

Representasjon av tekst

Vi vet hvordan vi kan representere heltall og reelle tall. Hvordan kan vi bruke dette til å representere tekst?

I datamaskinen er alt lagret som sekvenser av 0 og 1. Når dette skal leses må (datamaskinen) der vi vite hvordan det skal tolkes.

Tekst representeres ved at hvert tegn i tekstmengden får en heltallig kode. Det er denne koden som lagres når teksten skal lagres. I tillegg må vi ha en tabell som sier hvordan hvilken kode hvert tegn har.

ASCII - tabellen

Dette var den første standardiserte tabellen, og ble etterhvert grunnlaget for alle andre tabeller.

ASCII brukte 7 bit til å representere kodene - 128 forskjellige koder, og tegn

0 - 127.

6/62

ISO Latin

128 tegn er for lite. For eksempel er ikke œ , ø , å med. Løsningen var ô legge til et bit-til, totalt 8 bit som gir 256 tegn. De første 128 er ASCII, men i de siste 128 er det rom for nasjonale tegn, ISO Latin. Det fins flere varianter, for ~~en~~ norsk er ISO Latin1 den aktuelle.

Detto løser ikke grensproblem - vi kan ikke representere alle tegn.

Unicode

Unicode er en tabel som inneholder alle verdens tegn, i alt omtrent 100 000 tegn. Unicode har åpne plasser og kan dermed utvides. ~~Stan~~ Unicode inneholder så mange tegn at det ikke er effektivt å lagre alle med like mange bits. UTF-8, UTF-16, UTF-32 er forskjellige måter å kode "kodenumbere" på, altså plasser i Unicode, se seksjon 4.3.3 i kompendiet.