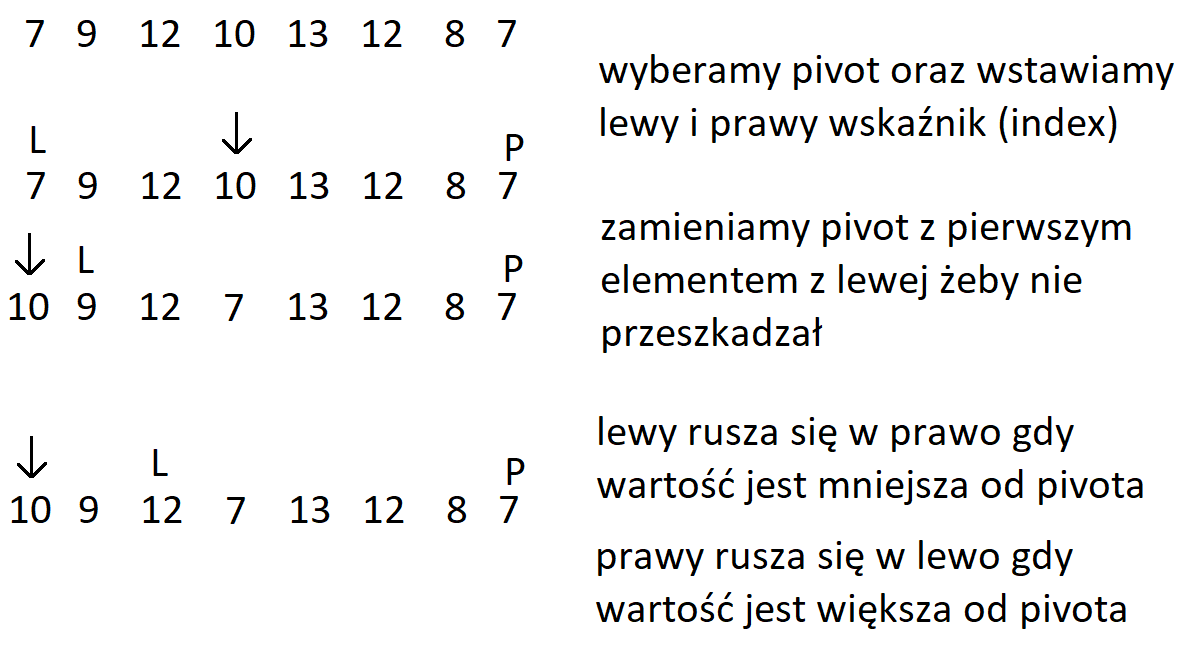
- w pesymistycznych sytuacjach złożoność obliczeniowa

- niestabilne

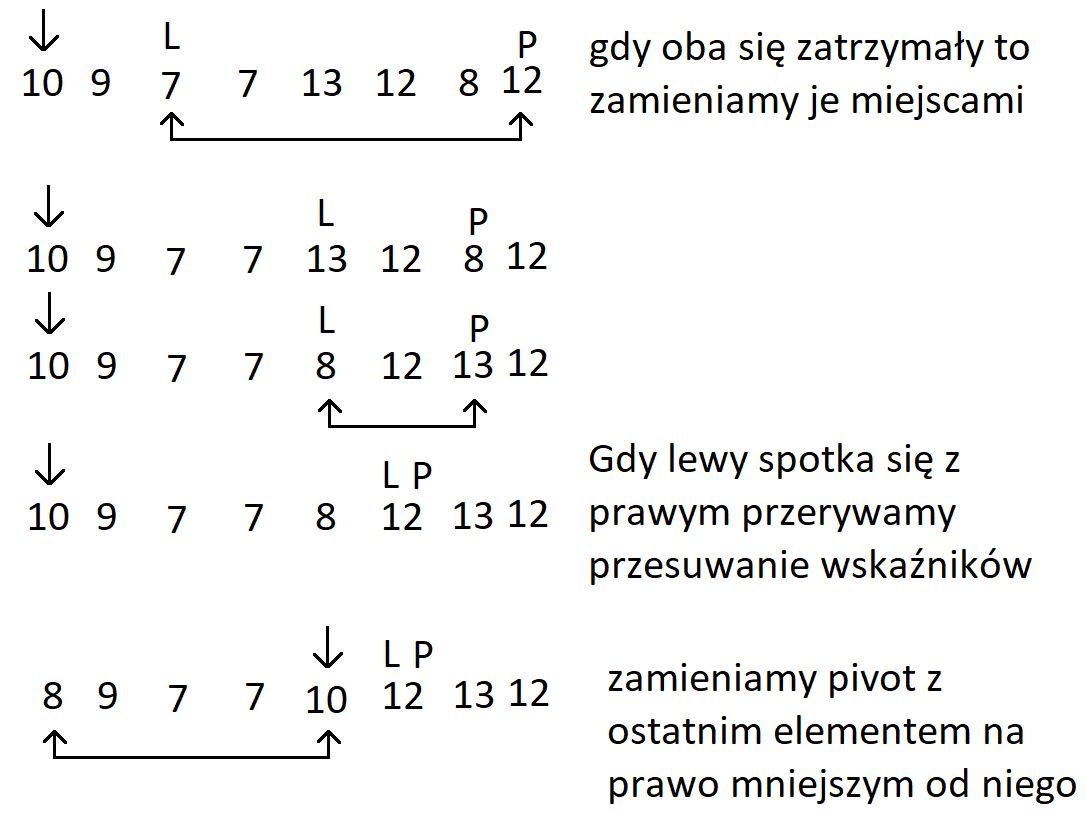
### Implementacja:



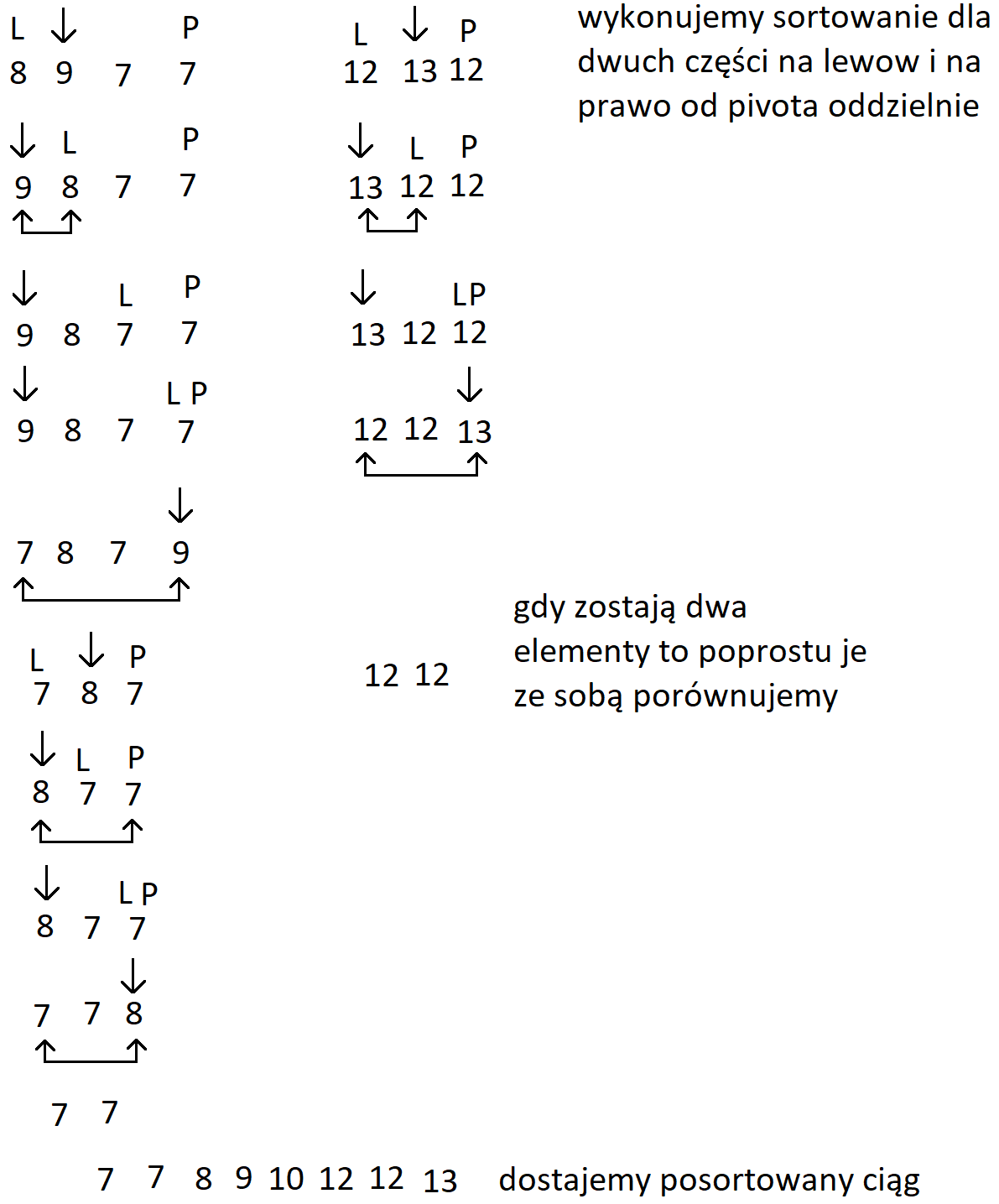
Rys. implementacja sortowania szybkiego



Rys. sortowanie szybki graficzne rozwiązanie cz1



Rys. sortowanie szybki graficzne rozwiązanie cz2



Rys. sortowanie szybki graficzne rozwiązanie cz3

## Podsumowanie:

### Tabela złożoności (szeregowanie algorytmów):

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Nazwa | optymistyczna | typowa | pesymistyczna | stabilność | w miejscu | Pamięć |
| Bąbelkowe | O(n2) | O(n2) | O(n2) | TAK | TAK | O(1) |
| Wstawianie | O(n) | O(n2) | O(n2) | TAK | TAK | O(1) |
| Wybieranie | O(n2) | O(n2) | O(n2) | NIE | TAK | O(1) |
| Zliczanie | O(k + n) | O(k + n) | O(k + n) | TAK | NIE | O(n + k) |
| Scalanie | O(n\*logn) | O(n\*logn) | O(n\*logn) | TAK | NIE | O(n) |
| Szybkie | O(n\*logn) | O(n\*logn) | O(n2) | NIE | TAK | O(1) |

### Algorytmy stabilne:

- bąbelkowe

- przez wstawianie

- przez scalanie

- przez zliczanie

- kubełkowe

- pozycyjne

### Algorytmy niestabilne:

- przez wybieranie

- Shella

- szybkie

- przez kopcowanie

## Definicje:

### Sortowanie w miejscu:

Algorytm nie potrzebuje dodatkowej pamięci na obliczenia. Ewentualnie niewielką stałą niezależną od ilości danych.

### Sortowanie stabilne:

Gdy wśród elementów do sortowania występują takie, które się powtarzają to po sortowaniu jest zachowana ich pierwotna kolejność. Czyli ta która wystąpiła jako pierwsza nadal będzie pierwsza.

# Złożoności obliczeniowe:

* 1
* log log n
* log n
* n
* n log log n
* n log n

# Stos

Stos jest interfejsem, w którym mamy określone funkcję. Możemy podglądać tylko element, który położyliśmy, jako pierwszy.

push() – dokładamy nowy element do stosu.

pop() - która pozwala na zdjęcie i odczytanie ostatnio położonego elementu.

isEmpty() – sprawdza czy stos jest pusty

pick() – pobranie wartości bez ściągania jej ze stosu. (opcjonalnie)

size() – zwraca rozmiar stosu. (opcjonalnie)

Nie mylić stosu z listą czy kolejką. Stos może być zrealizowany na dowolnej strukturze danych (tablica, tablica dynamiczna, lista dwukierunkowa) i może być dla dowolnego typu (int, float char).

# Odwrotna notacja polska

1) AB\*CD\*E/-

(A\*B)-((C\*D)/E)

2) 6 5\*3 7+8 4-2/\*+

Gdy natrafiamy na liczbę dokładamy ją do stosu, gdy na trafimy na działanie ściągamy dwie liczby i wykonujemy na nich działanie, a wynik wrzucamy na stos.

|  |  |  |
| --- | --- | --- |
| Lp. | wejście | stos |
| 1 | 6 | 6 |
| 2 | 5 | 6 5 |
| 3 | \* | 30 |
| 4 | 3 | 30 3 |
| 5 | 7 | 30 3 7 |
| 6 | + | 30 10 |
| 7 | 8 | 30 10 8 |
| 8 | 4 | 30 10 8 4 |
| 9 | - | 30 10 4 |
| 10 | 2 | 30 10 4 2 |
| 11 | / | 30 10 2 |
| 12 | \* | 30 20 |
| 13 | + | 50 |

(6 \* 5) + ((3 + 7) \* ((8 – 4) / 2)) = 50

# Kolejka

Kolejka podobnie jak stos nie jest strukturą danych tylko interfejsem. Z kolejki możemy, jeśli jest jednostronna to zabierać i odczytywać element, który został położony, jako pierwszy, a dokładamy elementy na sam koniec. W dwustronnej możemy zabierać elementy jeszcze z góry, ale nie ze środka.

Init(q) – opróżnia kolejkę i przygotowuje pamięć

isEmpty(q) – sprawdza czy kolejka jest pusta

Enqueue(q, x) – dokłada element **x** na koniec kolejki **q**

Dequeue(q) – zabiera element z początku kolejki

# Haszowanie

## Haszowanie liniowe i obsługa kolizji

Mamy tablicę 8-elementową:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 13 | 0 | 7 | 6 | 8 | 10 | 12 | 5 |

Zakładamy że funkcja haszująca to . Zwraca ona wartości od 0 do 9, więc tworzymy 10-elementową tablicę haszującą. Zwrócona wartość oznacza nr docelowego indeksu w tablicy.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  | **13** |  |  |
| *0* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **0** |  |  |  |  |  | **13** |  |  |
| *0* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **0** |  |  |  | **7** |  | **13** |  |  |
| *0* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **0** |  | **6** |  | **7** |  | **13** |  |  |
| *0* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* |

Występuje konflikt, więc przesuwamy się w prawo tak długo aż znajdziemy wolne miejsce.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **0** |  | **6** |  | **7** |  | **13** | **8** |  |
| *0* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* |

Występuje konflikt, więc przesuwamy się w prawo tak długo aż znajdziemy wolne miejsce.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **0** | **10** | **6** |  | **7** |  | **13** | **8** |  |
| *0* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* |

Występuje konflikt, więc przesuwamy się w prawo tak długo aż znajdziemy wolne miejsce.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **0** | **10** | **6** |  | **7** | **12** | **13** | **8** |  |
| *0* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* |

Występuje konflikt, więc przesuwamy się w prawo tak długo aż znajdziemy wolne miejsce.

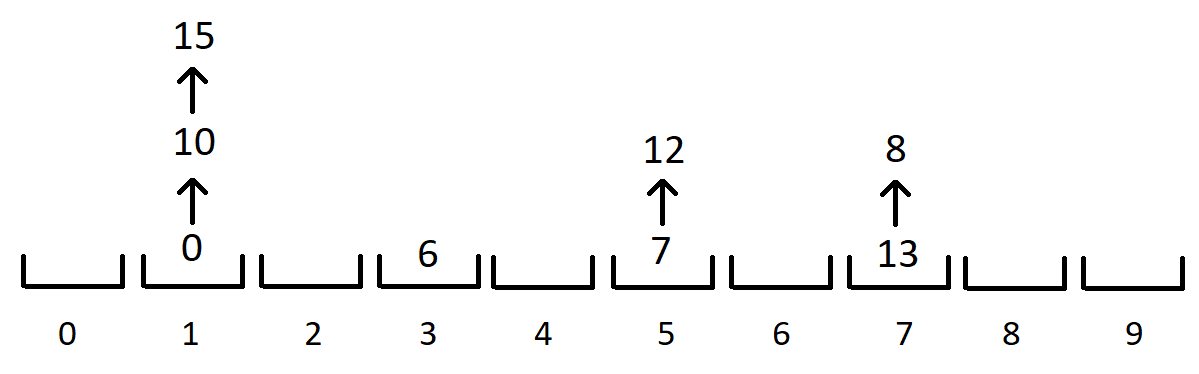
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **0** | **10** | **6** | **5** | **7** | **12** | **13** | **8** |  |
| *0* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* |

## Haszowanie łańcuchowe

Dla ułatwienia i pokazania różnic użyjemy tablicy z poprzedniego zadania

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 13 | 0 | 7 | 6 | 8 | 10 | 12 | 15 |
| 7 | 1 | 5 | 3 | 7 | 1 | 5 | 1 |

Drugi rząd to już obliczone pozycje do tablicy za pomocą tego samego wzoru co w poprzednim przykładzie



Haszowanie łańcuchowe zamiast nie jest tablicą przechowującą dane wartości tylko tablicą list.

Te które były jako pierwsze są niżej na liście. Nie ma różnicy czy narysujemy to pionowo (tak jak tu) czy poziomo (jak było na prezentacji).

## Haszowanie kwadratowe

Jeszcze raz rozważmy te same dane oraz tej samej funkcji haszującej

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 13 | 0 | 7 | 6 | 8 | 10 | 12 | 15 |
| 7 | 1 | 5 | 3 | 7 | 1 | 5 | 1 |

Rozważmy liczby 0, 10 oraz 15 ponieważ mają kolizje na pozycji 1

Zastąpmy naszą liczbę 1 w funkcji przez i:

Przy każdej kolizji używamy następującego wzoru poczym wykonujemy ++i

Każda następna kolizja jest przesuwana o kwadrat danej liczby, a nie o jeden jak w liniowym

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **0** |  | **6** |  | **7** |  | **13** |  |  |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

Pierwsza kolizja to 8 bo chcemy ją postawić na indeks 7 który jest zajęty przez liczbę 10

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **8** | **0** |  | **6** |  | **7** |  | **13** |  |  |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

Druga kolizja to 10

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **8** | **0** |  | **6** | **10** | **7** |  | **13** |  |  |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

Trzecia kolizja to 12

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 8 | 0 |  | 6 | 10 | 7 |  | 13 | 12 |  |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

Trzecia kolizja to 15

Tu następuje kolejna kolizja więc dodajemy do jeszcze 1

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **8** | **0** |  | **6** | **10** | **7** |  | **13** | **12** | **15** |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

## Pojęcia związane z haszowaniem

O(1) wyszukiwanie w optymistycznym przypadku

O(n/m + 1) wyszukiwanie w średnim przypadku

**n** – ilość elementów w ciągu **m** – długość haszowanej tablicy

O(n) wyszukiwanie w pesymistycznym przypadku

współczynnik przeładowania (load factor)

oczekiwana liczba prób do znalezienia oczekiwanej wartości

# Drzewo

Drzewo (wolne): graf spójny acykliczny nieskierowany. Jeśli graf jest nieskierowany i acykliczny, ale niekoniecznie spójny, to nazywamy go lasem.

Drzewo z korzeniem (drzewo ukorzenione): drzewo wolne z jednym wyróżnionym wierzchołkiem, tzw. korzeniem (oznaczamy go zwykle przez *r*).

Drzewo uporządkowane: drzewo ukorzenione, w którym dzieci każdego węzła są uporządkowane.

Regularne drzewo binarne: każdy węzeł ma 0 lub 2 dzieci.

Pełne drzewo binarne: wszystkie liście są na tej samej głębokości *h*(dla pewnego h), a wszystkie węzły wewnętrzne mają stopień 2.(Innymi słowy, pełne drzewo binarne jest regularne.)

## przeszukiwanie

> inorder

\* lewe dziecko

\* węzeł

\* prawe dziecko

> preorder

\* węzeł

\* lewe dziecko

\* prawe dziecko

> postorde

\* lewe dziecko

\* prawe dziecko

\* węzeł

## Podstawowe pojęcia

**Liść -** Węzeł drzewa, z którego nie wychodzi żadna krawędź. Nie ma dzieci

**Korzeń -** Węzeł początkowy drzewa znajdujący się, jako jedyny na wysokości 0 i nie ma rodzica

**Bracia –** Węzły wychodzące z tego samego rodzica

**Wysokość -** Odległość między korzeniem, a najodleglejszym liściem. Do wysokości drzewa nie wlicza się korzenia (jest na głębokości 0).

**Głębokość -** Odległość miedzy korzeniem, a konkretnym wierzchołkiem

**Drzewo binarne -** Drzewo, w którym stopień każdego wierzchołka jest nie większy od 3.

**Drzewo BST (przeszukiwań binarnych) -** To takie drzewo, w którym prawe dzieci są większe od rodzica, a lewe mniejsze od niego.

# Grafy

## Graf skierowany (zorientowane)

To para G = (V, E) zbiorów skończonych gdzie V to zbiór węzłów (wierzchołków), a E to zbiór uporządkowanych par wierzchołków (krawędzi, łuków, pętli).

## Graf nieskierowaniy (niezorientowane)

To para G = (V, E) zbiorów skończonych gdzie V to zbiór węzłów (wierzchołków), a E to zbiór dwuelementowych zbiorów wierzchołków (krawędzi, łuków, pętli).

## Macierze sąsiedztwa

Dwuwymiarowa (V2), jest jedna dla całego grafu. Wypełniona 0 (brak połączenia) lub 1 (połączenie)

Dla grafów skierowanych: diagonalna nie musi być z samych 0.

Dla grafów nieskierowanych: diagonalna jest z samych 0, oraz jest symetryczna względem diagonalnej.

Diagonalna to ta przekątna od góry po lewo do dołu po prawo.

## Listy sąsiedztwa

Jednowymiarowa (V), przechowuje informację o tym, do których wierzchołków prowadzi droga z danego wierzchołka. Jest listą wskaźników.

## Macierze incydencji

Dwuwymiarowa (V\*E) prostokątna, niesymetryczna i przyjmuje wartości -1, 0, 1 oraz 2

-1 oznacza krawędź w grafie skierowanym w przeciwnym kierunku do wierzchołka (wychodzi z niego), 0 to brak połączenia, 1 to krawędź wchodząca do danego wierzchołka, a 2 to pętla.

## Przeszukiwanie BFS (w szerz)

- wrzucamy wierzchołek startowy do kolejki

- sprawdzamy wszystkich sąsiadów i wrzucamy do kolejki każdego, który nie został jeszcze odwiedzony.

- po dorzuceniu wierzchołków wierzchołka na dole kolejki wyrzucamy go

Lub

- Na początku wszystkie wierzchołki są białe.

- Wierzchołki szare są przechowywane w kolejce FIFO.

- Po jej opuszczeniu kolorujemy je na czarno.

## Przeszukiwanie DFS (w głąb)

- badamy wszystkie krawędzie ostatnio odwiedzonego wierzchołka v

- gdy wszystkie krawędzie v są zbadane przechodzimy wierzchołka v, z którego przeszliśmy do obecnego

- kontynuujemy do momentu, gdy wszystkie wierzchołki osiągalne z wierzchołka startowego zostaną odwiedzone

## Złożoność pamięciowa grafu

Złożoność pamięciowa grafu jest

**Graf spójny -** gdy do każdego wierzchołka prowadzi krawędź i da się dojść od niego do każdego innego. Graf spójny może być słabo lub silnie spójny.

**Graf słabo spójny -** gdy do każdego wierzchołka prowadzi krawędź i da się dojść od niego do każdego innego z tym, że gdy graf jest skierowany to by trzeba było iść pod prąd.

**Graf silnie spójny -** gdy zawsze da się dojść z każdego do dowolnego innego wierzchołka uwzględniając kierunki dróg.

# Kompresja danych

## Podstawowe dane

Kompresja bezstratna: (możemy odczytać dane w niezmienionej)

- tekst

- programy

- baza danych

Kompresja stratna: (część informacji jest tracona)

- muzyka

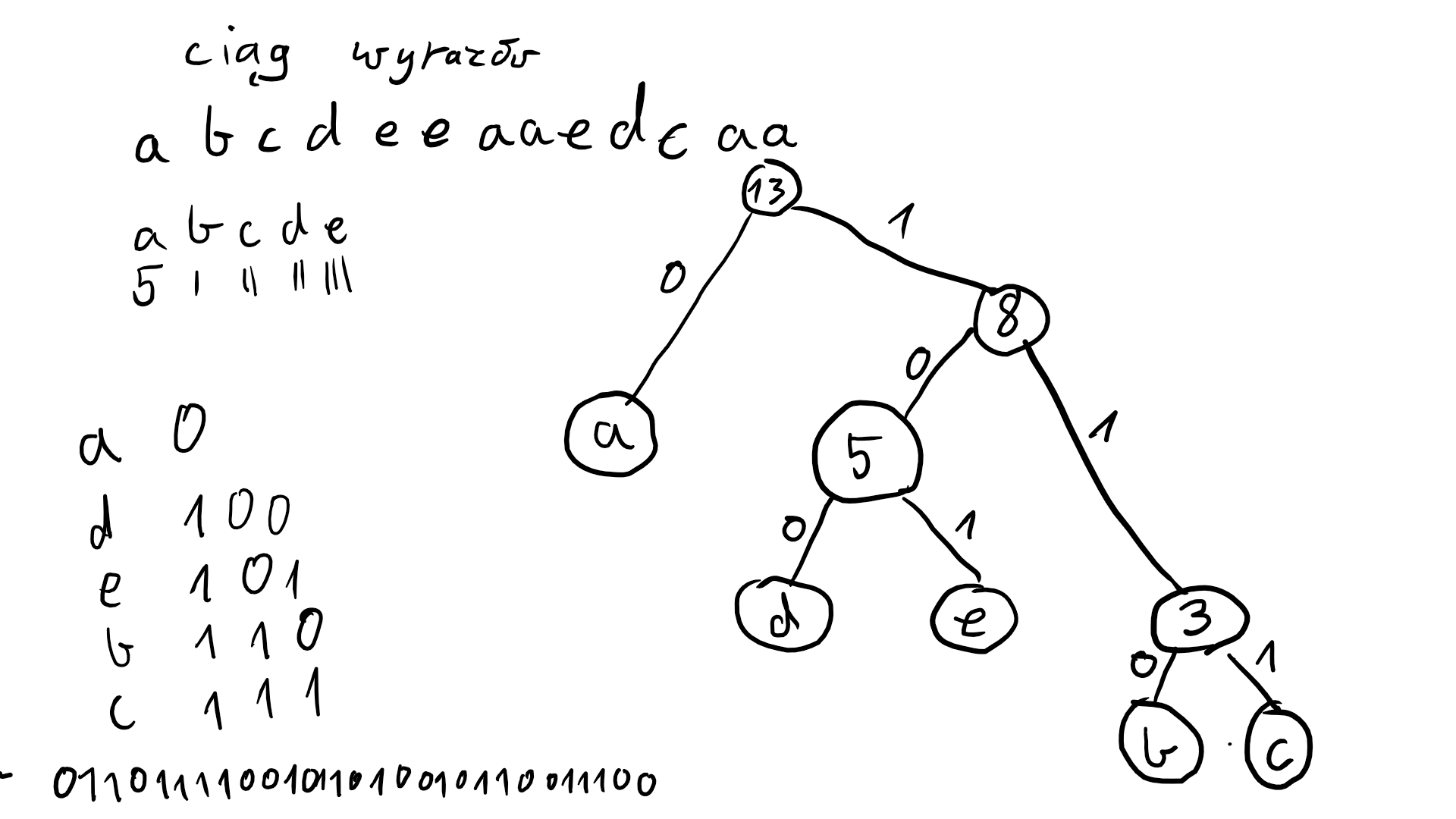
- filmy

- multimedia

## Algorytm Huffmana

Jest to algorytm kompresji danych, który polega na jak największym skróceniu danych które najczęściej występują.

W dowolnym ciągu liczymy ilość wystąpienia poszczególnych znaków i budujemy na ich podstawie drzewo tak by na samym dole znajdowały się najrzadziej występujące, a im wyżej tym częściej występujące.



Typowa litera zajmuje pamięć 8 bitów czyli 1 bajta

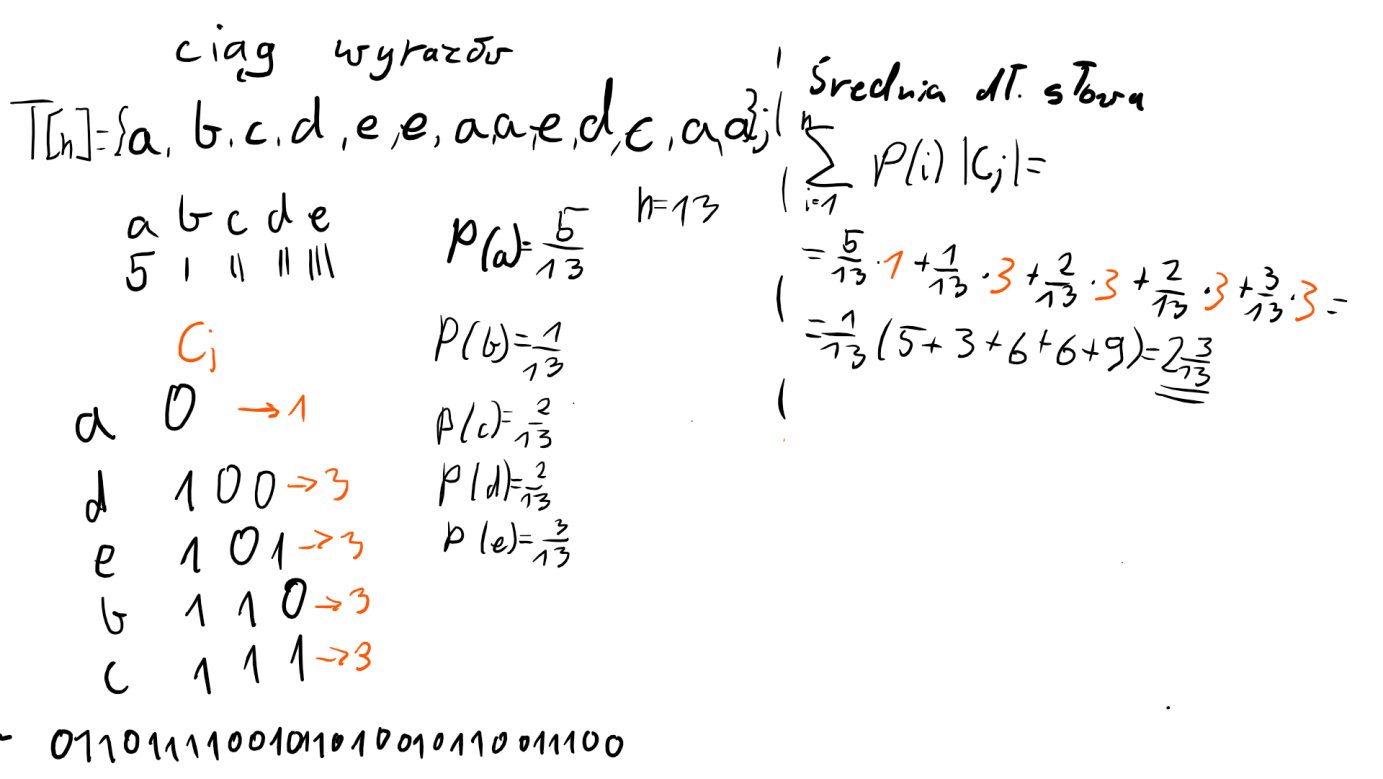
Zauważmy, że tu poszczególne litery zajmują od 1 do 3 bitów, a kod jest jednoznacznie odczytywalny

np.: jeśli litera zacznie się od **0** to od razu wiemy, że to **a**, zaś gdy od **11** to będzie to **b** lub **c**

Algorytm bardzo dobrze działa dla danych o **małej ilości różnych znaków** w tym przypadku było tylko 5 różnych liter

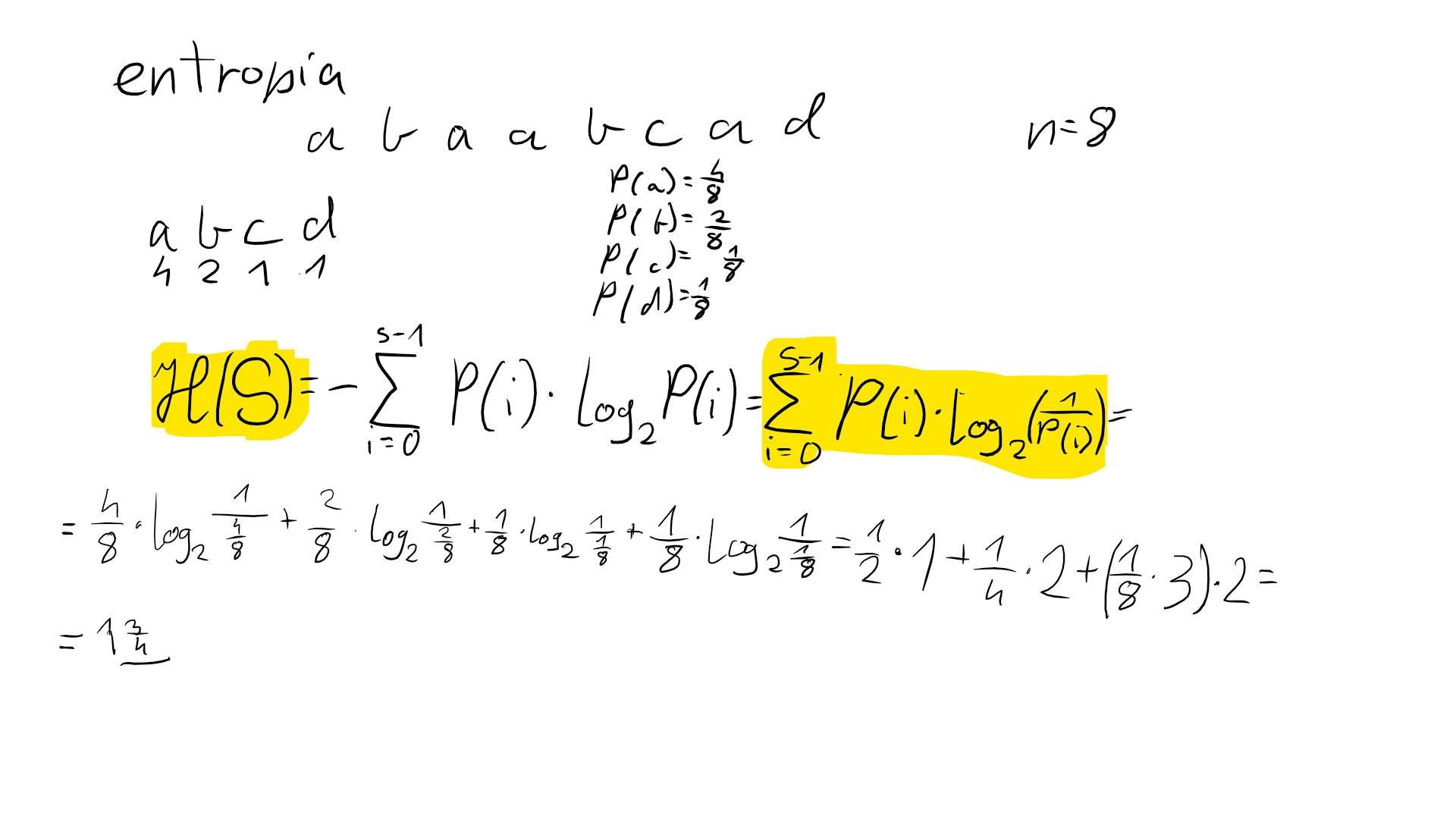
Do algorytmu jest potrzebna dodatkowa pamięć przetrzymująca drzewo, oprócz samej sekwencji

## Średnia długość słowa



Sposób na obliczenie średniej długości słowa na podstawie przykładu z algorytmu Huffnama

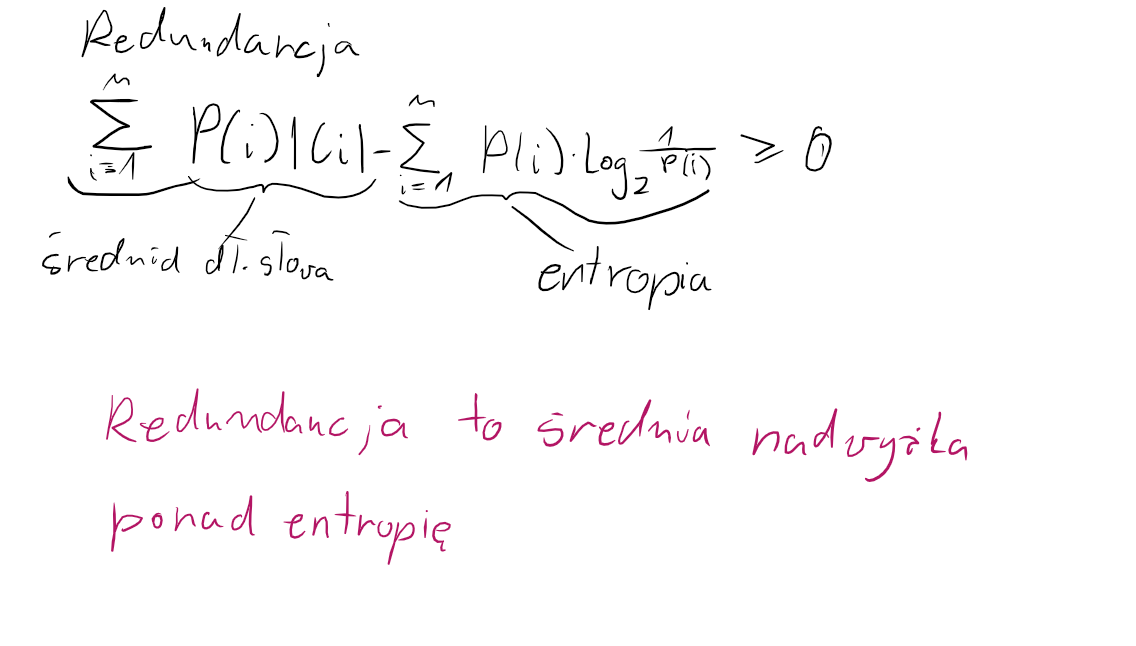
Entropia



Entropia jest miarą rozrzutu danych. Określa ona czy warto kompresować badane dane.

Im mniejsza entropia tym dane powinny dać się lepiej skompresować

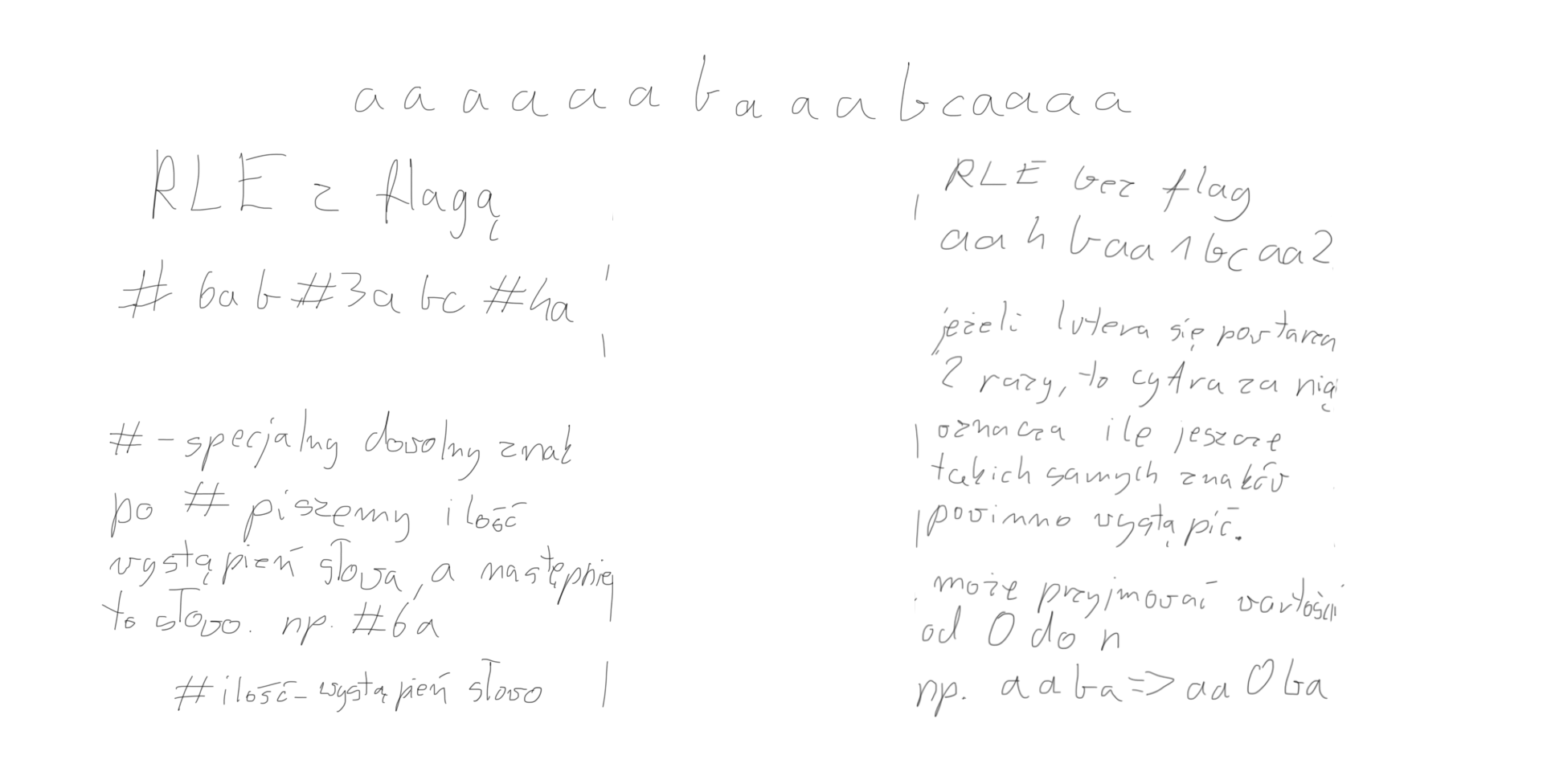
## Redundancja



Redundancja jest miarą przeciwną do entropii jest to wartość tego jak dużo jest nadmiarowych danych

Im wyższa redundancja bym dane powinny się lepiej kompresować

## Kompresja RLE



Run-length encoding to algorytm kompresji bezstratnej gdzie znaki zapisujemy zliczając ich ilość i pisząc np. 10g zamiast gggggggggg. W praktyce mało przydatny.

Dobry dla danych, w których jest dużo powtarzających się **znaków następujących po sobie**

## LZ77

Algorytm kompresji, w którym bierzemy ciąg który się powtarza mi zamiast go pisać drugi raz zapisujemy gdzie on wystąpił i jakiej jest długości.

Przykład

Wczoraj wieczorem Hamming garał w golfa z Hammingway’em

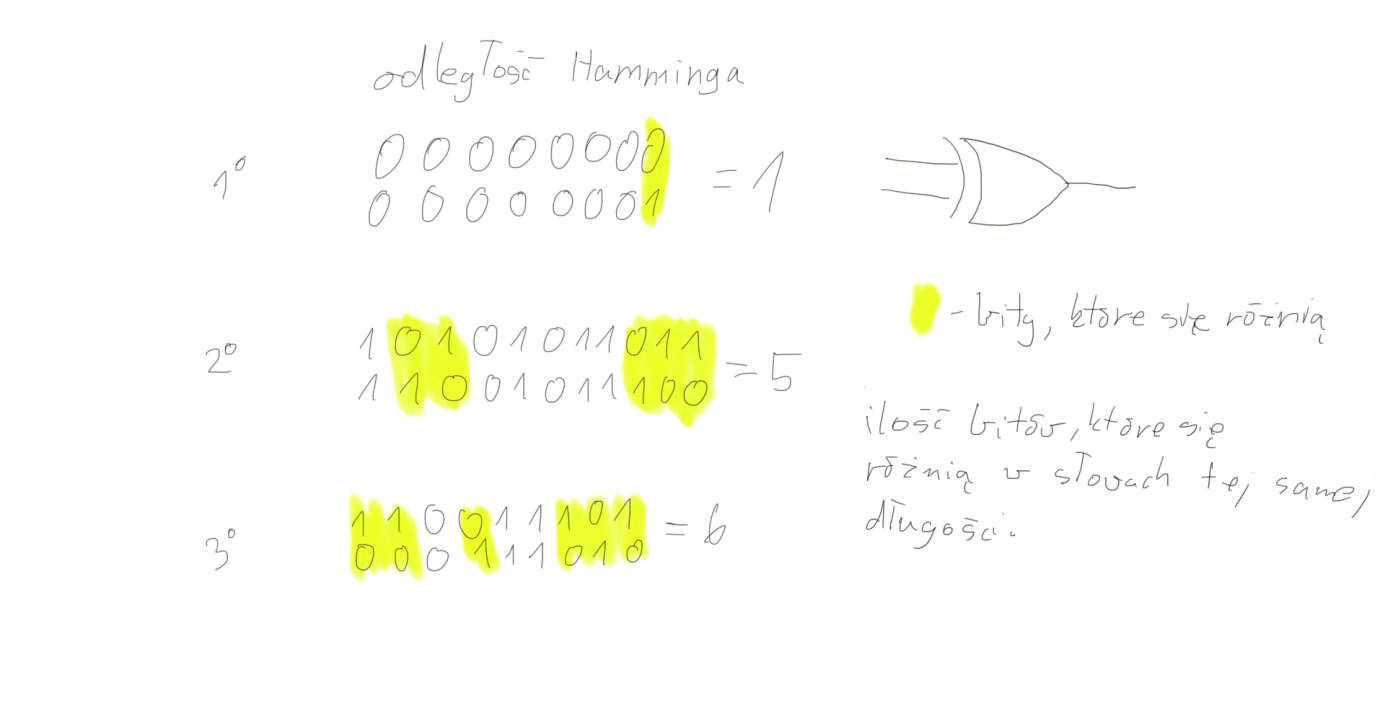
Wczoraj wieczorem Hamming garał w golfa z <24, 7>way’em

Oznacza to, że 24 znaki wcześniej jest siedmioliterowe powtórzenie

Dobry dla długich ciągów znaków, które się powtarzają. Przydatny także w praktyce.

# Stringologia:

## Długość Hamminga



## Bit parzystości

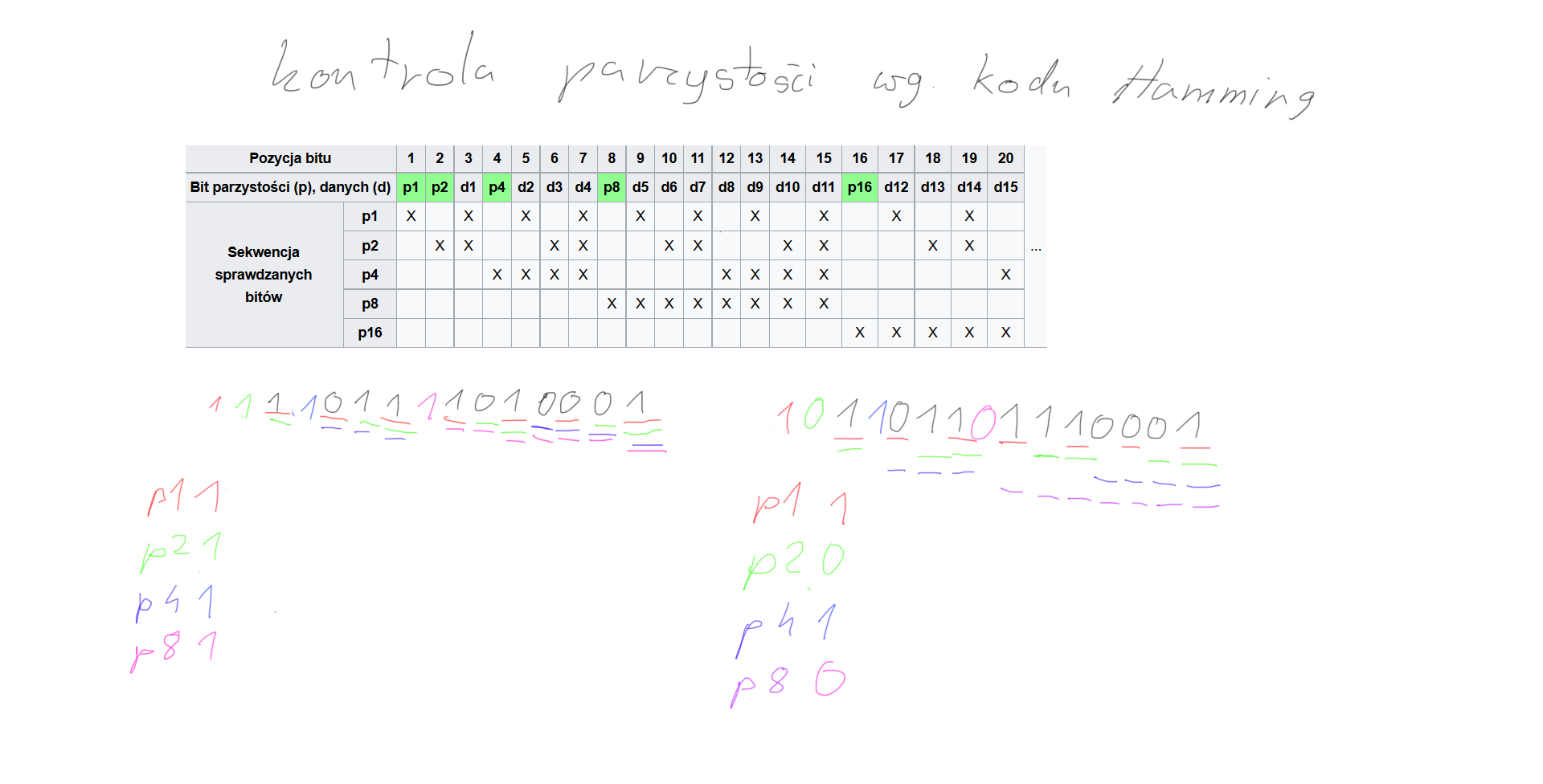
By sprawdzić czy dane są poprawne często dodaje się do nich sztuczne dane które pokazują, że Informacje są zgodne z tym co się przesłało.

Na początku lub na końcu dodaje się jeden bit którego wartość oznacza ilość jedynek w całym pozostałym ciągu.

0 to nieparzysta ilość jedynek

1 to parzysta ilość jedynek

## Kontrola parzystości



Gdy wysyłamy dane przez sieć i nie chcemy by nam jakieś dane się zgubiły lub co gorsze przekłamały.

Do policzenia ilości bitów potrzebnych do korekcji bierzemy ilość wartości, które chcemy wysłać.

Liczbę 7 można zapisać jako gdy r = 3, bo

Więc potrzeba 3 bitów, by można było skorygować

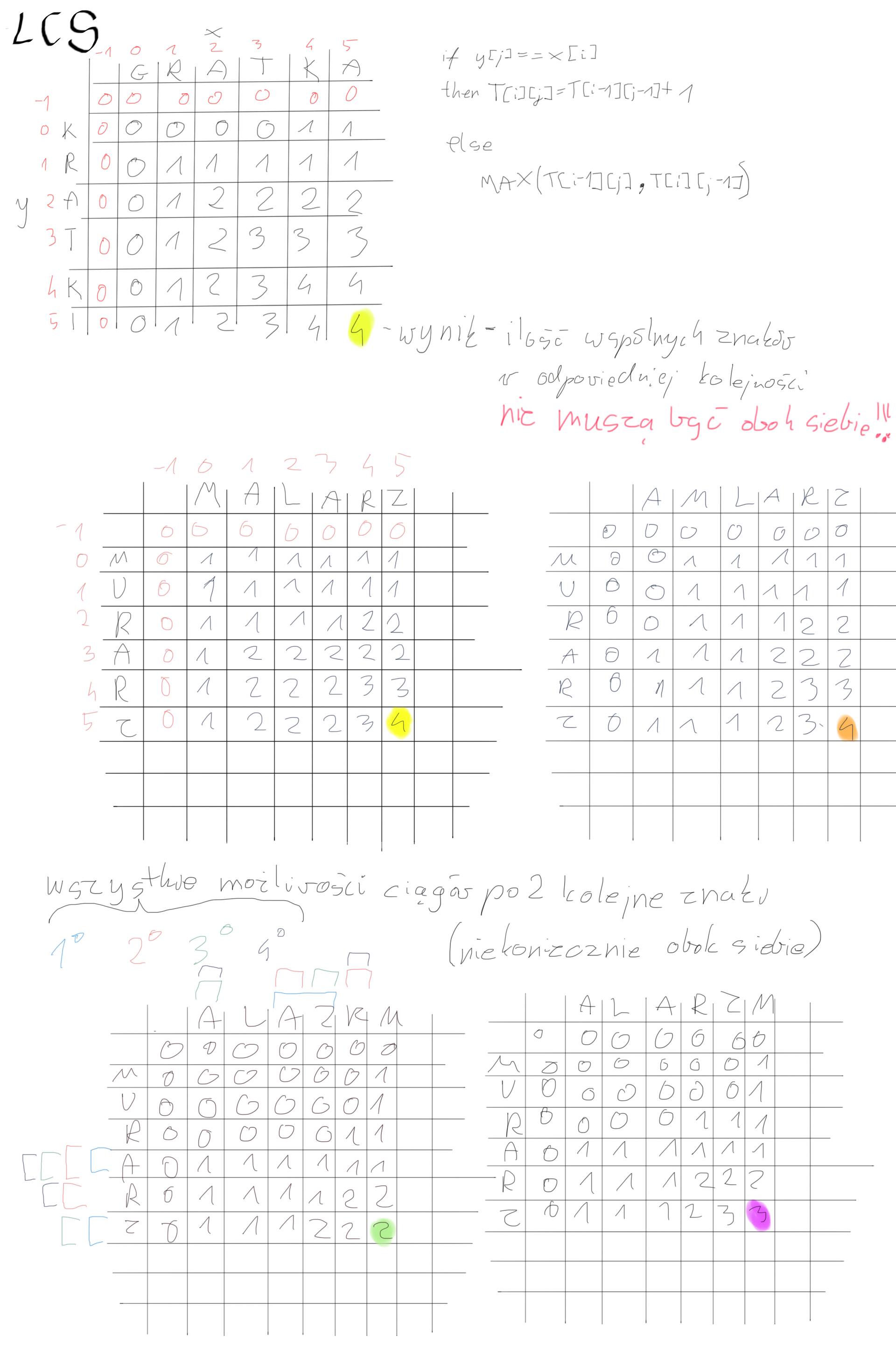
Liczbę 9 można zapisać jako gdy r = 5, bo

Więc potrzeba 4 bitów, by można było skorygować. W tym przypadku trzeba dać trochę nadmiaru.

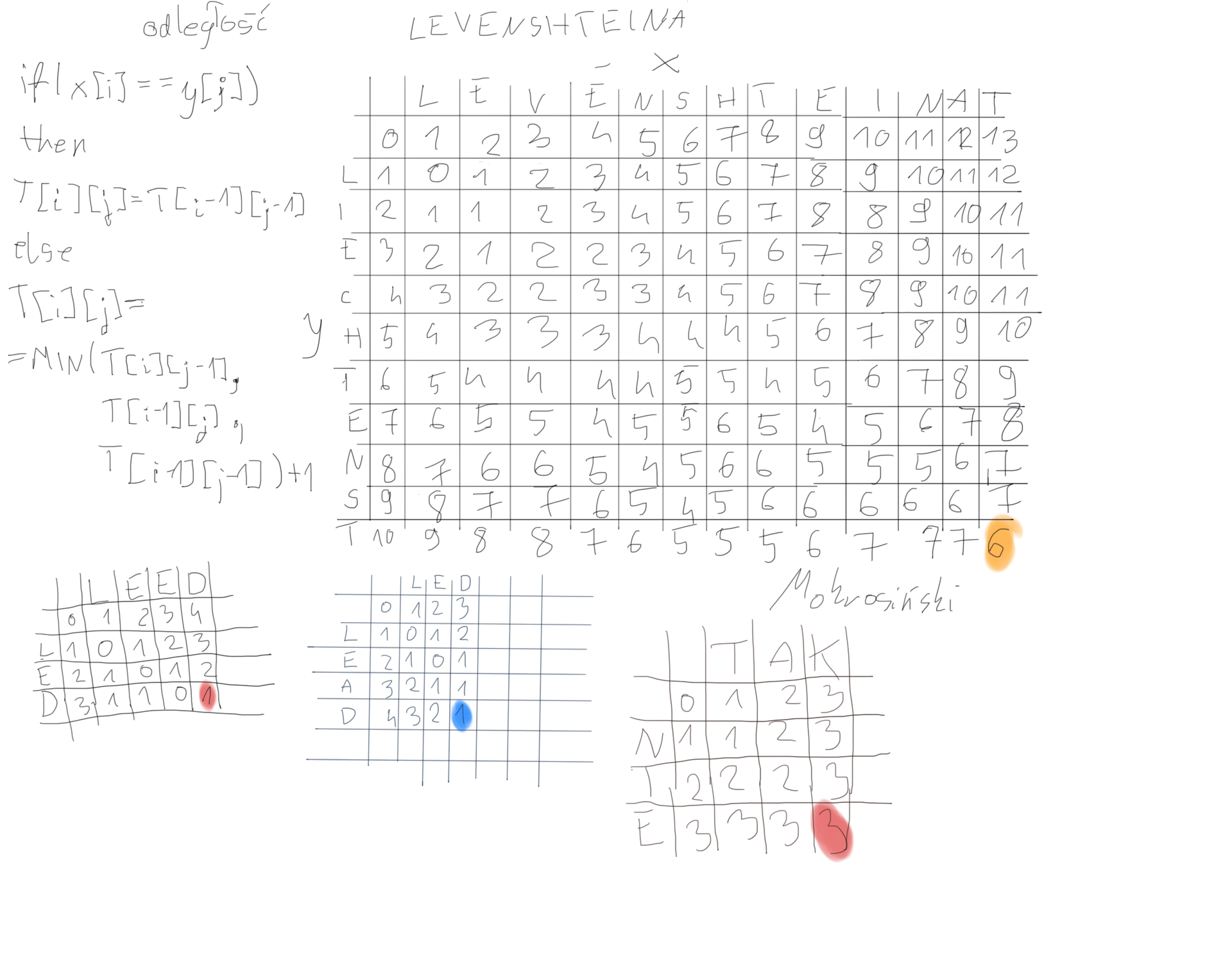
## LCS – Longest common subsequence

Najdłuższa wspólna podsekwencja

Algorytm jest stosowany np.: w repozytoriach do sprawdzania różnic między poszczególnym wersjami plików.



## Odległość Levenshteina (edycyjna)



Służy do wyznaczania różnić między poszczególnymi wyrazami

Mówi nam o tym ile minimalnie znaków musimy **usunąć, przestawić** lub **zastąpić,**

by jeden wyraz zamienić w drugi.

Algorytm ten może służyć jako podpowiedzi w słownikach np.: gdy zrobimy jakąś literówkę zjemy jakąś jedną literę możemy sobie obliczyć do jakich wyrazów ze słownika jest podobny ten, w którym się pomyliliśmy.