

Flex - erste Schritte

Karin Haenelt, 09.11.2004

1 Flex - Erkennung von Textmustern und Ausführung von Aktionen

Flex (Fast LEXical Analyzer) ist ein Werkzeug, das zur Erstellung von Programmen zur Erkennung von Textmustern dient. Flex ermöglicht eine komfortable Spezifikation von Textmustern und übersetzt diese Spezifikation dann in ein Programm (C, C++; JLex übersetzt in Java). Ein Flex-Programm besteht im Kern aus Paaren aus regulären Ausdrücken und aus C-Code (oder C++). Die regulären Ausdrücke beschreiben die Textmuster, die erkannt werden sollen, der C-Code beschreibt die Aktionen, die ausgeführt werden sollen, wenn ein beschriebenes Textmuster erkannt wurde. Abbildung 1 zeigt ein solches Paar. Im Beispiel wird ein „u“ erkannt und ein „x“ ausgegeben.

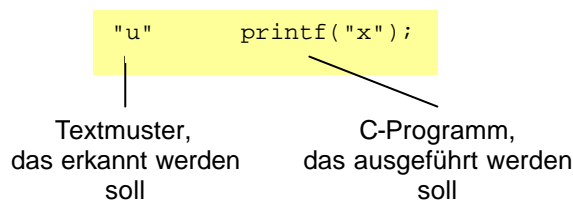


Abbildung 1 : ein einfaches Beispiel

2 Flex - Schritte der Verwendung

Die Verwendung von Flex erfolgt in folgenden Schritten

- 1) In einem Flex-Programm werden die Paare aus regulären Ausdrücken und C-Code spezifiziert
- 2) Das Flex-Programm wird mittels eines Übersetzers in ein C-Programm übersetzt
- 3) Das C-Programm wird kompiliert wie ein normales C-Programm
- 4) Das C-Programm wird ausgeführt wie ein normales C-Programm.

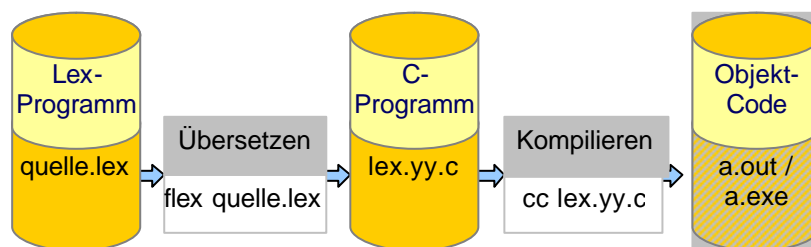


Abbildung 2: Schritte der Verwendung eines Lex-Programms

3 Flex - ein erstes Beispiel

Das Programm des ersten Beispiels ist in Abbildung 3 gezeigt.

```
/* *****  
/* ein x für ein u vormachen */  
/* *****  
%option noyywrap  
%%  
"u"      printf("x");  
%%  
  
main () {  
    yylex();  
}
```

Abbildung 3: Flex-Beispiel-Programm

Bei den Römern hatten die Buchstaben U und V beide das gleiche Schriftzeichen V. Das Schriftzeichen V steht auch für die Zahl 5. Mancher betrügerische Gläubiger machte sich den Umstand zunutze, daß das Zahlzeichen für die 10, das X, durch einfache Verlängerung der beiden Striche des V entstehen kann - aus einer Schuld von 5 Sesterzen wurde dann die doppelte Summe. Daher rührt die Redensart „ein X für ein U vormachen“, womit solcher Nepp gemeint ist.
(Kurt Krüger-Lorenzen, Das geht auf keine Kuhhaut Deutsche Redensarten und was dahinter steckt.)

Abbildung 4: Beispiel-Text zur Verarbeitung mit dem Beispiel-Programm

Die einzelnen Schritte zur Verwendung dieses Programms sind:

1. Schreiben des Flex-Programms

- Anlegen einer Datei namens `Beispiel01.lex` mit dem Beispiel aus Abbildung 3

2. Übersetzung des Flex-Programms in ein C-Programm

```
> flex Beispiel01.lex
```

erzeugt ein C-Programm namens `lex.yy.c`

3. Kompilation des C-Programms

```
> cc lex.yy.c (je nach C-Compiler auch gcc lex.yy.c)
```

erzeugt ein Programm namens `a.out` (unter UNIX) bzw. `a.exe` (unter Windows)

4. Ausführen des Programms mit einer Testdatei

unter UNIX:

```
> a.out < xu.txt oder
```

```
> ./a.out < xu.txt
```

unter Windows:

```
> a < xu.txt
```

Wenn alles funktioniert hat, erscheint nun auf dem Bildschirm folgender Text:

```
Bei den Römern hatten die Buchstaben X und V beide das
gleiche Schriftzeichen V. Das Schriftzeichen V steht auch
für die Zahl 5. Mancher betrügerische Gläubiger machte
sich den Umstand zunutze, daß das Zahlzeichen für die 10,
das X, durch einfache Verlängerung der beiden Striche
des V entstehen kann - aus einer Schuld von 5 Sesterzen
wurde dann die doppelte Summe. Daher rührt die Redensart
„ein X für ein U vormachen“, womit solcher Nepp gemeint
ist.
(Kurt Krüger-Lorenzen, Das geht auf keine Kuhhaut
deutsche Redensarten und was dahinter steckt.)
```

4 Weitere Beispiele

Weitere Übungsbeispiele und Erläuterungen (auch zu JLex) stehen unter

<http://kontext.fraunhofer.de/haenelt/kurs/Lex/index.html>

5 Software

Zur Verwendung von Lex braucht man Flex und einen C-Compiler bzw. JLex und einen Java-Compiler. Die Software steht im Internet frei zur Verfügung.

Zur Zeit gibt es folgende Quellen:

5.1 Flex

Programmdokumentation. [Flex](#)

(Bestandteil der gnu-Tools, verfügbar unter Unix, Linux, Windows: [Flex-Homepage](#))

Wilbur Streett (1996)

[Flex und Bison für WIN32](#)

Cygwin

eine Unix-artige Umgebung für Windows, die viele Unix-Komponenten enthält,

u.a. verschiedene C-Compiler, Flex und den Emacs-Editor;

das Paket ist sehr groß, und man sollte nur die Komponenten installieren,

die man auch braucht. <http://www.cygwin.com>

5.2 JLex

Berk, Elliot (1997ff)

[JLex - A lexical analyzer generator for Java](#). Department of Computer Science, Princeton University. (Dokumentation, Benutzungsanleitung und Quellcode)