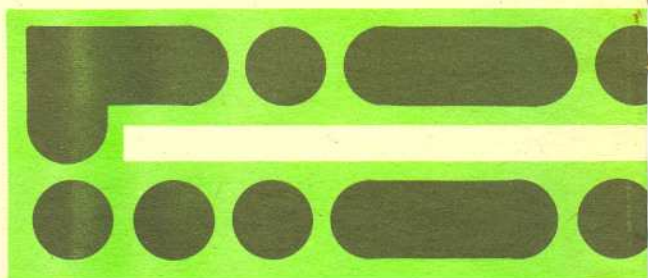
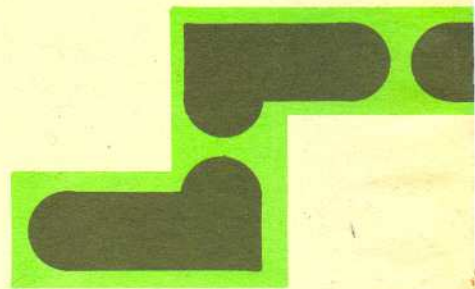


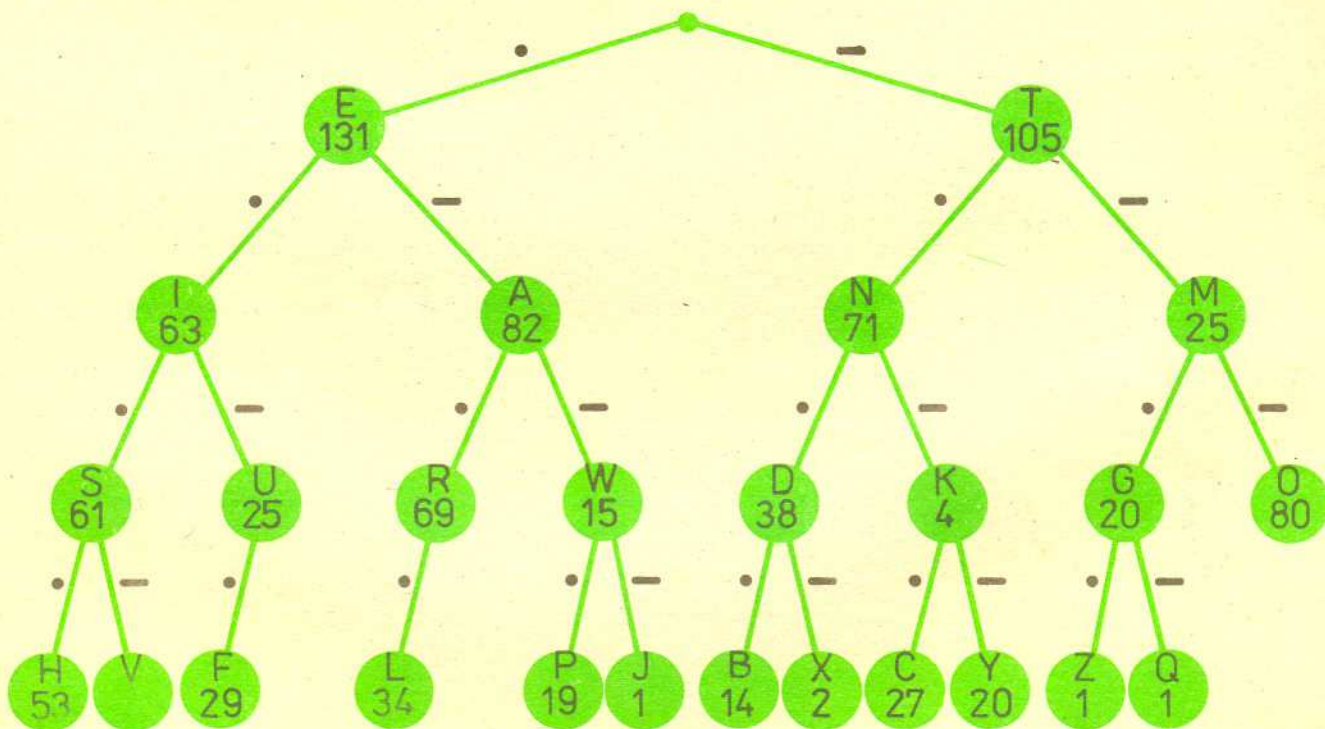
delta



Alfabet Morse'a

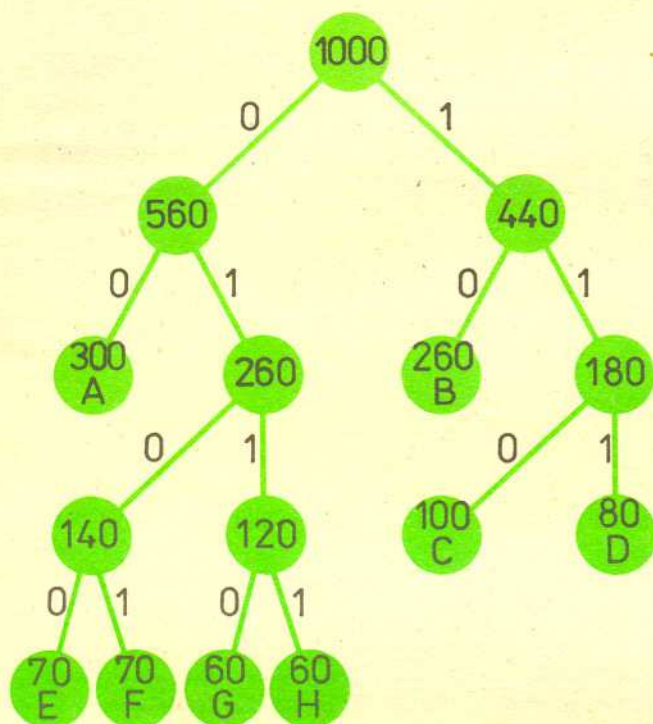
został opracowany w roku 1838 w celu przekazywania wiadomości telegrafem. Każdej literze odpowiada pewna kombinacja „kropek” i „kresek” — krótkich i długich sygnałów. Co ciekawe, literom częściej występującym w języku angielskim odpowiadają krótkie kody, a rzadziej występującym — dłuższe. E, które pojawia się najczęściej (średnio 131 razy na 1000 liter), to pojedyncza kropka. Z występuje średnio 1 raz na 1000 liter i ma kod „— — —”. Nietrudno zauważyć, że taki sposób kodowania jest oszczędny. Przekazanie wiadomości złożonej z 1000 liter wymaga średnio 2500 znaków i 1000 odstępów między kodami poszczególnych liter (są to odpowiedniej długości pauzy). Gdybyśmy natomiast chcieli kodować tekst w sposób prymitywny, dla każdej z 26 liter trzeba by było pięciu znaków (dlatego, że jest tylko $2^4 = 16$ ciągów złożonych z czterech znaków; dopiero $2^5 = 32$). Moglibyśmy tu zrezygnować z odstępów między literami, niemniej jednak 1000 liter wymagałoby 5000 znaków.

Alfabet Morse'a można przedstawić w postaci drzewa. W węzłach znajdują się litery wraz z częstościami. Krok w lewo to kropka, w prawo — kreska. Widać tu jednak pewne odchylenia od optymalnego wyboru kodów. Zastanawia niska pozycja O (— — —) i wysoka M (— — —). Może wzięto pod uwagę łatwość uczenia się alfabetu? Z sześciu liter: E, T, I, A, N, M, których kody składają się z co najwyżej dwóch znaków, można pewnie ułożyć więcej słów do ewentualnych „wprawek”, niż z sześciu liter najczęściej występujących: E, T, I, A, N, O. Na dodatek litery E, T, I, A, N, M stanowią i tak prawie 50% tekstu.



Co będzie, jeśli zechcemy kodować nie pojedyncze litery, ale pary lub trójki liter czy wręcz całe słowa? Okazuje się, że taki kod jest jeszcze bardziej oszczędny, ma jednak poważną wadę: procesowi kodowania podołałby chyba tylko komputer. Pierwotny pomysł Samuela Morse'a był właśnie taki. Kombinacje kropek i kresek oznaczały numery słów w księdze kodowej, nad którą Morse pracował w latach 1832—1837.

Przedstawimy teraz inny sposób kodowania, pochodzący od Huffmana. Dla prostoty przypuścimy, że w przekazywanym tekście występują symbole A, B, C, D, E, F, G, H z częstościami odpowiednio 300, 260, 100, 80, 70, 70, 60, 60 na 1000 znaków. Budujemy teraz drzewo od dołu: najpierw znajdujemy dwa znaki o najmniejszych częstościach. Łączymy je węzłem i piszemy w węzle sumę częstości. Zapominamy teraz o częstościach znaków poniżej węzła (węzłów) i powtarzamy operację, dopóki się da. (Drzewo na rysunku zostało już uporządkowane.) Na gałęziach drzewa prowadzących w lewo piszemy 0, w prawo — 1. Możemy teraz odczytać kody: E to 0100, B — 10, A — 00. Po chwili zastanowienia widać, że nie potrzeba tu dodatkowych znaków dla odstępów między kodami poszczególnych symboli. Taki kod jest najoszczędniejszy, jeśli chcemy kodować pojedyncze symbole.



Na zakończenie zauważmy, że kodem służącym do przekazywania wiadomości jest język naturalny. Procesy optymalizacyjne zachodzą w nim między innymi dzięki naszemu lenistwu. Oto kilka przykładów: w sklepie mówimy raczej „kilo” zamiast „kilogram”, a „sześćdziesiąt” w potocznej wymowie brzmi „szesiąt”. Przypomnijmy szkolne „panie psorze” i staropolskie „Waszmość” i „Waść”, pochodzące od „Wasza Miłość” oraz liczne zdrobnienia imion. W rezultacie słowa częściej używane są krótsze, a uczące się mówić dziecko najpierw przyswaja sobie słowa „mama”, „tata”, „kot” i tym podobne, zanim przyjdzie kolej na tasiemce, takie jak „antyimperialistyczny”, „kontrpropozycja” i „rewaloryzacja”.



Małą Deltę przygotował Rafal SZTENCEL