

MATERIAŁY DLA UCZNIA

Algorytm wyszukiwania w zbiorze danych

Informatyka

Klasy VII i VIII

45 minut

Lekcja otwarta

Uczeń zapisuje algorytm wyszukiwania jako ciąg jednoznacznych kroków i porównuje liczbę kroków szukania po kolei oraz przez połowienie zakresu na tym samym zbiorze danych.

Czego się nauczysz

Zapis algorytmu w krokach

Umiesz zapisać rozwiązanie jako ponumerowane kroki, które da się wykonać bez zgadywania, i wskazać w nich warunek zakończenia.

Wyszukiwanie po kolei i przez połowienie

Potrafisz policzyć, ile kroków zajmuje znalezienie danej dwiema metodami, i powiedzieć, kiedy połowienie zakresu się opłaca, a kiedy nie.

Kiedy zgadujesz liczbę od 1 do 1000, a druga osoba mówi tylko „za mało” albo „za dużo”, zwykle trafiasz po kilkunastu pytaniach, choć liczb jest tysiąc. To nie jest szczęście, tylko konkretna metoda, którą da się zapisać tak dokładnie, żeby wykonał ją ktoś zupełnie nieświadomy sprytu, na przykład komputer. Z tego materiału dowiesz się, czym różni się algorytm od pomysłu i dlaczego przy milionie danych połowienie zakresu potrzebuje około dwudziestu kroków.

Co odróżnia algorytm od zwykłego pomysłu

Pomysł brzmi „pytaj o środek i zawężaj”. To wystarczy dla człowieka, który zna zasadę gry, ale nie wystarczy dla wykonawcy, który nie wie nic. Algorytm to skończony ciąg kroków na tyle jednoznacznych, że wykonawca nie musi niczego się domyślać ani niczego wybierać na oko, i który przy tych samych danych zawsze daje ten sam wynik.

Jednoznaczność sprawdza się prostym testem: dajesz zapis drugiej osobie i każesz wykonać dosłownie, bez dopowiadania. Krok „sprawdź środek” ten test oblewa, bo przy zakresie od 1 do 16 środek może wypaść na 8 albo na 9. Krok „ustaw środek na wynik

dzielenia sumy dolnej i górnej przez 2, zaokrąglony w dół” test przechodzi, bo nie zostawia miejsca na wybór.

Druga rzecz, bez której zapis nie jest algorytmem, to warunek zakończenia. Algorytm musi umieć zakończyć się także wtedy, gdy szukanej danej w zbiorze nie ma, i podać wtedy sensowną odpowiedź. Bez takiego kroku wykonanie kręci się w kółko, bo żaden warunek go nie przerywa.

Zapisując algorytm, numeruj kroki i używaj tych samych nazw przez cały czas. Jeśli w kroku drugim mówisz „dolna”, nie pisz w kroku szóstym „początek zakresu”, bo wykonawca nie ma prawa wiedzieć, że chodzi o to samo.

SPRÓBUJ SAM

Napisz instrukcję wiązania butów w ponumerowanych krokach i daj ją komuś, kto wykona ją dosłownie, bez poprawiania Twoich niedomówień. Zaznacz każdy krok, przy którym musiał się czegoś domyślić. To są dokładnie te miejsca, w których Twój zapis nie był jeszcze algorytmem.

Szukanie po kolei

Najprostsza metoda wyszukiwania polega na sprawdzaniu pozycji jedna po drugiej, od pierwszej do ostatniej. Zapisana jako algorytm wygląda tak: 1. Ustaw pozycję na 1. 2. Porównaj numer na tej pozycji z szukany. 3. Jeśli równy, wypisz pozycję i zakończ. 4. Jeśli pozycja to ostatnia, wypisz „nie ma” i zakończ. 5. Zwiększ pozycję o 1 i wróć do kroku 2.

Zwróć uwagę, że kroki 3 i 4 to dwa różne zakończenia: jedno dla trafienia, drugie dla wyczerpania listy. Oba są potrzebne. Gdyby zabrakło kroku 4, algorytm szukałby danej, której nie ma, poza końcem listy.

Ta metoda ma dwie ważne zalety. Po pierwsze, działa na zbiorze w dowolnej kolejności, bo niczego nie zakłada o uporządkowaniu. Po drugie, jest krótka w zapisie i trudno się w niej pomylić. Ma jednak jedną poważną wadę: liczba kroków rośnie tak samo szybko jak zbiór. Przy 16 pozycjach to do 16 porównań, przy tysiącu do tysiąca, przy milionie do miliona.

Jest jeszcze jeden szczegół, który przyda się dalej. Liczba kroków szukania po kolei zależy od tego, gdzie leży szukana dana. Jeśli jest piąta od początku, wystarczy 5 kroków, niezależnie od tego, czy lista ma 16 pozycji, czy tysiąc. Najgorszy przypadek to dana na samym końcu albo dana, której w zbiorze nie ma.

SPRÓBUJ SAM

Weź listę 16 numerów inwentarza: 1042, 1087, 1130, 1156, 1199, 1204, 1233, 1290, 1315, 1342, 1388, 1401, 1455, 1470, 1512, 1560. Policz, ile kroków zajmie szukanie po kolei numeru 1199, numeru 1560 i numeru 1305, którego na liście nie ma. Zapisz trzy liczby i sprawdź, która z nich jest największa.

Połowienie zakresu

Druga metoda korzysta z jednej dodatkowej informacji: lista jest uporządkowana rosnąco. Dzięki temu porównanie z jedną pozycją mówi coś o wszystkich pozostałych. Jeśli numer ze środka jest mniejszy od szukanego, cała lewa połowa listy odpada, bo wszystko tam jest jeszcze mniejsze. Jednym porównaniem odrzucasz połowę zbioru.

Zapis wygląda tak: 1. Ustaw dolną na 1, górną na ostatnią pozycję. 2. Jeśli dolna jest większa od górnej, wypisz „nie ma” i zakończ. 3. Ustaw środek na wynik dzielenia sumy dolnej i górnej przez 2, zaokrąglony w dół. 4. Porównaj numer ze środka z szukanym. 5. Jeśli równy, wypisz środek i zakończ. 6. Jeśli numer ze środka jest mniejszy, ustaw dolną na środek plus 1, w przeciwnym razie ustaw górną na środek minus 1. 7. Wróć do kroku 2.

Najważniejszy i najczęściej psuty jest krok 6. Dolną ustawiasz na środek plus 1, a nie na środek, i górną na środek minus 1, a nie na środek. Powód jest praktyczny: pozycję ze środka właśnie sprawdziłeś, więc nie ma sensu trzymać jej w zakresie. Bez tych jedynek zakres przestaje się kurczyć, gdy zostaną dwie pozycje, i algorytm nigdy się nie kończy.

Sprawdźmy to na liście z poprzedniej sekcji dla numeru 1401. Zakres 1 do 16, środek 8, tam stoi 1290, czyli mniej, więc dolna idzie na 9. Zakres 9 do 16, środek 12, tam stoi 1401, trafione. Dwa kroki na szesnastu pozycjach. Dla numeru 1305, którego na liście nie ma, po czterech krokach dolna przekracza górną i algorytm kończy się odpowiedzią „nie ma”, dokładnie tak, jak przewiduje krok 2.

SPRÓBUJ SAM

Wykonaj połowienie zakresu na tej samej liście 16 numerów dla wartości 1199 i 1560, zapisując po każdym kroku trzy liczby: dolną, górną i środek. Porównaj wynik z liczbą kroków szukania po kolei z poprzedniej sekcji. Przy jednym z tych numerów przewaga połowienia będzie zaskakująco mała i warto sprawdzić, na której pozycji leży ten numer.

Ile kroków przy większym zbiorze

Skoro każdy krok zmniejsza zakres o połowę, to pytanie o liczbę kroków jest pytaniem o to, ile razy trzeba przepołować zbiór, żeby zostać z jedną pozycją. Najłatwiej policzyć to od drugiej strony: ile razy trzeba podwoić liczbę 1, żeby dojść do wielkości zbioru. 1, 2, 4, 8, 16 to cztery podwojenia, więc zbiór 16 pozycji wymaga do 5 kroków razem z ostatnim sprawdzeniem.

Dalej ta sama arytmetyka daje wyniki, które na początku wydają się nieprawdopodobnie. Tysiąc pozycji to do 10 kroków. Milion pozycji to do 20 kroków. Miliard to do 30. Podwojenie wielkości zbioru dokładnie dokładnie jeden krok, a nie podwaja pracy, i to jest cała siła tej metody.

Porównaj to z szukaniem po kolei. Przy milionie pozycji jedna metoda potrzebuje do miliona porównań, a druga do dwudziestu. Ta różnica decyduje o tym, czy wyszukiwarka albo system biblioteczny odpowiada natychmiast, czy po długiej chwili. Nie chodzi o to, że komputer jest wolny. Chodzi o to, ile pracy mu zadałeś.

Warto też zauważyć, że liczby te dotyczą najgorszego przypadku. Połowienie może trafić w pierwszym kroku, jeśli szukana dana stoi dokładnie w środku. Szukanie po kolei może trafić w pierwszym kroku, jeśli dana stoi na początku. Metody porównuje się jednak po najgorszym przypadku, bo to on decyduje, czy system zadziała zawsze, a nie tylko czasem.

SPRÓBUJ SAM

Wypisz ciąg podwojeń: 1, 2, 4, 8, 16, 32 i dalej, aż przekroczysz 1000. Policz, ile razy podwoiłeś liczbę, czyli ile strzałek między kolejnymi liczbami zapisałeś, a nie ile liczb. Powinno wyjść 10. Potem zrób to samo do miliona i sprawdź, czy wynik zgadza się z liczbą 20. Na koniec odpowiedz: o ile kroków wzrośnie praca, gdy zbiór urośnie z miliona do dwóch milionów.

Kiedy połowienie nie pomaga

Połowienie ma jeden twardy warunek: zbiór musi być uporządkowany. Cała metoda opiera się na założeniu, że wszystko na lewo od środka jest mniejsze, a wszystko na prawo większe. Na liście w przypadkowej kolejności to założenie nie obowiązuje, więc porównanie ze środkiem nie mówi, po której stronie szukać. Algorytm wykona się do końca i poda wynik, tylko że ten wynik może być błędny, a to gorsze niż brak wyniku.

Dlatego przed wyborem metody warto dopisać sobie krok zerowy: sprawdź, czy zbiór jest uporządkowany. Jeśli nie jest, masz dwa wyjścia. Możesz szukać po kolei albo najpierw uporządkować zbiór, a porządkowanie też kosztuje czas, i to zwykle więcej niż jedno szukanie po kolei.

Stąd praktyczny wniosek: jeśli szukasz w zbiorze tylko raz i zbiór jest krótki, prostsze i szybsze bywa szukanie po kolei. Połowienie opłaca się wtedy, gdy zbiór jest duży, uporządkowany i przeszukiwany wiele razy. Dokładnie tak działają systemy biblioteczne czy sklepowe: dane trzyma się w uporządkowanej postaci, bo szuka się w nich bez przerwy.

Jest też przypadek pośredni, który łatwo przeoczyć. Jeśli szukana dana leży blisko początku listy, szukanie po kolei potrafi wygrać nawet na uporządkowanym zbiorze. Na liście szesnastu numerów szukanie piątej pozycji zajmuje po kolei 5 kroków, a połowieniem 4, więc przewaga niemal znika. Metoda nie jest lepsza sama w sobie, jest lepsza dla konkretnych danych i konkretnego sposobu użycia.

SPRÓBUJ SAM

Weź listę 16 numerów i przestaw je w przypadkowej kolejności. Wykonaj na tak przestawionej liście algorytm połowienia dla numeru 1130, wykonując kroki dosłownie, bez zaglądania na całą listę. Zapisz, jaką odpowiedź podała algorytm i czy jest zgodna z prawdą.

Słowniczek

algorytm

Skończony ciąg jednoznacznych kroków, który przy tych samych danych zawsze daje ten sam wynik i da się wykonać bez domyślania.

warunek zakończenia

Krok, który przerywa wykonanie algorytmu, także wtedy, gdy szukanej danej w zbiorze nie ma.

krok

Jedno porównanie szukanej wartości z wartością ze zbioru; jednostka, w której mierzymy pracę algorytmu wyszukiwania.

szukanie po kolei

Metoda sprawdzająca pozycje jedna po drugiej od początku zbioru, działająca także na zbiorze nieuporządkowanym.

połowienie zakresu

Metoda odrzucająca po każdym porównaniu połowę zbioru, wymagająca zbioru uporządkowanego.

zbiór uporządkowany

Zbiór ustawiony rosnąco albo malejąco, w którym pozycja wartości mówi coś o wszystkich pozostałych.

najgorszy przypadek

Największa liczba kroków, jakiej metoda może potrzebować dla danego zbioru; po nim porównuje się metody.

Częste pomyłki

- Stosowanie połowienia do listy w przypadkowej kolejności. Algorytm poda wtedy wynik, który bywa błędny, a to trudniej wykryć niż brak wyniku. Zawsze sprawdź uporządkowanie, zanim wybierzesz metodę.
- Pominięcie kroku „jeśli dolna jest większa od górnej, wypisz nie ma i zakończ”. Bez niego szukanie danej spoza zbioru nie ma jak się skończyć i wykonanie kręci się w kółko.
- Ustawianie dolnej na środek zamiast na środek plus 1. Przy zakresie od 5 do 6 środek zawsze wypada na 5, więc zakres przestaje się kurczyć i algorytm pyta w nieskończoność.
- Zaokrąglanie środka raz w górę, raz w dół. Te same dane dają wtedy różne liczby kroków i nie da się porównać wyników. Zapisz regułę zaokrąglania obok algorytmu i trzymaj się jej do końca.

Sprawdź, czy rozumiesz

1. Na uporządkowanej liście 16 pozycji szukasz wartości leżącej na pozycji trzeciej. Która metoda wypadnie lepiej i czy Twoja odpowiedź zmieni się, gdy lista urośnie do tysiąca pozycji?
2. Kolega zapisał algorytm połowienia i ustawia dolną na środek zamiast na środek plus 1. Na jakim najmniejszym zakresie pokażesz mu, że jego zapis nigdy się nie kończy?
3. Masz zbiór miliona nieuporządkowanych danych i musisz znaleźć w nim jedną wartość dokładnie raz. Czy opłaca się najpierw go uporządkować i od czego zależy Twoja odpowiedź?

Wyzwanie dla chętnych**PONAD PROGRAM**

Zbuduj w Scratchu wersję odwrotną: to Ty zgadujesz liczbę od 1 do 1000 pomyślaną przez komputer (blok „losuj od 1 do 1000”), a program liczy Twoje próby i odpowiada „za mało” albo „za dużo”. Potem zagraj dziesięć razy i zapisz największą liczbę prób, jaką potrzebowałeś.

Jak sprawdzisz, że Ci wyszło: Jeśli za każdym razem pytasz o środek zakresu, żadna z dziesięciu gier nie przekroczy 10 prób. Jeśli któraś przekroczyła, wróć do jej zapisu i znajdź pytanie, które nie było o środek.

Zadanie do domu na trzy poziomy

Wybierz jeden poziom. Możesz zacząć od łatwego i wrócić po trudniejszy.

POZIOM: ŁATWE

Ułóż 8 karteczek z liczbami 3, 8, 15, 21, 27, 34, 40, 48 rosnąco, liczbami do dołu, i znajdź liczbę 34 połowieniem, odwracając tylko wskazaną karteczkę. Zapisz, które karteczki odwróciłeś i ile ich było. Powinny wyjść dokładnie 2.

POZIOM: ŚREDNIE

Na liście 16 numerów z lekcji policz kroki obu metod dla numerów 1042, 1342 i 1600 (tej ostatniej na liście nie ma). Zapisz wyniki w tabeli i podkreśl numer, przy którym szukanie po kolei wygrało.

POZIOM: TRUDNE

Zbuduj w Scratchu skrypt zgadujący liczbę od 1 do 1000 i dopisz sprawdzenie uczciwości: jeśli dolna przekroczy górną, a gracz nie powiedział „trafione”, program wypisuje, że odpowiedzi były sprzeczne. Wyjaśnij w dwóch zdaniach, dlaczego ten sam warunek w algorytmie wyszukiwania oznacza „nie ma”.

Samoocena

Zaznacz, co już potrafisz. To tylko dla Ciebie, nikt tego nie ocenia.

- ☐ Potrafię zapisać algorytm jako ponumerowane kroki, które kolega wykona bez domyślania.
- ☐ Potrafię wskazać w algorytmie krok, który kończy działanie, gdy szukanej danej nie ma.
- ☐ Potrafię policzyć kroki szukania po kolei i połowienia dla podanego numeru na liście 16 pozycji.
- ☐ Potrafię powiedzieć, kiedy połowienie się nie opłaca albo w ogóle nie działa.