

## 7. Übung zu Programmierung 1

**Abgabe: KW 51**

### 1. Aufgabe

Schreiben Sie ein Java-Programm, das aus HTML-Dateien die Hypertextlinks herausfiltert und ausgibt. Da wir noch keine Dateiverarbeitung besprochen haben, ist es am einfachsten, diese Dateien über `java.util.Scanner` einzulesen. Übergeben Sie dazu den Namen der zu lesenden Datei über die Kommandozeile oder in BlueJ als Parameter der `main`-Methode.

Start des Programms in der Shell:

```
java LinkFilter test1.html
```

Die Datei kann dann wie folgt geöffnet werden:

```
Scanner input = new Scanner(new File(args[0]));
```

Danach kann die Datei ganz normal gelesen werden wie bisher die Standardeingabe. Die Hypertextlinks werden durch das `a`-Element definiert (siehe dazu auch <https://wiki.selfhtml.org/wiki/HTML/Textauszeichnung/a>)

Wenn die HTML-Datei z. B. wie folgt aussieht:

test1.html:

```
.  
<li><a href="http://www.htwsaar.de/">Hochschule für Technik und Wirtschaft</a>  
.   
<li><a href="http://www.google.com">Google</a>  
<a href="http://www.heise.de">Heise Online</a>  
.
```

Dann sollte die Ausgabe in etwa so aussehen:

Hochschule für Technik und Wirtschaft:	<a href="http://www.htwsaar.de">http://www.htwsaar.de</a>
Google:	<a href="http://www.google.com">http://www.google.com</a>
Heise Online:	<a href="http://www.heise.de">http://www.heise.de</a>
....	

### Anmerkungen:

- Ihr Programm sollte kein Dialogprogramm sein.
- Wahlweise können Sie auch gerne mehrere Dateinamen als Parameter übergeben und dann mehrere Dateien nacheinander auswerten.
- Bauen oder suchen Sie sich geeignete Beispiel-HTML-Dateien.
- Beim Filtern der Datei(en) können Sie sich gerne auch mit regulären Ausdrücken "austoben".
- Der Einfachheit halber kann von Folgendem ausgegangen werden:
  - die Hypertextlinks sind immer nur eine Zeile lang
  - die Texte in den Tags sind immer klein geschrieben
- Wenn Sie mehr über HTML erfahren wollen, dann ist die folgende Seite sehr zu empfehlen: <https://selfhtml.org>