







"bad awk": a powerful text processing language



Ashley An, Christine Hsu, Melanie Sawyer, Victoria Yang PLT Fall 2018

Motivation

- Robust text processing language with intuitive C-like syntax
- Make it easy to analyze, read, and write to files
- Data-driven
- More verbose than awk
- Abstract away boilerplate code that repeatedly executes same actions over lines of a file
- Addition of mutable multidimensional arrays, easily mutable configuration variables

Tutorial – Run a bawk Program

```
hello.bawk
BEGIN {}
LOOP {
    print($0);
}
END {}
input.txt
hello
world
world
```

./bawk.sh hello.bawk input.txt ./bawk.sh [.bawk file] [input file] al@numel:~/plt/PLT-f18\$./bawk.sh test.bawk hello.txt
hello

_

world

Tutorial – Program Structure

```
BEGIN {
    # function declarations and global variable declarations
LOOP {
    # loop over each line of a file; execute these statements for each
line
END {
    # execute these statements after we're done with the file
CONFIG {
           # optional
    # set the field (word) separator & record (line) separator
```

Tutorial

Tutorial

Functions & Control Flow

```
int function (int a, int b) {
     while (a != b) {
          if (a > b) {
               a = a - b;
          else {
               b = b - a:
     return a;
```

Control Flow

```
int i = 0;
arr = [1, 2, 3, 4, 5];

for ( i=0; i < 10; i++) {
    print(int_to_string(arr[i]));
}</pre>
```

 "if" statements do not require matching "else" blocks

Tutorial

Other Special Keywords

- NF Number of Fields
- RS Record Separator
- FS Field Separator

Built-in Functions

type conversion functions

```
e.g. int_to_string
```

array functions

```
insert, delete, contains, length, index_of
```

- print
- nprint

Key Features – File Looping

```
LOOP {
    # everything in here is executed
    # once for each line of the file
}
```

- Continues looping until entire file is read through
- CONFIG block sets how the file will be looped through
 - Line separators are set with "RS"
 - Field separators are set with "FS"

Key Features – Field Access (\$)

Access a specified field of a line

Sample Line: Another layer of indirection

Set in CONFIG block:

- FS = Field Separator
 - o FS = ","
- RS = Record Separator
 - \circ RS = "\r\n"

```
print($0):
```

>> Another layer of indirection

print(\$1):

>> Another

print(\$2):

>> layer

Key Features – Infinitely nested mutable arrays

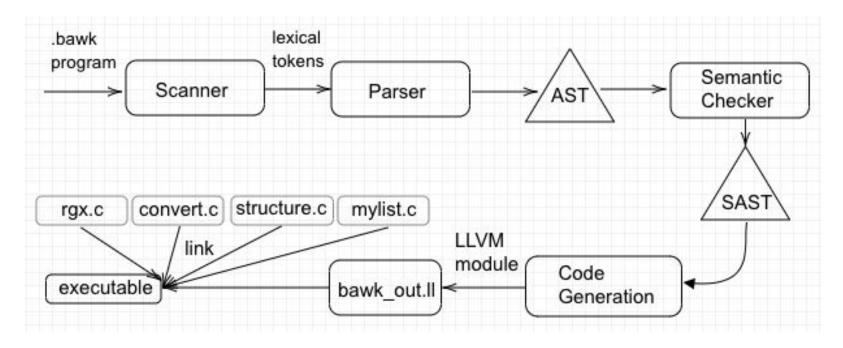
```
 \begin{aligned} &\inf \left[ \left[ \left[ 1, 2 \right], \left[ 3, 4 \right] \right], \left[ \left[ 5, 6 \right], \left[ 7, 8 \right] \right] \right]; \\ &m = \left[ \left[ \left[ 1, 2 \right], \left[ 3, 4 \right] \right], \left[ \left[ 5, 6 \right], \left[ 7, 8 \right] \right] \right]; \\ &m = \left[ \left[ \left[ 0, 2 \right], \left[ 3, 4 \right] \right], \left[ \left[ 5, 6 \right], \left[ 7, 8 \right] \right] \right]; \\ &delete(m, 1); & \# m = \left[ \left[ \left[ 0, 2 \right], \left[ 3, 4 \right] \right] \right] \\ &insert(m, 1, \left[ \left[ 9, 10 \right], \left[ 11, 12 \right] \right]); & \# m = \left[ \left[ \left[ 0, 2 \right], \left[ 3, 4 \right], \left[ \left[ 9, 10 \right], \left[ 11, 12 \right] \right] \right]; \end{aligned}
```

Key Features – Regex

- POSIX regex pattern matching with wrapper functions
- Allows text filtering and expression comparisons

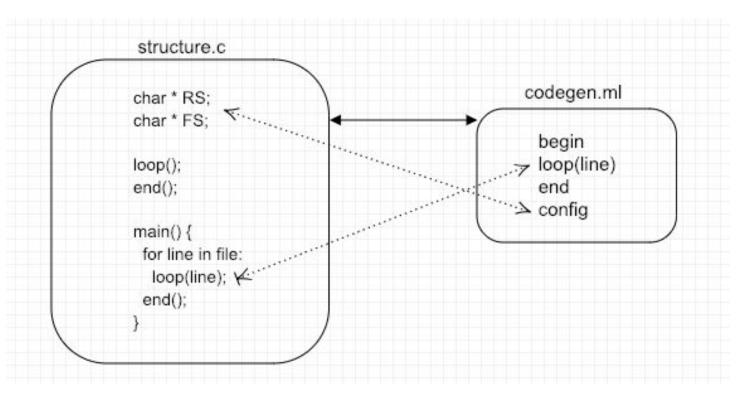
```
pattern = 'i .[a-zA-Z]* plt';
if (feeling ~ pattern) {
    print(feeling);
}
would match on "I love plt", "I hate plt", "I despise plt", "I fear plt", "I enjoy plt"
would not match on "I plt", "I do not love plt"
```

System Architecture



C libraries implement arrays, built-in conversion functions, regex, and main function

System Architecture



Testing

- Pass and fail tests for each stage of development
 - Lexer, parser, semantic checking, code generation
- Aim to pinpoint every feature of our language
- Check that the correct output / error messages are being generated
- Range from small tests (ex: basic operations) to larger tests (ex: file reading)
- Use bawk.sh [./bawk file] [input file] to run single test
- Use testall.sh to run all tests -> to automate running over 150 tests

Testing

[al@numel:~/bawk\$./bawk.sh tests/pass-helloworld.bawk input.txt
Hello World!

```
pass-dollar1...OK
                                                                                pass-ops3...OK
                                                                                                     pass-strarr6...0K
                                                                                                                              fail-blocks1...OK
                                                     pass-global2...OK
al@numel:~/bawk$ ./testall.sh
                                                                                                                                                    fail-func1...OK
                                                                                                     pass-strarr7...OK
                                                                                                                              fail-conversion1...OK
                              pass-dollar2...OK
                                                     pass-global3...OK
                                                                               pass-ops4...OK
pass-add1...0K
                                                                                                                                                    fail-func2...OK
                              pass-dynamicarr1...OK pass-helloworld...OK
                                                                                                                              fail-conversion2...OK
                                                                               pass-print...OK
                                                                                                     pass-strarr8...0K
pass-add2...OK
                                                                                                                                                    fail-func3...OK
                              pass-dynamicarr2...OK pass-helloworldloop...OK
                                                                                                                              fail-conversion3...OK
                                                                               pass-printbegin...OK
pass-arith1...OK
                                                                                                     pass-strarr9...OK
                                                                                                                                                    fail-func4...OK
                              pass-dynamicarr3...OK pass-helloworldloopend...OK
                                                                                                                             fail-conversion4...OK
                                                                                pass-rgx1...0K
                                                                                                     pass-string1...OK
pass-arith2...OK
                                                                                                                                                    fail-func5...OK
                              pass-dynamicarr4...0K pass-if1...0K
                                                                                                                             fail-conversion5...OK
pass-arith3...OK
                                                                                pass-rgx2...OK
                                                                                                     pass-string2...OK
                                                                                                                                                    fail-func6...OK
                                                                                                                              fail-dead1...OK
                              pass-dynamicarr5...OK pass-if3...OK
                                                                               pass-rgx3...OK
pass-arith4...OK
                                                                                                     pass-string3...OK
                                                                                                                                                    fail-func7...OK
                                                                                                                              fail-decl1...OK
                                                                               pass-rgx4...OK
                                                                                                                                                    fail-func8...OK
pass-arith5...OK
                              pass-dynamicarr6...OK pass-if4...OK
                                                                                                     pass-string4...OK
                                                                                                                              fail-decl2...OK
                                                                               pass-rgx5...OK
pass-arrayliterals...OK
                                                                                                                                                    fail-func9...OK
                              pass-dynamicarr7...OK pass-if5...OK
                                                                                                     pass-while1...OK
                                                                                                                              fail-decl3...OK
                                                                                                                                                    fail-global1...OK
                                                                               pass-rgx6...0K
pass-bool1...OK
                              pass-fib...OK
                                                     pass-if6...OK
                                                                                                     pass-while2...OK
                                                                                                                             fail-dollarbegin...OK fail-global2...OK
                                                                               pass-rgxarr1...0K
pass-bool2...OK
                                                     pass-if7...OK
                              pass-for1...OK
                                                                                                     fail-array1...0K
                                                                                                                             fail-dynamicarr1...OK fail-helloworldbegin...OK
pass-boolarr1...OK
                                                                               pass-rgxarr2...0K
                                                     pass-intarr1...OK
                              pass-for2...OK
                                                                                                     fail-array2...OK
                                                                                                                             fail-expr1...OK
                                                                                                                                                    fail-if1...OK
                                                     pass-intarr2...OK
                                                                               pass-rgxarr3...0K
pass-boolarr2...OK
                              pass-func1...OK
                                                                                                     fail-array3...OK
                                                                                                                              fail-expr2...OK
                                                                                                                                                    fail-if2...OK
                                                     pass-intarr3...OK
                                                                               pass-rgxarr4...0K
pass-boolarr3...OK
                              pass-func2...OK
                                                                                                     fail-arrayassign1...OK fail-expr3...OK
                                                                                                                                                    fail-if3...OK
                                                     pass-intarr4...OK
                                                                               pass-rgxarr5...OK
pass-boolarr4...OK
                              pass-func3...OK
                                                                                                     fail-arrayassign2...OK fail-expr4...OK
                                                                                                                                                    fail-length...OK
                                                     pass-intarr5...OK
pass-boolarr5...OK
                                                                               pass-rgxarr6...0K
                                                                                                     fail-arrayassign3...OK fail-expr5...OK
                                                                                                                                                    fail-print1...OK
                              pass-func4...OK
                                                     pass-intarr6...OK
                                                                               pass-rgxarr7...0K
pass-boolarr6...OK
                                                                                                                                                    fail-print2...OK
                                                     pass-intarr7...OK
                                                                                                                             fail-for1...OK
                              pass-func5...OK
                                                                                                     fail-assign1...OK
pass-boolarr7...OK
                                                                               pass-rgxarr8...0K
                                                                                                                                                    fail-return1...OK
                                                     pass-intarr8...OK
                                                                                                                             fail-for2...OK
                                                                                                     fail-assign2...OK
                              pass-func6...OK
pass-boolarr8...OK
                                                                               pass-rgxarr9...0K
                                                                                                                                                    fail-RS...OK
                                                     pass-intarr9...OK
                                                                                                                             fail-for3...OK
                                                                                                     fail-assign3...OK
                              pass-func7...OK
pass-boolarr9...OK
                                                                               pass-strarr1...OK
                                                     pass-local1...OK
                                                                                                                                                    fail-scope1...OK
                                                                                                                             fail-for4...OK
                              pass-func8...OK
                                                                                                     fail-assign4...OK
pass-comment...OK
                                                                               pass-strarr2...OK
                                                                                                                                                    fail-structure1...OK
                                                     pass-local2...OK
                                                                                                                              fail-for5...OK
pass-config1...OK
                              pass-func9...OK
                                                                                                     fail-assign5...OK
                                                                               pass-strarr3...0K
                                                                                                                                                    fail-structure2...OK
                                                     pass-nf1...OK
                                                                                                                              fail-for6...OK
pass-config2...OK
                              pass-gcd...OK
                                                                                                     fail-assign6...OK
                                                     pass-ops1...OK
                                                                               pass-strarr4...OK
                                                                                                                                                    fail-while1...OK
                                                                                                                              fail-for7...OK
pass-config3...OK
                              pass-global1...OK
                                                     pass-ops2...OK
                                                                                                     fail-assign7...OK
                                                                                                                                                    fail-while2...OK
                                                                               pass-strarr5...0K
                                                                                                                              fail-FS...OK
```

Demo

./bawk.sh demo/demo.bawk demo/shuffled.txt

:+/:://:/++co+ayys//-:
-:::://+++/+ayyyydimho::://:
:::+ocossydm############/:
:///+expyydmid00000mid00000ddyo+::::
:-:::+acopyhdddd#800000Hear:y/x:yf000000Mmy+///:
://oxyyhhddd#######:okhyfth-d#####yo+//:::
:/+coxyythd8000000/-Miy./Mist.s0000000my+//////-:
://cccsyhdaeeeeen,yesta.heest,+eeeeeeeests+++/+//
:/+coxxyd#00000004yf#00xd#0040yf#0000000004::
//+posta####################################
!!!//oyaff0588hoo//!!!!//exuhaff00008ay+oo+++/!!
-:::-:/oxmbay/::::/+::+//:::::/oybabbby+///:-::
//::/ohms:-,-/+ohsoy+/oo:/so+/yssos+::::++sddo::
-:1/++++//:1/ymmho/:/:1//+ohmnyhho/
::///::/:/hithdusossythyso/':/oyhdddhihyygyydittidy/
::::::+syMMm/+oo++dmdy/.'-:+ssyys+yddhyyssyhddo
······································
/a/nyyssssyyyy+///+yd#####!/:/:://:://:
://+ymmhhhdy/:+xx/syha:/+xxyhhmnd::/-::/-:
th++rat68dci:=addry//yh+:/+o++cosy+/:++
+hs' .shddhs++shdddd: 'sdo' '!/++/!/+++, ''//'/+
/hs' :00+//:-/sso/.''od+ '''::/:/:://:://'''.:/(:::+:/+/:::''
ys. +o-+coy//.+s:yyco:dd'''':::::-:://::/o+:,-/csys+/:::::::+o/:/+++/::'''
nd '//+ h/y:-hy:d:.hodnhm/yd.''':/+++oo//:/+oso/::/++++++//osoo+/:'
hd' :+/'s+. /+'as'-y+heydodh' '''''::!/++/////:'://+as+o//++//+assao//::.'
+d+ '-:-s+'-s:/o-::+mo ''''://:/os/:-:+yh/::::-/:+shhy/-::-/ox/+ssooo++//:
"sid/"ohfm.:+/+////-"":ho.' '.''://+ydy+:+::'".::oo+::shyo+o+++/:"
'sdo:++, 'ss/++,-sy/, '.+yo,'''
':ydo:'-/o:oy-'+///+yy-''''/osds##sy//ofdddy+///:
:oyys+/:/:-/+syho:/ohds#888y/ -:/oyddshs+//:-
:oyys+/:/-/+syho:/ohds888y/:/oyddshs+//: -/+++o+/:/ohmss88by:+ohd88hpo+/
:::::-:.:::::::::::::::::::::::::
://+coccassasytyc+///:::/ammmbbus/cyda000884dsc+/:
-:///coccessessylyto+///::::/smmm88ms/oyd899988dso+/: -:://////+++coccssylythdddddryso+/:+x488mds/::::://oxdm9988888ms+//:
://+cocassassy/byo+///:::-, /.smmmddns- /-/cydd00000cis-;-, -://////+++cocassy/bhddddryso+/:
-://-cocususus/yo////:::/namm#les-/-/cyde9998dco-/:://////-+
-://-occusions/hyo-///:::/smmm88es/o/x889886dxo-/::://////+occusyyfrhddddryso-/:htd88estx/::://occus9888888estx//:::::////+occusyyfrhddddmy/ds8eles8es8888888dx/:/:::::///occusyyfrhddddi/:ohm8899888es88888dx/:/::::///occusyyyfryddddi/:ohm8899888es88888dx/:/:
-://-cocasassy/nyo-///:::/namm#blus/cycleP0048dsco-/::://////-+cocassy/nyo-//::
-://-cocassassy/nyo////::://ammmMdes-/-/cyclePSM4dsch/:/:!//////++-cocassy/nyo///::
-://-coccusions/phyce///:::/namm##es-/cycle@99846co+/::!/////-+
-://-coccusions/y/or///:::-, //mmm88es-, /o/x289988dco+/:-, -://////
-://-coccusions/phyce///:::/namm##es-/cycle@99846co+/::!/////-+
-://-coccusions/y/or///:::-, //mmm88es-, /o/x289988dco+/:-, -://////
-://-coccusions/phyce///:::/nammidies-/cycle/99846to+/::://////-+
-://-coccusions/phyce///:::/nammidies-/cycle/99846to+/::://////-+
-://-coccusions/y/or///:::/nummibles/o/gd@9886dco+/:
-://-coccusions/phyce///:::/nammillions/cyde99998dico+/::://////
-://-coccusions/yryor///:::/nummibles
-://-coccusions/phyce///:::/nammbles/cycle/9984dco+/::://////
-://-coccusions/phyce///:::
-://-coccusions/phyce///:::/namm#less-/cock@edictor/:/cock@edictor/:/ -:://////
-://-coccusions/yryor///:::/nammibles
-://-coccusions/yryor///:::/nammibles
-://-coccusions/y/yor///:::
-://-coccusions/phyce///:::/summillions-/-/cycle/998884cs-/::://////
-://
-:://
-://-coommunication/co////::/nummibles://///-coommunication/co//-colabeleddom/:::://commiblesdom/co//:://////
-:://

