



NTNU

Kunnskap for en bedre verden

TDT4105 Informasjonsteknologi grunnkurs:

Uke 42 – Strenger og strenghåndtering

Anders Christensen
anders@idi.ntnu.no

Rune Sætre
satre@idi.ntnu.no

Læringsmål og pensum

- Læringsmål
 - Skal kunne forstå og programmere med tekststrenger
 - Funksjoner: `blanks`, `double`, `char`, `deblank`, `ischar`, `isletter`, `isspace`, `length`, `lower`, `num2str`, `size`, `str2num`, `strcat`, `strcmp`, `strfind`, `strncmp`, `strrep`, `strtok`, `strtrim`, `upper`, ...
- Pensum
 - Matlab, Kapittel 7

Tekststrenger

- Består av et hvilket som helst antall bokstaver (tegn) omsluttet av enkel apostrof:

```
'dette er en streng'
```

```
'2a@.'
```

```
'c'
```

- Behandles som vektorer der hver enkelt bokstav er et element
- Mange av vektoroperasjonene kan derfor anvendes på strenger

```
- >>length('hus')
```

- Ans=

```
3
```

Behandling av tekst i Matlab

- Alle tegn som kan benyttes i forbindelse med behandling av data er representert i datamaskinens **tegentabell**

- De første 128 tegnene i tegntabellen utgjør ASCII – tabellen.
- Her finner vi de tegnene vi vanligvis trenger
- Lenke til ASCII – tabellen:

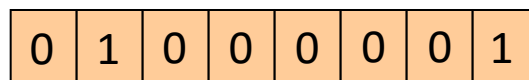
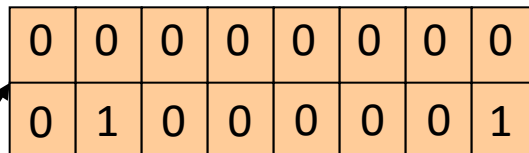
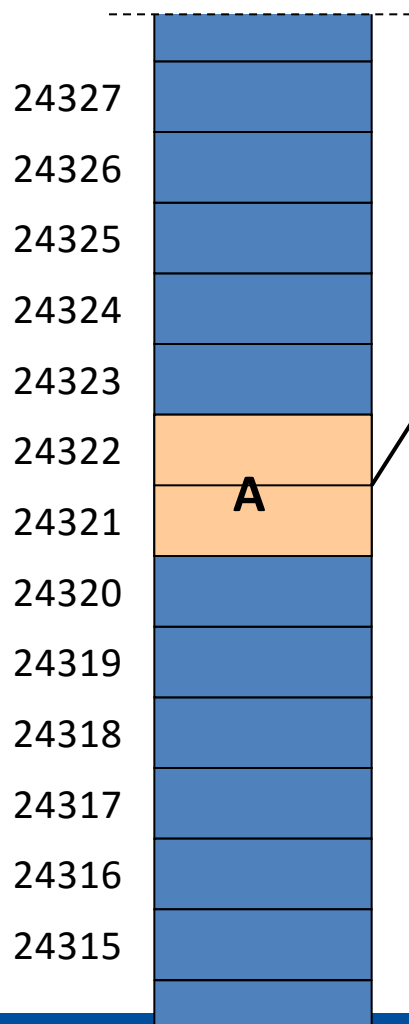
<http://www.asciitable.com/>

- Fra tabellen kan vi lese at tegnet **A** ligger på plass **65** i ASCII-tabellen
- Ved sortering av tekst er det bokstavenes plass i ASCII – tabellen (datamaskinens tegn-tabell) som avgjør rekkefølge
- Problemet med norsk tekst er Æ, Ø, Å, æ, ø og å

| Dec | Hx | Oct | Html | Chr |
|-----|----|-----|-------|-----|
| 64 | 40 | 100 | @ | @ |
| 65 | 41 | 101 | A | A |
| 66 | 42 | 102 | B | B |
| 67 | 43 | 103 | C | C |
| 68 | 44 | 104 | D | D |
| 69 | 45 | 105 | E | E |
| 70 | 46 | 106 | F | F |
| 71 | 47 | 107 | G | G |
| 72 | 48 | 110 | H | H |
| 73 | 49 | 111 | I | I |
| 74 | 4A | 112 | J | J |
| 75 | 4B | 113 | K | K |
| 76 | 4C | 114 | L | L |
| 77 | 4D | 115 | M | M |
| 78 | 4E | 116 | N | N |
| 79 | 4F | 117 | O | O |
| 80 | 50 | 120 | P | P |
| 81 | 51 | 121 | Q | Q |
| 82 | 52 | 122 | R | R |
| 83 | 53 | 123 | S | S |
| 84 | 54 | 124 | T | T |
| 85 | 55 | 125 | U | U |

Behandling av tekst i Matlab

Tegnet **A** lagret i primærlager (RAM)



0 1 0 0 0 0 0 1



$$0 \cdot 2^7 + 1 \cdot 2^6 + 0 \cdot 2^5 + 0 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 = 65$$

Tegnets **desimale** tegnkode

Hver byte er bygd opp av åtte bits. Matlab benytter to byte for å lagre et tegn (Octave 1 byte)

Den siste byten gir ASCII-kode

Tegnets **binære** tegnkode (tall i to-talls systemet)

ASCII - tabellen

$$01000001 = 0 \cdot 2^7 + 1 \cdot 2^6 + 0 \cdot 2^5 + 0 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 = 65$$

| | |
|----|---|
| 69 | E |
| 68 | D |
| 67 | C |
| 66 | B |
| 65 | A |
| 64 | @ |
| 63 | ? |
| 62 | > |
| 61 | = |

← Utsnitt av datamaskinens tegntabell

Tegnet **A** sin binære tegnkode : 01000001

Tegnet **A** sin desimale tegnkode : 65

Tegnkoden angir tegnets **plass** i tegntabellen
(tabellen starter med element 0 på første plass)

Tre Forskjellige Variabeltyper (så langt)

- `bryter=true`
- `tall = 5`
- `(talltekst='5')`
- `tekst = 'Fem'`

The screenshot shows the MATLAB interface. In the Command Window, the following commands have been executed:

```
>> bryter = true
bryter =
     1
>> tall=5
tall =
     5
>> talltekst='5'
talltekst =
     5
>> tekst='Fem'
tekst =
    Fem
fx >>
```

The Workspace window displays the following table of variables:

| Name | Value | Min | Max |
|-----------|-------|-----|-----|
| bryter | 1 | | |
| tall | 5 | 5 | 5 |
| talltekst | '5' | | |
| tekst | 'Fem' | | |

Strengoperasjoner

- Konkatenering (sammensetting)

- horisontalt:

```
>>start = 'heia';
```

```
>>slutt = 'RBK';
```

```
>>[start slutt]
```

```
ans=
```

```
heiaRBK
```

- vertikalt:

```
>>char( start, slutt )
```

```
ans=
```

```
heia
```

```
RBK
```


Flere funksjoner for generering av tekststrenger

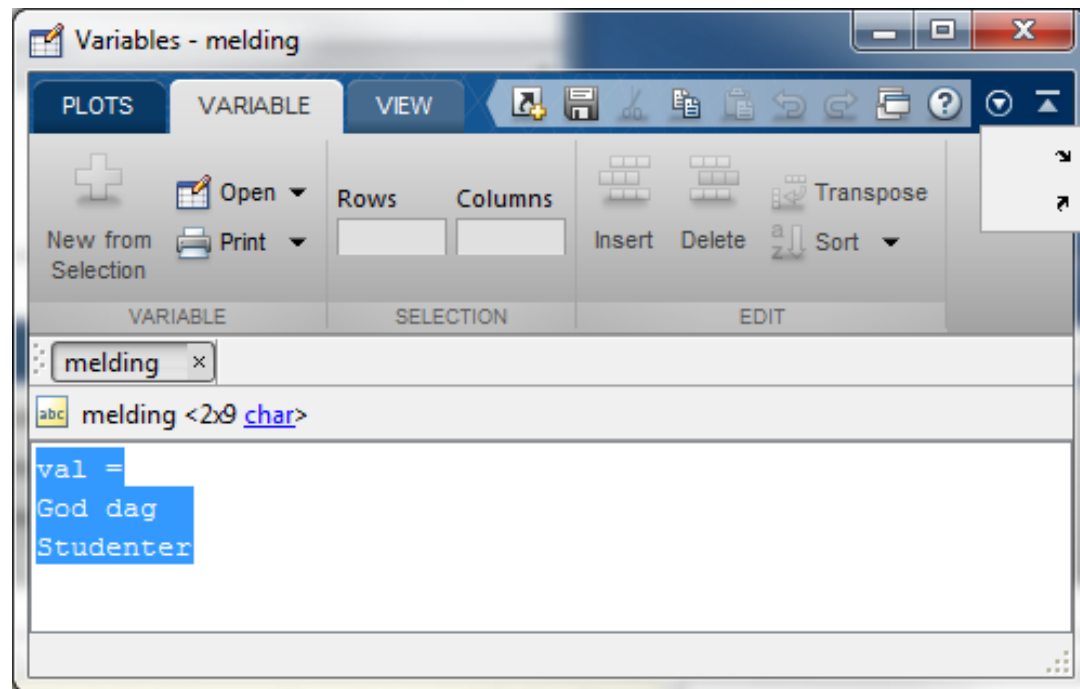
- Med blank streng: blanks

```
>>[start blanks(5) slutt]
Ans=
heia      RBK
```
- lage matrise av strenger: char

```
>> melding = char( ...
    'God dag', 'Studenter )
melding =
God dag
Studenter
>> size(melding)
ans =
2 9
```
- skrive til en streng: sprintf

```
>> tekst = sprintf('Tilfeldig tall er %d', randi(9) );

>> strcat(tekst, '!!!!')
ans = Tilfeldig tall er 8!!!!
```



Noen andre nyttige strengfunksjoner

```
>>melding = '  AbCdeFgHi  '
```

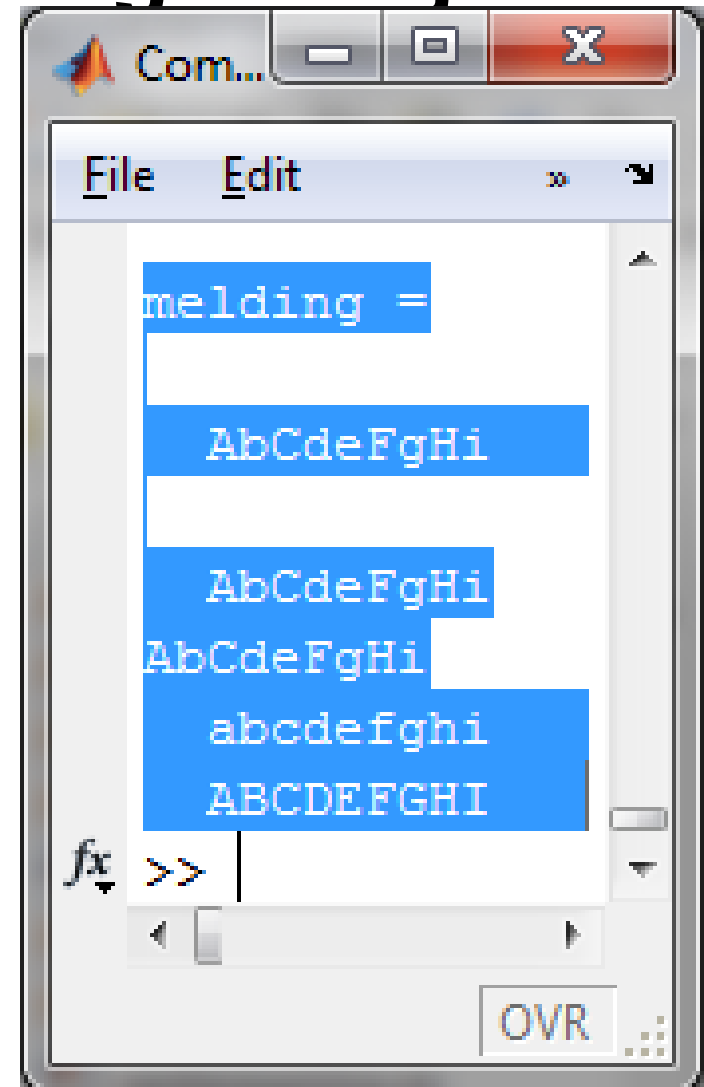
- Fjerne blanke i slutten av strenger: `deblank`

```
>>deblank( melding )
ans =
  AbCdeFgHi
```
- Fjerne blanke i både start og slutt av streng: `strtrim`

```
>>strtrim( melding )
ans =
  AbCdeFgHi
```
- Forandre til små bokstaver `lower`

```
>> lower( melding )
ans=
  abcdefghi
```
- Forandre til store bokstaver `upper`

```
>>upper( melding )
ans =
  ABCDEFGHI
```



Sammenligning av tekst

- Sammenligning av tegn i strenger: strcmp

```
>>ord1= 'hus' ;
```

```
>>ord2= 'hud' ;
```

```
>>ord3= 'hustru' ;
```

```
>>strcmp(ord1,ord1)
```

```
ans=
```

```
1
```

```
>>strcmp(ord1,ord2)
```

```
ans=
```

```
0
```

```
>>strncmp( ord1, ord3, 3 )
```

```
ans=
```

```
1
```

Finn, erstatt og splitt tekst

- Finn den 2. strengen i den 1., og returner indeksen(e) / posisjonen(e)

```
>>strfind( 'abcd abcd', 'cd')  
ans = 3 8
```

- Merk rekkefølge på parameterne:
 - strfind(høystakk, nål)
- Ingen forekomster -> [] returneres

- Erstatte tekst i en streng:

```
>>strrep('abcd abcd', 'a', 'x')  
ans=  
xbcd xbcd
```

- Splitt tekst i to :

```
>>[start rest] = strtok( 'heia RBK' )  
start=  
heia  
rest=  
RBK
```

Konvertering mellom tekst og tall

- Tekst til tall: `str2num`

```
>> nummer = str2num('123.456')
```

```
nummer =
```

```
123.456
```

```
>> nummer2 = str2num('1 2 3');
```

```
nummer2 =
```

```
1 2 3
```

```
>> num2str(123)
```

```
ans =
```

```
123
```

Funksjonene `is...`

- Returner en `boolsk` vektor med samme lengde som strengen og som indikerer om `is`-betingelsen er oppfylt:

```
>>isletter('rbk23')
ans=
  1  1  1  0  0
```

```
>>isspace('RBK 23')
ans=
  0  0  0  1  0  0
```

```
>>vektor = 'RBK'
>>ischar(vektor)
ans=
  1
```

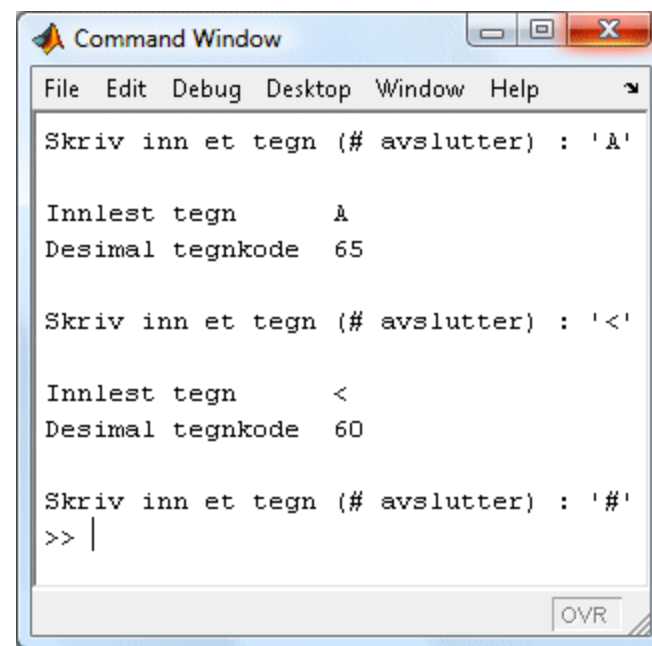
Quiz etter pausen

- Les på strengfunksjoner 😊

Eksempel 1

Lag et program i Matlab som leser inn et tegn og som skriver ut tegnets desimale tegnkode

Funksjonen **double()** mottar tegn og returnerer tegnkode (10-tallsystem)



```
Command Window
File Edit Debug Desktop Window Help
Skriv inn et tegn (# avslutter) : 'A'

Innlest tegn      A
Desimal tegnkode  65

Skriv inn et tegn (# avslutter) : '<'

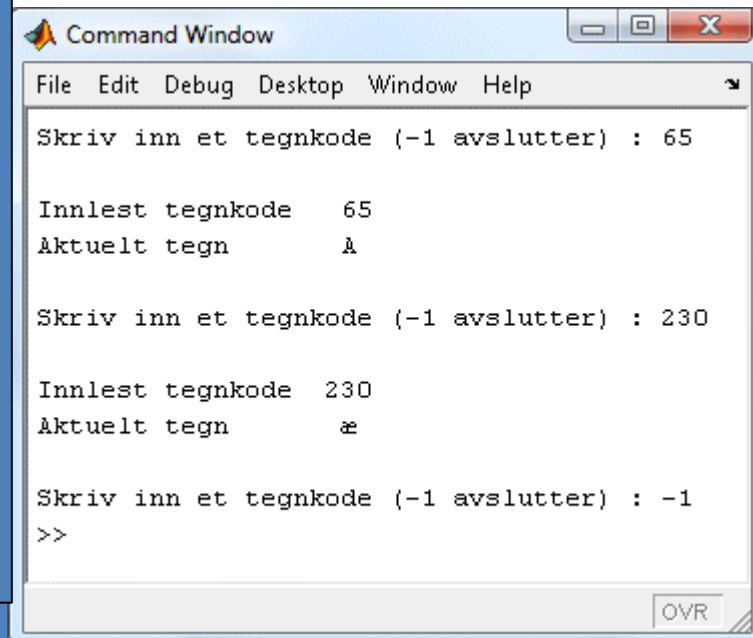
Innlest tegn      <
Desimal tegnkode  60

Skriv inn et tegn (# avslutter) : '#'
>> |
```


Eksempel 2

Lag et program i Matlab som leser inn en desimal tegnkode og som skriver ut det tegn som ligger på tegnkodens plass i tegntabellen

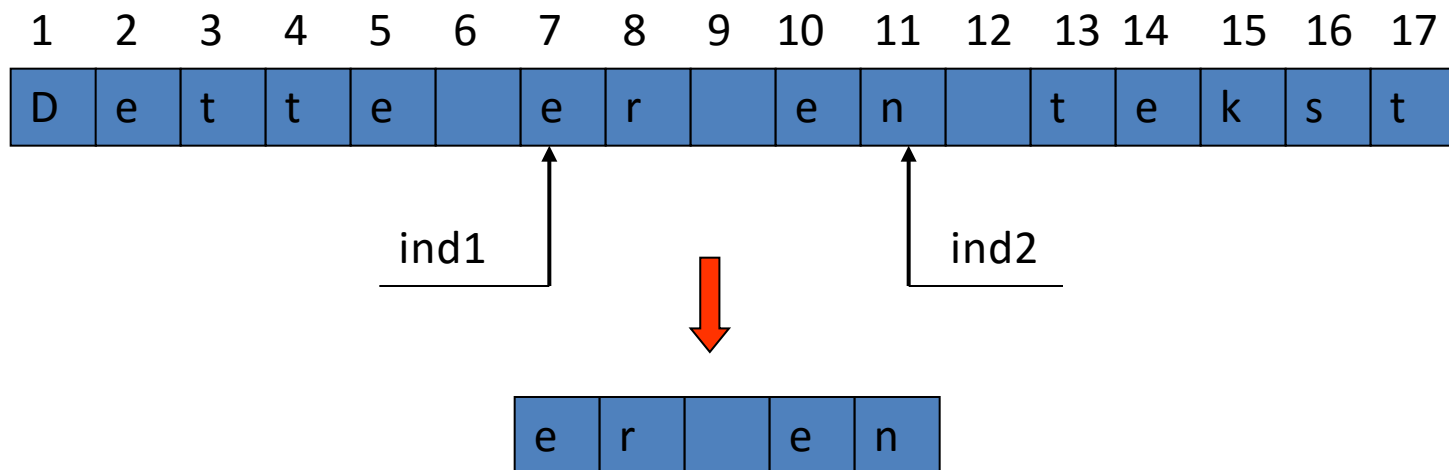
Funksjonen **char()** mottar en tegnkode og returnerer det tilhørende tegn



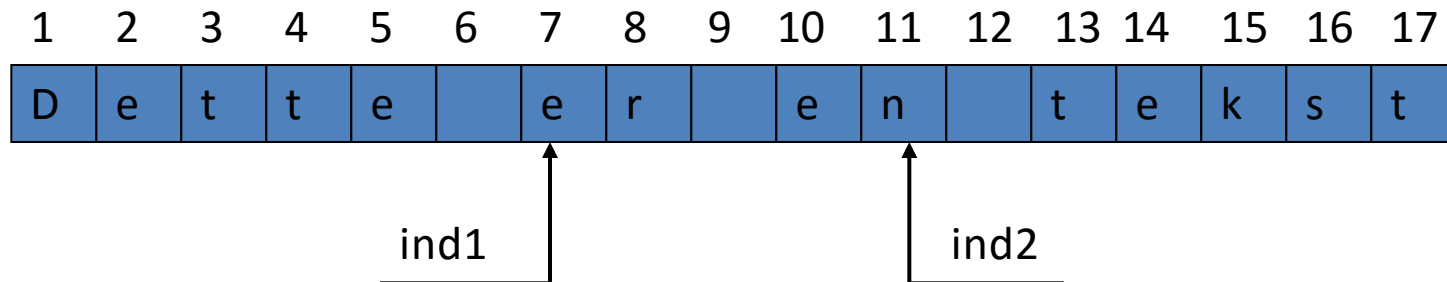
```
Command Window
File Edit Debug Desktop Window Help
Skriv inn et tegnkode (-1 avslutter) : 65
Innlest tegnkode      65
Aktuelt tegn          A
Skriv inn et tegnkode (-1 avslutter) : 230
Innlest tegnkode      230
Aktuelt tegn          æ
Skriv inn et tegnkode (-1 avslutter) : -1
>>
```

Eksempel 3

- Lag et program som leser inn en tekst fra tastaturet
- Les også inn to tall som angir start- og sluttindeks for en deltekst i teksten
- Til slutt skrives delteksten ut på skjerm.



finnDelStreng()



```
delStreng = tekst( ind1:ind2 );
```

```
delStreng = [ tekst(ind1) tekst(ind1+1) ... tekst(ind2) ]
```

Logisk tolking

```
while ( strcmp( svar, 'ja' ) )
```

Funksjonen **strcmp()** mottar to tekststrenger.

Dersom de er identiske returneres **true**,
ellers returneres verdien **false**

```
C:\Users\satre\Dropbox\ITGK\2012\Foiler\Foiler_Matlab\Roger\2012\Matlabfiler...
File Edit Text Go Cell Tools Debug Desktop Window Help
- 1.0 + ÷ 1.1 x % % %
1 %=====  
2 %=== finnkDelStreng( ) ===  
3 %=====  
4 svar = 'ja';  
5 while (strcmp(svar,'ja'))  
6 %-----  
7 %--- Leser inn verdier ---  
8 %-----  
9 tekst = input('Skriv inn en tekst : ');  
10 ind1 = input('Startindeks..... : ');  
11 ind2 = input('Sluttindeks..... : ');  
12 %-----  
13 %--- Finner delstreng ---  
14 %-----  
15 delStreng = tekst( ind1 : ind2 );  
16 %-----  
17 %--- Skriver ut resultat ---  
18 %-----  
19 fprintf('\n');  
20 fprintf('Innlest tekst.....: %s\n',tekst);  
21 fprintf('Innlest indeks1.....: %3d\n',ind1);  
22 fprintf('Innlest indeks2.....: %3d\n',ind2);  
23 fprintf('Deltekst.....: %s\n',delStreng);  
24 fprintf('\n');  
25 %-----  
26 svar = input('Flere tekster ('ja', 'nei') : ');  
27 end  
28 %endprog  
29  
script Ln 22 Col 42 OVR
```

```
Command Window  
File Edit Debug Desktop Window Help  
Skriv inn en tekst : 'Dette er en tekst'  
Startindeks..... : 7  
Sluttindeks..... : 11  
  
Innlest tekst.....: Dette er en tekst  
Innlest tekst.....: 7  
Innlest tekst.....: 11  
Deltekst.....: er en  
  
Flere tekster (ja, nei) : 'ja'  
Skriv inn en tekst : 'Dette er en tekst'  
Startindeks..... : 13  
Sluttindeks..... : 17  
  
Innlest tekst.....: Dette er en tekst  
Innlest tekst.....: 13  
Innlest tekst.....: 17  
Deltekst.....: tekst  
  
Flere tekster (ja, nei) : 'nei'  
>> |  
OVR
```

Kunnskap for en bedre verden

Eksempel 4

- Lag et program som leser inn en (hoved-) tekst fra tastaturet.
- Deretter leses det inn en ny (del-) tekst.
- Programmet skal sjekke om delteksten finnes i hovedteksten.
- Melding skal skrives ut.

| | | | | | | | | | | | | | | | | | |
|---------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| tekst = | D | e | t | t | e | | e | r | | e | n | | t | e | k | s | t |

| | | | | | |
|------------|---|---|--|---|---|
| deltekst = | e | r | | e | n |
|------------|---|---|--|---|---|

length(tekst)

Funksjonen **length()** mottar en tekst og returnerer lengden av denne (antall tegn i teksten)

| | | | | | | | | | | | | | | | | | |
|---------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| tekst = | D | e | t | t | e | | e | r | | e | n | | t | e | k | s | t |

| | | | | | |
|------------|---|---|--|---|---|
| delTekst = | e | r | | e | n |
|------------|---|---|--|---|---|

```

resultat = 'nei';
for (m = 1 : lengdeTekst - lengdeDelTekst + 1)
    delStreng = tekst( m : m+lengdeDelTekst-1 );
    if ( strcmp(delTekst, delStreng) )
        resultat = 'ja';
    end
end

```

Bygger delStreng som er like lang som **delTekst**, fra indeks m

Sjekker om nåværende delStreng er lik den **delTekst** som vi leter etter

```

1  %=====
2  %=== sjekkDelStreng( )
3  %=====
4 - svar = 'ja';
5 - while (strcmp(svar,'ja') == 1)
6     %-----
7     %--- Leser inn verdier
8     %-----
9 -     tekst    = input('Skriv inn en tekst : ');
10 -    delTekst = input('Skriv inn deltekst : ');
11     %-----
12     %--- Finner tekstlengder
13     %-----
14 -    lengdeTekst    = length(tekst);
15 -    lengdeDelTekst = length(delTekst);
16     %-----
17     %--- Finner delstreng
18     %-----
19 -    resultat = 'nei';
20 -    for (m = 1 : lengdeTekst - lengdeDelTekst + 1)
21 -        delStreng = '';
22 -        for (n = m : m+lengdeDelTekst-1)
23 -            delStreng = [delStreng tekst(n)];
24 -        end
25 -        if (strcmp(delTekst, delStreng) == 1)
26 -            resultat = 'ja';
27 -        end
28 -    end

```

```

Command Window
File Edit Debug Desktop Window Help
Skriv inn en tekst : 'Dette er en tekst'
Skriv inn deltekst : 'er en'

Innlest tekst.....: Dette er en tekst
Innlest deltekst.....: er en
Lengde tekst.....: 17
Lengde deltekst.....: 5

Tekst inneholder deltekst...: ja

Flere tekster (ja, nei) : 'ja'
Skriv inn en tekst : 'Dette er en deltekst'
Skriv inn deltekst : 'tekst'

Innlest tekst.....: Dette er en deltekst
Innlest deltekst.....: tekst
Lengde tekst.....: 20
Lengde deltekst.....: 5

Tekst inneholder deltekst...: ja

Flere tekster (ja, nei) : 'nei'
>> |

```

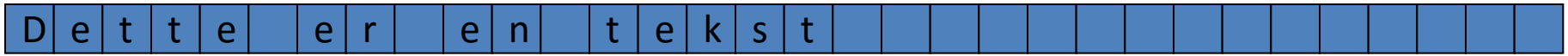
```

29  %-----
30  %---  Skriver ut resultat                      ---
31  %-----
32 -  fprintf(1, '\n');
33 -  fprintf(1, 'Innlest tekst.....: %-32s\n', tekst);
34 -  fprintf(1, 'Innlest deltekst.....: %-32s\n', delTekst);
35 -  fprintf(1, 'Lengde tekst.....: %3.0f\n', lengdeTekst);
36 -  fprintf(1, 'Lengde deltekst.....: %3.0f\n', lengdeDelTekst);
37 -  fprintf(1, '\n');
38 -  fprintf(1, 'Tekst inneholder deltekst...: %3s\n', resultat);
39 -  fprintf(1, '\n');
40  %-----
41 -  svar = input('Flere tekster (ja, nei) : ');
42 -  end
43  %endprog

```

Formatet : **%-32s** reserverer 32 plasser for en tekst. Minustegnet gjør at teksten venstre-justeres innen plassen.

Dette er en tekst



Eksempel 5

Vi skal lage to funksjoner for «hemmelig kommunikasjon» i Matlab

- `krypter()` leser inn en tekst og returnerer en kryptert versjon (ASCII +1)
- `dekrypter()` leser inn en kryptert melding, og returnerer den opprinnelige teksten (ASCII -1)
- **Eksempel på innlest tekst og kryptering**

```
>> krypter('Jeg heter Rune Sætre')
```

```
ans =
```

```
Kfh!ifufs!Svof!Tçusf
```

- **Eksempel på dekryptering**

```
>> dekrypter(ans)
```

```
ans =
```

```
Jeg heter Rune Sætre
```

Binær-koding?

