

Einführung in die Computerlinguistik – Einführung in Perl (2) – reguläre Ausdrücke

Dozentin: Wiebke Petersen

31.5.2010

Perl: In- und Output in Dateien

Einlesen aus einer Datei

open(INPUT, "<file.ext"); öffnet die Datei `file.ext` zum Lesen.

Mit `$line=<INPUT>`; schreiben Sie die erste Zeile der Datei in die Variable `$line`.

Mit **close**(INPUT); schließen Sie die Datei.

Schreiben in eine Datei

open(OUTPUT, ">file.ext"); öffnet die Datei `file.ext` zum Schreiben.

Mit **print** OUTPUT "text"; schreiben Sie `text` in die Datei.

Mit **close**(OUTPUT); schließen Sie die Datei.

reguläre Ausdrücke in Perl

```
1  #!/perl -w
2  print "Please␣enter␣password:␣";
3  $password = <STDIN>;
4  chomp($password);
5  # While the entered line doesn't contain "42"
6  while ($password !~ /42/) {
7    print "Access␣denied.\n";
8    print "Please␣enter␣password:␣";
9    $password = <STDIN>;
10   chomp($password);
11  }
12  print "Correct␣password!␣Welcome.␣";
```

- reguläre Ausdrücke in Perl werden wie folgt eingeschlossen: `"/regex/`
- `!~` fordert, dass der reguläre Ausdruck nirgendwo im String matcht
- `=~` fordert, dass er irgendwo im String matcht
- Schreiben Sie ein Programm, das Passwörter akzeptiert, die mindestens drei Großbuchstaben enthalten.

Schalter und Ersetzen

Schalter

- `/regex/g` führt die Suche nach `regex` wiederholt bis zum Ende des Strings aus.
- `/regex/i` ignoriert Groß- und Kleinschreibung.

Ersetzen

`s/regex1/regex2/` ersetzt `regex1` durch `regex2`.

```
1 $text = "The meat costs 10 Mark, the fish costs 15 Mark. ";
2 $text =~ s/Mark/Euro/;
3 # "The meat costs 10 Euro, the fish costs 15 Mark."
```

```
1 $text = "The meat costs 10 Mark, the fish costs 15 Mark. ";
2 $text =~ s/Mark/Euro/g;
3 # "The meat costs 10 Euro, the fish costs 15 Euro."
```

Gruppieren und Speichern

- `/(text)/` findet und speichert `text`
- Der erste Klammerinhalt wird in `$1`, der zweite in `$2` u.s.w. gespeichert.

```
1 # Substitution and grouping
2 $sumM = 0;      # initializing the variable with zero
3 $sumE = 0;      # initializing the variable with zero
4 $text = "The meat costs 10 Mark, the fish costs 15 Mark.";
5 while ($text =~ /(\d+) Mark/g) {
6   $sumM = $sumM + $1;
7   $sumE = $sumE + $1/2;
8 }
9 print "total: $sumM Mark or $sumE Euro";
```

- 1 Überlegen Sie sich, wie das Programm `namen.pl` funktioniert (wählen Sie dazu `namen.txt` als Inputdatei) und ändern Sie es so, daß es die Namen in der Form “Initial Nachname” ausgibt.
- 2 Schreiben Sie ein Programm, das die Zahl der unbestimmten Artikel, die in einem Text vorkommen, zählt.
- 3 Schreiben Sie ein Programm, das alle Strophen des Liedes “Auf der Mauer auf der Lauer” aus der ersten Strophe generiert.

Dialogsystem: Eliza (Weizenbaum 1966)

User: Men are all alike

ELIZA: IN WHAT WAY

User: They're always bugging us about something or other

ELIZA: CAN YOU THINK OF A SPECIFIC EXAMPLE?

User: Well, my boyfriend made me come here

ELIZA: YOUR BOYFRIEND MADE YOU COME HERE

User: He says I'm depressed much of the time

ELIZA: I AM SORRY TO HEAR THAT YOU ARE DEPRESSED

Implementierungen:

- <http://www.hsweis.schule.neu-ulm.de/eliza/eliza3.html>
- <http://bs.cyty.com/menschen/e-etzold/archiv/science/rat.htm>
- <http://folk.uib.no/hhiso/eliza/index.htm>

Dialogsystem: Eliza (Weizenbaum 1966)

① Wechsel von 1. zur 2. Person:

```
1 s/\ bl( 'm|_am)\ b_/you_ are/g
2 s/\ bmy\ b_/your/g
3 s/\ bmine\ b_/yours/g
```

② Generierung von Antworten:

```
1 s/. * you are (depressed|sad) .*/i am sorry to hear you are $1/
2 s/. * you are (depressed|sad) .*/why do you think you are $1/
3 s/. * all .*/in what way/
4 s/. * always .*/can you think of a specific example/
```


Hausaufgabe (Abgabe: 10.6.2010)

- 1 Schreiben Sie ein Programm, das ein Paßwort abfragt und testet ob dieses die folgenden 3 Merkmale aufweist (verwenden Sie verschachtelte if-else Anweisungen):
 - es ist zwischen 5 und 8 Zeichen lang und
 - es endet auf einen Großbuchstaben und
 - es enthält mindestens zwei Ziffern.
- 2 ELIZA: Das Skript `eliza.pl` ist ein simples Eliza-Programm. Überlegen Sie sich, wie man das Programm verbessern könnte (z.B. durch die Hinzunahme weiterer Antwortmöglichkeiten), und nehmen Sie mindestens eine Verbesserung vor. Bitte dokumentieren Sie in ihrem Programm, worin Ihre Verbesserung besteht und wodurch Sie sie erreichen.
- 3 Bearbeiten Sie Aufgabe 2 von Folie 6.

Weiterführende Informationen

- Über den Turingtest
<http://plato.stanford.edu/entries/turing-test/>
- Zum Loebner Preis
<http://www.loebner.net/Prizef/loebner-prize.html>
- Kritik am Loebner Preis von Stuart Shieber
<http://www.eecs.harvard.edu/shieber/Biblio/Papers/loebner-rev-html/loebner-rev-html.html>
- Chatbot eines der Preisträger:
www.elbot.de