Podstawowe pojęcia i narzędzia informatyki

Informacja i sposoby jej reprezentacji w pamięci komputera

Informacja i dana

Informacja to obiekt abstrakcyjny, który w postaci zakodowanej (jako dana) może być **przechowywany**, **przesyłany**, **przetwarzany** i może **służyć do sterowania**.

Dana to formalna reprezentacja określonej treści (informacji) nadająca się do przechowywania, przesyłania i przetwarzania (np. imię, nazwisko, wzrost, waga, data ur.).

Sposoby przekazywania informacji

- dane liczbowe (różne formaty danych)
- dane tekstowe (ciągi znaków)
- obrazy (rysunki, schematy, mapy, fotografie czarno-białe lub kolorowe)
- animacje
- filmy i nagrania video
- dźwięki

Redundancja, kod zwarty, algorytm Huffmana

Jeśli komunikatom, które występują z prawdopodobieństwami p_i (i=1, ..., n), przypisano słowa kodowe o długościach N_i (i=1,..,n), to wielkość: $L = \sum_{i=1}^{n} p_i N_i$ nazywamy **średnią długością słowa kodowego**.

Jeśli źródło nadaje *n* różnych komunikatów z prawdopodobieństwami p_i (i=1,..., n), to *średnia ważona ilość informacji* w komunikatach z tego źródła wyrażana wzorem: $H = \sum_{i=1}^{n} p_i \log_2 \left(\frac{1}{p_i}\right)$ nosi nazwę **entropii informacyjnej źródła informacji**.

Różnicę **R = L - H** nazywamy **redundancją kodu (nadmiarowością)**, a wielkość **1- H/L** nazywamy **względną redundancją kodu**.

Dla każdego zbioru komunikatów można zbudować różne kody. Szczególne znaczenie mają kody:

- **jednoznaczny**, w którym żaden ciąg kodowy nie jest początkiem innego ciągu
- **zwarty**, to kod jednoznaczny, o minimalnej redundancji
- **o równej długości słów kodowych**

Algorytm Huffmana pozwala wyznaczyć kod zwarty. Jest kodem prefiksowym (oznacza to, że żadne słowo kodowe nie jest początkiem innego słowa), a średnia długość słowa kodowego jest najmniejsza spośród kodów prefiksowych.

Systemy pozycyjne

W matematyce liczby zapisywane są w **systemie pozycyjnym**, tzn. ciąg znaków $a_n a_{n-1} \dots a_1 a_0 . a_{-1} a_{-2} \dots a_{-m}$ jest zapisem liczby $\sum_{i=-m}^{n} a_i q^i$, gdzie q jest podstawą systemu; a_i = 0, 1, …, q-1 (i = -m, -m+1, …, 0, 1, …, n) są cyframi. System nazywa się pozycyjnym, ponieważ wartość (waga) każdej cyfry zależy od miejsca, czyli pozycji w ciągu.

Zapis stałopozycyjny

0	0	0	0	1	0	1	0	1	1	1	0	1	0	1	1

bit znaku

Stosowany dla liczb całkowitych (pierwszy bit przeznaczony jest na zapis znaku: 0 to +, 1 to -. Przedstawienie liczb jest dokładne.

Zapis zmiennopozycyjny

0	1	0	0	1	0	1	0	1	1	1	0	0	1	1	1

bit znaku | mantysa | cecha

Mantysa – liczba ułamkowa
Cecha – liczba całkowita

Algorytmy

Algorytm

Algorytmem nazywać będziemy opis obiektów łącznie z opisem czynności, które należy wykonać nad/z tymi obiektami w ściśle określonej kolejności, aby w skończonej liczbie kroków osiągnąć określony cel (rozwiązać zadanie).

Do opisu algorytmu można używać:

- języka naturalnego
- schematów blokowych
- języka formalnego (języka programowania)

Analiza algorytmu

- analiza danych (wejściowych), ich zakresu i dokładności oraz określenie klasy spodziewanych wyników
- poprawność algorytmu (testowanie i weryfikacja)
- złożoność algorytmu
- złożoność obliczeniowa, pamięciowa, optymalność rozwiązania
- analiza uzyskanych wyników: dokładność, numeryczna poprawność, stabilność algorytmu

Poprawność algorytmu: eksperymentalne testowanie algorytmu dla wybranych zestawów danych nie gwarantuje, że dla innych zestawów danych algorytm działa poprawnie. Testowanie może wykazać istnienie błędów, ale nigdy nie wykaże ich braku.

Częściowa i całkowita poprawność algorytmu

- Algorytm A jest **częściowo poprawny** względem warunków początkowych P i końcowych Q, gdy dla każdego zestawu danych spełniających warunki P, jeżeli algorytm zakończy obliczenia, to wyniki spełniają warunki końcowe Q.
- Algorytm A jest **całkowicie poprawny** względem warunków początkowych P i końcowych Q, gdy dla każdego zestawu danych spełniających warunki P, algorytm zakończy obliczenia i wyniki spełniają warunki końcowe Q (tzn. że algorytm jest **częściowo poprawny i posiada tzw. własność stopu).**

Analiza algorytmu

Złożoność obliczeniowa algorytmu to ilość zasobów komputera potrzebnych do wykonania tego algorytmu (czas, pamięć, ...). Złożoność obliczeniową liczymy zwykle jako ilość charakterystycznych dla danego algorytmu operacji elementarnych lub ilość zajętych komórek (słów) pamięci, wyrażając ją jako funkcję pewnej charakterystycznej wielkości (n) zwanej **rozmiarem zadania** (jest to najczęściej rozmiar danych wejściowych). Dla algorytmów sortowania jest to długość ciągu, dla operacji na macierzach ich rozmiar itp.

Rząd złożoności

Niech LA(n) oznacza ilość operacji elementarnych, jaką należy wykonać (w najgorszym przypadku), aby rozwiązać zadanie algorytmem A dla dowolnych danych rozmiaru n.

Zwykle zamiast funkcji $L_A(n)$ poszukujemy pewnej prostszej funkcji ϕ(n) takiej, że: $\lim_{n\to\infty} \frac{L_A(n)}{\varphi(n)} = const \neq 0$ i mówimy wówczas, że algorytm A ma złożoność rzędu O(ϕ(n)).

Metoda zstępująca, wstępująca i rekurencja

Metoda zstępująca

- rozpoczynamy od zdefiniowania problemu, który chcemy rozwiązać (jest to metoda analityczna)
- problem dzielimy na główne kroki – pod problemy
- pod problemy są dzielone na drobniejsze kroki tak długo, aż rozwiązania kolejnych pod pod problemów stają się łatwe. Nazywamy to stopniowym uszczegółowianiem.

Metoda wstępująca

- zaczynamy od najprostszych operacji i wzbogacamy je nowymi, bardziej złożonymi, dopóki nie będzie można wyrazić ostatecznego rozwiązania problemu (jest to metoda syntetyczna)
- w pewnym stopniu odwrócenie metody zstępującej.

Rekurencja

Rekurencja to zdolność algorytmu do wywoływania samego siebie (zwykle ze zmienionymi parametrami). Teoretyczną podstawą tej techniki programowania jest teoria funkcji rekurencyjnych (Kleene).

Algorytm rekurencyjny można w wielu przypadkach zastąpić równoważnym mu, algorytmem iteracyjnym. Wadą rekurencyjności jest możliwość przepełnienia stosu pamięci, co ogranicza zakres stosowalności. Rekurencja jest narzędziem silnym i użytecznym, ale należy ją stosować z rozwagą.

Metoda „dziel i zwyciężaj”, metoda zachłanna

Metoda „dziel i zwyciężaj”

Dzielimy problem na mniejsze części tej samej postaci, co problem pierwotny – pod problemy dzielimy dalej na coraz mniejsze, używając tej samej metody, aż rozmiar problemu stanie się tak mały, że rozwiązanie będzie oczywiste lub będzie można użyć jakiejś innej efektywnej metody rozwiązania

Rozwiązania wszystkich pod problemów muszą zostać połączone w celu utworzenia rozwiązania całego problemu. Metoda ta jest zazwyczaj implementowana z zastosowaniem rekurencji.

Przykłady: sortowanie przez scalanie, wyszukiwanie binarne.

Metoda zachłanna

W celu wyznaczenia optymalnego rozwiązania globalnego, metoda zachłanna dokonuje w każdym kroku zachłannego, tj. najlepiej rokującego w danym momencie wyboru rozwiązania częściowego.

Algorytm zachłanny w kolejnych krokach dokonuje decyzji lokalnie optymalnej, tzn. dokonuje wyboru wydającego się w danej chwili najlepszym.

Przykład: algorytm Kruskala wyznaczający minimalne drzewo rozpinające dla spójnego grafu nieskierowanego.

Oprogramowanie

Oprogramowanie

Oprogramowanie (software) to całość informacji w postaci zestawu instrukcji, zaimplementowanych interfejsów i zintegrowanych danych przeznaczonych dla komputera do realizacji wyznaczonych celów. Celem oprogramowania jest przetwarzanie danych w określonym zakresie.

Podział oprogramowania

- oprogramowanie systemowe
- oprogramowanie narzędziowe i użytkowe
- oprogramowanie aplikacyjne

System operacyjny

System operacyjny (operating system) to zbiór programów:

- umożliwiających uruchamianie programów użytkownika
- przeznaczonych do dynamicznego zarządzania zasobami komputera i umożliwiających wykorzystanie tych zasobów

Zasób systemu komputerowego (resource) to każdy, niekoniecznie fizyczny, środek, o który mogą ubiegać się użytkownicy i ich programy, zwykle niezbędny do wykonania przez komputer postawionego mu zadania.

Proces to obiekt opisujący pracę systemu komputerowego, któremu przydzielony został procesor, własny obszar pamięci operacyjnej oraz zbiór innych, potrzebnych zasobów.

Wątek to część programu (procesu), która może być wykonywana asynchronicznie, podczas gdy główny kod programu zajmuje się innymi zadaniami.

Jądro systemu – wykonuje główne zadania systemu

Powłoka – pośredniczy w komunikacji użytkownika z systemem operacyjnym i sprzętem

System plików – określa sposób zapisu danych na nośnikach

Warstwowy model systemu

- **Jądro**
  - ma bezpośredni dostęp do wszystkich zasobów komputera
  - zawiera tzw. **program szeregujący** i sterowniki urządzeń
  - zarządza pamięcią operacyjną
- **Biblioteki**
  - zestawy podprogramów wykonujących różne, często stosowane operacje
  - biblioteki są dołączane do programów na etapie konsolidacji (faza po kompilacji)
- **Powłoka**
  - warstwa ta oddziela wewnętrzną część systemu operacyjnego od użytkownika
  - zawiera interpreter poleceń, który umożliwia komunikację z użytkownikiem
  - interpreter poleceń uruchamia polecenia systemu operacyjnego oraz programy użytkownika
- **Programy**
  - procesy uruchamiane przez użytkownika, zarządzane przez program szeregujący jądra
  - każdy program ma przydzielony odpowiedni obszar pamięci i priorytet
  - jeśli proces użytkownika próbuje dostać się do cudzego obszaru pamięci, zostaje przerwany, a system wyświetla odpowiedni komunikat

Sieci komputerowe, Internet

Sieci komputerowe, typy sieci

Sieć komputerowa to zespół systemów komputerowych (sprzętu i oprogramowania), tzw. węzłów sieci, połączonych ze sobą wspólnym medium komunikacyjnym umożliwiającym wymianę informacji i korzystających ze wspólnych zasobów (sprzęt, programy, dane).

Typy sieci komputerowych

- **prywatne** (**PAN** Private Area Network) - zazwyczaj o niewielkim zasięgu, używane w środowisku domowym lub biurowym
- **lokalne** (**LAN** Local Area Network) - zajmują stosunkowo niewielki, ograniczony obszar, np. biuro, zakład pracy, duża szybkość przesyłania informacji, korzystają z okablowania typu skrętka lub kabel koncentryczny
- **kampusowe** - sieci akademickie, miejskie (**MAN** Metropolitan Area Network) - pośrednie między LAN i WAN sieć tego typu może tworzyć szereg sieci LAN na terenie danego miasta połączonych mostami i routerami, medium transmisji to zwykle światłowód
- **rozległe** (**WAN** Wide Area Network) - zajmują dowolnie duży obszar, tworzą je lokalne sieci komputerowe lub sieci MAN połączone ze sobą na duże odległości np. za pomocą łącz telefonicznych, światłowodów lub łącz satelitarnych. Największą i najpopularniejszą siecią rozległą jest **Internet**

Topologia sieci

- **gwiazda** (star) - jeden centralny komputer zarządzający, do którego osobną magistralą są połączone inne
- **pierścień** (ring) - każdy węzeł ma połączenie z dwoma sąsiednimi (może nie mieć centralnego zarządzania)
- **magistrala** (bus) - wszystkie komputery podłączone są do jednej magistrali, nie ma elementów uprzywilejowanych
- **drzewo** (tree) - struktura hierarchiczna
- **siatka** (mesh) - każdy komputer znajdujący się w sieci jest połączony z każdym innym (rozwiązanie niezwykle kosztowne ale niezawodne)
- **mieszane** (hybrid) - łączy cechy różnych topologii
- **sieć szkieletowa** (backbone network) - łączy zwykle mniejsze sieci

Model ISO/OSI

- **warstwa fizyczna** (najniższa) – odpowiada za media fizyczne łączące hosty w sieci
- **warstwa łącza danych** - odpowiada za sterowanie przepływem przesyłanych informacji oraz synchronizację transmisji, określa dostęp do mediów oraz adresuje porty z wykorzystaniem adresów MAC, na tym poziomie tworzona jest i przesyłana **ramka**, czyli sekwencja bitów zawierających dane i elementy kontrolne
- **warstwa sieciowa** - odpowiada za wybór drogi w sieci (routing) na podstawie warunków sieciowych, priorytetu usługi i wielu innych czynników, tworzy podstawowy obiekt - **pakiet** (pakiet ma adres nadawcy i odbiorcy)
- **warstwa transportowa** - po ustaleniu połączenia (dokładnego adresu) nadzoruje transmisję między dwoma użytkownikami, odpowiada za segmentowanie danych oraz zarządzanie niezawodną komunikacją
- **warstwa sesji** - odpowiada za rozpoczęcie i zakończenie, a także sterowanie konwersacją między dwoma aplikacjami poprzez sieć, zapewnia bezpieczeństwo, rozpoznawanie nazw, haseł i logowanie się
- **warstwa prezentacji** - interpretuje i formatuje dane do wyświetlenia i drukowania, przekazuje informacje użytkownikowi, dokonuje konwersji formatów, szyfrowania, kompresji, obsługuje kody sterujące, znaki specjale itp.
- **warstwa zastosowań** - warstwa aplikacji, obsługuje polecenia systemowe systemu operacyjnego oraz programy użytkowe

Protokoły

Protokół TCP (Transmission Control Protocol) steruje ruchem pakietów w sieci, umożliwia sprawdzenie, czy dane dotarły do adresata nieuszkodzone, porządkuje i scala pakiety w takiej kolejności, w jakiej zostały wysłane. Wymaga nawiązania połączenia nadawca – odbiorca i zapewnia niezawodność transmisji.

Protokół IP (Internet Protocol) jest protokołem bezpołączeniowym, nie zapewnia niezawodności transmisji, odpowiada za prawidłowe adresowanie oraz dostarczenie do miejsca przeznaczenia datagramów
Protokoły TCP i IP łącznie zarządzają przepływem danych przez sieć w obu kierunkach. IP przesyła pakiety bez ich rozróżniania, natomiast zadaniem TCP jest gwarancja ich dotarcia.

UDP (User Datagram Protocol) identyfikuje aplikację docelową, oferuje bezpołączeniową usługę transportową umożliwiającą szybką, lecz zawodną metodą dostarczania danych, nie wymaga potwierdzenia odebrania danych i nie ponawia wysłania danych, w przypadku ich utraty lub uszkodzenia, jest używany przez aplikacje wysyłające dane do wielu komputerów

ARP (Address Resolution Protocol) jest odpowiedzialny za rozwiązanie adresów dla wychodzących pakietów (adresy IP są mapowane do adresów MAC)

Usługi w sieci

Protokoły rodziny TCP/IP

DHCP (Dynamic Host Configuration Protocol) - protokół umożliwiający komputerom uzyskanie od serwera danych konfiguracyjnych, np. adresu IP hosta, adresu IP bramy sieciowej, adresu serwera DNS, maski sieci, w szczególności dynamiczne (automatyczne) przyznawanie tymczasowych adresów IP

Telnet to usługa (program) pozwalająca na zdalne połączenie komputera (terminala) użytkownika z oddalonym od niego komputerem (serwerem) przy użyciu sieci, wykorzystując protokół telnet TCP/IP oraz standardowo przypisany jej **port 23**. Umożliwia to zdalną, interakcyjną pracę (dostęp do zasobów – aplikacji i danych) na dowolnym komputerze w sieci, na którym użytkownik ma konto lub zaloguje się jako gość (jeśli jest to dopuszczalne).

SSH (**Secure SHell**, bezpieczna powłoka) jest usługą odpowiadającą usłudze telnet, dodatkowo rozszerzoną o możliwość szyfrowania połączenia pomiędzy klientem i serwerem. Po stronie użytkownika musi być włączony klient SSH, po stronie serwera – serwer SSH. Protokół SSH wykorzystuje **port 22**.

FTP (**File Transfer Protocol**) to usługa typu klient-serwer, która umożliwia przesyłanie plików (tekstowych i binarnych) z i na serwer FTP poprzez sieć TCP/IP.

Protokoły rodziny TCP/IP

- **HTTP** (Hypertext Transfer Protocol) - protokół warstwy aplikacji, jest odpowiedzialny za przesyłanie dokumentów hipertekstowych (stron WWW), w tym informacji z formularzy
- **HTTPS** (Hypertext Transfer Protocol Secure) to szyfrowana za pomocą technologii SSL wersja protokołu HTTP
- **SSL** (Secure Sockets Layer) protokół warstwy transportowej składający się z dwóch podprotokołów: protokołu uzgodnienie parametrów transmisji i protokołu odpowiadającego za utajnianie transmisji.

  Protokół SSL zapewnia trzy podstawowe atrybuty bezpieczeństwa:

  - uwierzytelnienie - weryfikacja serwera i klienta WWW
  - poufność - szyfrowanie przesyłanych informacji
  - integralność - niedopuszczenie do zmiany zawartości komunikatów

Protokoły pocztowe rodziny TCP/IP

- **SMTP** (Simple Mail Transfer Protocol) - podstawowy protokół warstwy aplikacji, służy do wysyłania poczty elektronicznej, brak mechanizmu weryfikacji nadawcy
- **SMTP-AUTH** - rozszerzenie protokołu SMTP o mechanizmy uwierzytelniania
- **POP3** (Post Office Protocol) - protokół warstwy aplikacji pozwalający na odbiór poczty elektronicznej ze zdalnego serwera do lokalnego komputera poprzez połączenie TCP/IP, komunikacja może zostać zaszyfrowana z wykorzystaniem technologii SSL
- **IMAP** (Internet Message Transfer Protocol) - protokół pozwalający na odbiór poczty elektronicznej ze zdalnego serwera, nowszy w stosunk

Cloud computing

Cloud computing (przetwarzanie w chmurze) model przetwarzania oparty na użytkowaniu usług dostarczonych przez zewnętrzne organizacje:

- funkcjonalność jest tu rozumiana jako usługa oferowana przez dane oprogramowanie (oraz konieczną infrastrukturę)
- eliminacja konieczności zakupu licencji czy konieczności instalowania i administrowania oprogramowaniem
- konsument płaci za użytkowanie określonej usługi, np. za możliwość korzystania z arkusza kalkulacyjnego
- nie kupuje sprzętu, ani oprogramowania
- termin cloud computing związany jest z pojęciem wirtualizacji. Idea cloud computing polega na działaniu wszystkich aplikacji w Internecie, niezbędne dla użytkownika są jedynie przeglądarka i szybkie łącze internetowe.