

REGULAR EXPRESSIONS AND AWK

ŚRODOWISKO URUCHOMIENIOWE

1. Download file *SimpleAWKforWindows.zip* (AWK95)
(<http://www.cs.put.poznan.pl/mantczak/teaching/itc/SimpleAWKforWindows.zip>).
2. Extract the zip file into the working directory (During classes it is probably C:\Temp).
3. Important files used during compilation:
 - a. *awk95.exe* – executable file with interpreter of AWK language,
 - b. *in.txt* – text file with input data (text that should be processed),
 - c. *out.txt* – output text file with results (the result of processing of data from the input file made on the base of input file),
 - d. *prog.awk* – script's source code in AWK language,
 - e. *run.bat* – contains the command that executes *prog.awk* script, that reads data from the *in.txt* file and saves the result to *out.txt* file.

```
awk95 -f Skurce_code.awk <input_stream >output_stream
```

for example:

```
awk95 -f prog.awk <in.txt >out.txt
```

ZADANIA

ZAD. 1 (*). Run all programs presented during the lecture.

ZAD. 2. All operations in AWK represents values of input fields as:

- a) numbers,
- b) strings compared in lexicographical order,
- c) differently, depending on the field value.

ZAD. 3. The following input file is given:

```
FName:Alek SName:Gor Salary 700  
FName:Jurek SName:Busz Salary 585  
FName
```

Write the program that:

- a) Finds the longest line in the input file, outputs it and its length.

Output file

```
The longest record is FName:Jurek SName:Busz Salary 585  
Its size is 33
```

- b) Outputs every second line from the input file.

Output file

```
FName:Alek SName:Gor Salary 700
```

(*) asterisk marks problems which are not solved during exercise classes and should be solved as a homework

FName

- c) Output all fields building the row in reversed order.

Output file

```
700 Salary SName:Gor FName:Alek
585 Salary SName:Busz FName:Jurek
FName
```

- d) Outputs the count of lines that contains „SName“.

Output file

2

ZAD. 4. Let the input file contains only integer numbers, for example:

```
1   -2 3  -5 2
0 10 -12 -2 3  1
```

Write the program that:

- a) Outputs the absolute value of numbers.

Output file

```
1 2 3 5 2
0 10 12 2 3 1
```

- b) For each line outputs the sum of numbers in this line.

Output file

```
-1
0
```

- c) Outputs the count of numbers in each line:

Output file

```
5
6
```

- d) Outputs each row in such a way that its center is on 30th character of the output.

Output file

```
      1  -2 3  -5 2
    0 10 -12 -2 3  1
```

ZAD. 5. The following input file is given:

```
000902|Beavis|Theodore|333-242-2222|149092
000901|Jones|Bill|532-382-0342|234023
```

Write a program that::

- a) Outputs the phone number of employee whose name is *Jones*.

Output file

```
532-382-0342
```

- b) Outputs the maximal and the minimal value of first field from all lines in input file.

Output file

```
Max = 000902   Min = 000901
```

ZAD. 6. For each line output the number of fields in it, “:” character and unchanged line.

Input file

```
1, -2, 3, -5, 2
0, 10, -12, -2, 3, 1
```

Output file

```
5:1, -2, 3, -5, 2
6:0, 10, -12, -2, 3, 1
```

ZAD. 7. Write a program that removes from the beginning of each line spaces and tabs.

Input file

```
      1      -2 3      -5 2
0 10  -12 -2 3      1
```

(*) asterisk marks problems which are not solved during exercise classes and should be solved as a homework

Output file

```
1      -2 3  -5 2
0 10  -12 -2 3   1
```

ZAD. 8. Write a program that will replace some occurrences of sequence „foo” with sequence „bar”. Below there are detailed descriptions which sequences should be replaced.

Input file

```
foo afoo befoo allafoo bellafoo fooella
foo afoo befoo allafoo bellafoo fooella bazella bazolla ollabaz
bazo obaz
```

- a) Only first occurrence.

Output file

```
bar afoo befoo allafoo bellafoo fooella
bar afoo befoo allafoo bellafoo fooella bazella bazolla ollabaz
bazo obaz
```

- b) All occurrences.

Output file

```
bar abar bebar allabar bellabar barella
bar abar bebar allabar bellabar barella bazella bazolla ollabaz
bazo obaz
```

- c) All occurrences but only in lines that does not contain sequence „baz”.

Output file

```
bar abar bebar allabar bellabar barilla
```

- d) Output all lines in reversed order.

Output file

```
bazo obaz
foo afoo befoo allafoo bellafoo fooella bazella bazolla ollabaz
foo afoo befoo allafoo bellafoo fooella
```

- e) (*) For each line output it but before outputting replace the order of two first fields.

Output file

```
afoo foo befoo allafoo bellafoo fooella
afoo foo befoo allafoo bellafoo fooella bazella bazolla ollabaz
obaz bazo
```

- f) (*) For each line output all its fields except the second.

Output file

```
foo befoo allafoo bellafoo fooella
foo befoo allafoo bellafoo fooella bazella bazolla ollabaz
bazo
```

- g) Output all lines in reversed order and additionally reverse the order of fields in each line.

Output file

```
obaz bazo
ollabaz bazolla bazella fooella bellafoo allafoo befoo afoo foo
fooella bellafoo allafoo befoo afoo foo
```

ZAD. 9(*). Write a program that outputs all lines in unchanged form but if the line ends with backslash then this character is removed and no new line character is printed at the end of line.

Input file

```
foo afoo befoo allafoo bellafoo fooella
foo afoo befoo allafoo bellafoo fooella bazella bazolla ollabaz\
bazo obaz
```

Output file

```
foo afoo befoo allafoo bellafoo fooella
foo afoo befoo allafoo bellafoo fooella bazella bazolla ollabazbazo obaz
```

(*) asterisk marks problems which are not solved during exercise classes and should be solved as a homework

ZAD. 10(*). Following input file is given:

Input file

ala
bala
ala
bala
bola
bela

Write a program that:

- a) Outputs all unique lines.

Output file

ala
bala
bola
bela

- b) Groups following 5 lines and outputs them in the same line separated by comma.

Output file

ala,bala,ala,bala,bola
bela

- c) Outputs only the first line.

Output file

ala

- d) Outputs only two last lines from the input file in the reversed order.

Output file

bola
bela

- e) Outputs each line that is before some line that contains sequence „bola”.

Output file

bala