

# **Πανεπιστήμιο Θεσσαλίας**

## **Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών Τηλεπικοινωνιών και Δικτύων**

### **Διπλωματική Εργασία**

**Γραφική αναπαράσταση , επεξεργασία και  
προσομοίωση Αφηρημένου Συντακτικού  
Δέντρου.**

**Σταύρος Αλχατζίδης**

**Επιβλέποντες καθηγητές:  
Γιώργος Δημητρίου,  
Γιώργος Σταμούλης**

**Σεπτέμβριος 2008**



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ  
ΒΙΒΛΙΟΘΗΚΗ & ΚΕΝΤΡΟ ΠΛΗΡΟΦΟΡΗΣΗΣ  
ΕΙΔΙΚΗ ΣΥΛΛΟΓΗ «ΓΚΡΙΖΑ ΒΙΒΛΙΟΓΡΑΦΙΑ»**

Αριθ. Εισ.: 6678/1  
Ημερ. Εισ.: 13-01-2009  
Δωρεά: Συγγραφέα  
Ταξιθετικός Κωδικός: ΠΤ - ΜΗΥΤΔ  
2008  
ΑΛΧ

## Σύντομη περίληψη

Σε αυτή την εργασία παρουσιάζουμε ένα εργαλείο παραγωγής, επεξεργασίας και προσομοίωσης *Ενδιάμεσης Αναπαράστασης (Intermediate Representation)* σε μορφή *Αφηρημένου Συντακτικού Δέντρου (Abstract Syntax Tree - AST)*. Το εργαλείο αποτελείται από τρία κομμάτια: το εμπρόσθιο μέρος, την γραφική αναπαράσταση του συντακτικού δέντρου και τον προσομοιωτή. Το όραμα μας για το εργαλείο είναι η δημιουργία μιας πλατφόρμας μεταγλώττισης και προσομοίωσης σε επίπεδο ΑΣΔ η οποία να προσφέρει αλληλεπίδραση, εποπτικότητα και σχετική, με τα αποτελέσματα, ακρίβεια.

## Abstract

In this project we present a tool which produces, edits and simulates an Intermediate Representation (IR). The IR of preference is the Abstract Syntax Tree. The tool consists of three parts: the frontend, the graphical representation of the AST and the simulator. Our vision for this tool is to become a platform for compilation and simulation at the AST level which can offer interactivity , the better understanding which goes with visualization, and a relative level of accuracy in order for the results to be of use.

## 1. Εισαγωγή

Το *Αφηρημένο Συντακτικό Δέντρο* αποτελεί μια μορφή *Ενδιάμεσης Αναπαράστασης* κατά την μεταγλώττιση κώδικα από μια γλώσσα υψηλού επιπέδου προς τη γλώσσα μηχανής. Αποτελεί ένα σύνηθες παράγωγο του *Εμπρόσθιου Τμήματος* ενός τυπικού μεταγλωττιστή και συμπυκνώνει την χρήσιμη πληροφορία που φέρει ο κώδικας σε μια μορφή δομημένη και συνεπώς εύκολη προς επεξεργασία. Επιπλέον, είναι ικανό να φέρει την πληροφορία με έναν τρόπο ουδέτερο ως προς την γλώσσα υψηλού επιπέδου, εάν μετασχηματιστεί σε μια απομακρυσμένη από τον κώδικα μορφή.

Ως μια τέτοια αφαιρετική αναπαράσταση, το ΑΣΔ αποτελεί μια μορφή της πληροφορίας του κώδικα χρήσιμη σε πολλές εφαρμογές. Εμείς στην παρούσα εργασία έχουμε προσπαθήσει να αποσυνδέσουμε την επεξεργασία του κώδικα από την γλώσσα μηχανής (ή γενικότερα το *Οπίσθιο Τμήμα*) δημιουργώντας ένα εργαλείο επεξεργασίας και προσομοίωσης του κώδικα στο επίπεδο του ΑΣΔ. Μια τέτοια αφαίρεση μπορεί να δώσει αποτελέσματα πιο γρήγορα, πιο γενικά αλλά και πιο εποπτικά για την δομή του κώδικα.

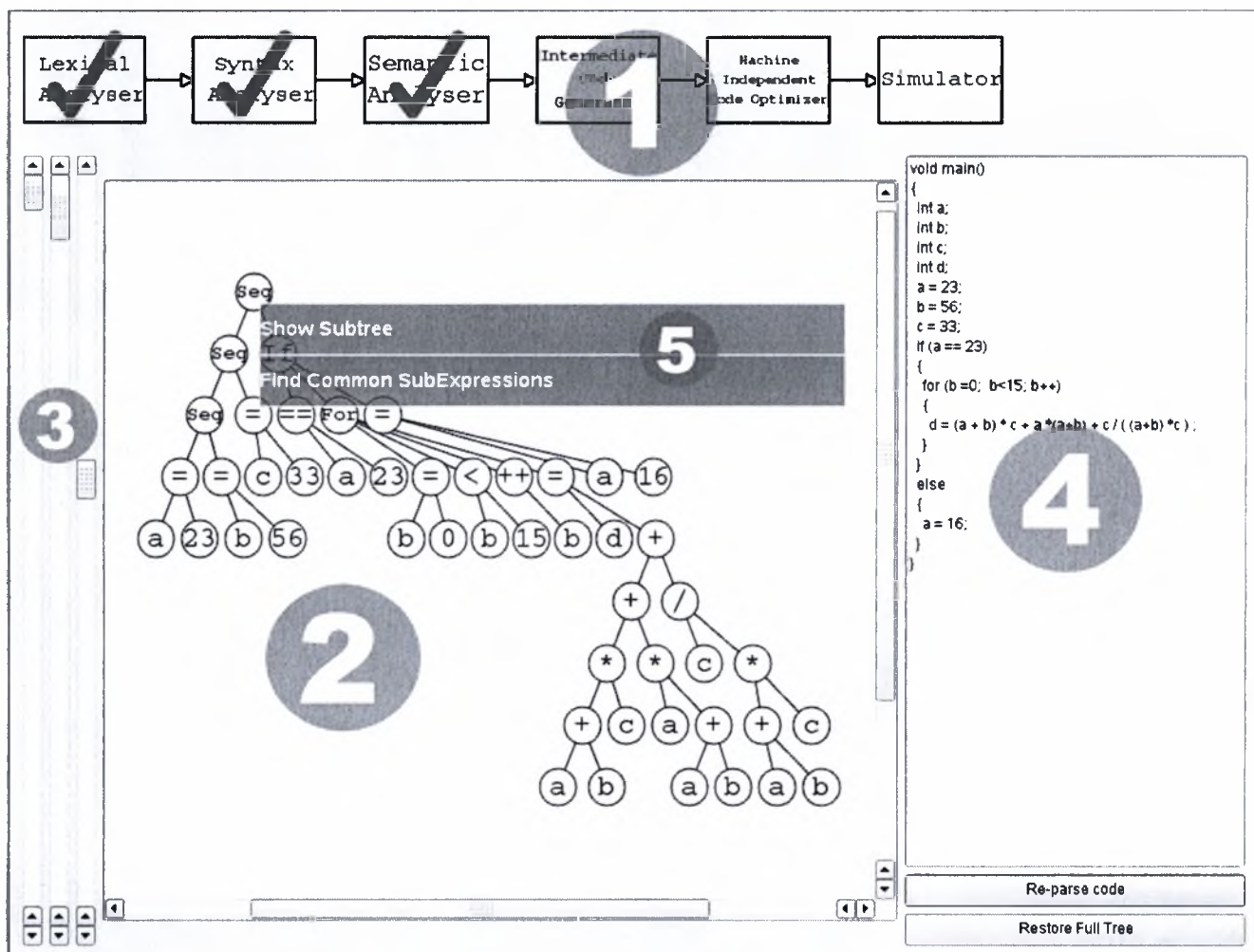
## 2. Περιγραφή του εργαλείου

### 2.1. Το Εμπρόσθιο Μέρος

Το κομμάτι αυτό είναι υπεύθυνο για την λεκτική και συντακτική ανάλυση καθώς και για την δημιουργία της πρώτης μορφής της ενδιάμεσης αναπαράστασης που απεικονίζει το επόμενο μέρος του εργαλείου. Είναι γραμμένο σε γλώσσα C με την καθιερωμένη χρήση και διασύνδεση των συνηθισμένων για αυτό τον σκοπό εργαλείων, *lex/yacc* (εδώ χρησιμοποιούνται στις *open-source* εκδόσεις τους *flex/bison*). Η γλώσσα που αναλύεται είναι ένα υποσύνολο της *c++*.

### 2.2 Το Γραφικό Περιβάλλον

Το γραφικό περιβάλλον έχει προγραμματιστεί σε *C++* με χρήση της βιβλιοθήκης *Qt* στην έκδοση 4.2.1. Το γραφικό περιβάλλον έχει σαν σκοπό να απεικονίσει και να διευκολύνει την γραφική επεξεργασία του Συντακτικού Δέντρου. Αποτελείται από τέσσερα υποπαράθυρα:



Εικόνα 2.1: Ένα στιγμιότυπο του γραφικού περιβάλλοντος. Αριθμούνται τα υποπαράθυρα της εφαρμογής.

### 2.2.1 Η μπάρα προόδου

Σκοπό έχει να βοηθήσει στην συνολική εποπτεία της διαδικασίας ώστε ο χρήστης να καταλαβαίνει το στάδιο στο οποίο βρίσκεται. Το σχήμα είναι σύμφωνο με αυτό που απεικονίζεται στους (Aho κ.α. 2007: 5).

### 2.2.2 Το υποπαράθυρο απεικόνισης του δέντρου

Οι κόμβοι του δέντρου τυπώνονται κατά επίπεδα (level-order διαπέραση της δομής) χρησιμοποιώντας τον εξής αλγόριθμο:



### **2.2.3 Οι κυλιόμενες μπάρες επιλογών προβολής**

Όταν το δέντρο είναι πολύ μεγάλο και οι κοινές υποεκφράσεις είναι διάσπαρτες λόγω ύπαρξης μεταβλητών με μεγάλους χρόνους ζωής, η ελοπτεία του είναι δύσκολη. Σαν βοήθεια προς τον χρήστη υπάρχουν τρεις μπάρες οι οποίες ρυθμίζουν:

1. Το οριζόντιο περιθώριο μεταξύ των κόμβων.
2. Το κάθετο περιθώριο μεταξύ των κόμβων.
3. Το πόσο κοντά θα προβάλλεται το δέντρο (zoom).

### **2.2.4 Το υποπαράθυρο προβολής κώδικα**

Εδώ τυπώνεται ο κώδικας στον οποίο αντιστοιχεί το συντακτικό δέντρο. Ο χρήστης μπορεί να τον μεταβάλλει και πατώντας το “Repars Code” να ξανακαλέσει το εμπρόσθιο τμήμα να παράγει ένα νέο συντακτικό δέντρο.

### **2.2.5 Το μενού προσαρμοσμένων ενεργειών**

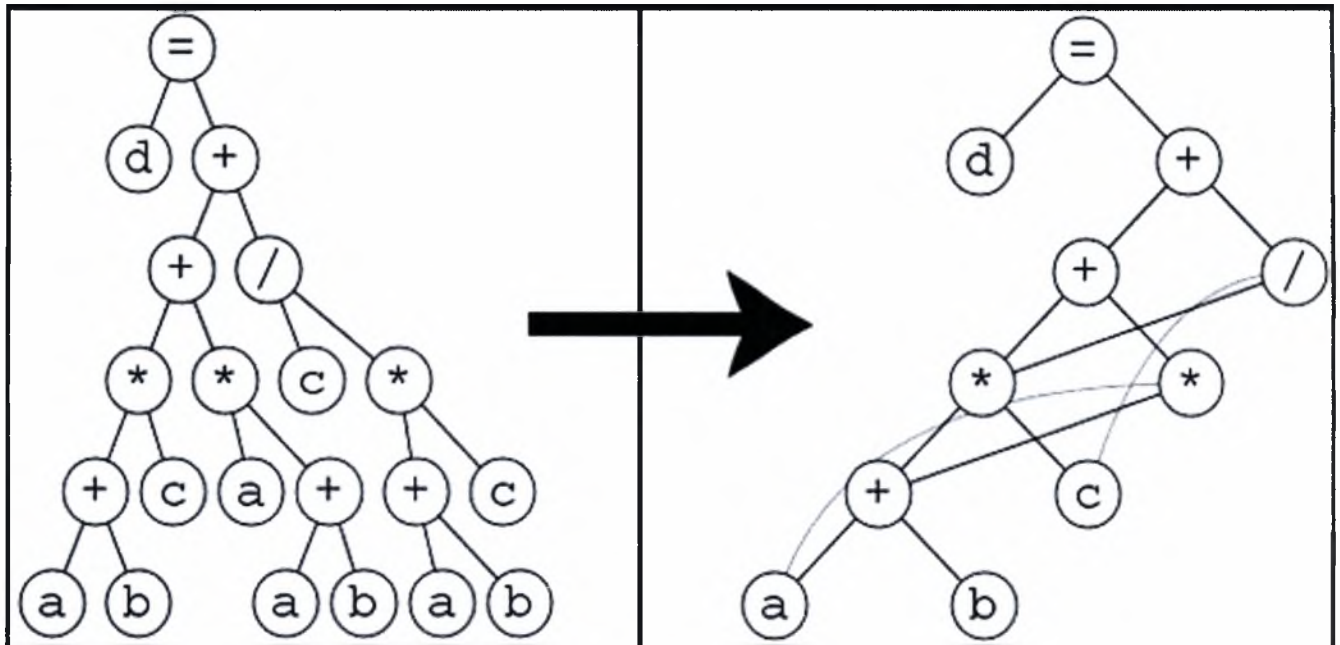
Το μενού αυτό τυπώνεται με δεξί κλικ πάνω σε κάθε κόμβο και εμφανίζει μια σειρά από εντολές για εφαρμογή πάνω στον κόμβο. Κάθε επιλογή εκτελεί μια συνάρτηση callback με όρισμα των κόμβο με χρήση του μηχανισμού signals and slots της βιβλιοθήκης Qt. Αυτό μας δίνει ένα σύστημα με μεγάλη ευελιξία ως προς την προσαρμοσμένη επεξεργασία του ΑΣΔ. Οι εντολές που μπορούν να προστεθούν ποικίλουν: από εντολές για την χειροκίνητη επεξεργασία του δέντρου (πρόσθήκη κόμβου ως παιδιού, αφαίρεση κόμβου, επεξεργασία περιεχομένων κόμβου) μέχρι εφαρμογή σε υποδέντρα γενικών μετασχηματισμών.

## **2.3 Μετασχηματισμοί**

Οι μετασχηματισμοί μεταβάλλουν το δέντρο σε μορφές πιο αποδοτικές προς την εκάστοτε εφαρμογή διατηρώντας όμως πάντα την σημασιολογία του. Συνήθως αμβλύνουν τις συνέπειες από την μεταγλώττιση μιας γλώσσας υψηλού επιπέδου η οποία για λόγους καθαρότητας, ελοπτικότητας ή/και συντήρησης του κώδικα, μπορεί να εισάγει πλεονασμούς στην δομή του δέντρου.

Στην τωρινή μορφή του εργαλείου έχει υλοποιηθεί η μετατροπή του ΑΣΔ σε Κατευθυνόμενο Άκυκλο Γράφο (ΚΑΓ), μέσω της αφαίρεσης των κοινών υποεκφράσεων. Ο μετασχηματισμός αυτός αποτελεί βάση για την εφαρμογή άλλων μετασχηματισμών, αλλά ο λόγος που επιλέχτηκε μεταξύ άλλων είναι ότι μεταβάλλει την γραφοθεωρητική μορφή του ενδιάμεσου κώδικα και επανακαθορίζει συνεπώς τα μέσα της γραφικής του αναπαράστασης.

Όταν το δέντρο έχει μετασχηματιστεί σε ΚΑΓ αφενός δεν αποτελεί πλέον δέντρο, αφετέρου μπορεί ακόμα να τυπωθεί σαν δέντρο λόγω της διατήρησης κάποιων από την πρότερη δομή του και της έλλειψης κύκλων η οποία μας επιτρέπει μια εύκολη ανάγνωση. Δηλαδή, μπορούμε ακόμα να διαβάζουμε το δέντρο μέσω μιας pre-order διαπέρασης απλώς ακολουθώντας κάποιες ακμές που συνδέουν απομακρυσμένους κόμβους. Έτσι (Εικόνα 2.3) ο γράφος τυπώνεται με τον αλγόριθμο που περιγράψαμε στην 2.2.2, ενώ οι επιπλέον ακμές που δημιουργούνται τυπώνονται ως εξής:



*Εικόνα 2.3: Μετασχηματισμός Συντακτικού Δέντρου έκφρασης σε Κατευθυνόμενο Άκυκλο Γράφο. Παρατηρήστε τις καμπύλες bezier που συνδέουν κόμβους που απέχουν πάνω από ένα επίπεδο.*

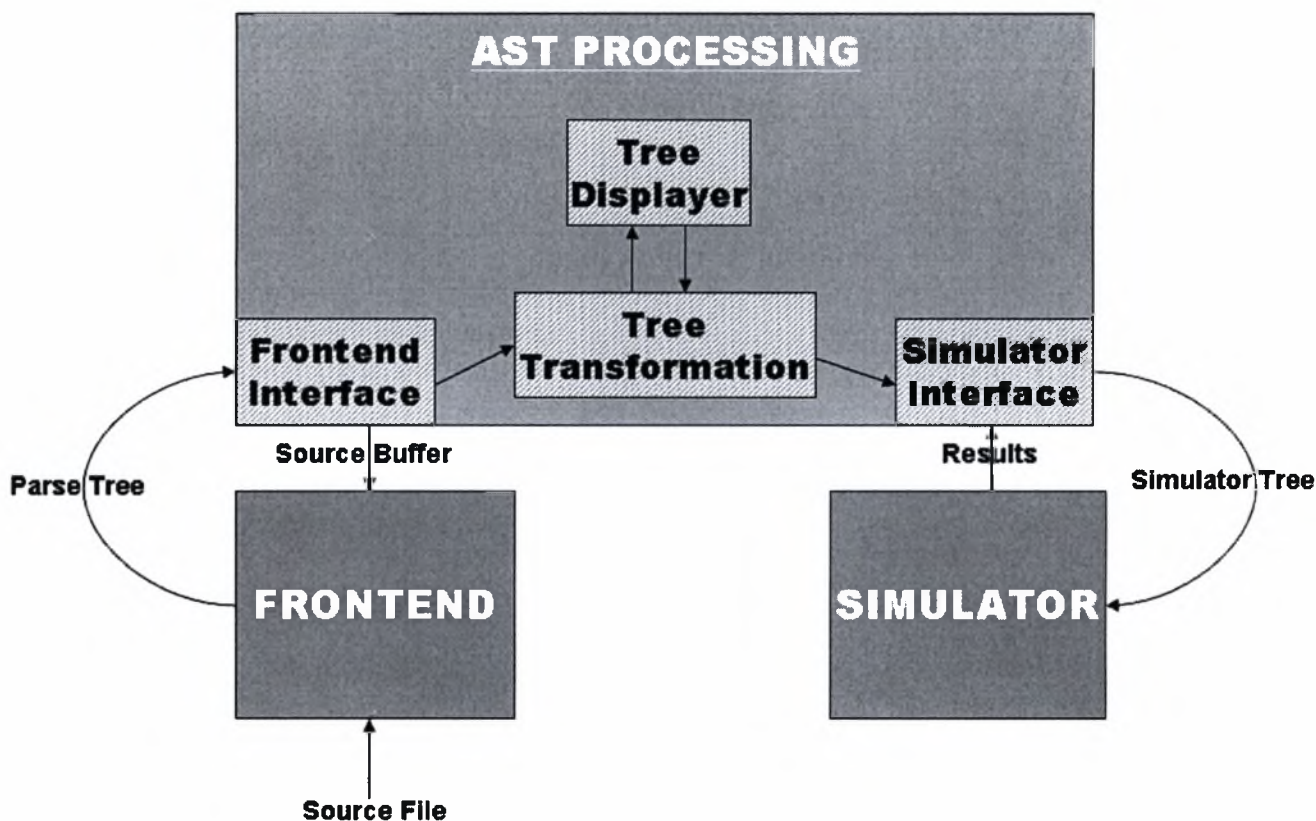
1. Ακμές μεταξύ κόμβων με διαφορά ενός επιπέδου τυπώνονται σαν κανονικές ακμές δέντρου.
2. Ακμές μεταξύ κόμβων με διαφορά άνω του ενός επιπέδου τυπώνονται σαν καμπύλες bezier με χρώμα
  - a. Κόκκινο, εάν πρόκειται για αριστερότερο παιδί
  - b. Πράσινο, εάν πρόκειται για δεξιότερο παιδί
  - c. Διαβαθμίσεις των δύο χρωμάτων εάν πρόκειται για ενδιάμεσο παιδί.

## 2.4 Ο Προσομοιωτής

Ο προσομοιωτής είναι το κομμάτι που θα πρέπει να προσαρμόζεται ώστε να υλοποιεί τις μετρικές της εκάστοτε εφαρμογής. Έτσι π.χ εάν έχουμε κάποιο εργαλείο το οποίο μετράει ισχύ στο επίπεδο του ΑΣΔ θα έπρεπε να έχουμε έναν προσομοιωτή που θα αντιστοιχεί κάποια κατανάλωση σε κάθε εντολή με στατιστικό τρόπο ή θα διασυνδέει τις εντολές με ένα δυναμικό σύστημα προσομοίωσης της κατάστασης του υλικού.

Ο προσομοιωτής που χρησιμοποιείται είναι ο csense και έχει γραφτεί σε γλώσσα C. Στην δική μας περίπτωση η λειτουργικότητα που υλοποιεί (πέραν του ρόλου του στη δομή μας) είναι η επικύρωση της ορθότητας του δέντρου μας μέσω της παραγωγής αποτελεσμάτων.





Εικόνα3.1: Γραφική αναπαράσταση της δομής του κώδικα (Η αγγλική ονοματολογία χρησιμοποιείται για την άμεση αναφορά στις κλάσεις του προγράμματος)

### 3.Περιγραφή του κώδικα

Θα ακολουθήσει μια περιγραφή των κλάσεων του προγράμματος ώστε να γίνουν πιο κατανοητές οι δομές του και η λειτουργία του (Εικόνα 3.1):

#### 3.1 Frontend Interface

Η κλάση αυτή είναι υπεύθυνη για την χρήση του εκάστοτε εμπρόσθιου τμήματος από το πρόγραμμα. Σκοπός της είναι να προμηθεύσει το υπόλοιπο πρόγραμμα με ένα Συντακτικό Δέντρο κατανοητό προς αυτό μετατρέποντας την αντίστοιχη μορφή ενδιάμεσης αναπαράστασης που παράγει το εμπρόσθιο τμήμα. Επίσης πρέπει να μπορεί να προμηθεύει το εμπρόσθιο τμήμα με τον ανανεωμένο κώδικα του υποπαραθύρου 2.2.4.

#### 3.2 Tree Transformation

Η κλάση αυτή διαχειρίζεται τους μετασχηματισμούς και την κατάσταση του καθενός ξεχωριστά αλλά και της συνολικής διαδικασίας του μετασχηματισμού του δέντρου. Οι μέθοδοι των μετασχηματισμών καλούνται όταν ζητηθεί από το γραφικό περιβάλλον και επιστρέφουν το δέντρο στην μετασχηματισμένη του μορφή.

### 3.3 Tree Display

Η κλάση αυτή διαχειρίζεται την απεικόνιση του δέντρου και επικοινωνεί με το γραφικό περιβάλλον για την ανανέωση ή αλλαγή της.

### 3.4 Simulator Interface

Η κλάση αυτή είναι υπεύθυνη για την χρήση του προσομοιωτή από το πρόγραμμα. Πρέπει να μπορεί να μεταβάλλει την δομή του ΑΣΔ που διαχειρίζεται εσωτερικά το πρόγραμμα σε μια μορφή κατανοητή από τον προσομοιωτή. Επίσης πρέπει να μπορεί να λαμβάνει τα αποτελέσματα της προσομοίωσης και να τα επιστρέφει στο πρόγραμμα.

## 4.Λειτουργικότητα

Η εργασία σαν σκοπό έχει να καταδείξει και να αναδείξει, όσο το δυνατόν περισσότερο, μια δομή επεξεργασίας και προσομοίωσης στο επίπεδο του ΑΣΔ. Δυστυχώς, ο χρόνος και η εμβέλεια της δεν επέτρεψαν περισσότερη λειτουργικότητα. Συνεπώς, ο παραδοτέος κώδικας πρέπει να αντιμετωπιστεί ως μια εργασία σε εξέλιξη παρά ως ένα τελειωμένο πρόγραμμα. Ορισμένοι περιορισμοί στην λειτουργικότητα τους οποίους θα πρέπει να γνωρίζει ο μελλοντικός ασχολούμενος είναι:

1. Ενώ το εμπρόσθιο τμήμα είναι γραμμένο για μια αντικειμενοστραφή γλώσσα, λαμβάνουμε μόνο το υποδέντρο της συνάρτησης `main()` προς περαιτέρω επεξεργασία
2. Το `csense` είναι ένας πλήρης προσομοιωτής της γλώσσας `c`, σε επίπεδο ΑΣΔ, όμως χρησιμοποιούμε μόνο τις δυνατότητες της γλώσσας που έχουμε εντάξει στο δεύτερο τμήμα επεξεργασίας του ΑΣΔ.
3. Το `csense` δεν υποστηρίζει την προσομοίωση Κατευθυνόμενων Άκυκλων Γράφων και έτσι προσομοιώνει μια προηγούμενη μορφή του ΑΣΔ.

## 5.Παρατηρήσεις

1. Για την παραγωγή εγγράφων τεκμηρίωσης του κώδικα σε HTML χρησιμοποιήθηκε το πρόγραμμα ***doxygen***.
2. Ο σχολιασμός του κώδικα έχει γίνει στα **αγγλικά**.
3. Ο κώδικας του εμπρόσθιου μέρους βασίζεται σε εργασία των **Αλχατζίδη Σ., Καρανικολόπουλου Γ., Κουκά Ι., Παναγόπουλου Γ.**

## ΠΑΡΑΡΤΗΜΑ – Εργαλεία που χρησιμοποιήθηκαν

### A. Doxygen



Το doxygen είναι ένα open-source σύστημα τεκμηρίωσης που υποστηρίζει ένα μεγάλο αριθμό από γλώσσες. Μπορεί να δημιουργήσει αρχεία τεκμηρίωσης προσαρμοσμένα για πολλά και διαφορετικά πρότυπα αρχείων. Λειτουργεί κατά τα πρότυπα άλλων αντίστοιχων εργαλείων (το πιο γνωστό το Javadoc της Sun) δημιουργώντας τα αρχεία τεκμηρίωσης από τα αρχεία του κώδικα, μετατρέποντας έτσι την τεκμηρίωση σε μια διαδικασία συνδεδεμένη με την συντήρησή και αναβάθμιση του κώδικα (θέση που πρέπει να έχει η τεκμηρίωση σε κάθε σύγχρονο project).

Υποστηρίζει από πολύ απλά σχολιασμένο κώδικα μέχρι την χρήση ειδικών εντολών για την καλύτερη απεικόνιση της τεκμηρίωσης στα διάφορα πρότυπα. Δημιουργεί αυτόματα διαγράμματα UML, μπορεί να περιέχει μια μηχανή αναζήτησης στην HTML μορφή, ενώ αναγνωρίζει πλήρως της επιπλέον δομές της βιβλιοθήκης Qt (κάτι που αποδείχτηκε χρήσιμο για την εργασία μας).

### B. Βιβλιοθήκη Qt



Η Qt είναι μια βιβλιοθήκη που διευκολύνει την ανάπτυξη γραφικών περιβαλλόντων (widget library). Είναι φτιαγμένη από την Trolltech Inc. με ένα ιδιότυπο καθεστώς δικαιωμάτων (ανοιχτού-κώδικα για την χρήση στην ανάπτυξη λογισμικού ανοιχτού-κώδικα, με προστασία δικαιωμάτων για χρήση στην ανάπτυξη λογισμικού κλειστού-κώδικα). Έχει κυκλοφορήσει σε εκδόσεις που την ενσωματώνουν σε διάφορες γλώσσες και λειτουργικά συστήματα.

Αν και φέρει τον τίτλο της widget library, η βιβλιοθήκη της Qt αποτελεί ένα σύστημα με περισσότερη λειτουργικότητα. Για την έκδοση της σε C++ χρησιμοποιεί έναν σύνθεστο προεπεξεργαστή, τον οποίο ονομάζει moc (Meta-Object Compiler), για την επέκταση της γλώσσας. Συγκεκριμένα μέσω του προεπεξεργαστή προσθέτει στην C++ μεταξύ άλλων :

1. Συναρτήσεις callback με έλεγχο τύπων.
2. Ένα σύστημα ιδιοτήτων αντικειμένων μη εξαρτώμενο από τον μεταγλωττιστή.
3. Ενσωμάτωση στον κώδικα της μετάφρασης του κειμένου του προγράμματος σε άλλες γλώσσες (internationalization).
4. Ένα νέο τρόπο μετατροπής τύπων (dynamic\_cast) για δυναμικά φορτωμένες βιβλιοθήκες.

## Βιβλιογραφία

- [1] Μποζάνης Π.Δ. (2003), *Δομές Δεδομένων : Ταξινόμηση και Αναζήτηση με Java*, Τζιόλα
- [2] Παπασπύρου Ν.Σ., Σκορδαλάκης Ε.Σ. (2002), *Μεταγλωττιστές, Συμμετρία*
- [3] Aho A.V., Lam M.S., Sethi R. , Ullman J.D. (2007) *Compilers: Principles, Techniques and Tools 2<sup>nd</sup> Edition*, Pearson International Edition
- [4] *Doxygen Manual*, <http://www.stack.nl/~dimitri/doxygen/manual.html>
- [5] *Qt Reference Documentation (Open Source Edition)*, <http://doc.trolltech.com/4.2.2/>

