

Temat 5

20 pytań – *Teoria informacji*

Streszczenie

Ile informacji znajduje się w tysiącstronicowej książce? Czy więcej informacji znajduje się w książce telefonicznej, na 1000 stron tradycyjnych wydruków kartek A4, czy we *Władcy Pierścieni* Tolkiena? Jeśli będziemy potrafili to oszacować, będziemy potrafili też oszacować ile miejsca potrzeba do zapisania jakiegokolwiek innej informacji.

Czy potrafisz odczytać zapisane poniżej zdanie?

W tm zdn brkj smgłsk.

Bardzo możliwe, że tak. W samogłoskach nie ma zawartych zbyt wielu „informacji”. W ramach tych zajęć przedstawiona będzie metoda pomiaru „zawartości” informacyjnej.

Wiek

✓ 10 i więcej

Środki dydaktyczne

✓ Żadne środki dydaktyczne nie będą używane.

Do dodatkowego zadania będzie potrzebna karta pracy: Drzewa decyzyjne (s. 40)

20 pytań

Dyskusja

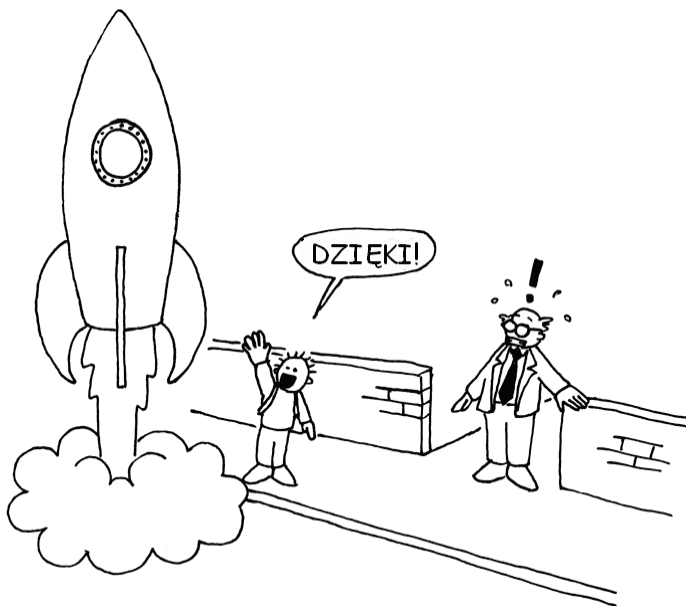
1. Zapytaj dzieci o to, czym według nich jest informacja.
2. W jaki sposób moglibyśmy zmierzyć ilość informacji, która jest zawarta w jakiejś książce? Czy ważne są np. liczba stron albo liczba wyrazów? Czy jedna książka może zawierać więcej informacji niż inna? Co w przypadku książki bardzo nudnej lub szczególnie interesującej? Czy książka, w której na 400 stronach powtarza się fraza „bla, bla, bla” ma mniej czy więcej informacji niż np. książka telefoniczna?

Informatycy-naukowcy mierzą ilość informacji według stopnia jej prawdopodobieństwa. Usłyszenie czegoś, co jest już Ci znane – na przykład słów „Przyszedłem dziś do szkoły pieszo” od kolegi, który tak robi codziennie – nie wiąże się to z otrzymaniem właściwie żadnej informacji, ponieważ nie jest to czymś nieprawdopodobnym (zaskakującym). Gdyby jednak kolega powiedział „Przyleciałem dziś do szkoły helikopterem”, to byłoby to zaskakujące, i dlatego wiązałoby się z przekazaniem większej ilości informacji.

W jaki sposób można zmierzyć to, jak nieprawdopodobna jest informacja?

Jednym ze sposobów jest sprawdzenie, czy łatwo odgadnąć daną informację. Twój kolega mówi: „Spróbuj zgadnąć, w jaki sposób dotarłem dziś do szkoły!”. Jeśli dotarł pieszo lub autobusem, prawdopodobnie od razu odgadniesz. Nie przyjdzie Ci na myśl, by wymienić helikopter czy statek kosmiczny.

Ilość informacji, jaka jest zawarta w danej wiadomości, mierzy się tym, jak łatwo lub trudno je odgadnąć. Następująca gra pozwoli lepiej zrozumieć ten temat.



Gra w 20 pytań

Posłużymy się adaptacją dość znanej gry w 20 pytań: Wybrane dziecko ustala w myślach liczbę a pozostałe dzieci zadają pytania (aż do skutku), na które może ono odpowiadać tylko „tak” lub „nie”.

Propozycje pytań:

Pomyśl o:

- ✓ liczbie między 1 i 100
- ✓ liczbie między 1 i 1000
- ✓ liczbie między 1 i 1 000 000
- ✓ dowolnej liczbie całkowitej
- ✓ szczególnym ciągu sześciu liczb (np. 2, 4, 6, 8, 10)

Zlicz liczbę pytań, które zostały zadane. To będzie miara wartości „informacji”.

Dyskusja

Jakich strategii używaliście? Które z nich były najlepsze?

Podkreśl, że do znalezienia liczby między 1 i 100 powinno wystarczyć siedem prób odgadnięcia, jeśli za każdym razem dzielimy przedział na pół. Dla przykładu:

Czy to liczba mniejsza od 50?	Tak.
Czy to liczba mniejsza od 25?	Nie.
Czy to liczba mniejsza od 37?	Nie.
Czy to liczba mniejsza od 43?	Tak.
Czy to liczba mniejsza od 40?	Nie.
Czy to liczba mniejsza od 41?	Nie.
To musi być liczba 42!	Tak!

Ciekawe jest to, że dziesięciokrotne rozszerzenie zakresu liczb (do 1000) nie wymaga dziesięciokrotnego zwiększenia nakładu pracy – liczba potrzebnych pytań wzrasta co najwyżej o trzy! Za każdym razem, gdy zakres liczb podwajamy, liczba pytań potrzebnych do znalezienia odpowiedzi zwiększa się tylko o jedną.

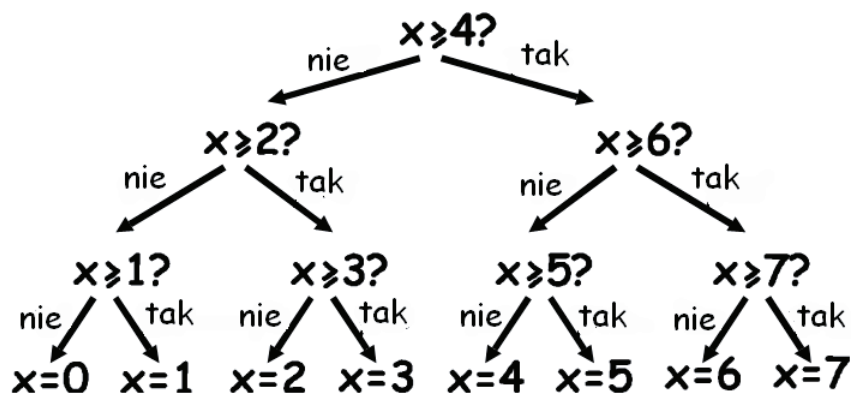
Rozwinięcie: Ile informacji jest w tej wiadomości?

Informatycy posługują się metodą domniemywania (odgadywania) nie tylko w przypadku liczb – również np. w przypadku liter starają się określić tę najbardziej prawdopodobną, która może pojawić się jako następna w wyrazie czy zdaniu.

Możecie spróbować gry w odgadywanie krótkiego zdania, składającego się z 4-6 wyrazów. Litery muszą być odgadywane we właściwej kolejności od pierwszej do ostatniej. Niech ktoś zapisuje odgadnięte litery oraz informację o tym, ilu prób to wymagało. Można używać różnych pytań, na które można odpowiedzieć „tak” lub „nie”. Na przykład: „Czy to litera *t*?”, „Czy to samogłoska?”, „Czy litera to występuje w alfabecie przed literą *m*?”. Odstęp między wyrazami także traktujemy jako „literę” do odgadnięcia”. Na końcu zamieńcie się rolami. Które części wiadomości najłatwiej wykryć?

Karta pracy: Drzewa decyzyjne

Oto schemat zwany drzewem decyzyjnym, który może służyć odgadnięciu liczby z zakresu od 0 do 7:



Których pytań użyć do „odgadnięcia” liczby 5?

Ile pytań trzeba zadać, aby „odgadnąć” określoną liczbę?

Liczby 0, 1, 2, 3... z ostatniego rzędu zapisz w postaci numeracji binarnej.
Dostrzegasz coś ciekawego?

Przyjrzyj się dokładnie drzewu. „Nie” = 0 i „tak” = 1. Co można zauważyć?
(Zakodowany zerojedynkowo, czyli w numeracji dwójkowej, optymalny ciąg „zgadywań”.)

Zaprojektuj drzewo decyzyjne dla „odgadywania” liczb od 0 do 15.

Zadanie dodatkowe:

Jak wyglądałoby drzewo służące do odgadywania wieku pyтанej osoby?
Jak wyglądałoby drzewo, które służyłoby odgadywaniu następnej litery w zdaniu?

O co w tym wszystkim chodzi?

Znany amerykański matematyk (żongler oraz unicyklista) Claude Shannon przeprowadził wiele eksperymentów z opisaną w scenariuszu grą. Mierzył ilość informacji za pomocą bitów – każda odpowiedź „tak” lub „nie” jest odpowiednikiem bitów 1 lub 0. Odkrył, że ilość „informacji” zawartej w jakiejś wiadomości zależy od tego, co wcześniej jest już znane. Czasami można zadać takie pytanie, które eliminuje konieczność stawiania wielu innych. W takim wypadku zawartość informacyjna wiadomości (komunikatu) jest niska. Dla przykładu: informacja dotycząca wyniku rzutu monetą jest jednobitowa: orzeł lub reszka. Ale jeśli moneta okaże się być niesymetryczną i orzeł wypada średnio 9 razy na 10 rzutów, to wówczas miarą ilości informacji dotyczącej rzutu taką monetą jest mniej niż 1 bit. Musimy postawić odpowiednie pytanie dotyczące dwóch kolejnych rzutów. Należy użyć pytania: „Czy wynikami kolejnych dwóch rzutów monetą jest orzeł?”. W przypadku rzutu taką niesymetryczną monetą odpowiedź „tak” pojawi się w ok. 80% przypadków. W ok. 20% przypadków, kiedy pojawi się odpowiedź „nie”, potrzebne będzie zadanie dwóch dalszych pytań. Jednak średnio (statystycznie) liczba pytań będzie mniejsza niż jedno pytanie na jeden rzut monetą!



Shannon nazwał zawartość informacyjną wiadomości „entropią”. Entropia zależy nie tylko od liczby możliwych wyników – np. dwóch w przypadku rzutu monetą – ale także od prawdopodobieństwa ich wystąpienia. Nieprawdopodobne zdarzenia, czy zaskakujące informacje, wymagają postawienia o wiele większej liczby pytań prowadzących do odgadnięcia odpowiedzi, ponieważ zawierają więcej informacji, których wcześniej nie znaliśmy. Dokładnie tak było w przypadku historii, wspomnianego w czasie zajęć, ucznia, który dotarł do szkoły helikopterem.

Entropia jest ważnym pojęciem dla informatyki. Niemożliwym jest skompresowanie (upakowanie) wiadomości tak, aby zajmowała mniej miejsca niż wartość jej entropii. Najlepsze systemy kompresji są, używając obrazowego porównania, odpowiednikami wyżej opisanej zgadywanki (gry w odgadywanie). Wówczas „zgadywanie” jest wykonywane przez program komputerowy. (Lista pytań może być odtworzona później.) Plik skompresowany to ciąg bitów, które są odpowiedziami w „zgadywance”. Tak długo jak odpowiedzi (bity) są przechowywane, możemy odtworzyć informację! Najlepsze systemy kompresji potrafią zmniejszyć rozmiar plików tekstowych do ok. ¼ ich rozmiaru oryginalnego – oznacza to dużą oszczędność przestrzeni dyskowej!

Podobna metoda „odgadywania” może być zastosowana podczas projektowania komputerowego interface’u. Aby przewidzieć kolejne czynności użytkowników, zwłaszcza tych szczególnych – osób niepełnosprawnych, dla których np. pisanie na klawiaturze stanowi dużą trudność. Komputer sugeruje wówczas np. najbardziej prawdopodobne (najczęściej używane) słowa, a użytkownicy tylko wskazują to właściwe. Dobrze zaprojektowany system wymaga średnio tylko dwóch pytań i odpowiedzi typu „tak” lub „nie”, w celu „odgadnięcia” kolejnej litery. System tego rodzaju jest również używany w niektórych telefonach komórkowych (tzw. metoda słownikowa).

Rozwiązania i wskazówki

Odpowiedź „tak” lub „nie” na postawione pytanie odpowiada dokładnie jednemu bitowi informacji –niezależnie od tego, czy jest to proste pytanie typu „Czy jest to liczba większa od 50?” czy bardziej złożone jak “Czy jest to liczba większa od 20 i mniejsza od 60?”.

W przypadku gry dotyczącej odgadywania liczb, niezależnie od sposobu zadawania pytań, ciąg odpowiedzi jest po prostu liczbą zapisaną w postaci numeracji binarnej. Liczba trzy to 011 i odpowiada ciągowi odpowiedzi „Nie, tak, tak” w drzewie decyzyjnym.

Drzewo, którego można by użyć do odgadywania wieku pytanej osoby powinno być drzewem niezrównoważonym (tzn. w jego korzeniu powinno pojawić się pytanie dot. stosunkowo małej liczby lat).

Decyzja co do kolejnej litery w zdaniu mogłaby być uzależniona od poprzedniej litery