JFlex Regular Expressions

Lecture 17
Section 3.5, JFlex Manual

Robb T. Koether

Hampden-Sydney College

Wed, Feb 25, 2015

Outline

The JAR Files

- JFlex uses a number of Java class.
- These classes have been compiled and are stored in the Java archive file `flex-1.6.0.jar`.
- Assignment 6 will have instructions on how to download and install this file.

Running JFlex

- The lexical analyzer generator is the `Main` class in the JFlex folder.
- To create a lexical analyzer from the file *`filename`*`.flex`, type

```
java jflex.Main filename.flex
```

- This produces a file `Yylex.java` (or whatever we named it), which must be compiled to create the lexical analyzer.

Running the Lexical Analyzer

Example (Using the `Yylex` Class)

```
InputStreamReader isr = new InputStreamReader(System.in);
BufferedReader br = new BufferedReader(isr);
Yylex lexer = new Yylex(br);
token = lexer.yylex();
```

- To run the lexical analyzer, a `Yylex` object must first be created.
- The `Yylex` constructor has one parameter, specifying a `Reader`.
- We will convert standard input, which is an input stream, to a buffered reader.
- Then call the function `yylex()` to get the next token.

Outline

JFlex Rules

- Each JFlex rule consists of a regular expression and an action to be taken when the expression is matched.
- The associated action is a segment of Java code, enclosed in braces `{ }`.
- Typically, the action will be to return the appropriate token.

JFlex Regular Expressions

- Regular expressions are expressed using ASCII characters (32 - 126).
- The following characters are metacharacters.

```
? * + | ( ) ^ $ . [ ] { } " \
```

- Metacharacters have special meaning; they do not represent themselves.
- All other characters represent themselves.

JFlex Regular Expressions

Regular Expression	Matches
r	One occurrence of r
$r?$	Zero or one occurrence of r
$r*$	Zero or more occurrences of r
$r+$	One or more occurrences of r
$r \mid s$	r or s
rs	r concatenated with s

- r and s are regular expressions.

JFlex Regular Expressions

- Parentheses are used for grouping.
- The expression

  ```
  ("+"|"-")?
  ```

  represents an optional plus or minus sign.
- If a regular expression begins with `^`, then it is matched only at the beginning of a line.
- If a regular expression ends with `$`, then it is matched only at the end of a line.
- The dot `.` matches any non-newline character.

JFlex Regular Expressions

- Brackets `[ ]` match any single character listed within the brackets.
- For example,
  - `[abc]` matches `a` or `b` or `c`.
  - `[A-Za-z]` matches any letter.
- If the first character after `[` is `^`, then the brackets match any character *except* those listed.
  - `[^A-Za-z]` matches any nonletter.

JFlex Regular Expressions

- A single character within double quotes `" "` or after `\` represents itself, except for `n`, `r`, `b`, `t`, and `f`.
- Metacharacters lose their special meaning and represent themselves when they stand alone within single quotes or follow `\`.

JFlex Escape Sequences

Escape Sequence	Matches
`\n`	newline (LF)
`\r`	carriage return (CR)
`\b`	backspace (BS)
`\t`	tab (TAB)
`\f`	form feed (FF)

- The character `c` is matched by `c`, `"c"`, and `\c`.
- The character `n` is matched by `n` and `"n"`, but not `\n`.
- The character `?` is matched by `"?"` and `\?`, but not `?`.

Outline

JFlex Example

Example (JFlex Example)

Design a JFlex file that will recognize

- Identifiers
- Integers
- Floating-point numbers
- String literals
- Comments
- Keywords

Outline

Identifiers

Example (Identifiers)

```
letter = [A-Za-z]
digit  = [0-9]
id     = {letter}({letter}|{digit}|"_")*
```

Outline

Integers

Example (Integers)

```
digit   = [0-9]
sign    = "+"|"-"
num     = {sign}?{digit}+
```

Integers

Example (Integers)

```
digit   = [0-9]
sign    = "+"|"-"
num     = {sign}?{digit}+
```

- What about octal numbers?

Integers

Example (Integers)

```
digit   = [0-9]
sign    = "+"|"-"
num     = {sign}?{digit}+
```

- What about octal numbers?
- What about hexadecimal numbers?

JFlex Example

Example (Floating-Point Numbers)

```
digit   = [0-9]
sign    = "+"|"-"
numpart = {digit}*({digit}.|.{digit}){digit}*
exp     = E{sign}?{digit}+
fpnum   = {sign}?{numpart}{exp}?
```

Outline

JFlex Example

Example (String Literals)

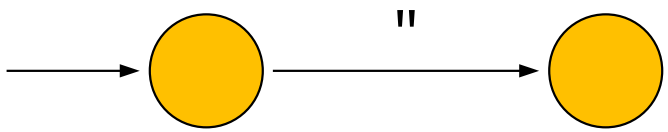
- What regular expression would describe string literals?
- `"hello"` represents `hello`.
- `"\"hello\""` represents `"hello"`.
- `"\h\e\l\l\o"` represents `hello`.
- String literals may not include newlines.

JFlex Example

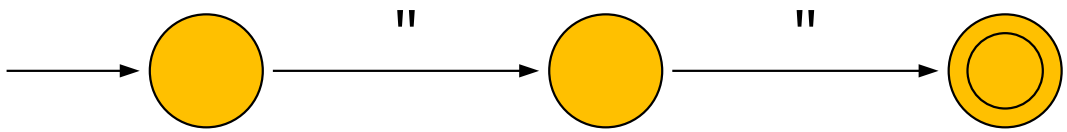
Example (String Literals)

- What regular expression would describe string literals?
- `"hello"` represents `hello`.
- `"\"hello\""` represents `"hello"`.
- `"\h\e\l\l\o"` represents `hello`.
- String literals may not include newlines.
- Draw a transition diagram for a DFA and then use it to write the regular expression.

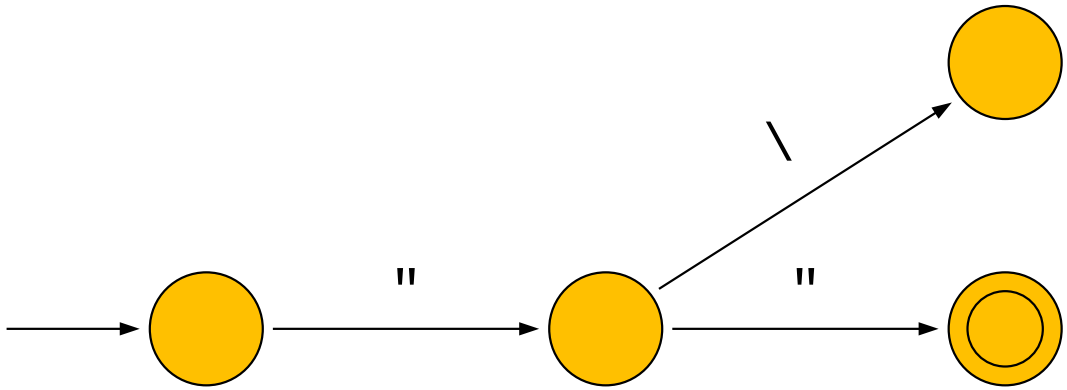
JFlex Example



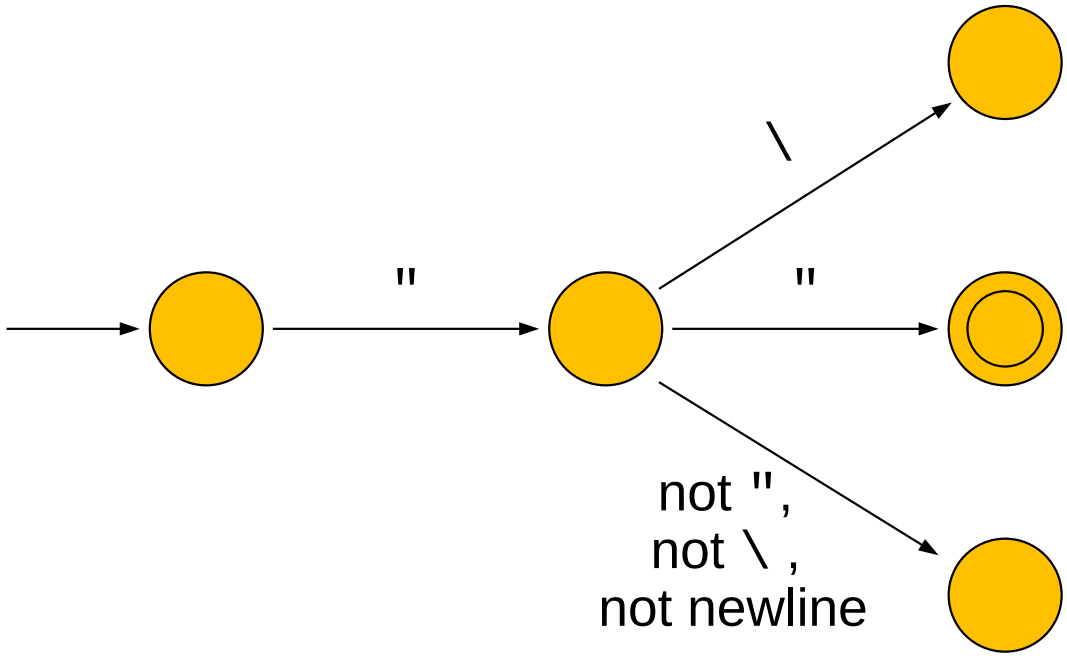
JFlex Example



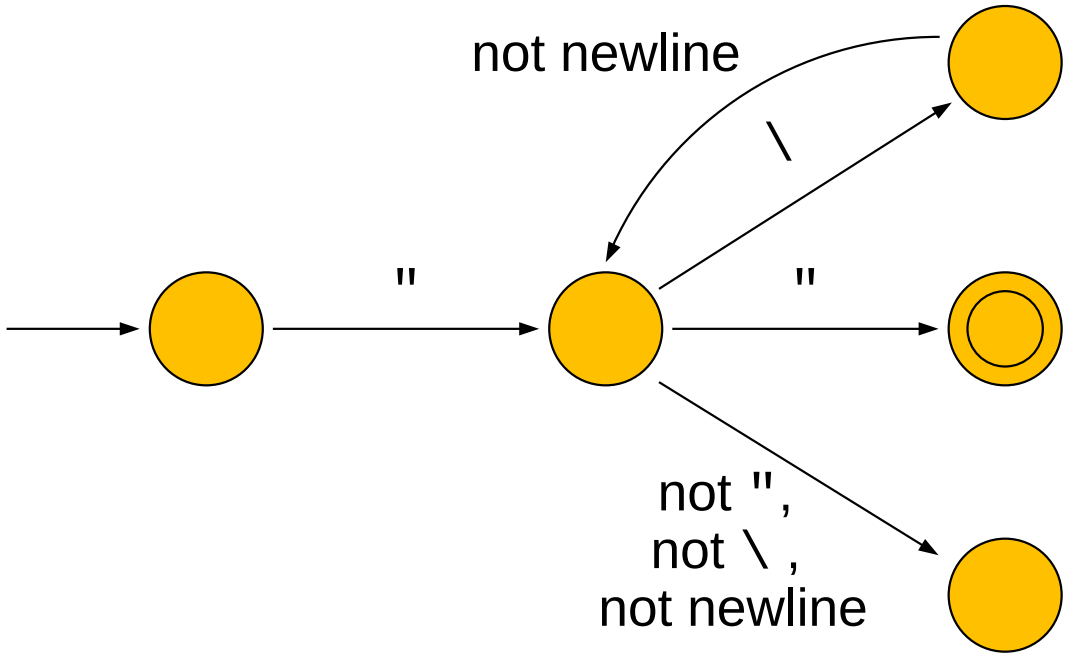
JFlex Example



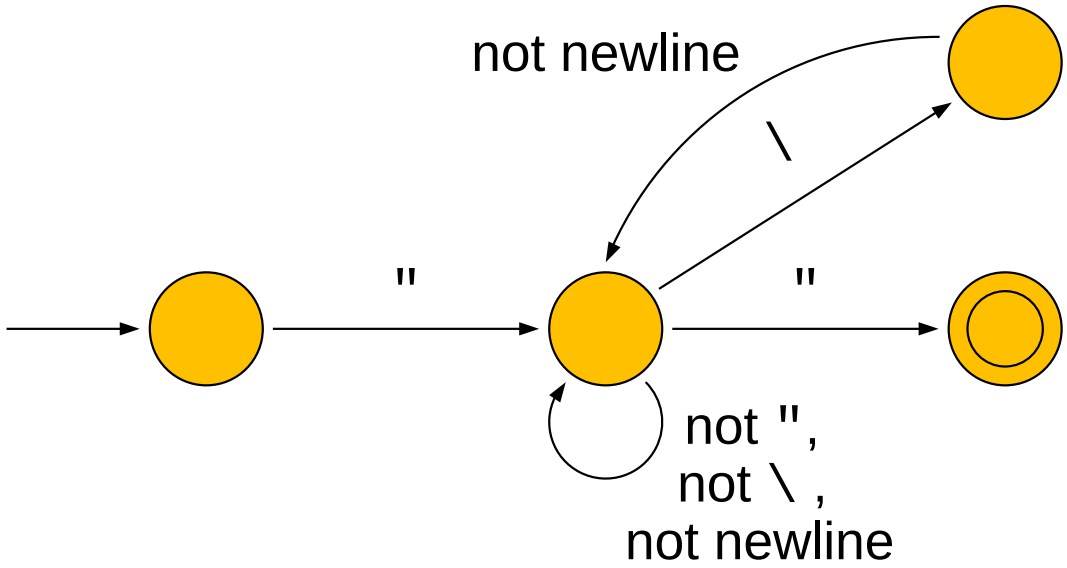
JFlex Example



JFlex Example



JFlex Example



Outline

JFlex Example

Example (Inline Comments)

- What regular expression would describe inline comments?
- Inline comments begin with `//` and extend up to, but not including, the next newline (or EOF).

JFlex Example

Example (Multiline Comments)

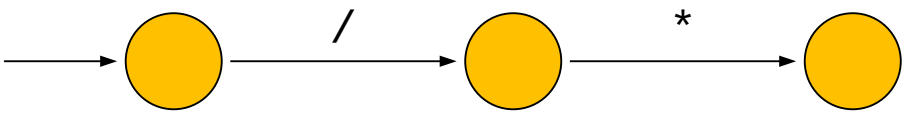
- What regular expression would describe multiline comments?
- Multiline comments begin with `/*` and end with `*/`.
- In between, there may occur any characters, including newlines.

JFlex Example

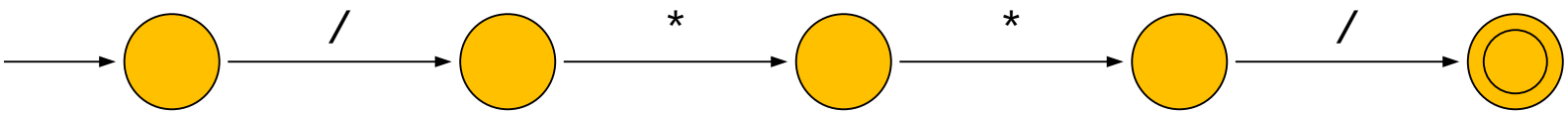
Example (Multiline Comments)

- What regular expression would describe multiline comments?
- Multiline comments begin with $/*$ and end with $*/$.
- In between, there may occur any characters, including newlines.
- Draw a transition diagram for a DFA and then use it to write the regular expression.

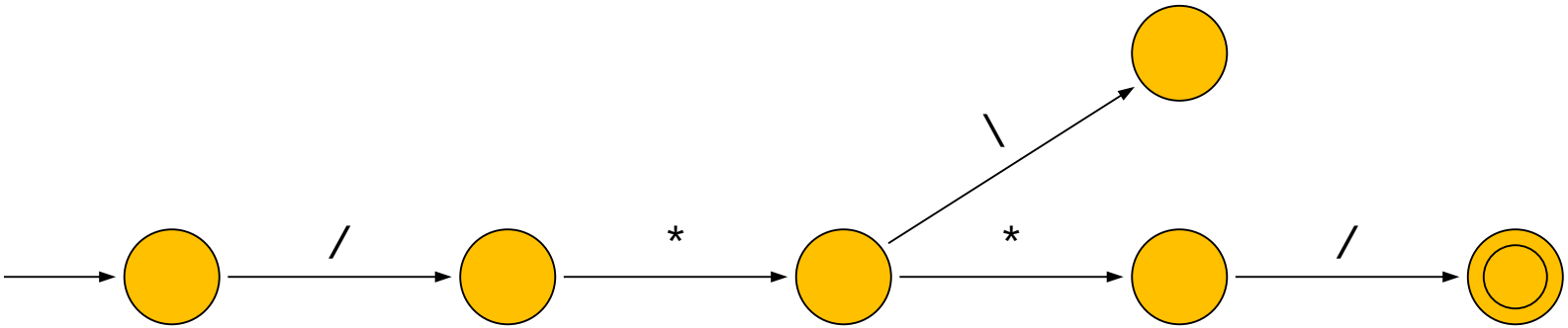
JFlex Example



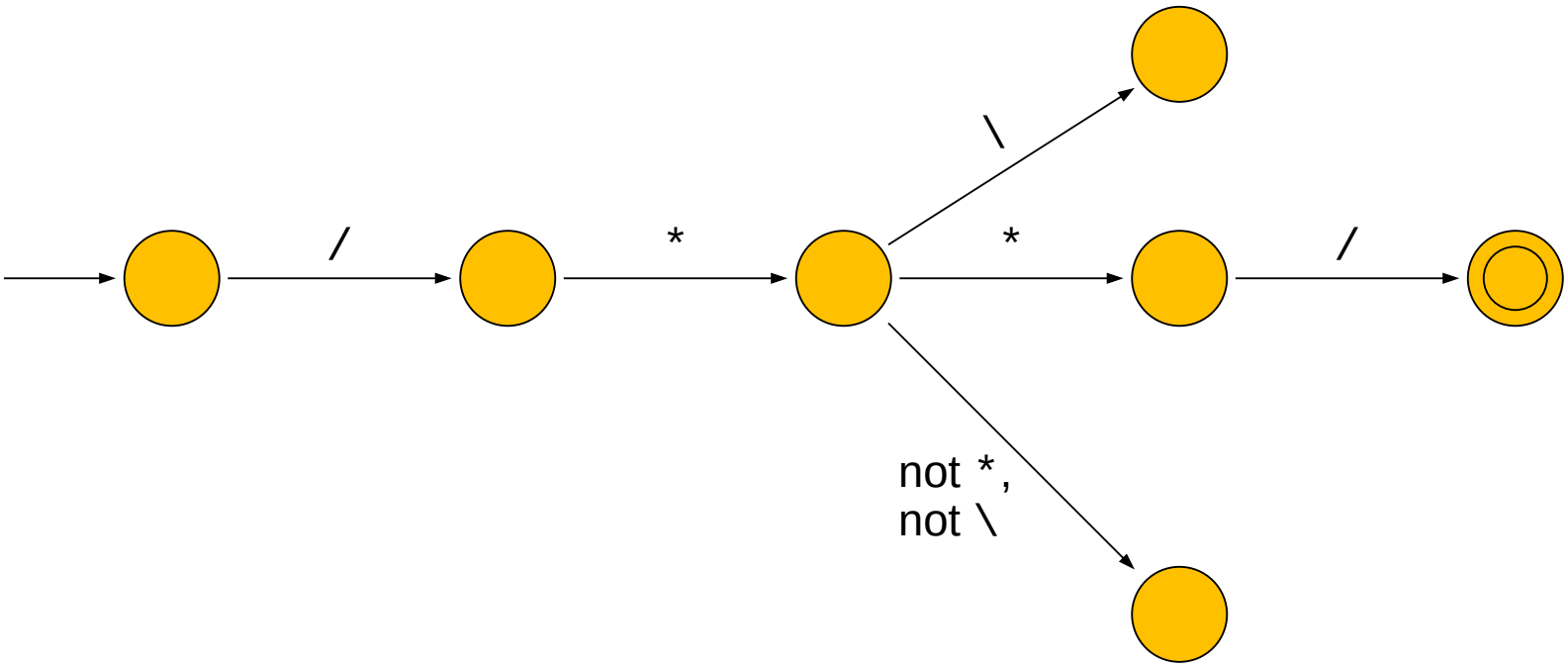
JFlex Example



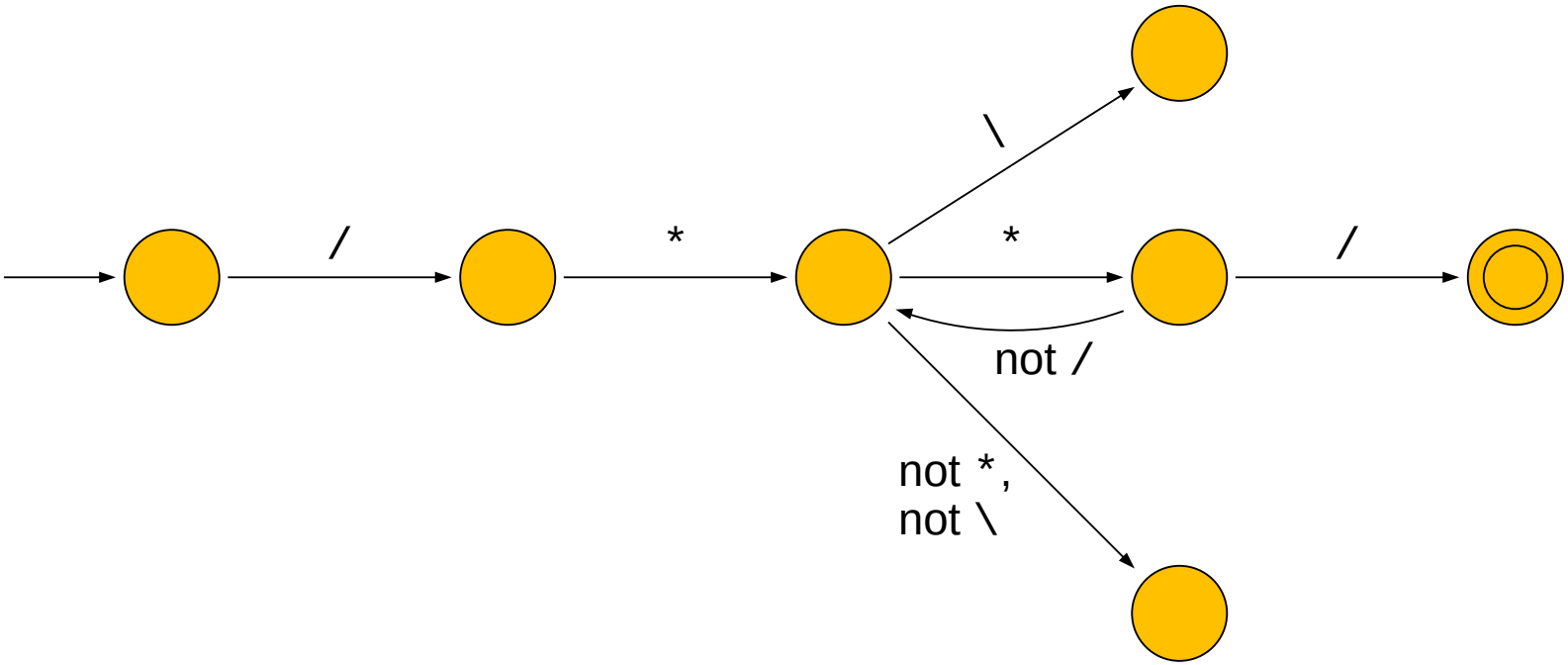
JFlex Example



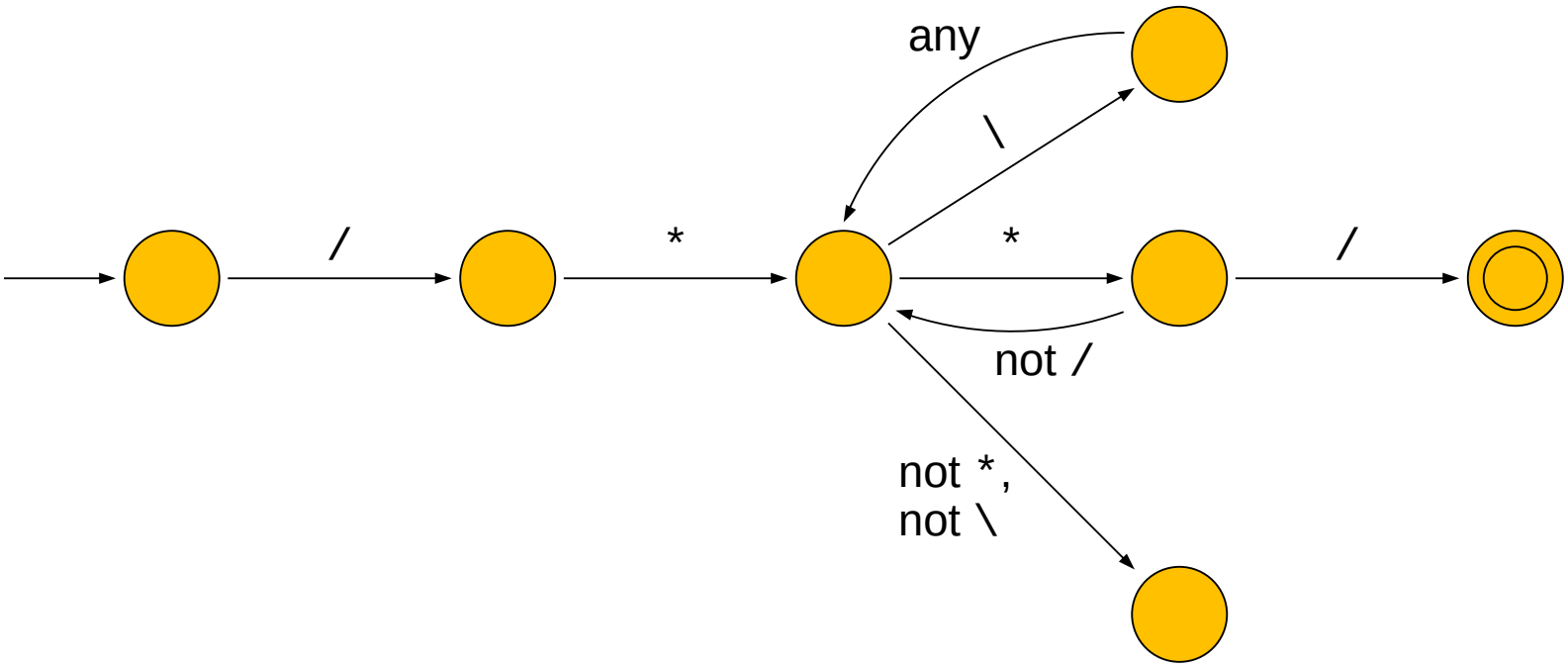
JFlex Example



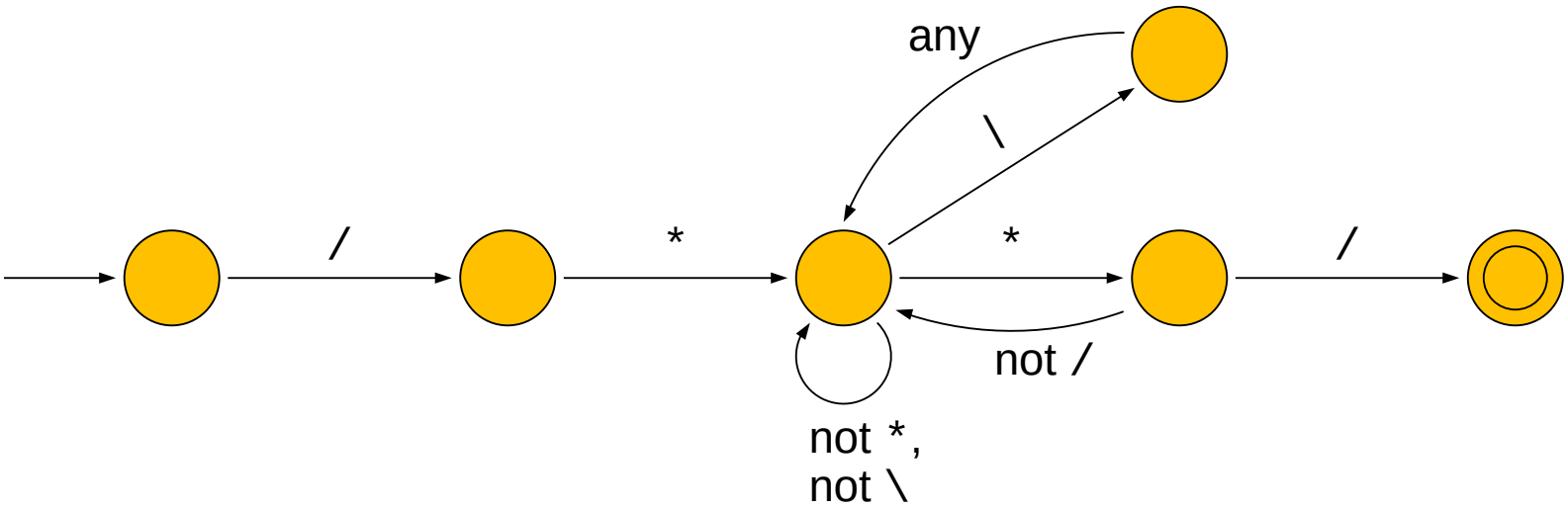
JFlex Example



JFlex Example



JFlex Example



Outline

JFlex Example

- The regular expression for a keyword is the very same sequence of characters.
- For example, the regular expression for **`if`** is `if`.
- How is JFlex to distinguish between keywords and identifiers?

Outline

Assignment

Assignment

- Read Section 3.5, which is about lex, not JFlex, but they are very similar.