

HMMY ΕΜΠ – Εαρινό Εξάμηνο 2020
Μεταγλωττιστές
Εξαμηνιαία Εργασία – Αναφορά

Ονοματεπώνυμο: Γιώργος Γκανάς
ΑΜ: 03116095

Ως μέρος της διεξαγωγής του μαθήματος των Μεταγλωττιστών στο εαρινό εξάμηνο του 2020, υλοποιήθηκε μεταγλωττιστής για τη γλώσσα Tony, όπως αυτή περιγράφεται στο ακόλουθο αρχείο: [Περιγραφή Εργασίας και Γλώσσας Tony](#).

Για τον σκοπό αυτό, χρησιμοποιήθηκε η γλώσσα OCaml σε συνδυασμό με τα εργαλεία ocamllex και ocamlyacc, καθώς και LLVM, το εργαλείο μεταγλώττισης ocamlbuild και ο διαχειριστής πακέτων opam. Περαιτέρω έγινε χρήση του OCaml πακέτου Cmdliner. Χρησιμοποιήθηκε επίσης ο βοηθητικός κώδικας που δόθηκε για την υλοποίηση πίνακα συμβόλων και τη διαχείριση σφαλμάτων καθώς και η βιβλιοθήκη χρόνου εκτέλεσης του Αχιλλέα Μπενετόπουλου για την υποστήριξη των έτοιμων συναρτήσεων.

Ο κώδικας είναι οργανωμένος και η εργασία πραγματοποιήθηκε σε τέσσερα βασικά μέρη, συγκεκριμένα τη Λεκτική Ανάλυση, Συντακτική Ανάλυση, Σημασιολογική Ανάλυση και Παραγωγή Ενδιαμέσου Κώδικα. Το τελευταίο στάδιο της μεταγλώττισης (όταν είναι επιθυμητό) πραγματοποιείται από ένα ξεχωριστό σενάριο Bash το οποίο καλεί το μεταγλωττιστή του LLVM (καθώς και το πρόγραμμα OCaml) για αυτόν το σκοπό, διότι τα bindings του LLVM για Ocaml δεν επιτρέπουν κάτι τέτοιο με χρήση της γλώσσας στο ίδιο πρόγραμμα.

Έτσι, τα αρχεία Lexer.mll, Parser.mly, Sem.ml και IRep.ml αντίστοιχα υλοποιούν τα παραπάνω ενώ επίσης σημαντικό είναι το Ast.ml για την επικοινωνία του σημασιολογικού αναλυτή με τα επόμενα στάδια, χτίζοντας σε συνδυασμό με αυτόν το αφηρημένο συντακτικό δέντρο. Το Main.ml συντονίζει τη λειτουργία των παραπάνω, ενώ τα υπόλοιπα αρχεία έχουν βοηθητικό ρόλο προς αυτά. Ο κώδικας του τελικού σεναρίου Bash βρίσκεται αποθηκευμένος στο αρχείο .aux.

Η λειτουργία του βασικού προγράμματος καθορίζεται από το αρχείο Main.ml, το οποίο επίσης με χρήση του Cmdliner module κτίζει τη διεπαφή γραμμής εντολών του προγράμματος και είναι η εξής: καλείται πρώτα ο συντακτικός αναλυτής, ο οποίος με τη σειρά του καλεί τον λεκτικό αναλυτή. Ο δεύτερος με τη χρήση του Lexing module και του εργαλείου ocamllex αναγνωρίζει τις λεκτικές μονάδες του προγράμματος-εισόδου και τις περνάει στον συντακτικό αναλυτή. Αυτός, χρησιμοποιώντας το εργαλείο ocamlyacc ελέγχει αν το πρόγραμμα ανήκει στη γραμματική που ορίζει η γλώσσα και κτίζει το συντακτικό δέντρο. Το δέντρο αυτό ύστερα περνιέται στο σημασιολογικό αναλυτή ο οποίος εκτελεί έλεγχο τύπων και άλλους ελέγχους διαφόρων ειδών (σημασιολογικά σφάλματα). Χρησιμοποιείτε εκεί ο πίνακας

συμβόλων που μας δόθηκε. Όταν επιβεβαιωθεί ότι το πρόγραμμα είναι ορθό, στέλνεται τότε το συντακτικό δέντρο στο επόμενο στάδιο για παραγωγή ενδιαμέσου κώδικα. Ο ενδιαμέσος κώδικας είναι στη γλώσσα LLVM και χρησιμοποιούνται για αυτό τον σκοπό τα bindings του LLVM για OCaml. Επίσης, για την αντιστοίχιση ονομάτων σε τιμές του LLVM γίνεται επαναχρησιμοποίηση του πίνακα συμβόλων με μερικές τροποποιήσεις ώστε να μπορεί να αποθηκεύει και lvalues. Τέλος, παράγεται σε αυτό το στάδιο ο κώδικας του LLVM module που χτίστηκε, είτε σε μορφή κειμένου είτε σε μορφή bytecode. Στη δεύτερη περίπτωση, χρησιμοποιείται ο μεταγλωττιστής llc (μέσα από το σενάριο Bash tony, το οποίο αποτελεί το τελικό εκτελέσιμο) για την περαιτέρω μεταγλώττιση σε assembly.

Πιο συγκεκριμένα χρησιμοποιήθηκαν τα ακόλουθα για την υλοποίηση της εργασίας:

- OCaml 4.02.3
- ocambuild 4.02.3
- ocamlfind 1.5.5
- oram 1.2.2
- LLVM 6.0.1
- Το oram πακέτο llvm (έκδοση 6.0.0, που είναι και η ανώτερη διαθέσιμη έκδοση στο oram 1.2.2)
- Το oram πακέτο cmdliner (έκδοση 1.0.2)
- Το oram πακέτο camlp5 (έκδοση 7.06)
- Bash 4.3.48

Ο μεταγλωττιστής αναπτύχθηκε και ελέγχτηκε σε περιβάλλον Linux Mint 18.3 Sylvia (64-bit).

Για την εγκατάσταση και χρήση του μεταγλωττιστή πρέπει να ακολουθηθούν τα εξής βήματα (αν δεν έχουν γίνει ήδη):

- Εγκατάσταση των OCaml, ocaml-findlib και oram στις παραπάνω εκδόσεις (είναι οι αυτόματες χρησιμοποιώντας το εργαλείο apt στο παραπάνω σύστημα).
- Εγκατάσταση του LLVM 6.0.1 (πακέτο llvm-6.0 καθώς και όσα πακέτα εγκαθίστανται αυτόματα μαζί με αυτό), αφού πρώτα προστεθεί το κατάλληλο αποθετήριο (deb <http://apt.llvm.org/xenial/> llvm-toolchain-xenial-6.0 main) στις πηγές λογισμικού, αν γίνει με χρήση apt.
- Εγκατάσταση των πακέτων llvm, cmdliner, camlp5 με χρήση της εντολής oram install.
- Κατέβασμα και αποσυμπίεση του πηγαίου κώδικα του μεταγλωττιστή
- Τρέχοντας την εντολή make, θα παραχθούν δύο αρχεία το .tonycompiler (OCaml πρόγραμμα) και tony (Bash script).
- Τρέχοντας το σενάριο tony, παράγεται ο ενδιαμέσος ή τελικός κώδικας και με χρήση της εντολής clang -o [OUT] [IN] lib.a (μπορεί να γίνει χρήση και άλλων μεταγλωττιστών όπως gcc), παράγεται και ο εκτελέσιμος κώδικας (μπορεί να παραχθεί και αυτόματα προσθέτοντας την παράμετρο -e στο tony).

Γενικά, το εκτελέσιμο `tony` υποστηρίζει τις παραμέτρους:

- `-O` (προσθέτει βελτιστοποιήσεις)
- `-i` (διαβάζει από `stdin`, γράφει ενδιάμεσο κώδικα στο `stdout`)
- `-f` (διαβάζει από το `stdin`, γράφει τελικό κώδικα στο `stdout`)
- `-e` (παράγει εκτελέσιμο πρόγραμμα με κατάληξη `.out`)