

**1. Fragen zu utf-8:**

1.1. Welche zwei Methoden gibt es, um STDIN und STDOUT auf utf-8 zu setzen.

Man kann STDIN und STDOUT als erste Möglichkeit beim Programmstart mit „perl -C“ auf utf-8 setzen. Außerdem kann man im Programm mit den Anweisungen „binmode(STDIN, „:utf8“);“ und „binmode(STDOUT, „:utf8“);“ zur Laufzeit auf utf-8 kodieren.

1.2. Was bedeutet die Anweisung `use utf8;` in einem PERL Programm?

Die Codierung des Programms wird auf utf8 gestellt. Es erlaubt zum Beispiel auch Umlaute in Variablennamen.

**2. Codierung/ UNICODE:**

2. Was geben die folgenden PERL Fragemente aus, wenn die Zeilen der folgenden Liste mit den regulären Ausdrücken getestet werden.

2.1.

Das Programm gibt das folgende aus:  
Feb. und Heute ist der 6.  
Klausur2 Klausur1

2.2.

Das Programm gibt das folgende aus:  
ist e  
schreiben r  
hoffentlich d

2.3.

Das Programm gibt das folgende aus:  
r schreiben

3. In wievielen nachfolgenden Bytes wird die Zeichenkette *Straße*, bestehend aus den 6 Buchstaben *S*, *t*, *r*, *a*, *ß* und *e* in einer Datei gespeichert, wenn die Datei im Format a) Iso-Latin und im Format b) utf-8 gespeichert ist. Geben Sie die Anzahl der Bytes an.

a) 6 Byte

b) 7 Byte

4. Geben Sie die fehlenden Zahlendarstellungen in den entsprechenden Codierungen an.

Dezimal	Binär	Oktal	Hexadezimal
4	100	4	4
161	10100001	241	A1

### 3. Reguläre Ausdrücke:

5. Was wird bei folgenden substitute Befehlen ausgegeben? Warum?

5.1.

Die Ausgabe lautet: Das ist XXX die Aufgabe richtig löst.  
Das Programm ersetzt alles vom ersten großen Der bis zum zweiten großen Der mit XXX.

5.2.

Die Ausgabe lautet: Das ist xxx die Aufgabe richtig löst.  
Das Programm ist jetzt non-greedy eingestellt und würde die erste Möglichkeit auswählen um von Der bis Der zu matchen. Da es nur eine Möglichkeit gibt ist das Ergebnis dasselbe.

5.3.

Die Ausgabe lautet: Das ist xxx letzte, Der die Aufgabe richtig löst.  
Das Programm ignoriert jetzt Groß- und Kleinschreibung und ersetzt global. Da es immer noch non-greedy sucht matched es die erste Möglichkeit von Der bis der. Da es keine zweite Möglichkeit gibt hört es an der Stelle auf.

6. Sie arbeiten mit Wörtern aus einem deutschen Text, die in der Liste @woerter gespeichert sind. Wie lauten die regulären Ausdrücke (ausschließlich UNICODE Properties), die folgende Wörter in der Liste @woerter finden.

6. 1.

`$wort =~ /lich$/`

6. 2.

`$wort =~ /[0123]?p{N}\.[01]?p{N}\.p{N}{4}/`

6. 3.

`$wort =~ /^Be\p{L}{0,5}ung$/`

6. 4.

`$wort =~ /\p{Lu}+ \p{N}{5} \sp{L}+$/`

6. 5.

\$wort =~ /[0123]? \p{N} V [01]? \p{N} \p{N}{4}/

6. 6.

\$wort =~ /a/i

6. 7.

\$wort =~ /^[01]+\$ /

6. 8.

\$wort =~ /\p{Lu}+\$/

#### 4. Kleine Programme und Subroutinen:

7. Schreiben Sie eine Subroutine *sub minimum (\$\$)* die zwei ganzzahlige Werte als Argument bekommt und den Wert des kleineren Elements zurückgibt.

```
sub minimum($$) {
    my $zahl1 = $_[0];
    my $zahl2 = $_[1];

    if ($zahl1 < $zahl2) {
        return $zahl1;
    } else {
        return $zahl2;
    }
}
```

8. Schreiben Sie eine Subroutine, die mit einer *while* Schleife alle Zahlen zwischen 0 und 100 ausgibt. Bei Zahlen, die durch 10 teilbar sind, gibt die Subroutine zusätzlich den Text „durch 10 teilbar“ auf der Konsole aus.

```
sub zahlen() {
    my $zahl = 0;

    while ($zahl <= 100) {
        if (($zahl % 10) == 0) {
            print „$zahl durch 10 teilbar \n“;
        } else {
            print „$zahl \n“;
        }

        $zahl++;
    }
}
```

**9. Rekursion****9.1. Was berechnet das folgendes Programm?**

Das Programm multipliziert zwei Zahlen miteinander, indem es die erste Zahl solange zu sich selbst addiert bis die zweite Zahl 0 ist.

9.2.

Das Ergebnis lautet 5.

9.3.

Das Programm gibt kein Ergebnis da es nicht terminiert. Es verringert die negative Zahl -5 immer weiter und kommt so nie in den Abbruchfall  $n == 0$ , da die Zahl sich immer weiter von 0 entfernt.

**5. Erstellen einer Frequenzliste:**

10. Von einer utf-8 Datei sz.txt, in der deutscher Text gespeichert ist, sollen die Anzahl der Wörter und die 5 seltensten Wörter ausgegeben werden, die maximal 5 Buchstaben lang sind.

Die Datei ist zu öffnen, soll zeilenweise gelesen, die Anzahl aller Wörter gezählt werden und die gesuchten Wörter in einen HASH eingetragen werden. Der HASH soll nach der Häufigkeit der Wörter sortiert werden und die sortierten Keys in einer Liste gespeichert werden. Aus dieser Liste sollen die 5 seltensten Wörter auf dem Terminal ausgegeben werden und in der Datei unimportant.txt gespeichert werden. Es soll die Anzahl aller Wörter ausgegeben werden.

```
#!/usr/bin/perl
use strict;
use utf8;
use locale;
{
    my($line, $selem, %freq, @woerter, @sortierte_woerter, $anzahl);

    open(INP, "sz.txt") or die "file not found!";
    open(AUS, ">unimportant.txt");

    while ($line = <INP>) {
        chomp $line;
        @woerter = split(/\p{L}/, $line);

        foreach $selem (@woerter) {
            if ($selem =~ /\p{L}{0,5}$/) {
                $freq{$selem}++;
            }
        }
    }
}
```

```

@sortierte_woerter = (sort{$freq{$a} <=> $freq{$b}} keys %freq);
$anzahl = scalar (@sortierte_woerter);
print "Im Text kommen $anzahl Wörter vor.\n";

my $i = 0;
foreach $elem(@sortierte_woerter) {
    if ($i < 5) {
        print „$elem = $freq{$elem} \n“;
        print AUS „$elem = $freq{$elem} \n“;
        $i++;
    }
}

close INP,
close AUS;
}

```

## 6. UNIX:

11. Ausgehen von folgenden Informationen auf Ihrer Konsole sollen sie mit UNIX-Befehlen folgende Aufgabe lösen.

11.1. Wechseln Sie bitte in Ihr Homeverzeichnis.

```
cd
```

11.2. Legen Sie dort ein Verzeichnis mit dem Namen backup an!

```
mkdir backup
```

11.3. Kopieren Sie foto.jpg in das Verzeichnis backup

```
cp ~/projekt/foto.jpg backup/
```

12. Sie haben eine Datei aus dem Internet geladen. Wie lauten die UNIX-Befehle?

12.1. Packen sie diese Datei vollständig aus.

```
tar -xz textarchiv.tar.gz
```

12.2. Lassen Sie sich anzeigen wieviel Speicherplatz in Ihrem Arbeitsverzeichnis noch frei ist.

```
df-h .
```

13. Erklären Sie kurz den grundlegenden Unterschied zwischen den Befehlen `scp` und `ssh`.

SSH stellt eine Verbindung zum Server her. Mit SCP kann man Dateien kopieren.

Erklären Sie alle Bestandteile des Befehlsaufrufs: `scp urlaubsbilder.zip studiosus@remote.cip.ifi.lmu.de: privat/`

Die Datei `urlaubsbilder.zip` wird vom eigenen PC auf den Cip-Pool in den Ordner `privat` kopiert.

Erklären Sie den Unterschied des folgenden Befehlsaufrufs zum vorherigen: `scp urlaubsbilder.zip studiosus@remote.cip.ifi.lmu.de: /privat/`

Hier wird das Verzeichnis erst angelegt, davor ist es schon vorhanden.

14. Erstellen Sie aus einer Datei `wortliste.txt` eine Datei `frequenzliste.txt`. Nutzen Sie dazu die Befehle `cat`, `sort`, `tr` und `uniq` unter Anwendung von Pipes in einer Befehlszeile.

```
cat wortliste.txt | tr " " "\n" | sort -f | uniq -ic | sort -n > frequenzliste.txt
```