

Praktikum aus Softwareentwicklung 2

WS 2006/2007, LVA 365.009

Abgabetermin: Fr. 24. 11. 2006, 17:00

Erreichbare Punkte: 24

Übung 3 – Network Programming

Aufgabe 1: Http-Picture-Crawler

16 Punkte

Entwickeln Sie ein Programm, welches ausgehend von einer Html-Startseite, alle Bilder bis zu einer bestimmten Tiefe herunterlädt. Die Bilder sollen lokal in einem Order gespeichert werden und identische Bilder soll nur einmal vorhanden sein. Gleiche Bilder besitzen den gleichen Pfad, Namen und die gleiche Filegröße (natürlich können sie auch die Bytes vergleichen oder eine MD5 Summe über die Bytes bilden, aber es wird nicht verlangt). Haben 2 Bilder zwar den selben Pfad und Namen, aber ungleiche Dateigrößen, benennen Sie eines der Bilder entsprechend um.

Für das herausparsing der Bilder und verlinkten Html-Seiten treffen Sie folgende Einschränkungen:

- Als Links zu Bilder sind nur Image-Tags relevant.
- Für weiterführende Links sind nur die A-Href Tags relevant und Sie können sich auf Seiten mit den Endungen html und htm beschränken.
- Javascript oder andere komplizierte Konstrukte brauchen nicht berücksichtigt werden.
- Groß- Kleinschreibung ist bei den Pfaden nicht relevant, jedoch sollte sie bei den Tags berücksichtigt werden.
- Folgende Bilder sollten gefunden werden:

- `<IMG SRC="http://www.ifs.uni-linz.ac.at/ifs/images/butusefl.gif">`
 - `<IMG src="/ifs/images/butusefl.gif">`
 - `<img Src='../butusefl.gif'>`

Die ersten beiden Bilder hätten denselben Pfad.

Die Tiefe des Herunterladens ergibt sich wie folgt: Tiefe 0 → nur die Bilder der Startseite werden heruntergeladen, 1 → die Bilder der Startseite und alle Bilder der auf der Startseite verlinkten Seiten, usw.. Damit man ein wenig in die Tiefe gehen kann, gibt es noch eine wichtige Einschränkung: Es werden nur Html-Seiten geladen, welche am selben Server wie die Startseite liegen. Für das Zerlegen eines Urls in seine Teile (Server, Port, Pfad..) verwenden Sie die Klasse: `java.net.URL`

Beim Starten des Programms, sollen 3 Parameter angegeben werden: Die Startpage, die Tiefe des Herunterladens und das Ausgangsverzeichnis der Bilder.

```
java -cp ./bin; ueb3.PictureCrawler http://www.ifs.uni-linz.ac.at 3 C:/temp
```

In dem Ausgangsverzeichnis können Sie die Bilder auch in Unterordner speichern oder andere Mechanismen für die Eindeutigkeit der Bilder verwenden. Achten Sie auf entsprechende Fehlerbehandlung und eine schöne Programmstruktur und Klassenaufteilung. Zur Kontrolle: Der Mozilla Firefox zeigt Ihnen mit dem „Webdeveloper“ Tool alle Bilder und Links einer Html-Seite an.

Aufgabe 1: Smtplib Client

8 Punkte

Implementieren Sie einen einfachen Email Client für die Kommandozeile, der zumindest folgende Funktionalität bietet:

- Neue Emails erstellen (mit Empfänger, CC, Subject und Text; *keine* Attachments)
- Einstellungen (IP-Adresse + Port für SMTP-Server, Name und Email-Adresse des Benutzers usw.)

Überlegen Sie sich auch ein entsprechendes User Interface.

Hinweis: Die meisten SMTP-Server verwenden den Port 25. Die RFCs finden Sie unter:

- SMTP: <http://www.ietf.org/rfc/rfc0821.txt>

Beispiel für eine SMTP-Session:

```
C:\ telnet email.uni-linz.ac.at 25

220 alijku01.edvz.uni-linz.ac.at ESMTP
HELO alijku01.edvz.uni-linz.ac.at
250 alijku01.edvz.uni-linz.ac.at Hello aaa.bbb.ccc [...], pleased to
meet you
MAIL FROM:eugen@ifs.uni-linz.ac.at
250 2.1.0 eugen@ifs.uni-linz.ac.at... Sender ok
RCPT TO:eugen@ifs.uni-linz.ac.at
250 2.1.5 eugen@ifs.uni-linz.ac.at... Recipient ok
DATA
354 Enter mail, end with "." on a line by itself
Date: 24 Dec 02 10:00:00
From: Eugen <eugen@ifs.uni-linz.ac.at>
Subject: Frohe Weihnachten
To: Eugen K. <eugen@ifs.uni-linz.ac.at>

Frohe Weihnachten
.
250 2.0.0 g9N6FltZ039996 Message accepted for delivery
QUIT
221 2.0.0 alijku01.edvz.uni-linz.ac.at closing connection

C:\
```

Tip: Verwenden Sie telnet, um die Kommunikation bzw. die Kommandos mit dem SMTP-Server zu testen.