

Overføring og koding av informasjon

To problem:

Alt for mange megabyte

Alt for mange feil i overføringa

To enkle idear

Informasjonsmengd kan målast

Mesteparten av den informasjonen me sender og mottek er overflødig

Telegrammet frå onkel Rolv:

ROLV KJEM MÅNDAG

3 ord

Alternativ versjon per brev:

Kjære Solveig. Som du hugsar, vart me samde om at eg skulle koma og vitja dykk ein gong i sumar. Eg hev no endeleg funne eit høveleg tidspunkt, og kjem oppover no på måndag. Beste helsing Rolv

37 ord

Eigentleg informasjonsmengd:

Informasjon er opplysingar som mottakaren ikkje veit frå før.

Informasjonsmengda i ein tekst = det minste talet på ord me kunne greia oss med utan at meininga forsvinn.

Altso: informasjonsinnhaldet i onkel Rolv sitt brev er på berre 3 ord, ikkje 37.

Overflødig informasjon kallast *redundans*.

Redundansen i brevet vert altso $37 - 3 = 34$ ord eller 92%.

Kode:

Eit sett omskrivingsreglar slik at:

- Avsendaren kan omskriva ei melding.
- Mottakaren kan rekonstruera den opphavslege meldinga frå den omskrivne

Norsk frekvensordliste

Dei 10000 mest brukte
norske orda:

| | |
|-------|------------|
| 1 | i |
| 2 | og |
| 3 | det |
| 4 | er |
| 5 | på |
| 6 | til |
| 7 | som |
| | |
| 9994 | brukbar |
| 9995 | begravelse |
| 9996 | analysert |
| 9997 | viken |
| 9998 | sektorer |
| 9999 | rønningen |
| 10000 | phoenix |

Middelordlengd:

Middelverdien til ordlengda er ca. 4.3 bokstavar per ord.

Dvs. ei norsk tekst på 1000 ord vil utgjera ca. 4300 bokstavar.

Vil det vera råd å omkoda ein slik tekst slik talet på bokstavar vert mindre - men innhaldet intakt?

Kode for datakompresjon

| | | | |
|-------|-----------|---|-----|
| 1 | i | → | A |
| 2 | og | → | B |
| 3 | det | → | C |
| ... | | | |
| 28 | vil | → | Ø |
| 29 | kan | → | Å |
| 30 | ble | → | AA |
| 31 | skal | → | AB |
| ... | | | |
| 870 | bedt | → | ÅØ |
| 871 | delta | → | ÅÅ |
| 872 | gruppen | → | AAA |
| | | | |
| 9999 | rønningen | → | PDE |
| 10000 | phoenix | → | PDG |

Middelordlengd: $\underbrace{\hspace{10em}}$ 4.3 bokstavar

$\underbrace{\hspace{10em}}$ 1.8 bokstavar

Telegrammet frå onkel Rolv igjen

ROLV KJEM MÅNDAG → SST GRF DGX
14 bokstavar 9 bokstavar

Det er altso mogeleg å gjera meldinga
endå kortare utan at informasjon
gjeng tapt!

Kva stend det her?

Tre av fire spurte sier de tror den
rødgrønne regjeringen må gå av etter
valget neste høst

Tre av fire spurte sier de tror den
rødgrønne regjeringen må gå av etter
valget neste høst

Norsk språk er robust mot trykkfeil!

RBLV KJEX MÅNDRG tre feil

ROLV KJEX MÅNDRG to feil

ROLV KJEM MÅNDRG ein feil

Kva med den komprimerte versjonen?

ROLV KJEM MÅNDAG

→ SST GRF DGX

SRT GRP AGX

→ KAMPANE VARETEKT BODØ

SST GRP AGX

→ ROLV VARETEKT BODØ

SST GRF AGX

→ ROLV KJEM BODØ

Kvifor er det mogleg å retta feil i den ukoda teksten
men ikkje i den koda?

To eigenskapar til ukoda tekst:

1. Ikkje alle moglege bokstakombinasjonar gjev
meningsfull tekst.
2. Mottakaren hev evne til å gissa kva for ein
meningsfull (lovleg) tekst som er mest sannsynleg
når han ser ein meningslaus (ulovleg)
bokstavkombinasjon.

Viktig prinsipp:

I ein tekst utan redundans er alle bokstavkombinasjonar lovlege.

Dermed er det uråd for ein mottakar å sjå skilnad på ein bokstavkombinasjon med feil og ein feilfri.

Dermed: Redundans er naudsynt for å gjera ei digital melding robust mot feil!

Omkoding bokstavar til tal

| | | |
|-----|---|----|
| A | → | 1 |
| B | → | 2 |
| C | → | 3 |
| ... | | |
| Ø | → | 28 |
| Å | → | 29 |
| – | → | 30 |

ROLV KJEM MÅNDAG →

18 15 12 22 30 14 10 5 13 30 13 29 14 4 1 7

Koding for feilkorreksjon

fyrste steg:

Ordnar fyrst tala i eit rektangulært mønster

| | | | |
|----|----|----|----|
| 18 | 15 | 12 | 22 |
| 30 | 14 | 10 | 5 |
| 13 | 30 | 13 | 29 |
| 14 | 4 | 1 | 7 |

Koding for feilkorreksjon

neste steg:

Finn summen av tala i kvar rad og kvar søyle:

| | | | | |
|----|----|----|----|----|
| 18 | 15 | 12 | 22 | 67 |
| 30 | 14 | 10 | 5 | 59 |
| 13 | 30 | 13 | 29 | 95 |
| 14 | 4 | 1 | 7 | 26 |
| 75 | 63 | 36 | 63 | |

Før på radsummene og søylesummene som haleheng:

| | | | | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 18 | 15 | 12 | 22 | 30 | 14 | 10 | 5 | 13 | 30 | 13 | 29 | 14 |
| 4 | 1 | 7 | 67 | 59 | 95 | 26 | 75 | 63 | 36 | 63 | | |

Har no 24 tal istaden for 16. Dvs. me har innført redundant informasjon.

Koding for feilkorreksjon

dekoding:

Mottakaren kan no rekna ut dei tilsvarande summane:

| | | | | | | |
|----|----|------|----|----|----|------|
| 18 | 15 | 12 | 22 | 67 | 67 | OK |
| 30 | 14 | 10 | 5 | 59 | 59 | OK |
| 13 | 30 | 11 | 29 | 93 | 95 | FEIL |
| 14 | 4 | 1 | 7 | 26 | 26 | OK |
| 75 | 63 | 34 | 63 | | | |
| 75 | 63 | 36 | 63 | | | |
| OK | OK | FEIL | OK | | | |

Mottakaren oppdagar at to av summane har ein feil på 2.
Dvs. at talet 11 skal vera $11+2=13$ - som er rett!

Konklusjon

Målet: *effektiv og påliteleg* kommunikasjon

- *Effektiv* kommunikasjon oppnår me ved å fjerna redundant informasjon.
- *Påliteleg* kommunikasjon oppnår me ved å leggja til redundant informasjon.

Poenget: Den redundante informasjonen må vera av eit slag som mottakaren kan utnytta til å korrigera feil.

Utfordringa: Å representera informasjonen slik at maskiner (algoritmer) kan utføra feilkorreksjonen