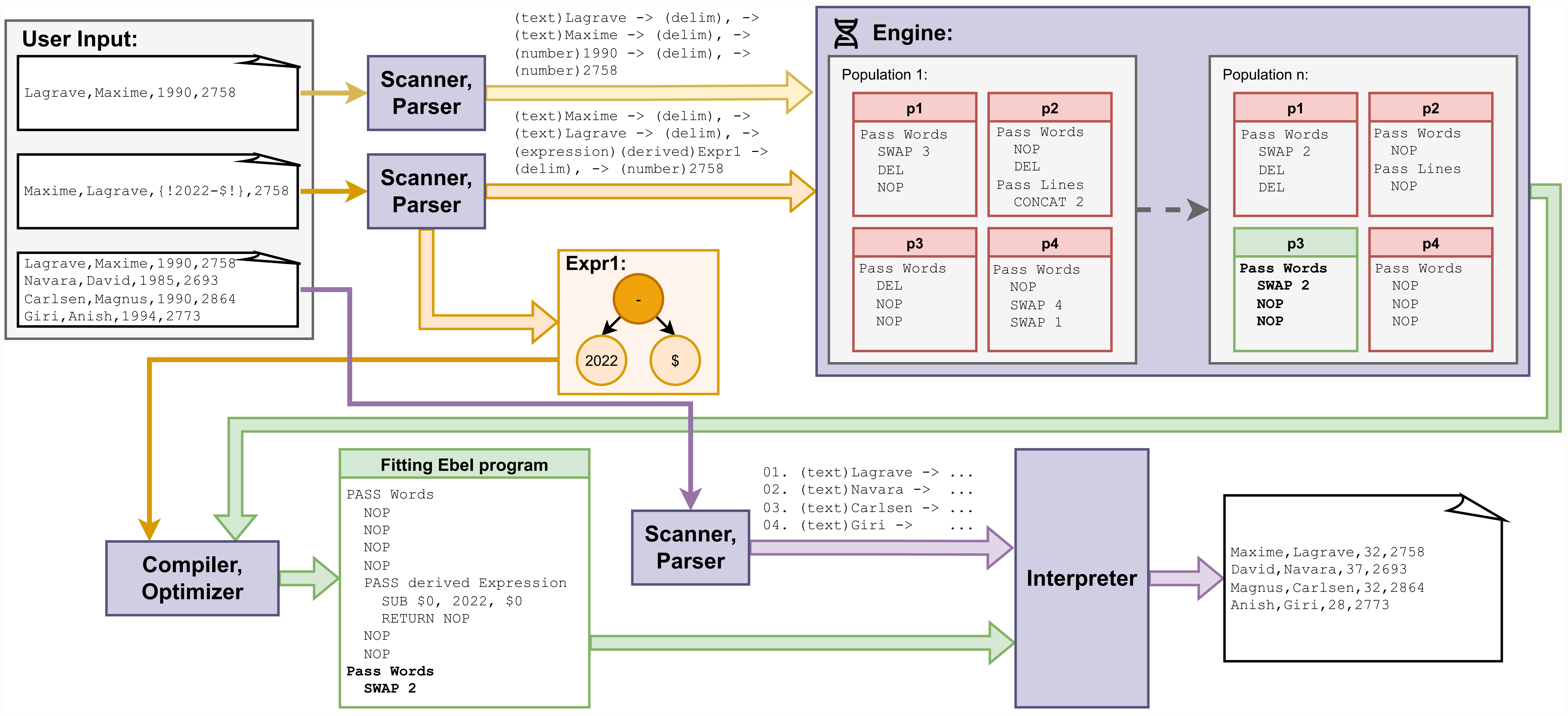
Ebe – A tool for automated file editing using genetic programming

Bc. Marek Sedláček

Supervisor: prof. Ing. Lukáš Sekanina, Ph.D.

Faculty of Information Technology, Brno University of Technology



```
Lagrave,Maxime,1990,2758
```

```
Maxime,Lagrave,{!2022-$!},2758
```

```
Lagrave,Maxime,1990,2758
Navara,David,1985,2693
Carlsen,Magnus,1990,2864
Giri,Anish,1994,2773
```

```
(text)Lagrave -> (delim), ->
(text)Maxime -> (delim), ->
(number)1990 -> (delim), ->
(number)2758
```

```
(text)Maxime -> (delim), ->
(text)Lagrave -> (delim), ->
(expression)(derived)Expr1 ->
(delim), -> (number)2758
```

Population 1:

p1
```
Pass Words
  SWAP 3
  DEL
  NOP
```

p2
```
Pass Words
  NOP
  DEL
Pass Lines
  CONCAT 2
```

p3
```
Pass Words
  DEL
  NOP
  NOP
```

p4
```
Pass Words
  NOP
  SWAP 4
  SWAP 1
```

Population n:

p1
```
Pass Words
  SWAP 2
  DEL
  DEL
```

p2
```
Pass Words
  NOP
Pass Lines
  NOP
```

p3
```
Pass Words
  SWAP 2
  NOP
  NOP
```

p4
```
Pass Words
  NOP
  NOP
  NOP
```

Fitting Ebel program
```
PASS Words
  NOP
  NOP
  NOP
  NOP
  PASS derived Expression
    SUB $0, 2022, $0
    RETURN NOP
  NOP
  NOP
Pass Words
  SWAP 2
```

```
01. (text)Lagrave -> ...
02. (text)Navara ->  ...
03. (text)Carlsen -> ...
04. (text)Giri ->    ...
```

```
Maxime,Lagrave,32,2758
David,Navara,37,2693
Magnus,Carlsen,32,2864
Anish,Giri,28,2773
```

Figure 1: Ebe's processing pipeline.

Introduction

Editing files is a process that many people in everyday computer work have to do. For programmers this might mean writing some script in language like AWK or Python, running it and checking its correctness. For those less skilled in programming this might mean a tedious work with editing the file by hand in an editor. The latter approach is in no way efficient and is due for an upgrade. The first approach can always be made less prone to errors and faster.

Ebe – Edit by Example

Ebe is a free open-source program (compiler and interpreter) for editing files just from given examples.
All the user has to do is create a snipped of file before and after the desired edits (input and output example), give this to Ebe and it will try to find a fitting algorithm (program in *Ebel* language), which would do the desired transformations from input example to the output example. This program then can be interpreted using Ebe even on multiple files with similar structure at once.
Since Ebe is aimed also at people with little to no programming knowledge, Ebe's philosophy is to not cause exceptions and errors as long as it is not necessary. Meaning that, when an incorrect instruction or input is encountered, rather than exiting the execution with error, only a warning is printed and the instruction or input is ignored. Ebe is very verbal about this and will notify about any problems, but this philosophy is quite handy when interpreting multiple files, which might slightly differ (a division by 0 might be encountered or different data type at some position), but one Ebel will work on all of them.
One of Ebe's strengths is that it generates Ebel instructions, which in most cases care only about the word's position rather than its type or even value. This means that the example provided can only have the same structure as the actual file that needs to be edited. This gives the option to provide someone an editing script without requesting the actual data that needs to be edited, which could otherwise be a problem for security or other reasons. This approach can also be useful in cases, where the output is ambiguous to the input because of the values.
Ebe can be used with very little knowledge of it, but at the same time offers additional power to those more skilled with it. Ebe contains heuristic, which set up all the parameters for genetic programming, but this heuristic is product of experimenting and is not all-knowing. For these reasons and to allow for experimenting, the user can at compile time setup many different parameters and attributes to find the best configuration or try out different approaches.

Genetic programming as the programmer

Genetic programming (*GP*) is a process, which imitates evolution in nature. In case of Ebe, the evolutionary process has a population of Ebel programs (see more info bellow) and these are being evolved until the program does exactly the transformation of the example input to the example output. One key feature for all of GP approaches is to have a fitness function, which ranks each entity that is being evolved. Ebe's fitness function is how similar is the interpreted example input to the example output, the more similar the better fitness is given to an evolved program.
Finding expressions is a very difficult task for genetic programming that would require more than just one example. For this reason Ebe allows the user to specify the exact expression to use, such expression is then considered by Ebe as always correct.

Ebel – Ebe language

Ebel is an imperative, case insensitive, programming language designed for file editing and to work well with genetic programming.
Ebel resembles a bytecode (example can be seen in figure 1) and was designed this way to allow for quick parsing and execution in the interpreter. It is interpreted over a file, where the Ebel code can be thought of as a pipeline of instructions through which the file objects (lexemes) go and get modified by.
Ebel is composed of multiple sections called passes. A pass defines in which way the input file is read. Each pass is then composed of instructions which take as an input objects its parent pass parses (word or line).

Summary and conclusions

Ebe displays a promise as an editing tool for not only non-programmers. It requires minimum knowledge of the tool, finds the algorithm for the user and even allows to edit multiple similar files at once.
Although Ebe does not yet contain all the wanted features and optimizations, it still allows for some advanced edits, which it can find in a short time and most importantly it can do large simple edits, which might be lot of times needed more than the complex ones. With the addition of user defined expressions Ebe can also tackle more complex problems.
The goal is to get Ebe to the point, where it can do more complicated edits faster than using other file editing approaches or at least in a similar time frame without any additional help from the user.