

Informatikai alapismeretek

Az információ

Alapismeretek: információ, adat, jel

A tágabb értelemben vett informatika az információ keletkezésével, továbbításával, tárolásával és feldolgozásával foglalkozik, tehát hatalmas területet fed le.

Az információ olyan új ismeret, amely megszerzője számára szükséges, és korábbi tudása alapján értelmezhető. Szokták ezt úgy is fogalmazni, hogy az információ valamely meglévő bizonytalanságot szüntet meg. Egy középiskolás tanuló részére nem információ, hogy mikor volt a mohácsi vész, hiszen számára ez nem új ismeret. Ugyanez egy afrikai bennszülött számára szintén nem információ, mert nem tudja értelmezni.

A szűkebb értelemben vett informatikán a számítógépes információfeldolgozást értik. Világos, hogy itt az információ speciális módon jelenik meg. **A számítástechnikai eszközökkel rögzített, azokkal feldolgozható és megjeleníthető információt adatnak nevezünk.**

Az információ hordozója a jel. A jel igen tág fogalom, a készülék ebéd illata éppúgy információt közöl, mint egy közlekedési jelzőtábla. **Egy jelet analógnak nevezünk, ha két érték között tetszőleges értéket felvehet.** Ilyen például az ember magassága vagy tömege. **A digitális jel csak előre meghatározott számú véges sok értéket vehet fel.** Ilyen például egy felnőtt ember gyerekeinek a száma, vagy a boltban kapható pólók mérete. Ugyanez az információ megjelenhet akár mindkét módon is, gondoljunk pl. az analóg és a digitális órákra.

Az analóg jel mindig digitálissá alakítható (*digitalizálható*). Ilyenkor az analóg jel értékét csak bizonyos lépésközzönként rögzítik (*mintavételezés*), majd a kapott számokat előre megadott értékekre kerekítik (*kvantálás*). Természetesen fontos, hogy a lépésközi kisebb legyen érzékszerveink (pl. kép esetén szemünk) érzékenységénél.



*A minta kedvéért a képet igen nagy lépésközzel digitalizáltuk.
Mit jelent ebben az esetben a mintavételezés és a kvantálás?*

Az információ egysége

Mivel az információ áruként jelenik meg, vagyis fizetnünk kell érte, nagyon fontos az információ mérése. Az információ mérése azon alapul, hogy ma gyakorlatilag minden informatikai eszköz speciális digitális jeleket, ún. bináris jeleket használ.

A bináris jelek csupán kétféle értéket vehetnek fel. Ennek megfelelően a **digitális eszközök** csak kétféle értéket továbbíthatnak, például van áram/nincs áram, vagy a mérhető elektromos feszültség $0\text{ V}/5\text{ V}$. Az ilyen eszközök **az információt bináris jelsorozatként kezelik**, pl. $0\text{ V} - 5\text{ V} - 5\text{ V} - 0\text{ V}$. Logikusnak tűnik, hogy az információt a jelsorozat hosszával mérjük, így ez a jelsorozat 4 egységnyi információt tartalmaz.

Az információ egységét a *bit*nek nevezték el (*binary unit*). **Egy bináris jelsorozat minden tagja 1–1 információt hordoz, maga a jelsorozat pedig annyi bit információt tartalmaz, ahány tagból áll.**

A digitális eszközök általában nagyon egyszerűek és rendkívül hibátűrőek. Például ha az információ továbbítása során az előző jelsorozat torzul, és $0,5\text{ V} - 4,8\text{ V} - 4,9\text{ V} - 0,2\text{ V}$ lesz belőle, akkor ebből az eredeti értékek még igen könnyen (lényegében automatikusan) visszaállíthatóak. Ezért a számítógépek és a legtöbb informatikai eszköz (DVD-lejátszó, mobiltelefon stb.) binárisan kezelik az adatokat.

A bináris jelek két állapotának általában a 0-t és az 1-et feleltetik meg. Ezzel a jelöléssel az előző bináris jelsorozat például így is megadható: 0110. Matematikailag ez a jelsorozat egy kettes számrendszerbeli szám, így a bináris jelek leírására a matematika kész eszközökkel rendelkezik. Érdeemes megjegyezni, hogy a kettes számrendszerbeli számjegyeket is bitnek nevezik (*binary digit*).

Láttuk, hogy ha a jelsorozat egy tagból áll (vagyis 1 bites), akkor az kétféle értéket vehet fel: 0 vagy 1. Ha a jelsorozat két tagból áll, akkor az négyféle lehet: 00, 01, 10, 11. Hasonló módon egy 3 bites jelsorozat már 8-féle lehet, mert mindhárom bitje kétféle értéket vehet fel, és $2 \cdot 2 \cdot 2 = 2^3 = 8$. Általában is **belátható, hogy n biten 2^n lehetőség ábrázolható.**

A valóságban a kérdés általában fordítva jelentkezik: hány bit szükséges pl. egy hónap sorszámának a továbbításához? Mivel 3 biten csak $2^3 = 8$ eset ábrázolható (ami kevés), így legalább 4 bit szükséges. Ez azonban pazarló, hiszen $2^4 = 16$, vagyis nem használunk ki 4 lehetőséget.

Vizsgáljuk meg, hány bit szükséges a betűk tárolásához! Ehhez meg kell határoznunk a tárolandó karakterek számát. Az angol ábécét feltételezve 24 kis- és 24 nagybetűt kell tárolnunk. Ehhez jönnek a számjegyek (10 db), az írásjelek (pont, vessző, szóköz stb.), ami szintén kb. 10 db, a műveleti jelek és zárójelek (kb. 10 db.), s máris 78 lehetőségünk van. Ennek tárolásához 7 bit szükséges, mert $2^6 = 64$, ám 7 biten 128 lehetőségünk van, tehát további jeleket is tárolhatunk. Ha viszont a leggyakoribb nem angol betűket is hozzávesszük, a jelek száma 128 fölé nő, tehát 8 bit szükséges. Ez viszont $2^8 = 256$ lehetőséget ad, ami már tényleg elegendő.

A bit rendkívül kicsi egység, hiszen már egy betű kódolásához is 8 bit szükséges. Kézenfekvő volt ezért többszöröseként a 8 bites egységet választani: **1 bájt = 8 bit**. (Angolul a bájtot byte-nak írják, és a *by eight*, azaz nyolcasával kifejezésből származik.) Ennek megfelelően **egy formázatlan szöveg annyi bájtos, ahány karaktert tartalmaz.**

A bájt többszöröseire kétféle rendszer alakult ki:

Az SI mértékrendszerben használt szokásos többszörösök (kilo-, mega-, giga-, ...) rendre ezerszeres szorzót jelentenek. 1000 bájt = 1 kilobájt, 1000 kilobájt = 1 megabájt, 1000 megabájt = 1 gigabájt, a megfelelő rövidítések pedig rendre B, KB, MB, GB stb. Ezt elsősorban a hardvergyártók alkalmazzák, pl. a merevlemezek kapacitásának megadásánál (1 GB = 1 000 000 000 B).

Mivel a számítógép kettes számrendszerben számol, számára nem a tíz, hanem a kettő hatványai természetesekek. Mivel $2^{10} = 1024$, így ezt a váltószámot használva mind az ember, mind a gép elboldogul. Ebben az esetben már nem használhatók az SI rendszer elnevezései, ezért újakat vezettek be: 1024 bájt = 1 kibibájt, 1024 kibibájt = 1 mebibájt, 1024 mebibájt = 1 gibibájt, a megfelelő rövidítések pedig rendre B, KiB, MiB, GiB stb.

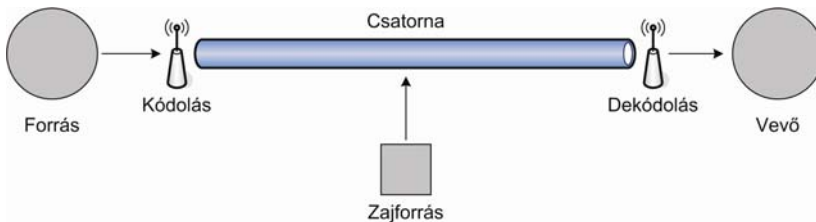
A „kibi”, „mebi”,... elnevezéseket a Nemzetközi Elektrotechnikai Bizottság (IEC) vezette be 1999-ben. (A „bi” szótöredék értelemszerűen a kettes számrendszerre utal.) Ez az elnevezés lassan terjed, sokan ma is előszeretettel használják az SI többszörösöket 1024-es szorzóval.

Az információ továbbítása

Az információátadás az informatikai eszközök esetében (de lényegében az emberek esetében is) leegyszerűsítve a következőképpen épül fel.

A **forrás** üzenetet vagy üzenetek egész sorát állítja elő, melyet továbbítani kíván a **vevőnek**. Ez az üzenet lehet pl. szöveg, kép, hang stb.

Az üzenet továbbítása a **csatorna** segítségével történik. Ehhez az üzenetet a forrás oldalán úgy kell átalakítani, hogy azt a csatorna továbbítani tudja (**kódolás**), majd a vevő oldalán vissza kell alakítani (**dekódolás**). A csatorna lehet kéteres kábel, lyukkártya, rádióhullám, nyomtatott papír stb.



Az információ továbbítása. Mutassuk be a folyamatot egy példán az ábra segítségével!

A csatornában a közlemény óhatatlanul sérül, az információhoz **zaj** adódik hozzá, gondoljunk pl. a vibráló képernyőre vagy a szaggató mobiltelefonra. A csatorna zajosságát nem a zaj mennyiségével, hanem a jelnek a zajhoz viszonyított értékével, a **jel/zaj aránnyal** mérik. A hányados – változatlan zaj mellett – jelentősen javítható a jel erősítésével, például mobiltelefonok esetén a térerő növelésével.

A közleményben meglévő, újabb információt már nem adó elemeket **redundanciának** nevezzük. A redundancia nem feltétlenül káros jelenség, mivel gyakran megkönnyíti a közlemény értelmezését, illetve lehetővé teszi ellenőrzését, a hibák javítását. A postai csekken például a feladott összeget betűkkel is ki kell írni: a szöveg ugyanis a számokkal ellentétben nehezen hamisítható.