ΜΕΤΑΦΡΑΣΤΗΣ CIMPLE

ΜΥΥ802 Μεταφραστές

1. Εισαγωγή

Αυτή η αναφορά έχει σκοπό να περιγράψει την ανάπτυξη, σε γλώσσα Python, ενός μεταφραστή για την γλώσσα προγραμματισμού *cimple*.

Η γλώσσα προγραμματισμού *cimple* είναι μια εκπαιδευτική γλώσσα προγραμματισμού που περιγράφεται στο έγγραφο «Η Γλώσσα Προγραμματισμού *cimple*».

Ο μεταφραστής που αναπτύχθηκε περιλαμβάνει τα εξής επιμέρους εργαλεία:

- Αναγνώστης κειμένου (file parser).
- Λεκτικός αναλυτής.
- Συντακτικός αναλυτής.
- Εργαλείο παραγωγής ενδιάμεσου κώδικα.
- Εργαλείο παραγωγής πίνακα συμβόλων.
- Εργαλείο παραγωγής τελικού κώδικα σε assembly.

Η λειτουργικότητα του μεταφραστή της *cimple* περιγράφεται εν συντομία στα επόμενα βήματα:

- Ο μεταφραστής διαβάζει ένα αρχείο κειμένου με πηγαίο κώδικα της γλώσσας *cimple*.
- Διατρέχει τον πηγαίο κώδικα και πραγματοποιεί λεκτική και συντακτική ανάλυση.
- Παράλληλα παράγει τον ενδιάμεσο κώδικα για κάθε μια από τις απαραίτητες εντολές πηγαίου κώδικα. Ο ενδιάμεσος κώδικας αποθηκεύεται στο αρχείο test. int.
- Δημιουργεί και κρατά στην μνήμη τον πίνακα συμβόλων για κάθε περιοχή (scope) του κώδικα. Για λόγους από-σφαλμάτωσης (debugging), ο πίνακας συμβόλων αποθηκεύεται στο αρχείο test. symb.
- Για λόγους από-σφαλμάτωσης, ο μεταφραστής παράγει από τον ενδιάμεσο κώδικα, ισοδύναμο πρόγραμμα σε γλώσσα c. Αποθηκεύεται στο αρχείο test. c.
- Τέλος, ο μεταφραστής παράγει τον τελικό κώδικα σε assembly για κάθε περιοχή (scope) ξεχωριστά όταν ολοκληρώνεται η αντίστοιχη μετάφραση της περιοχής. Ο τελικός κώδικας βασίζεται στον ενδιάμεσο κώδικα και στον πίνακα συμβόλων. Ο τελικός κώδικας αποθηκεύεται στο αρχείο test. asm.

Ομάδα Ανάπτυξης

Η ομάδα ανάπτυξης του μεταφραστή αποτελείται από τους εξής προγραμματιστές:

• Παναγιώτης Κολοκούρης, 4914

Δομή Αναφοράς

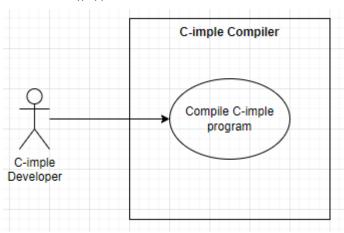
Η παρούσα αναφορά ακολουθεί την εξής δομής

- Κεφάλαιο 1. Εισαγωγή
- Κεφάλαιο 2. Περιγράφει το use case που ικανοποιεί ο μεταφραστής.
- Κεφάλαιο 3. Περιγράφει την ανάπτυξη του μεταφραστή.

2. Use Cases

Η παρούσα ενότητα περιγράφει το use case που καλείται να ικανοποιήσει ο μεταφραστής, ήτοι την μεταγλώττιση ενός προγράμματος cimple. Στην Εικόνα 2.1 παρουσιάζεται το διάγραμμα των use cases.

Εικόνα 2.1: Διάγραμμα Use Cases



Το ροή του use case παρουσιάζεται στον Πίνακας 2.1.

Πίνακας 2.1: Use Case - Μετάφραση Προγράμματος Cimple

Use case ID	UC1	
Actors	Cimple Developer	
Preconditions	n/a	
Main flow of events	 Ο προγραμματιστής καλεί από την γραμμή εντολών τον μεταφραστή παρέχοντας σαν όρισμα το όνομα του αρχείου, σε γλώσσα cimple, προς μετάφραση. Το σύστημα εκτελεί τον μεταφραστή. Το σύστημα εμφανίζει μήνυμα επιτυχίας στον προγραμματιστή. Το σύστημα παράγει και αποθηκεύει στον δίσκο (στην τοποθεσία από την οποία κλήθηκε) τα αρχεία test. int, test. symb, test. c και test. asm. 	
Alternative flow	Η μετάφραση αποτυγχάνει και το σύστημα εμφανίζει κατάλληλο	
	πληροφοριακό μήνυμα στον προγραμματιστή με την πιθανή αιτία του	
	λόγου αποτυχίας.	
Post conditions	n/a	

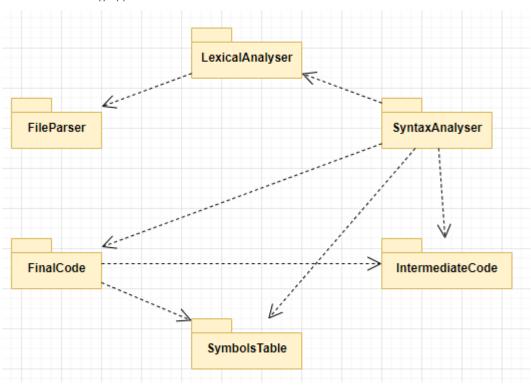
3. Ανάπτυξη Μεταφραστή

Η παρούσα ενότητα παρουσιάζει την λεπτομερή ανάπτυξη του μεταφραστή.

Αρχιτεκτονική

Η αρχιτεκτονική του μεταφραστή ως προς τα διακριτά πακέτα/εργαλεία παρουσιάζεται στην Εικόνα 3.1.

Εικόνα 3.1: Uml Διάγραμμα Πακέτων



Το κάθε πακέτο περιγράφεται αναλυτικά στις επόμενες ενότητες.

Αναγνώστης Κειμένου

Το διάγραμμα κλάσεων για το πακέτο του αναγνώστη κειμένου παρουσιάζεται στην Εικόνα 3.2. Απαρτίζεται μόνο από την κλάση FileParser.

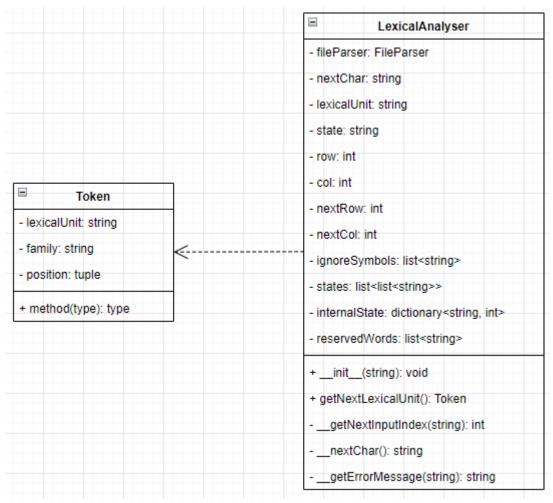
Εικόνα 3.2: Διάγραμμα κλάσεων για πακέτο αναγνώστη κειμένου

FileParser		
- filename: string		
- file: file stream		
+init(filename)		
+ getNextCharacter(): string		
del(): void		

Λεκτικός Αναλυτής

Το διάγραμμα κλάσεων για το πακέτο του λεκτικού αναλυτή παρουσιάζεται στην Εικόνα 3.3.

Εικόνα 3.3: Διάγραμμα κλάσεων για πακέτο λεκτικού αναλυτή



Συντακτικός Αναλυτής

Το διάγραμμα κλάσεων για το πακέτο του συντακτικού αναλυτή παρουσιάζεται στην Εικόνα 3.4.

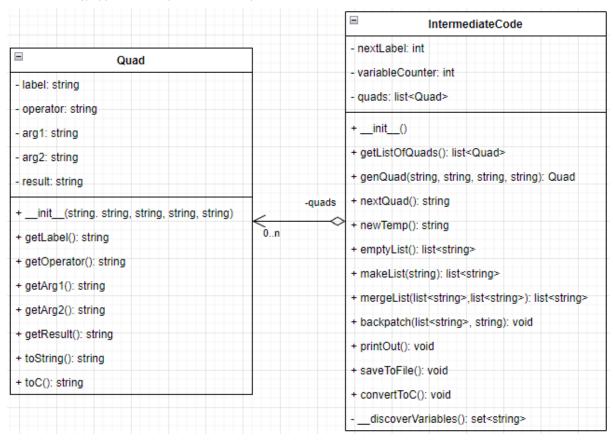
Εικόνα 3.4: Διάγραμμα κλάσεων για πακέτο συντακτικού αναλυτή

	SyntaxAnalyser
- filename	string
- lexicalA	nalyser: LexicalAnalyser
- token: T	oken
- REL_O	P: list <string></string>
- ADD_O	P: list <string></string>
- MUL_O	P: list <string></string>
- interme	diateCode: IntermediateCode
- symbols	Table: SymbolsTable
	le: FinalCode
	(string)
+ run(): v	oid
getNe	extToken(): void
progr	am(): void
progr	amBlock(string): void
decla	rations(): void
varlis	t(): void
subp	ograms(): void
subp	ogram(string): void
forma	lparlist(): void
forma	lparitem(): void
state	ments(): void
block	statements(): void
state	ment(): void
assig	nStat(): void
ifStat	(): void
elsep	art(): void
while	Stat(): void
switc	hcaseStat(): void
- forca	seStat(): void
	eStat(): void
	Stat(): void
	at(): void
	Stat(): void
	Stat(): void
	lparlist(): void
	Iparitem(): string, string
_	tion(): list <string>, list<string></string></string>
	erm(): list <string>, list<string></string></string>
-	actor(): list <string>, list<string></string></string>
	ssion(): string
): string
	(): string
idtail(string): string

Ενδιάμεσος Κώδικας

Το διάγραμμα κλάσεων για το πακέτο του ενδιάμεσου κώδικα παρουσιάζεται στην Εικόνα 3.5.

Εικόνα 3.5: Διάγραμμα κλάσεων για πακέτο ενδιάμεσου κώδικα



Πίνακας Συμβόλων

Το διάγραμμα κλάσεων για το πακέτο του πίνακα συμβόλων παρουσιάζεται στην Εικόνα 3.6.

SymbolsTable - scopes: list<Scope> scopeStrings: list<string> nestingLevel: int mainFrameLength: int - entities: list<Entity> offset: int + __init__() getTopScopeSize(): int + __init__(string, int) + addScope(string): void + getSize(): int + removeLastScope(): void + addProcedure(string): void + addProcedure(string): void + addFunction(string, string): void + addFunction(string, string): void + appendFormalParameter(Entity): void + addParameter(string, string, string): void + getLastFormalParameter(): Entity addVariable(string): void + addParameter(string, string, string): void + addTemporaryVariable(string): void + addVariable(string): void appendFormalParameterToCaller(): void + addTemporaryVariable(string): void + fillInStartingQuad(string): void + fillInStartingQuad(string): void + fillInFrameLength(int); void + setMainFrameLength(int): void + toString(): string + isVariableInScope(string): bool + getOffset(string); int printScopeTrace(): void + getPassingMode(string); string Entity + saveScopeTraceToFile(): void + getFrameLength(string): int + getStartingQuad(string): string + __init__(string) _getEntityInScope(string): Entity + toString(): string r isVariableInHigherScopeLowerThanMain(string): bool getOffsetFromMain(string): int + getOffset(): int egetStartingQuadFromMain(string): string getOffsetFromCurrentScope(string): int + getStartingQuad(): string getPassingModeFromCurrentScope(string): string getFrameLengthFromCurrentScope(string): intgetStartingQuadFromCurrentScope(string): string Variable FormalParameter Procedure getNumberOfLevelsAbove(string): int - datatype: string - datatype: string - startingQuad; string getOffsetFromHigherScope(string): int - offset: int mode: string frameLength: int getPassingModeFromHigherScope(string): string formalParameters: list<FormalParameter> + getFrameLengthFromHigherScope(string): int + __init__(string, string, int) + __init__(string, string, string) + init (string) + toString(); string + toString(); string + getStartingQuadFromHigherScope(string); string + getOffset(): int + toString2(): string + setStartingQuad(string): void isMainScopeCurrent(): bool + getPassingMode(): string + setFrameLength(int): void __getHigherScope(string): Scope + addFormalParameter(FormalParameter): void + toString(): string TemporaryVariable Parameter + getFrameLength(): int __init__(string, string, int) - offset: int + getStartingQuad(): string toString(): string + __init__(string, string, int) + getOffset(): int + toString(): string Function

+ toString2(): string

+ getOffset(): int

datatype: string

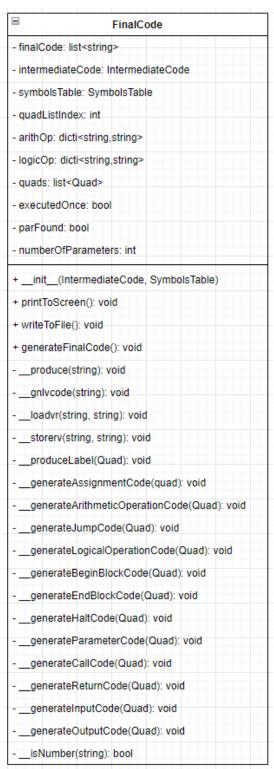
+ __init__(string, string)
+ toString(): string

Εικόνα 3.6: Διάγραμμα κλάσεων για πακέτο πίνακα συμβόλων

Τελικός Κώδικας

Το διάγραμμα κλάσεων για το πακέτο του τελικού κώδικα παρουσιάζεται στην Εικόνα 3.7.

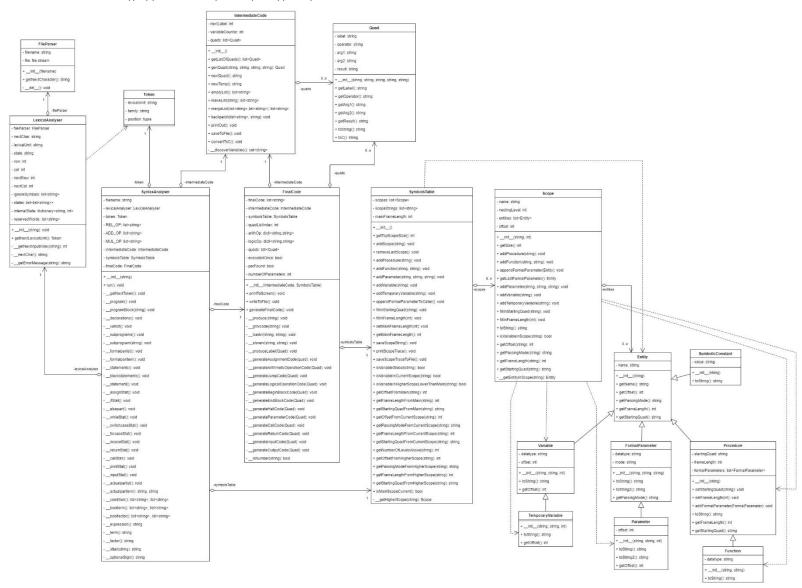
Εικόνα 3.7: Διάγραμμα κλάσεων για πακέτο τελικού κώδικα



Συνολικά

Τέλος, η συνολική αρχιτεκτονική του μεταφραστή περιγράφεται από το συγκεντρωτικό διάγραμμα κλάσεων στην Εικόνα 3.8.

Εικόνα 3.8: Συνολικό διάγραμμα κλάσεων για τον μεταφραστή



CRC Cards

Οι Python κλάσεις του μεταφραστή περιγράφονται με περισσότερη λεπτομέρεια στις παρακάτω ενότητες για κάθε πακέτο της αρχιτεκτονικής.

Αναγνώστης Κειμένου

Οι κλάση του πακέτου αναγνώστη κειμένου περιγράφεται στον Πίνακας 3.1.

Πίνακας 3.1: CRC card για την κλάση FileParser

Όνομα κλάσης: FileParser		
Αρμοδιότητα:	Εξαρτήσεις:	
 Άνοιγμα αρχείου πηγαίου κώδικα. 	καμία	
 Ανάγνωση αρχείου χαρακτήρα- χαρακτήρα. 		

Λεκτικός Αναλυτής

Οι κλάσεις του πακέτου λεκτικός αναλυτής καταγράφονται στους: Πίνακας 3.2 και Πίνακας 3.3.

Πίνακας 3.2: CRC card για την κλάση Token

Όνομα κλάσης: Token		
Αρμοδιότητα:		Εξαρτήσεις:
•	Δομή αποθήκευσης πληροφορίας για κάθε λεκτική μονάδα.	• καμία

Πίνακας 3.3: CRC card για την κλάση LexicalAnalyser

Όνομα κλάσης: LexicalAnalyser	
Αρμοδιότητα:	Εξαρτήσεις:
 Χρησιμοποιεί τον αναγνώστη κειμένου για να προσπελάσει το αρχείο πηγαίου κώδικα. 	FileParserToken
 Αναγνωρίζει και επιστρέφει την κάθε λεκτική μονάδα. 	
 Παράγει κατάλληλα πληροφοριακά μηνύματα και σταματάει την εκτέλεση όταν ανακαλύψει λεκτική μονάδα που δεν πληροί τους κανόνες της γλώσσας cimple. 	

Συντακτικός Αναλυτής

Οι κλάση του πακέτου συντακτικός αναλυτής παρουσιάζεται στον Πίνακας 3.4.

Πίνακας 3.4: CRC card για την κλάση SyntaxAnalyser

Όνομα κλάσης: SyntaxAnalyser

Αρμοδιότητα:

- Ελέγχει τον πηγαίο κώδικα για κάθε λεκτική μονάδα την φορά για να διαπιστώσει αν το πρόγραμμα αντιστοιχεί στην γλώσσα cimple.
- Καλεί μεθόδους της κλάσης του ενδιάμεσου κώδικα σε κατάλληλα σημεία ώστε να δημιουργήσει τις τετράδες του ενδιάμεσου κώδικα.
- Καλεί μεθόδους της κλάσης του πίνακα συμβόλων σε κατάλληλα σημεία ώστε να δημιουργήσει και να κρατά ενημερωμένο τον πίνακα συμβόλων ανάλογα με το σημείο της μετάφρασης.
- Καλεί μεθόδους της κλάσης του τελικού κώδικα σε κατάλληλα σημεία ώστε να δημιουργήσει τον τελικό κώδικα.

Εξαρτήσεις:

- LexicalAnalyser
- IntermediateCode
- SymbolsTable
- FinalCode

Ενδιάμεσος Κώδικας

Οι κλάσεις του πακέτου ενδιάμεσου κώδικα καταγράφονται στους: Πίνακας 3.5 και Πίνακας 3.6.

Πίνακας 3.5: CRC card για την κλάση Quad

Όνομα κλάσης: Quad		
Αρμοδιότητα:	Εξαρτήσεις:	
 Δομή αποθήκευσης πληροφορίας για κάθε τετράδα του ενδιάμεσου κώδικα. 	καμία	

Πίνακας 3.6: CRC card για την κλάση IntermediateCode

Όνομα κλάσης: IntermediateCode		
Αρμοδιότητα:	Εξαρτήσεις:	
 Παράγει τις τετράδες του ενδιάμεσου κώδικα παράλληλα με την συντακτική ανάλυση. 	Quad	
 Παρέχει μεθόδους για την εμφάνιση στην οθόνη αλλά και αποθήκευση στον δίσκο του τελικού κώδικα. 		
 Παρέχει μέθοδο για την μετατροπή του συνόλου των τετράδων σε ισοδύναμο πρόγραμμα c. 		

Πίνακας Συμβόλων

Οι κλάσεις του πακέτου πίνακα συμβόλων καταγράφονται στους: Πίνακας 3.7 έως Πίνακας 3.16.

Πίνακας 3.7: CRC card για την κλάση SymbolsTable.

Όνομα κλάσης: SymbolsTable	
Αρμοδιότητα:	Εξαρτήσεις:
 Δημιουργεί τον πίνακα συμβόλων υπό την μορφή μιας συλλογής περιοχών (scopes). 	ScopeEntity
 Κρατάει τον πίνακα συμβόλων ενημερωμένο ανάλογα με το σημείο της μετάφρασης του προγράμματος cimple. 	
 Παρέχει διεπαφή (ΑΡΙ) για τις κλάσεις του συντακτικού αναλυτή και τελικού κώδικα ώστε να πραγματοποιείται επικοινωνία με τις οντότητες του πίνακα συμβόλων. 	
 Παρέχει μεθόδους για την εμφάνιση στην οθόνη αλλά και αποθήκευση στον δίσκο του τελικού κώδικα. 	

Πίνακας 3.8: CRC card για την κλάση Scope.

Όνομα κλάσης: Scope Αρμοδιότητα: Εξαρτήσεις: Αποθηκεύει τις υπάρχουσες οντότητες **Entity** για κάθε περιοχή (scope) του υπό Variable μετάφραση προγράμματος. TemporaryVariable Παρέχει διεπαφή (ΑΡΙ) για την FormalParameter επικοινωνία με τις οντότητες. Parameter Procedure Function Symbolic Constant

Πίνακας 3.9: CRC card για την κλάση Entity.

Όνομα κλάσης: Entity		
Αρμοδιότητα:	Εξαρτήσεις:	
 Γενική (abstract) δομή αποθήκευσης πληροφορίας για κάθε οντότητα του πίνακα συμβόλων. 	• καμία	
 Αποτελεί την κορυφή στο δέντρο της ιεραρχίας των κλάσεων των υπολοίπων οντοτήτων. 		

Πίνακας 3.10: CRC card για την κλάση Variable.

Όνομα κλάσης: Variable	
Αρμοδιότητα:	Εξαρτήσεις:
 Δομή αποθήκευσης πληροφορίας για κάθε μεταβλητή του προγράμματος cimple που εισάγεται στον πίνακα συμβόλων. 	■ Entity

Πίνακας 3.11: CRC card για την κλάση TemporaryVariable.

Όνομα κλάσης: TemporaryVariable	
Αρμοδιότητα:	Εξαρτήσεις:
 Δομή αποθήκευσης πληροφορίας για κάθε προσωρινή μεταβλητή του προγράμματος cimple που εισάγεται στον πίνακα συμβόλων. 	 Variable

Πίνακας 3.12: CRC card για την κλάση FormalParameter.

Αρμοδιότητα:	Εξαρτήσεις:
 Δομή αποθήκευσης πληροφορίας για κάθε παράμετρο (κατά την δήλωση συνάρτησης) του προγράμματος cimple που εισάγεται στον πίνακα συμβόλων. 	■ Entity

Πίνακας 3.13: CRC card για την κλάση Parameter.

Όνομα κλάσης: Parameter		
Αρμοδιότητα:	Εξαρτήσεις:	
 Δομή αποθήκευσης πληροφορίας για κάθε παράμετρο (κατά την κλήση συνάρτησης) του προγράμματος cimple που εισάγεται στον πίνακα συμβόλων. 	■ FormalParameter	

Πίνακας 3.14: CRC card για την κλάση Procedure.

Όνομα κλάσης: Procedure		
Αρμοδιότητα:	Εξαρτήσεις:	
 Δομή αποθήκευσης πληροφορίας για κάθε διαδικασία του προγράμματος cimple που εισάγεται στον πίνακα συμβόλων. 	■ Entity	

Πίνακας 3.15: CRC card για την κλάση Function.

Όνομα κλάσης: Function		
Αρμοδιότητα:	Εξαρτήσεις:	
 Δομή αποθήκευσης πληροφορίας για κάθε συνάρτηση του προγράμματος cimple που εισάγεται στον πίνακα συμβόλων. 	■ Procedure	

Πίνακας 3.16: CRC card για την κλάση SymbolicConstant.

Όνομα κλάσης: SymbolicConstant		
Αρμοδιότητα:	Εξαρτήσεις:	
 Δομή αποθήκευσης πληροφορίας για κάθε σταθερά του προγράμματος cimple που εισάγεται στον πίνακα συμβόλων. (Στην cimple δεν υπάρχουν συμβολικές σταθερές) 	■ Entity	

Τελικός Κώδικας

Οι κλάση του πακέτου τελικού κώδικα περιγράφεται στον Πίνακας 3.17.

Πίνακας 3.17: CRC card για την κλάση FinalCode.

Όνομα κλάσης: FinalCode Αρμοδιότητα: Εξαρτήσεις: Παράγει εντολές τελικού κώδικα για IntermediateCode μια σειρά από τετράδες ενδιάμεσου SymbolsTable κώδικα που αντιστοιχούν στην πλήρη Quad μετάφραση μιας περιοχής (scope). Χρησιμοποιεί τον πίνακα συμβόλων για να αντλεί πληροφορίες για τις οντότητες και τις περιοχές (scopes) που αυτές ανήκουν. Παρέχει μεθόδους για την εμφάνιση στην οθόνη αλλά και αποθήκευση στον δίσκο του τελικού κώδικα.