

UNIVERZITET U BEOGRADU
MATEMATIČKI FAKULTET



Ivan Ristović

KREIRANJE ZAJEDNIČKE AST APSTRAKCIJE ZA RAZLIČITE PROGRAMSKE JEZIKE

master rad

Beograd, 2020.

Mentor:

doc. dr Milena VUJOŠEVIĆ-JANIČIĆ
Univerzitet u Beogradu, Matematički fakultet

Članovi komisije:

dr Ana ANIĆ
University of Disneyland, Nedodija

dr Laza LAZIĆ
Univerzitet u Beogradu, Matematički fakultet

Datum odbrane: _____

TODO zahvalnica

Naslov master rada: Kreiranje zajedničke AST apstrakcije za različite programske jezike

Rezime: Fijuče vetar u šiblju, ledi pasaže i kuće iza njih i gunda u odžacima. Nidžo, čežnjivo gledaš fotelju, a Đura i Mika hoće poziciju sebi. Ljudi, jazavac Džef trči po šumi glođući neko suho žbunje. Ljubavi, Olga, hajde pođi u Fudži i čut ćeš nježnu muziku srca. Boja vaše haljine, gospođice Džafić, traži da za nju kulućim. Hadži Đera je začutao i bacio čežnjiv pogled na šolju s kafom. Džabe se zec po Homolju šunja, čuvar Jožef lako će i tu da ga nađe. Odžaćar Filip šalje osmehe tuđoj ženi, a njegova kuća bez dece. Butić Đuro iz Foče ima pun džak ideja o slaganju vaših željica. Džajić odskoči u aut i izbeže đon halfa Pecelja i njegov šamar. Plamte odžaci fabrika a čađave guje se iz njih dižu i šalju noć. Ajšo, lepoto i čežnjo, za ljubav srca moga, dođi u Hadžiće na kafu. Hući šuma, a iza žutog džbuna i panja đak u cveću delje seji frulu. Goci i Jaćimu iz Banje Koviljače, flaša džina i žeđ padahu u istu uru. Džaba što Feđa čupa za kosu Milju, ona juri Živu, ali njega hoće i Daga. Dok je Fehim u džipu žurno ljubio Zagu Čadević, Cile se ušunjao u auto. Fijuče košava nad odžacima a Ilja u gunju žureći uđe u suhu i toplu izbu. Bože, džentlmeni osećaju fizičko gađenje od prljavih šoljica! Dočepaće njega jaka šefica, vođena ljutom srdžbom zlih žena. Pazi, gedžo, brže odnesi šefu taj đavolji ček: njim plaća ceh. Fine džukce ozleđuje bič: odgoj ih pažnjom, strpljivošću. Zamišljao bi kafedžiju vlažnih prstića, crnjeg od čađi. Daće, uštedu plaćaj žaljenjem zbog džinovskih cifara. Džikljaće žalfija između tog busenja i pešćanih dvoraca. Zašto gđa Hadžić leći živce: njena ljubav pred fijaskom? Jež hoće peckanjem da vređa ljubičastog džina iz flaše. Džej, ljubičast zec, laže: gađaće odmah pokvašen fenjer. Plašljiv zec hoće jeftinu dinju: grožđe iskamči džabe. Džak je pun žica: čućeš tad svađu zbog lomljenja harfe. Čuj, džukac Flop bez daha s gađenjem žvaće stršljena. Oh, zadnji šraf na džipu slab: muž gđe Cvijić ljut koči. Šef džabe zvižduće: mlađi hrt jače kljuca njenog psa. Odbaciće kavgadžija plaštom čađ u željezni fenjer. Deblji krojač: zgužvah smeđ filc u tanjušni džepić. Džo, zgužvaćeš tiho smeđ filc najdeblje krpenjače. Štef, bacih slomljen dečji zvrk u džep gđe Žunjić. Debljoj zgužvah smeđ filc — njen škrt džepčić.

Ključne reči: TODO

Sadržaj

1	Uvod	1
2	Pregled relevantnih pojmova	3
2.1	Apstraktna sintaksna stabla - AST	3
2.2	ANTLR	5
3	Zaključak	8
	Literatura	10

Glava 1

Uvod

Fijuče vetar u šiblju, ledi pasaže i kuće iza njih i gunđa u odžacima. Nidžo, čežnjivo gledaš fotelju, a Đura i Mika hoće poziciju sebi. Ljudi, jazavac Džef trči po šumi glođući neko suho žbunje. Ljubavi, Olga, hajde pođi u Fudži i čut ćeš nježnu muziku srca. Boja vaše haljine, gospođice Džafić, traži da za nju kulućim. Hadži Đera je začutao i bacio čežnjiv pogled na šolju s kafom. Džabe se zec po Homolju šunja, čuvar Jožef lako će i tu da ga nađe. Odžačar Filip šalje osmehe tuđoj ženi, a njegova kuća bez dece. Butić Đuro iz Foče ima pun džak ideja o slaganju vaših željica. Džajić odskoči u aut i izbeže đon halfa Pecelja i njegov šamar. Plamte odžaci fabrika a čađave guje se iz njih dižu i šalju noć. Ajšo, lepoto i čežnjo, za ljubav srca moga, dođi u Hadžiće na kafu. Hući šuma, a iza žutog džbuna i panja đak u cveću delje seji frulu. Goci i Jaćimu iz Banje Koviljače, flaša džina i žeđ padahu u istu uru. Džaba što Feđa čupa za kosu Milju, ona juri Živu, ali njega hoće i Daca. Dok je Fehim u džipu žurno ljubio Zagu Čadević, Cile se ušunjao u auto. Fijuče košava nad odžacima a Ilja u gunju žureći uđe u suhu i toplu izbu. Bože, džentlmeni osećaju fizičko gađenje od prljavih šoljica! Dočepaće njega jaka šefica, vođena ljutom srdžbom zlih žena. Pazi, gedžo, brže odnesi šefu taj đavolji ček: njim plaća ceh. Fine džukce ozleđuje bič: odgoj ih pažnjom, strpljivošću. Zamišljao bi kafedžiju vlažnih prstića, crnjeg od čađi. Đaće, uštedu plaćaj žaljenjem zbog džinovskih cifara. Džikljaće žalfija između tog busenja i peščanih dvoraca. Zašto gđa Hadžić leći živce: njena ljubav pred fijaskom? Jež hoće peckanjem da vredi ljubičastog džina iz flaše. Džej, ljubičast zec, laže: gađaće odmah pokvašen fenjer. Plašljiv zec hoće jeftinu dinju: grožđe iskamči džabe. Džak je pun žica: čućeš tad svađu zbog lomljenja harfe. Čuj, džukac Flop bez daha s gađenjem žvaće stršljena. Oh, zadnji šraf na džipu slab:

muž gđe Cvijić ljut koči. Šef džabe zvižduće: mlađi hrt jače kljuca njenog psa. Odbaciće kavgadžija plaštom čađ u željezni fenjer. Deblji krojač: zgužvah smeđ filc u tanjušni džepić. Džo, zgužvaćeš tiho smeđ filc najdeblje krpenjače. Štef, bacih slomljen dečji zvrk u džep gđe Žunjić. Debljoj zgužvah smeđ filc — njen škrt džepčić.

Glava 2

Pregled relevantnih pojmova

U ovom poglavlju će biti opisani koncepti i alati čije je razumevanje potrebno kako bi se razumeo opis dalje apstrakcije i implementacije samog programa. Umesto analize samog sadržaja izvornog koda analizira se *apstraktno sintaksno stablo* (eng. *Abstract Syntax Tree*, u daljem tekstu *AST*), opisano u odeljku 2.1. Alat koji je korišćen za generisanje parsera za proizvoljnu gramatiku jezika se zove *Another Tool For Language Recognition* [1], u daljem tekstu *ANTLR*, opisan u odeljku 2.2. Parser generiše AST specifičan za datu gramatiku i nema sličnosti u dobijenim apstrakcijama za različite jezike¹. Kako bismo poredili stabla različitih jezika, kreiramo reprezentaciju na višem nivou i specifični AST podižemo na taj nivo. Ta reprezentacija će biti opisana u narednim poglavljima, kao i načini kako se ona može analizirati. Takođe, pojmovi specifični za implementaciju će takođe biti opisani u ovom poglavlju.

2.1 Apstraktna sintaksna stabla - AST

Kako bi se od datoteke na fajl sistemu koja sadrži izvorni kod programa došlo do izvršivog programa, potrebno je izvršiti više koraka [3]:

- pretprocesiranje
- prevođenje
- asembliranje

¹Zanimljivo je da, čak iako intuitivno pomislimo da su im gramatike slične, C i C++ gramatike nemaju skoro nijedno istoimeno pravilo. Jedina sličnost je da obe gramatike koriste rekurzivna pravila da reše problem prioriteta operatora.

- linkovanje

Ovi koraci će biti opisani na jednom primeru. Pretpostavimo da želimo da kompajliramo kod pisan u programskom jeziku C prikazan na slici 2.1. Primetimo, da postoji greška u datom kodu - simbol `c` koji se koristi će biti prepoznat kao identifikator koji ne odgovara nijednoj promenljivoj. Ovo, doduše, nije sintaksna greška - `a + c` je sasvim validan izraz u programskom jeziku C. Problem će postati očigledan tek nakon parsiranja izvornog koda i provere ispunjenosti sintaksnih pravila.

```
1      #include<stdio.h>
2
3      #define T int
4
5      int main()
6      {
7          T a, b;
8          a = a + c;          // c nije deklarirano
9          printf("%d", a);
10         return 0;
11     }
```

Slika 2.1: Primer izvornog koda pisanog u programskom jeziku C.

U fazi pretprocesiranja se vrše samo tekstualne operacije kao što su brisanje komentara ili zamena makroa u jezicima kao što je C. Prvo mesto gde se vrši analiza sadržaja izvornog fajla je faza prevođenja. Tu analizu vrši program koji se naziva *pretprocesor*. Rezultat rada pretprocesora, za kod sa slike 2.1 bi izgledao kao na slici 2.2 ².

Prilikom faze prevođenja, kako prevodilac ne bi radio nad sirovim karakterima izvornog koda, potrebno je izvršiti pripremu istog. Prevodilac ima u vidu moguće elemente programskog jezika, tzv. *tokene*, koje treba prepoznati u datom fajlu - ključne reči, operatore, promenljive itd. Program koji radi *tokenizaciju* - prepoznavanje tokena u izvornom fajlu - se naziva *lekser*. Pojednostavljen primer

² U nekim implementacijama C standardne biblioteke, moguće je da se poziv funkcije `printf` zameni pozivom funkcije `fprintf` sa ispisom na `stdout`. U standardu se propisuje da funkcije kao što je `printf` mogu biti implementirane kao makroi. Izlaz na slici 2.2 je generisan od strane GCC 7.4.0 po C11 standardu.

```
1      int main()
2      {
3          int a, b;
4          a = a + c;
5          printf("%d", a);
6          return 0;
7      }
```

Slika 2.2: Rezultat rada pretprocesora za kod sa slike 2.1.

tokena koje lekser pokušava da prepozna se može videti na slici 2.3. Primer izlaza leksera za izlaz pretprocesora sa slike 2.2 se može videti na slici 2.4.

```
1      Identifier : IdentifierNondigit
2                  (IdentifierNondigit | Digit)*
3                  ;
4
5      IdentifierNondigit : Nondigit
6                          | UniversalCharacterName
7                          ;
8
9      Nondigit : [a-zA-Z_]
10             ;
11
12     Digit : [0-9]
13         ;
```

Slika 2.3: Primer delimične definicije tokena za ime promenljive po C11 standardu.

Počevši od kontekstno-slobodne gramatike [2], koja se može smatrati konkretnom sintaksom programskog jezika, može se

2.2 ANTLR

Fijuće vetar u šiblju, ledi pasaže i kuće iza njih i gunda u odžacima. Nidžo, čežnjivo gledaš fotelju, a Đura i Mika hoće poziciju sebi. Ljudi, jazavac Džef trči po šumi glođući neko suho žbunje. Ljubavi, Olga, hajde pođi u Fudži i čut ćeš nježnu muziku srca. Boja vaše haljine, gospodice Džafić, traži da za nju kulučim.

```

1      identifier 'main'      [LeadingSpace] Loc=<1.c:3:5>
2      l_paren '('          Loc=<1.c:3:9>
3      r_paren ')'          Loc=<1.c:3:10>
4      l_brace '{'          [StartOfLine] Loc=<1.c:4:1>
5      int 'int'            [StartOfLine] [LeadingSpace] Loc=<1.c:5:5>
6      identifier 'a'        [LeadingSpace] Loc=<1.c:5:9>
7      comma ','            Loc=<1.c:5:10>
8      identifier 'b'        [LeadingSpace] Loc=<1.c:5:12>
9      semi ';'             Loc=<1.c:5:13>
10     identifier 'a'        [StartOfLine] [LeadingSpace]
        Loc=<1.c:6:5>
11     equal '='            [LeadingSpace] Loc=<1.c:6:7>
12     identifier 'a'        [LeadingSpace] Loc=<1.c:6:9>
13     plus '+'             [LeadingSpace] Loc=<1.c:6:11>
14     identifier 'c'        [LeadingSpace] Loc=<1.c:6:13>
15     semi ';'             Loc=<1.c:6:14>
16     identifier 'printf'    [StartOfLine] [LeadingSpace]
        Loc=<1.c:7:5>
17     l_paren '('          Loc=<1.c:7:11>
18     string_literal '%"d"'  Loc=<1.c:7:12>
19     comma ','            Loc=<1.c:7:16>
20     identifier 'a'        [LeadingSpace] Loc=<1.c:7:18>
21     r_paren ')'          Loc=<1.c:7:19>
22     semi ';'             Loc=<1.c:7:20>
23     return 'return'       [StartOfLine] [LeadingSpace]
        Loc=<1.c:8:5>
24     numeric_constant '0'   [LeadingSpace] Loc=<1.c:8:12>
25     semi ';'             Loc=<1.c:8:13>
26     r_brace '}'          [StartOfLine] Loc=<1.c:9:1>
27     eof ''               Loc=<1.c:9:2>

```

Slika 2.4: Primer delimične definicije tokena za ime promenljive po standardu C11.

Hadži Đera je začutao i bacio čežnjiv pogled na šolju s kafom. Džabe se zec po Homolju šunja, čuvar Jožef lako će i tu da ga nađe. Odžaćar Filip šalje osmehe tuđoj ženi, a njegova kuća bez dece. Butić Đuro iz Foče ima pun džak ideja o slaganju vaših željica. Džajić odskoči u aut i izbeže don halfa Pecelja i njegov šamar. Plamte odžaci fabrika a čadave guje se iz njih dižu i šalju noć. Ajšo, lepoto i čežnjo, za ljubav srca moga, dođi u Hadžiće na kafu. Hući šuma, a iza žutog džbuna i panja đak u cveću delje seji frulu. Goci i Jaćimu iz Banje Koviljače, flaša džina i žeđ padahu u istu uru. Džaba što Feđa čupa za kosu Milju, ona juri

Živu, ali njega hoće i Daga. Dok je Fehim u džipu žurno ljubio Zagu Čadević, Cile se ušunjao u auto. Fijuće košava nad odžacima a Ilja u gunju žureći uđe u suhu i toplu izbu. Bože, džentlmeni osećaju fizičko gađenje od prljavih šoljica! Dočepaće njega jaka šefica, vođena ljutom srdžbom zlih žena. Pazi, gedžo, brže odnesi šefu taj đavolji ček: njim plaća ceh. Fine džukce ozleđuje bič: odgoj ih pažnjom, strpljivošću. Zamišljao bi kafedžiju vlažnih prstića, crnjeg od čađi. Đaće, uštedu plaćaj žaljenjem zbog džinovskih cifara. Džikljaće žalfija između tog busenja i pešćanih dvoraca. Zašto gđa Hadžić leći živce: njena ljubav pred fijaskom? Jež hoće peckanjem da vredi ljubičastog džina iz flaše. Džej, ljubičast zec, laže: gađaće odmah pokvašen fenjer. Plašljiv zec hoće jeftinu dinju: grožđe iskamči džabe. Džak je pun žica: čućeš tad svađu zbog lomljenja harfe. Čuj, džukac Flop bez daha s gađenjem žvaće stršljena. Oh, zadnji šraf na džipu slab: muž gđe Cvijić ljut koči. Šef džabe zvižduće: mlađi hrt jače kljuca njenog psa. Odbaciće kavgadžija plaštom čađ u željezni fenjer. Deblji krojač: zgužvao smeđ filc u tanjušni džepić. Džo, zgužvaćeš tiho smeđ filc najdeblje krpenjače. Štef, bacih slomljen dečji zvrk u džep gđe Žunjić. Debljoj zgužvao smeđ filc — njen škrt džepčić.

Glava 3

Zaključak

Fijuče vetar u šiblju, ledi pasaže i kuće iza njih i gunđa u odžacima. Nidžo, čežnjivo gledaš fotelju, a Đura i Mika hoće poziciju sebi. Ljudi, jazavac Džef trči po šumi glođući neko suho žbunje. Ljubavi, Olga, hajde pođi u Fudži i čut ćeš nježnu muziku srca. Boja vaše haljine, gospođice Džafić, traži da za nju kulućim. Hadži Đera je začutao i bacio čežnjiv pogled na šolju s kafom. Džabe se zec po Homolju šunja, čuvar Jožef lako će i tu da ga nađe. Odžaćar Filip šalje osmehe tuđoj ženi, a njegova kuća bez dece. Butić Đuro iz Foče ima pun džak ideja o slaganju vaših željica. Džajić odskoči u aut i izbeže đon halfa Pecelja i njegov šamar. Plamte odžaci fabrika a čađave guje se iz njih dižu i šalju noć. Ajšo, lepoto i čežnjo, za ljubav srca moga, dođi u Hadžiće na kafu. Hući šuma, a iza žutog džbuna i panja đak u cveću delje seji frulu. Goci i Jaćimu iz Banje Koviljače, flaša džina i žeđ padahu u istu uru. Džaba što Feđa čupa za kosu Milju, ona juri Živu, ali njega hoće i Daca. Dok je Fehim u džipu žurno ljubio Zagu Čadević, Cile se ušunjao u auto. Fijuče košava nad odžacima a Ilja u gunju žureći uđe u suhu i toplu izbu. Bože, džentlmeni osećaju fizičko gađenje od prljavih šoljica! Dočepaće njega jaka šefica, vođena ljutom srdžbom zlih žena. Pazi, gedžo, brže odnesi šefu taj đavolji ček: njim plaća ceh. Fine džukce ozleđuje bič: odgoj ih pažnjom, strpljivošću. Zamišljao bi kafedžiju vlažnih prstića, crnjeg od čađi. Đaće, uštedu plaćaj žaljenjem zbog džinovskih cifara. Džikljaće žalfija između tog busenja i peščanih dvoraca. Zašto gđa Hadžić leći živce: njena ljubav pred fijaskom? Jež hoće peckanjem da vredi ljubičastog džina iz flaše. Džej, ljubičast zec, laže: gađaće odmah pokvašen fenjer. Plašljiv zec hoće jeftinu dinju: grožđe iskamči džabe. Džak je pun žica: čućeš tad svađu zbog lomljenja harfe. Čuj, džukac Flop bez daha s gađenjem žvaće stršljena. Oh, zadnji šraf na džipu slab:

muž gđe Cvijić ljut koči. Šef džabe zvižduće: mlađi hrt jače kljuca njenog psa. Odbaciće kavgadžija plaštom čađ u željezni fenjer. Deblji krojač: zgužvah smeđ filc u tanjušni džepić. Džo, zgužvaćeš tiho smeđ filc najdeblje krpenjače. Štef, bacih slomljen dečji zvrk u džep gđe Žunjić. Debljoj zgužvah smeđ filc — njen škrt džepčić.

Fijuče vetar u šiblju, ledi pasaže i kuće iza njih i gundā u odžacima. Nidžo, čežnjivo gledaš fotelju, a Đura i Mika hoće poziciju sebi. Ljudi, jazavac Džef trči po šumi glođući neko suho žbunje. Ljubavi, Olga, hajde pođi u Fudži i čut ćeš nježnu muziku srca. Boja vaše haljine, gospođice Džafić, traži da za nju kulućim. Hadži Đera je začutao i bacio čežnjiv pogled na šolju s kafom. Džabe se zec po Homolju šunja, čuvar Jožef lako će i tu da ga nađe. Odžacar Filip šalje osmehe tuđoj ženi, a njegova kuća bez dece. Butić Đuro iz Foče ima pun džak ideja o slaganju vaših željica. Džajić odskoči u aut i izbeže đon halfa Pecelja i njegov šamar. Plamte odžaci fabrika a čađave guje se iz njih dižu i šalju noć. Ajšo, lepoto i čežnjo, za ljubav srca moga, dođi u Hadžiće na kafu. Hući šuma, a iza žutog džbuna i panja đak u cveću delje seji frulu. Goci i Jaćimu iz Banje Koviljače, flaša džina i žeđ padahu u istu uru. Džaba što Feđa čupa za kosu Milju, ona juri Živu, ali njega hoće i Daga. Dok je Fehim u džipu žurno ljubio Zagu Čadević, Cile se ušunjao u auto. Fijuče košava nad odžacima a Ilja u gunju žureći uđe u suhu i toplu izbu. Bože, džentlmeni osećaju fizičko gađenje od prljavih šoljica! Dočepaće njega jaka šefica, vođena ljutom srdžbom zlih žena. Pazi, gedžo, brže odnesi šefu taj đavolji ček: njim plaća ceh. Fine džukce ozleđuje bič: odgoj ih pažnjom, strpljivošću. Zamišljao bi kafedžiju vlažnih prstića, crnjeg od čađi. Đaće, uštedu plaćaj žaljenjem zbog džinovskih cifara. Džikljaće žalfija između tog busenja i peščanih dvoraca. Zašto gđa Hadžić leći živce: njena ljubav pred fijaskom? Jež hoće peckanjem da vređa ljubičastog džina iz flaše. Džej, ljubičast zec, laže: gađaće odmah pokvašen fenjer. Plašljiv zec hoće jeftinu dinju: grožđe iskamči džabe. Džak je pun žica: čućeš tad svađu zbog lomljenja harfe. Čuj, džukac Flop bez daha s gađenjem žvaće stršljena. Oh, zadnji šraf na džipu slab: muž gđe Cvijić ljut koči. Šef džabe zvižduće: mlađi hrt jače kljuca njenog psa. Odbaciće kavgadžija plaštom čađ u željezni fenjer. Deblji krojač: zgužvah smeđ filc u tanjušni džepić. Džo, zgužvaćeš tiho smeđ filc najdeblje krpenjače. Štef, bacih slomljen dečji zvrk u džep gđe Žunjić. Debljoj zgužvah smeđ filc — njen škrt džepčić.

Literatura

- [1] ANTLR4. <https://www.antlr.org/>.
- [2] Context-Free Grammars. https://www.cs.rochester.edu/~nelson/courses/csc_173/grammars/cfg.html.
- [3] William M. Waite and Gerhard Goos. *Compiler Construction*. 1995.

Biografija autora

Ivan Ž. Ristović rođen je 17.01.1995. godine u Užicu. Osnovnu školu, kao i prirodno-matematički smer Užičke gimnazije, završio je kao nosilac Vukove diplome. Tokom navedenog perioda školovanja isticao se u oblastima matematike, informatike, fizike, hemije i engleskog jezika, što potvrđuje veći broj nagrada na Državnim takmičenjima.

Smer Informatika na Matematičkom fakultetu Univerziteta u Beogradu upisuje 2014. godine. Na navedenom smeru je diplomirao 2018. godine, posle tri godine studija sa prosečnom ocenom 9,17. Master studije upisuje na istom fakultetu odmah nakon diplomiranja.

U avgustu 2018. biva izabran u zvanje „Saradnik u nastavi“ na Matematičkom fakultetu paralelno sa master studijama. Drži vežbe iz kurseva „Računarske mreže“, „Funkcionalno programiranje“, „Programske paradigme“ i „Objektno orijentisano programiranje“ na kasnijim godinama osnovnih studija.

Oblasti interesovanja uključuju pre svega razvoj i verifikaciju softvera, mikroservise i računarske mreže.