# 6 Wyszukiwanie elementów w liście oraz tablice haszujące

Algorytmy i struktury danych, dr A. Niemczynowicz

UWAGA: Wszystkie materiały są jedynie na użytek edukacyjny. Proszę ich nie rozpowszechniać poza grupą, a przede wszystkim publicznie!

Kolejnym tematem zajęć jest wyszukiwanie elementów w liście oraz struktury ułatwiające takie wyszukiwanie – tablice haszujące.

Wyszukiwanie, np. liczb w przypadku nieposortowanej tablicy polega na przejrzeniu każdego elementu tablicy. Tutaj niewiele możemy zrobić z przyśpieszeniem procesu.

Jednak gdy lista zawiera posortowane liczby, to możemy znacznie przyśpieszyć szukanie używając przeszukiwania binarnego, które obrazowo polega na potraktowaniu listy jako drzewa które przechodzimy wzdłuż gałęzi. To drzewo „spłaszcza” listę, a przez to mamy mniej elementów do przejrzenia, co przyśpiesza proces. Odbywa się to tylko dlatego, że dane zostały wcześniej uporządkowane i z tego uporządkowania korzystamy w algorytmie.

Aby przyśpieszyć szukanie możemy dane umieścić w tzw. **tablicy haszującej**. Działanie polega na wybraniu pewnej funkcji zwanej **funkcją haszującą** lub **funkcją skrótu**, która pobiera daną i zwraca liczbę (w ogólności pewien ciąg znaków) który mówi do której listy w tablicy haszującej (która jest zbiorem takich list – kubełków) dane trafią. Dzięki temu elementy mające tą samą wartość funkcji haszującej trafiają do tego samego kubełka. Jeżeli więc chcemy wyszukać element używając cechy którą (dobrze dobrana) funkcja haszująca uwypukla, wystarczy znaleźć kubełek i przeszukać go. Jest to zadanie szybsze, gdyż już na początku, przez skupienie się na konkretnym kubełku, ograniczamy liczbę elementów do przejrzenia.

**Sztuką jest dobranie funkcji haszującej** tak, aby dzieliła nasze dane na zbiory o cesze którą później będziemy wykorzystywać do szukania danych. W prezentacji **W5Wyszukiwanie** zaprezentowana jest najprostsza funkcja dzielenia modulo liczba kubełków. Jednak bardziej skomplikowane funkcje również są wykorzystywane. Niektóre funkcje haszujące (SHA1, MD5) są wykorzystywane przy podpisie elektronicznym. Proszę uzupełnić materiał dotyczący tego zagadnienia we własnym zakresie, jeśli wcześniej się nie pojawił.

Plan analizy załączonego materiału:

1) **W5Wyszukiwanie.pdf** oraz **Programy.tar.gz** - wykład i programy. Proszę to przeanalizować.

2)**1Szukanie.pdf** – tekst dokładniej opisujący istotę wyszukiwania. Proszę go również przeczytać.

3) **2PythonHashTable.pdf** - implementacja tablic haszujących w Pythonie. Materiał zawiera również znaczne rozszerzenie typów tablic haszujących z kompletnym kodem. Proszę zaimplementować te tablice w C/C++.

4) **3CormenHashTables.pdf** – teoretyczne podejście do tablic haszujących. Zawiera również teorię funkcji haszujących, więc osoby zainteresowane wyborem/konstruowaniem bardziej zaawansowanych funkcji haszujących znajdą tam informacje na temat tego, co jest dozwolone, a co zabronione przez matematykę w tym zagadnieniu.