

AWK Programming

Structure of an AWK Program

```
pattern { action }  
pattern { action }  
pattern { action }  
...
```

AWK Parsing

- AWK automatically reads every line of input
- AWK automatically divides each line into fields separated by whitespace
- AWK automatically assigns the following variables:

\$0	The entire line
\$1	The first field
\$2	The second field
\$3	The third field
$\$n$	The n th field
NR	The number of the current record (line)
NF	The number of fields on the current line

Patterns (1)

- Boolean expressions
 - `NR == 3`
 - `NR > 4 && NF == 2`
- Regular expressions enclosed in slashes
 - `/foo/` Matches all lines containing foo
 - `/^x/` Matches all lines starting with x
- Empty pattern
 - Matches all lines

Patterns (2)

- pattern1, pattern2
 - Range pattern
 - Starts matching with a line that matches pattern1
 - Stops matching with a line that matches pattern 2
 - Range is inclusive
- BEGIN
 - Matches the beginning of the input, before any lines are read
 - Any initialization code for the program goes here
- END
 - Matches the end of the input, after all lines are read
 - Used to print final results

Actions

0

- Syntax is similar to C, Java, etc
- `if, while, for, switch`
- `printf`
- Strings enclosed in double quotes
- The string concatenation operator is the empty space

String Functions

<code>length(s)</code>	Returns the length of s
<code>substr(s, i, n)</code>	Returns the substring starting at i of length n
<code>index(s, t)</code>	Returns the index where t starts in s
<code>gsub(r, s, t)</code>	Substitutes s for substrings matching regex r in t

Variables

- All variables are really strings
- Uninitialized variables are the empty string
- The empty string counts as 0 in arithmetic

Running AWK

<code>awk -f program file ...</code>	Runs AWK program on files
<code>awk -f program file ...</code>	Runs AWK program on stdin
<code>awk program file ...</code>	Runs AWK program in second argument on files
<code>awk program</code>	Runs AWK program in second argument on stdin
<code>awk -F sep program ...</code>	Runs AWK with field separator set to sep