Innlevering 2

Høst 2025

Om innleveringen

- Innleveringen skal leveres i innleverings- og tilbakemeldingssystemet Devilry innen **3. oktober klokken 23:59**.
- Dere skal *levere i grupper på to eller tre personer*, under forutsetning av at alle i gruppen både står for det som leveres og alene kan gjøre rede for alle delene av arbeidet.
- Gruppene fra første innlevering er videreført i Devilry. Ta kontakt ved behov for endringer.
- Dersom du tidligere har levert individuelt, kan du opprette gruppe for denne innleveringen.
- For hver oppgave skal innleveringen inneholde:
    - Et Java- eller Python-program for hver oppgave som ber om det.
        - Du kan dele et program i flere filer, men du må sørge for at programmet kan kjøres med én kommando.
    - Én `.pdf`-, `.md`- eller `.txt`-fil som skal inneholde kommandoen som brukes for å kjøre programmet, samt løsninger på oppgaver som skal besvares med pseudokode eller naturlig språk.
- Du skal ikke levere testfiler eller andre ressurser.
- Med mindre noe annet er spesifisert kan dere bruke hva dere vil fra Java eller Python sine standardbibliotek.
- Kommenter **«Klar for retting»** i Devilry hvis du leverer i god tid før fristen, og innleveringen er klar for retting.

Krav til innleveringen

- Alle oppgaver skal være besvart.
- Programmene må kompilere og kunne kjøres med kommandoen dere har oppgitt.
- Det som leveres skal være *lett forståelig, entydig og presist*.

Sortering

Ressurser for oppgaven:

https://github.uio.no/IN2010/sortering-ressursside

I denne oppgaven skal du implementere flere sorteringsalgoritmer!

Disse to sorteringsalgoritmene skal implementeres:

- Insertion sort
- Quicksort

I tillegg skal du velge *to andre* sorteringsalgoritmer som er nevnt på forelesning, hvor minst én er i $\mathcal{O}(n \cdot \log(n))$.

For den ekstra ivrige: Du kan også implementere sorteringsalgoritmer som er utenfor pensum. Dersom du velger sorteringsalgoritmer som er utenfor pensum må du selv sørge for at implementasjonen er forståelig for retteren, og kan ikke forvente hjelp fra gruppelærer i like stor grad.

For alle deloppgavene er input det samme. Det som skiller deloppgavene er output.

Input

Det består av n heltall. For hvert tall i er det slik at $0 \leq i \leq 2^{31} - 1$. Input skal leses fra fil, hvor filnavnet blir gitt som kommandolinjeargument.

Eksempelinput:

```
80
91
7
33
50
70
13
321
12
```

Korrekthet

For små input-filer skal programmet ditt skrive en fil for hver sorteringsalgoritme du implementerer. For større inputfiler kan det være lurt å kun kjøre de raskeste algoritmene. Hver av outputfilene skal inneholde resultatet av sorteringen på samme format som input. Fordi alle algoritmene skal oppnå samme resultat, er det forventet at alle outputfilene er like.

Navngi hver outputfil etter navnet på inputfilen, etterfulgt av sorteringsalgoritmen som ble brukt til å sortere input. Hvis inputfilen heter `example` skal outputfilen hete `example_alg.out`, der `alg` byttes ut med navnet til den aktuelle sorteringsalgoritmen. Algoritmenavnet skal bare inneholde små bokstaver (`a–z`). For eksempel kan outputfilen for insertion sort få navnet `example_insertion.out`.

Du skal redgjøre for hvilke sorteringsalgoritmer som er implementert, og hva som er testet for å sjekke at alle algoritmene gir rimelige svar.

Gitt eksempelinputet ovenfor bør vi få følgende output.

Eksempeloutput:

```
7
12
13
33
50
70
80
91
321
```

Sammenligninger, bytter og tid

Hver av sorteringsalgoritmene skal kunne telle *antall sammenligninger* og *antall bytter* som gjøres i løpet av en sortering. Det kan være lurt å skrive egne metoder for sammenligning og bytter, slik at det blir enkelt å telle. I tillegg skal den totale tiden brukt på sortering oppgis i *mikrosekunder*. I Java kan tiden på din sorteringsalgoritme (i mikrosekunder) måles slik:

```
long t = System.nanoTime();

// din sorteringsalgoritme kjøres her

long time = (System.nanoTime()-t)/1000;
```

Tilsvarende for Python:

```
import time

t = time.time_ns()

// din sorteringsalgoritme kjøres her

timeus = (time.time_ns() - t) / 1000
```

Du skal kun produsere én outputfil for en gitt inputfil, som en CSV-fil. Dersom inputfilen heter `example`, skal outputfilen hete `example_results.csv`.

Hvis inputfilen inneholder n heltall, skal du gjøre $n+1$ sorteringer for hver sorteringsalgoritme du tester. Første sortering gjøres på det tomme arrayet, andre sortering gjøres på et array som inneholder det første elementet, tredje sortering gjøres på et array som inneholder de to første elementene, og så videre, helt opp til n.

Den første linjen i output skal se slik ut:

```
n, alg1_cmp, alg1_swaps, alg1_time, alg2_cmp, alg2_swaps, alg2_time
```

Det samme mønsteret skal følges for alle algoritmene du tester. De etterfølgende linjene skal bestå av kommaseparerte heltall.

- Første tall skal angi hvor mange elementer som ble sortert
- Andre tall skal angi hvor mange sammenligninger `alg1` brukte
- Tredje tall skal angi hvor mange bytter `alg1` gjorde
- Fjerde tall skal angi hvor lang tid `alg1` brukte (i mikrosekunder)
- Femte tall skal angi hvor mange sammenligninger `alg2` brukte
- ...

Eksempeloutput:

```
n, insert_cmp, insert_swaps, insert_time, merge_cmp, merge_swaps, merge_time
0,          0,            0,           1,         0,           0,          0
1,          0,            0,           0,         0,           0,          1
2,          1,            0,           2,         5,           2,          5
3,          3,            2,           4,        10,           2,          6
4,          6,            4,           6,        21,           5,         11
5,          9,            6,           7,        28,           8,         15
6,         12,            8,          10,        29,           9,         16
7,         18,           13,          15,        40,          11,         21
8,         19,           13,          15,        44,          12,         23
9,         27,           20,          22,        45,          13,         23
```

Merk at mellomrom ignoreres.

Det er ikke forventet at deres algoritmer viser nøyaktig samme tall. Tallene er både avhengig av implementasjonen av sorteringsalgoritmen, og hvordan bytter og sammenligninger måles.

Det skal skrives kort om hvordan bytter og sammenligninger er målt.

Eksperimentér

Her skal du kun bruke programmet fra forrige deloppgave for å gjøre noen refleksjoner. Vi vil laste opp flere inputfiler på semestersiden som du kan bruke til å kjøre eksperimenter. Du bør kunne kjøre eksperimenter på ulike kombinasjoner av sorteringsalgoritmer (du trenger ikke lage et brukergrensesnitt for dette, men kan kommentere ut/inn kode utifra hva du trenger for å kjøre eksperimentene dine). Det er for eksempel lurt å kjøre eksperimenter hvor kun $\mathcal{O}(n \cdot \log(n))$ algoritmene kjøres, fordi da vil du kunne teste større inputfiler.

Filene som genereres er CSV-filer, og de kan åpnes i vanlige regnearkverktøy, som Excel, Numbers, etc. Her kan du lage enkle linjediagrammer, som kan gi deg innsikt i hvordan antall sammenligninger, bytter og tid henger sammen, samt hvor effektive de forskjellige sorteringsalgoritmene er.

Vi ønsker at du gjør noen egne refleksjoner. Du må ta stilling til disse spørsmålene:

- I hvilken grad stemmer kjøretiden overens med kjøretidsanalysene (stor $\mathcal{O}$) for de ulike algoritmene?
- Hvordan er antall sammenligninger og antall bytter korrelert med kjøretiden?
- Hvilke sorteringsalgoritmer utmerker seg positivt når n er veldig liten? Og når n er veldig stor?
- Hvilke sorteringsalgoritmer utmerker seg positivt for de ulike inputfilene?
- Har du noen overraskende funn å rapportere?

Alle svar må begrunnes. Legg gjerne ved skjermbilder eller lignende som gir grunnlag for påstandendene dine. Vær tydelig på hvordan eksperimentet ble utført, slik at det lar seg reprodusere av andre.