



## Algoritmer og datastrukturer

### A.5 – URL – ressurser på web

## A.5 URL – ressurser på web

### □ A.5.1 URL og InputStream

Klassen `URL` ligger under `java.net` og representerer en «Uniform Resource Locator», dvs. en referanse (adresse) til en «ressurs» på «the World Wide Web». En ressurs kan være en fil, en mappe (directory) eller et mer komplisert objekt. Siden dette står sammen med filbehandling på *bit*-, *byte*- og *char*-nivå, er det muligheten av å lese filer som vi skal se på.

Vi tar dette vedlegget, dvs. **Vedlegg A5**, som eksempel. Det har flg. url:

```
http://www.cs.hioa.no/~ulfu/appolonius/vedlegg/A/A5.html
```

`URL` har en konstruktør som tar i mot en url (i form av en tegnstring) som parameter:

```
URL url = new URL("http://www.cs.hioa.no/~ulfu/appolonius/vedlegg/A/A5.html");
```

En url kan deles opp i tre deler: protokoll, vert (host) og fil:

```
System.out.println(url.getProtocol()); // http
System.out.println(url.getHost());    // www.cs.hioa.no
System.out.println(url.getFile());    // ~/ulfu/appolonius/vedlegg/A/A5.html
```

`URL` har også en konstruktør med disse tre som parametere:

```
String protokoll = "http";
String vert = "www.cs.hioa.no";
String fil = "~/ulfu/appolonius/vedlegg/A/A5.html";

URL url = new URL(protokoll, vert, fil);

System.out.print(url); // http://www.cs.hioa.no/~ulfu/appolonius/vedlegg/A/A5.html
```

Konstruktørene for `URL` (f.eks. de to over) kaster en `MalformedURLException` hvis det er en feil i url-en. Dette er en subclasse av `IOException` og må derfor enten fanges opp eller bli videresendt. En mulig feil er at protokollen er feilstavet eller er ukjent. Det er flere aktuelle protokoller enn `http`, f.eks. `https`, `ftp` og `file`.

Av de øvrige metodene i `URL` er det først og fremst: `public InputStream openStream()` som er av interesse her. Den gjør at vi kan lese en fil på «the World Wide Web» på vanlig måte:

```
String url = "http://www.cs.hioa.no/~ulfu/appolonius/vedlegg/A/A5.html";
InputStream inn = new URL(url).openStream(); // oppretter en InputStream

int antallLinjer = 0, tegn; // 10 = LF (Line Feed)
while ((tegn = inn.read()) != -1) if (tegn == 10) antallLinjer++;
inn.close();

System.out.println(antallLinjer); // antall linjer på filen
```

*Programkode A.5.1 a)*

I den andre setningen i koden over gir metoden `openStream()` oss en `InputStream`. Men en `InputStream` er en abstrakt klasse. Dermed må den instansen vi får, være en instans av en av subklassene til `InputStream` og i dette tilfellet er det en `BufferedInputStream`. Det betyr at flg. kode da vil være «smør på flesk», dvs. unødvendig:

```
InputStream inn = new BufferedInputStream(new URL(ur1).openStream());
```

Det er også mulig å lese en vanlig fil, f.eks. på en harddisk eller en minnepinne, på denne måten. Da må vi bruke `file` som protokoll. Undetegnete har en kopi av dette vedlegget liggendes lokalt, dvs. på:

```
e:/www/appolonius/vedlegg/A/A5.html
```

Men dette utgjør kun fildelen i en url. I tillegg må vi ha `file` som protokoll og egentlig ingenting som vert (host). Da blir koden slik:

```
String url = "file:///e:/www/appolonius/vedlegg/A/A5.html";
InputStream inn = (new URL(url)).openStream();
// et eller annet
inn.close();
```

**Programkode A.5.1 b)**

Denne url-formen er også den vi finner i adressefeltet i en web-leser når den henter en lokal html-fil. Hvis vi skal lage en URL-instans ved å bruke protokoll, vert og fil, kan det gjøres slik:

```
URL url = new URL("file", "/", "e:www/appolonius/vedlegg/A/A5.html");
System.out.println(url); // Utskrift: file:///e:www/appolonius/vedlegg/A/A5.html
```

**Programkode A.5.1 c)**

## Oppgaver til A.5.1

- Sjekk at **Programkode A.5.1 a)** virker. Da må du importere `IOException`, `InputStream` og `URL` og ha `main(String... args) throws IOException`. Obs: Programmet finner antall linjer på selve html-filen. Men det er ikke det samme som antall linjer i web-leseren.
  - Sjekk hvilken feilmelding du får hvis du gjør en skrivefeil i protokoll delen, f.eks. `htp` istedenfor `http`.
- Lag et program av samme type som det i **Programkode A.5.1 b)**, men der du henter en lokal tekstfil (html, txt, java o.l.). Lag f.eks. kode som kopier filen.

