

Pamięć w komputerze

Tomasz IDZIASZEK

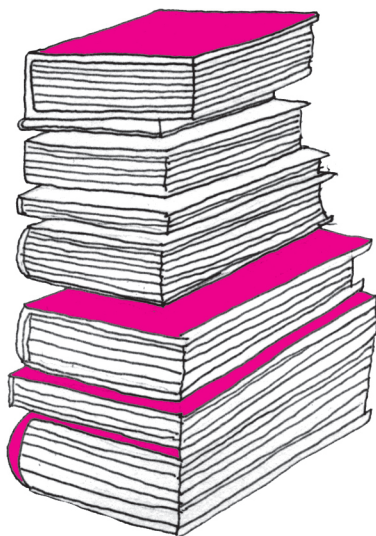
Wyobraźmy sobie ucznia, który ma napisać esej na temat gospodarki w południowych rejonach Brazylii. Powiedzmy, że historia ta dzieje się w czasach, gdy dostęp do Internetu nie był tak powszechny jak teraz. Uczeń w celu zgromadzenia potrzebnych materiałów udaje się do biblioteki, z której wypożycza rocznik statystyczny oraz kseruje potrzebne strony z encyklopedii. Następnie wraca do domu i zabiera się do pisania. Przyniesione materiały ma rozłożone na biurku, aby w każdej chwili mógł do nich sięgnąć. Co więcej, część co ważniejszych informacji z książek zapisał sobie na kartce-ściągawce. Raz na jakiś czas musi też sięgnąć do słownika ortograficznego, który stoi w domowej biblioteczce. Zapewne uczeń nie skończy pracy jednego dnia, przed pójściem spać schowa książki na półkę, skąd jutro będzie musiał je ściągnąć z powrotem.

Ale cóż ta historia ma wspólnego z komputerami – zapyta zniecierpliwiony Czytelnik. Ano tyle, że jest znakomitą analogią tego, jak zorganizowana jest pamięć w komputerze, i pozwoli nam wyjaśnić jej działanie.

Pomyślmy o naszym uczniu jak o procesorze, który jest centralnym elementem każdego komputera. Tak jak procesor realizuje wydawane mu przez programistę polecenia, nasz uczeń również dostał do wykonania pewne zadanie (napisanie eseju). Program, który uczeń ma wykonać, znajduje się już w jego głowie, ale do jego wykonania potrzebne są mu pewne dane (informacje na temat Brazylii). Dane te znajdują się w książkach, które przechowywane są w specjalnym budynku zwanym biblioteką. Uczeń może swobodnie korzystać z tych danych, wiąże się to jednak z pewnym kosztem (dojazd do biblioteki, skserowanie lub przyniesienie ciężkiej książki). W swoim pokoju ma biblioteczkę (mniejszy odpowiednik biblioteki), gdzie może przechowywać wypożyczane książki, posiada też kilka własnych. Zauważmy, że dostęp do książek w biblioteczce jest zdecydowanie szybszy, ale może ona przechowywać dużo mniej książek niż biblioteka. Aktualnie wykorzystywane książki uczeń trzyma na swoim biurku, aby nie musieć wstawać z krzesła i podchodzić do biblioteczki. Ponadto najważniejsze informacje wynotowuje sobie na kartce, żeby nie musieć wertować książek. Każdy z tych racjonalizatorskich pomysłów powoduje szybszy dostęp do wybranych danych, ale z oczywistych względów ilość tych wybranych danych jest w każdym przypadku istotnie mniejsza. Na końcu tego łańcucha miejsc, w których uczeń przechowuje dane, znajduje się pamięć krótkotrwała jego mózgu, zawierająca dane kluczowe w aktualnym momencie (jak choćby kolejne słowo eseju, które uczeń ma zapisać).

Procesor w komputerze postępuje podobnie. Powiedzmy, że wydaliśmy mu polecenie policzenia liczby słów w pewnym tekście zapisanym w pliku. Jak większość plików, znajduje się on na dysku twardym (ang. *hard disk*) komputera, zatem pierwszą czynnością programu będzie wczytanie tego pliku do pamięci operacyjnej (pamięci RAM, ang. *random access memory*). Następnie procesor będzie przeglądał kolejne słowa z tekstu, przenosząc je do swojej niewielkiej pamięci podręcznej (ang. *cache*). Najczęściej wykorzystywane wartości (jak choćby liczba znalezionych słów) procesor przechowuje w kilku wewnętrznych komórkach pamięci, zwanych rejestrami (ang. *register*).

Tak jak uczeń może korzystać z różnych rodzajów dostępu do danych (biblioteka, książka na biurku, kartka-ściągawka, mózg), tak procesor również ma różne rodzaje pamięci (dysk twardy, pamięć RAM, pamięć podręczna, rejestr), które razem tworzą tzw. *hierarchię pamięci w komputerze*. Każdy z tych rodzajów charakteryzuje się innymi wartościami parametrów, z których najważniejsze to pojemność (ang. *capacity*), opóźnienie (ang. *latency*), przepustowość (ang. *bandwidth*) i cena. Im bliżej procesora znajduje się dany rodzaj pamięci, tym ma on mniejsze opóźnienie i większą przepustowość (co nas cieszy), ale mniejszą pojemność i wyższą cenę (co cieszy już nas mniej).



Rozwiązanie zadania F 903.

Niech całkowita moc promieniowania Słońca wynosi L . W odległości R_{ZS} od Słońca na jednostkę powierzchni pada moc $L/(4\pi R_{ZS}^2)$. Oświetlona w tej odległości powierzchnia kuli o promieniu R_Z , współczynniku odbicia a_Z staje się wtórnym źródłem wysyłającym promieniowanie o mocy $a_Z \pi R_Z^2 L / (4\pi R_{ZS}^2)$ w kąt bryłowy 2π (oświetlona jest półkula), a więc w odległości równej odległości Księżyca od Ziemi przypadająca na jednostkę powierzchni moc światła odbitego od Ziemi, I_{ZK} wynosi:

$$I_{ZK} = \frac{a_Z L R_Z^2}{8\pi R_{ZS}^2 R_K^2}$$

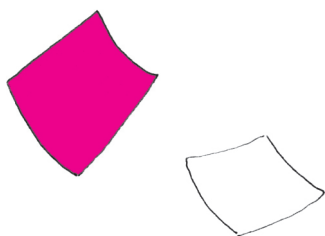
W analogicznym obliczeniu oświetlenia powierzchni Ziemi przez Księżyc, I_{KZ} należy zastąpić a_Z przez a_K i R_Z przez R_K , a R_{ZS} przez sumę odległości Ziemia-Słońce i Ziemia-Księżyc (podczas pełni Księżyc jest po przeciwnej stronie Ziemi niż Słońce), $R_{ZS} + R_{ZK} \approx R_{ZS}$. Ostatnie przybliżenie uwzględnia niewielką odległość Ziemia-Księżyc w porównaniu z odległością Ziemia-Słońce. Po podstawieniu odpowiednich wartości otrzymujemy:

$$\frac{I_{ZK}}{I_{KZ}} = \frac{a_Z}{a_K} \left(\frac{R_Z}{R_K} \right)^2 \approx 35.$$

| Pamięć | Pojemność | Przepustowość | Opóźnienie | $\times 2 \cdot 10^8$ | Koszt |
|------------------|----------------------|-----------------------|--------------------------------|-----------------------|----------------------|
| rejestr | 10^3 B (1 kB) | \approx TB/s | 1 ns | 0,2 s | \approx 1000 zł/GB |
| pamięć podręczna | 10^6 B (1 MB) | | 5 ns | 1 s | |
| pamięć RAM | 10^9 B (1 GB) | 10^4 MB/s (10 GB/s) | 50 ns | 10 s | 30 zł/GB |
| dysk SSD | 10^{11} B (100 GB) | 500 MB/s | $5 \cdot 10^4$ ns (50 μ s) | 3 h | 2 zł/GB |
| dysk talerzowy | 10^{12} B (1 TB) | 100 MB/s | 10^7 ns (10 ms) | 23 dni | 0,2 zł/GB |



O tym, jak pisać programy komputerowe, by efektywnie korzystały z pamięci podręcznej procesora, można przeczytać w artykułach *Pamięć cache w praktyce* (Delta 10/2009) oraz *Kolejność ma znaczenie* (Delta 9/2011).



Rozwiązanie zadania M 1493.
Nie!

Przypuśćmy przeciwnie i zauważmy, że jeśli $a > b$, to mamy $NWD(a, b) \leq b$ oraz $NWD(a, b) \leq a/2$. W takim razie $3 \cdot NWD(a, b) \leq a + b$. Ponieważ suma liczb na krawędziach jest równa sumie liczb w wierzchołkach, mamy

$$3 \cdot \sum_{a \in V} a = 3 \cdot \sum_{a, b \in E} NWD(a, b) \leq \sum_{a, b \in E} (a + b) = 3 \cdot \sum_{a \in V} a,$$

gdzie zbiory V i E odpowiadają zbiorom wierzchołków i krawędzi dwunastościanu. Dla każdej krawędzi o końcach a i b zachodzi więc równość

$$3 \cdot NWD(a, b) = a + b.$$

W takim razie każda krawędź ma na jednym końcu dwukrotność liczby z drugiego końca.

Ustalmy jeden z wierzchołków, z liczbą x . W każdym z trzech sąsiadujących z nim wierzchołków jest liczba $x/2$ lub $2x$, co jest sprzeczne z założeniem o przyporządkowaniu wierzchołkom parami różnych liczb.

W powyższej tabelce przedstawiliśmy przykładowe charakterystyki wybranych pamięci komputerowych. Oczywiście, konkretne liczby zależą od konkretnego sprzętu, ale rząd wielkości będzie podobny. Dla zobrazowania różnic między opóźnieniem dla różnych rodzajów pamięci wprowadziliśmy też kolumnę, w której czas ten został przemnożony przez $2 \cdot 10^8$. Daje to skalę szybkości naszego ucznia.

Rejestr procesora jest najszybszym rodzajem pamięci z czasem dostępu na poziomie jednego cyklu zegara, zatem dla procesora o szybkości 1 GHz (10^9 Hz) będzie to czas rzędu 1 ns (10^{-9} s). W zależności od rodzaju procesora takich rejestrów jest od kilku do kilkudziesięciu. W skali czasowej ucznia odpowiada to dostępowi do danych w mózgu, co zabiera ułamek sekundy.

Pamięć podręczna procesora zawiera dane ostatnio pobrane z pamięci operacyjnej, z czasem dostępu kilku cykli zegara. Zwykle pamięć podręczna jest dodatkowo podzielona na szybszą i mniejszą pamięć (około 32 kB) oraz większą i trochę wolniejszą (kilka MB). Dostęp do tego rodzaju pamięci odpowiada sprawdzaniu przez ucznia informacji na kartce-ściągawce.

Pamięć operacyjna jest głównym rodzajem pamięci w komputerze, stanowiącym rozsądny kompromis (ang. *trade-off*) między wielkością a szybkością. Jej rozmiar to kilka do kilkunastu gigabajtów. Dostęp do komórki pamięci, która nie została dotychczas skopiowana do pamięci podręcznej (ang. *cache miss*), zajmuje kilkadziesiąt nanosekund. Odpowiada to wyszukiwaniu informacji w książce leżącej na biurku.

Pamięć dyskowa została przedstawiona w dwóch wierszach tabelki ze względu na różnicę, jaka dzieli klasyczne dyski twarde (w których dane zapisywane są na talerzach pokrytych nośnikiem magnetycznym) od nowszych dysków SSD (ang. *solid-state disk*), zbudowanych w oparciu o pamięć flash. Pobranie danych z dysku SSD odpowiada wycieczce do biblioteki, podczas gdy pobranie danych z dysku talerzowego można porównać do sytuacji, gdy danej książki nie ma w lokalnej bibliotece i trzeba ją dopiero sprowadzić z dużego miasta.

Oczywiście, z uwagi na to, że rozmiar pamięci podręcznej jest dużo mniejszy niż rozmiar pamięci operacyjnej, a ten z kolei dużo mniejszy niż rozmiar pamięci dyskowej, podczas działania programu nowe dane będą musiały nadpisywać dane poprzednio znajdujące się w pamięci. Uczeń również ma ograniczone miejsce na kartce-ściągawce i biurku, zatem czasem będzie zmuszony wymazać gumką stare informacje lub odłożyć niektóre książki do biblioteczki, żeby zrobić miejsce na nowe.

Warto zwrócić uwagę na różnice między opóźnieniem a przepustowością pamięci. Ta pierwsza wartość mówi nam o średnim czasie, jaki jest potrzebny, aby pobrać z pamięci jeden losowo położony bajt danych. Natomiast druga mówi nam, jak szybko możemy pobierać dane, jeśli będziemy je czytać po kolei. Przykładowo dla dysku talerzowego dostęp do losowego bajtu może zająć 10 ms (w zależności od tego, jak daleko od miejsca zapisu znajduje się głowica dysku), zatem przeczytanie 100 milionów losowych bajtów może trwać nawet 10^6 s, czyli prawie 12 dni! Niemniej jednak, jeśli będziemy czytać kolejne bajty (zgodnie z ruchem głowicy), to każde kolejne 100 milionów przeczytamy średnio w sekundę. Mówimy tu, że dysk talerzowy jest pamięcią o dostępie sekwencyjnym (ang. *sequential access memory*).

Zauważmy, że jest to całkiem naturalne – znalezienie w książce ustępu traktującego o gospodarce Brazylii może zająć uczniowi trochę czasu, ale jeśli już go znajdzie, to kolejne słowa z paragrafu będzie czytał bardzo szybko.

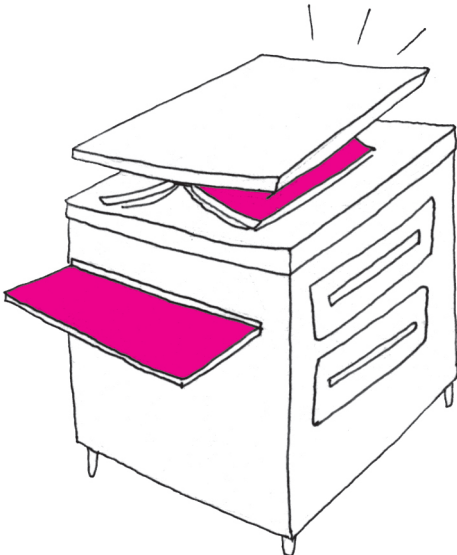


Rozwiązanie zadania F 904.

Każdy maszynista usłyszy wynik interferencji dźwięków dwu sygnałów: „własnej” lokomotywy i lokomotywy nadjeżdżającej z przeciwna. Częstota pierwszego z sygnałów nie ulegnie zmianie, a drugi, na skutek zjawiska Dopplera, będzie odbierany jako dźwięk o częstoci f :

$$f = \frac{c+v}{c-v} f_0.$$

Złożenie harmonicznych sygnałów o częstoci f i f_0 daje sygnał o częstoci $F_1 = (f + f_0)/2$ o amplitudzie modulowanej z częstocią $F_2 = (f - f_0)/2$. Podstawiając dane, otrzymujemy: $f \approx 477,5$ Hz, $F_1 \approx 458,7$ Hz i $F_2 \approx 18,7$ Hz.



Cechami charakterystycznymi obiektów astronomicznych mogą być, między innymi, ich rozmiary katowe, przesunięcie ku czerwieni, okres zmienności i jej amplituda, strumień promieniowania w zadanej długości fali, natężenie i szerokość linii widmowych, czy też różnica strumieni (zwana kolorem).

*Narodowe Centrum Badań Jądrowych w Warszawie

Co prawda, pamięć RAM, zgodnie ze swoją nazwą, jest pamięcią o bezpośrednim dostępie do dowolnej komórki, jednak istnieje jeszcze jeden mechanizm, który powoduje, że w praktyce również w tej pamięci dostęp do kolejnych komórek jest szybszy niż dostęp zupełnie losowy. Mianowicie, dane przesyłane są pomiędzy różnymi rodzajami pamięci nie jako pojedyncze bajty, a w blokach zawierających pewną liczbę kolejnych bajtów. Im większa pamięć, tym większe rozmiary bloków. Przykładowo, podczas pobierania danych z pamięci operacyjnej do pamięci podręcznej oprócz potrzebnego nam bajtu są też kopiowane 63 kolejne bajty (zatem dostęp do pozostałych będzie szybszy, bo w przyszłości będą znajdować się w pamięci podręcznej). Z kolei kopiowanie danych z dysku twardego odbywa się w blokach zawierających po 512 bajtów. W przypadku naszego ucznia odpowiada to kserowaniu całych stron książki, pomimo tego, że potrzebne są mu tylko niektóre informacje zawarte na danej stronie.

Oczywiście, w artykule nie wyliczyliśmy wszystkich możliwości, jakie ma uczeń, gdy chce zdobyć informacje na temat Brazylii (może przecież zapytać nauczyciela, zadzwonić do konsulatu, a w ostateczności może wsiąść w samolot i zdobyć informacje z pierwszej ręki). Podobnie nie omówiliśmy wszystkich sposobów, w jakie użytkownik komputera może dostarczyć mu dane. Wśród pominiętych rodzajów pamięci możemy wspomnieć o różnorodnych pamięciach zewnętrznych, takich jak dyskietki, dyski optyczne (CD, DVD), karty pamięci lub taśmy magnetyczne. Ponadto, aby przyspieszyć ich działanie, dyski twarde również mogą być zaopatrzone w dodatkową pamięć podręczną, obsługiwaną przez specjalny układ zwany kontrolerem dysku. A w dobie Internetu jedną z podstawowych metod pozyskiwania danych jest „ściągnięcie” ich z sieci (gdzie przy ograniczonej prędkości światła istotne jest, w którym miejscu na świecie znajduje się serwer, z którym się łączymy). Zachęcamy Czytelników do samodzielnego sprawdzenia, jakie charakterystyki mają te pamięci, i zaproponowania, jaka analogia ze świata ucznia im odpowiada.

Na koniec warto wspomnieć, że aktualnie kluczowym parametrem przy wyborze komputera biurowego nie jest wcale szybkość procesora, ale szybkość dysku twardego. Zatem to właśnie inwestycja w dysk SSD (lub dwa dyski: mały SSD zawierający partycję z programami oraz większy dysk talerzowy z danymi) może znacząco poprawić komfort pracy z komputerem.

Sztuczna inteligencja w astrofizyce

Aleksandra SOLARZ*, Katarzyna MAŁEK*

W swojej pracy astronomowie często korzystają z katalogów będących wynikiem wielkich przeglądów nieba. Zebrane w nich dane poddawane są różnorodnym analizom. Ważne jest przy tym, aby katalogi charakteryzowały się jak największą kompletnością i jednorodnością. Niejednokrotnie podczas analizy statystycznej dobrze poznanych typów obiektów wykrywane są źródła rzadko spotykane lub wykazujące niezwykle zachowanie. Dlatego też odpowiedni sposób klasyfikacji obserwowanych obiektów jest bardzo istotnym, o ile nie najistotniejszym krokiem, który należy wykonać przed przystąpieniem do zaawansowanego badania właściwości fizycznych obserwowanych obiektów niebieskich.

Cechy charakterystyczne. Przy tworzeniu przeglądów nieba, bądź tych ukierunkowanych na wykrywanie konkretnych klas obiektów, bądź tych zbierających wszystkie możliwe dane z danego obszaru nieba, bardzo ważna jest znajomość cech charakterystycznych obiektów, które spodziewamy się zaobserwować. Ułatwia to proces samej detekcji oraz klasyfikacji zebranych danych i, co za tym idzie, szybkiego wydobycia ważnych informacji z przeprowadzonego przeglądu do dalszej analizy. Często zdarza się, że unikalna klasa obiektów, choć obecna w przeglądzie, nie zostaje wykryta, ponieważ wykazuje cechy charakterystyczne podobne do innych źródeł licznie wykrytych przez przegląd.