



**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ**

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**Διαδικτυακό Περιβάλλον Στατιστικής Μορφοσυντακτικής  
Αναγνώρισης των Μονολεκτικών Ρηματικών Τύπων της  
Νέας Ελληνικής «VERBTAGGR++»**

**Ιωάννα Σ. Ζέρβα  
Βασιλική Ν. Κωλέτση  
Νικολέτα Γ. Μωραγιάννη**

**Επιβλέπουσες:**

**Αφροδίτη Τσαλαγατίδου, Αναπληρώτρια Καθηγήτρια ΕΚΠΑ  
Μαρία Γρηγοριάδου, Ομότιμη Καθηγήτρια ΕΚΠΑ**

**Συνεπιβλέπουσα:**

**Πηνελόπη Λεμπέση, Επιστημονική Συνεργάτις ΕΚΠΑ**

**ΑΘΗΝΑ  
ΜΑΡΤΙΟΣ 2016**

## **ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

Διαδικτυακό Περιβάλλον Στατιστικής Μορφοσυντακτικής Αναγνώρισης των  
Μονολεκτικών Ρηματικών Τύπων της  
Νέας Ελληνικής «VERBTAGGR++»

**Ιωάννα Σ. Ζέρβα**

**A.M.: 1115201000010**

**Βασιλική Ν. Κωλέτση**

**A.M.: 1115201000114**

**Νικολέτα Γ. Μωραγιάννη**

**A.M.: 1115201000184**

### **ΕΠΙΒΛΕΠΟΥΣΕΣ:**

**Αφροδίτη Τσαλαγιάδου**, Αναπληρώτρια Καθηγήτρια ΕΚΠΑ

**Μαρία Γρηγοριάδου**, Ομότιμη Καθηγήτρια ΕΚΠΑ

### **ΣΥΝΕΠΙΒΛΕΠΟΥΣΑ:**

**Πηνελόπη Λεμπέση**, Επιστημονική Συνεργάτις ΕΚΠΑ

## ΠΕΡΙΛΗΨΗ

Η παρούσα πτυχιακή εργασία πραγματεύεται την υλοποίηση ενός διαδικτυακού εργαλείου το οποίο επιτελεί μορφοσυντακτική αναγνώριση των μονολεκτικών ρηματικών τύπων της Νέας Ελληνικής και είναι διαθέσιμο στην ελληνική και την αγγλική γλώσσα. Τα γλωσσικά δεδομένα που χρησιμοποιεί προέρχονται από τη διδακτορική διατριβή της Πηνελόπης Λεμπέση με τίτλο "Στατιστική Μορφοσυντακτική Αναγνώριση και Λημματοποίηση του Ρηματικού Συνόλου της Νέας Ελληνικής" (Πανεπιστήμιο Marc Bloch, Στρασβούργο 2005). Η δημιουργία της βάσης δεδομένων της εφαρμογής καθώς και το λογισμικό που επεξεργάζεται και ελέγχει την ορθότητα των γλωσσικών δεδομένων και φροντίζει για την αυτοματοποιημένη εισαγωγή των δεδομένων αυτών στο εργαλείο προέρχεται από τη Διπλωματική εργασία της Σταυρούλας Κρουστάλλη, για το Μεταπτυχιακό Πρόγραμμα Σπουδών του Τμήματος Πληροφορικής και Τηλεπικοινωνιών του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών. Στο πλαίσιο της ανωτέρω Διπλωματικής εργασίας συμπεριλήφθηκαν στο εργαλείο μόνο τα δεδομένα που αφορούν στους τύπους που λήγουν σε άτονο '-α' και '-ν'. Στο πλαίσιο της συγκεκριμένης πτυχιακής εργασίας συμπεριλήφθηκαν όλοι οι υπόλοιποι ρηματικοί τύποι που ολοκληρώνουν το εύρος των μονολεκτικών ρηματικών τύπων της Νέας Ελληνικής. Για τη λημματοποίηση έχει ήδη δημιουργηθεί μία καινούρια βάση δεδομένων, η οποία ορίζει τις αλλαγές που πρέπει να γίνουν στο ρηματικό τύπο, ώστε να προκύψει το λήμμα. Ωστόσο, δεν έχουν συμπεριληφθεί οι σχετικοί γραμματικοί κανόνες, ώστε να υποστηρίζεται η λημματοποίηση στην παρούσα εφαρμογή. Παρόλο που το εργαλείο ασχολείται σήμερα μόνο με τη μορφοσυντακτική αναγνώριση των ρηματικών τύπων, έχει προβλεφθεί να επεκταθεί στο μέλλον έτσι ώστε να υποστηρίζεται και η λημματοποίηση των τύπων. Οι τεχνολογίες που χρησιμοποιήθηκαν είναι: Java (J2SE 1.4.2), JSP, Web Services, MySQL και Tomcat.

**ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ:** Υπολογιστική Γλωσσολογία

**ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ:** μορφοσυντακτική αναγνώριση ρηματικών τύπων, διαδικτυακή εφαρμογή, τεχνολογία web services, εξυπηρετητής διαδικτύου Tomcat, σύστημα διαχείρισης βάσεων δεδομένων MySQL

## ABSTRACT

The object of the present BA thesis is the implementation of a web-based tool which performs morphosyntactic recognition of the one-word Modern Greek verbal forms and is available in Greek and English. The linguistic data it uses originate in Penelope Lembessi's PhD dissertation entitled "*Statistical Morphosyntactic Disambiguation and Lemmatization of the Modern Greek Verbal Class*" (Marc Bloch University, Strasbourg, 2005). This thesis is based on Stauroula Kroustalli's thesis entitled "*Web application for morphosyntactic recognition of verb forms «VerbTagGr»*". Within the framework of Kroustalli's thesis, only the data concerning verb forms ending in unaccentuated '-α' and '-ν' had been included in the tool. Within the framework of this thesis, all verbs are included. Consequently, recognition is provided for all forms. Although the tool today concerns solely morphosyntactic recognition of verb forms, future expansion of it is foreseen, so that it will further support lemmatization of forms as well. Thus, for every input verb, the tool provides data originating in Lembessi's PhD dissertation which are necessary for the production of the canonical form (lemma). A first approach has been made for lemmatization, but in order for it to work, more steps should be made. Moreover, special software has been implemented, which processes and checks the correctness of the linguistic data. The technologies which have been used are: Java (J2SE 1.4.2), JSP, Web Services, MySQL and Tomcat.

**SUBJECT AREA:** Computational Linguistics

**KEYWORDS:** morphosyntactic recognition of verb forms, web application, web services technology, Tomcat web server, MySQL database management system

*ΣΤΙΣ ΟΙΚΟΓΕΝΕΙΕΣ ΜΑΣ*

## ΕΥΧΑΡΙΣΤΙΕΣ

Με την ολοκλήρωση της πτυχιακής μας εργασίας, η οποία υλοποιήθηκε στο τμήμα Πληροφορικής και Τηλεπικοινωνιών του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών, θα θέλαμε να ευχαριστήσουμε τους ανθρώπους που συνέβαλαν στην εκπόνησή της.

Οφείλουμε να εκφράσουμε τις θερμές μας ευχαριστίες στην Αναπληρώτρια Καθηγήτρια Αφροδίτη Τσαλαγατίδου και στην επιβλέπουσα Ομότιμη Καθηγήτρια Γρηγοριάδου Μαρία, οι οποίες μας προσέφεραν το ενδιαφέρον αυτό θέμα καθώς και για την εμπιστοσύνη που μας έδειξαν δίνοντάς μας τη δυνατότητα να υλοποιήσουμε την παρούσα πτυχιακή εργασία. Τις ευχαριστούμε επίσης για τις πολύτιμες γνώσεις και συμβουλές που μας παρείχαν καθ' όλη τη διάρκεια της εργασίας.

Ιδιαίτερες ευχαριστίες θα θέλαμε επίσης να απευθύνουμε στην Επιστημονική Συνεργάτιδα του ΕΚΠΑ Πηνελόπη Λεμπέση, για την αμέριστη βοήθειά και το αμείωτο ενδιαφέρον της. Επίσης, για την πολυετή ερευνά της και τις γνώσεις που μας προσέφερε, καθώς χωρίς αυτές δε θα ήταν δυνατή η ολοκλήρωση της εργασίας μας.

## ΠΕΡΙΕΧΟΜΕΝΑ

<b>ΠΡΟΛΟΓΟΣ .....</b>	<b>9</b>
<b>1. ΕΙΣΑΓΩΓΗ.....</b>	<b>10</b>
<b>2. ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΕΡΓΑΛΕΙΟΥ VERBTAGGR++.....</b>	<b>12</b>
<b>3. ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΡΗΜΑΤΙΚΟΥ ΣΥΣΤΗΜΑΤΟΣ.....</b>	<b>14</b>
<b>4. ΓΛΩΣΣΙΚΑ ΔΕΔΟΜΕΝΑ ΤΟΥ ΕΡΓΑΛΕΙΟΥ .....</b>	<b>20</b>
4.1 Γενικές Πληροφορίες.....	20
4.2 Δομή Αρχείων Γλωσσικών Δεδομένων.....	24
4.3 Έλεγχος των ήδη υπάρχοντων Γλωσσικών Δεδομένων .....	31
<b>5. ΠΕΡΙΓΡΑΦΗ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ ΑΝΑΓΝΩΡΙΣΗΣ ΡΗΜΑΤΙΚΩΝ ΤΥΠΩΝ .....</b>	<b>32</b>
5.1 Πίνακας ms_codeExplain.....	32
5.2 Πίνακας verb_info .....	33
5.3 Πίνακας equalsMscodes .....	37
<b>6. ΠΕΡΙΓΡΑΦΗ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ ΛΗΜΜΑΤΟΠΟΙΗΣΗΣ ΡΗΜΑΤΙΚΩΝ ΤΥΠΩΝ .....</b>	<b>38</b>
6.1 Πίνακας lemma .....	39
<b>7. ΥΛΟΠΟΙΗΣΗ ΕΡΓΑΛΕΙΟΥ VERBTAGGR++.....</b>	<b>41</b>
7.1 Λογισμικό Προεπεξεργασίας Γλωσσικών Δεδομένων .....	41
7.1.1 Πρόγραμμα dataProcess .....	42
7.1.2 Πρόγραμμα checkIntegers .....	42
7.1.3 Πρόγραμμα checkExceptionToForms.....	43
7.2 Τεχνολογία των Web Services .....	44
7.2.1 Ορισμός των Web Services .....	44
7.2.2 Μοντέλο των Web Services .....	45
7.2.3 Πρότυπα των Web Services .....	46
7.2.4 Κύκλος Ζωής των Web Services .....	48
7.2.5 Σύνοψη και Συμπεράσματα.....	49

7.3	Αλγόριθμος Μορφοσυντακτικής Αναγνώρισης Μονολεκτικών Ρηματικών Τύπων της Νέας Ελληνικής .....	51
7.4	Περιγραφή Web Service Μορφοσυντακτικής Αναγνώρισης Μονολεκτικών Ρηματικών Τύπων της Νέας Ελληνικής .....	54
<b>8.</b>	<b>ΠΕΡΙΓΡΑΦΗ ΔΙΑΔΙΚΤΥΑΚΗΣ ΕΦΑΡΜΟΓΗΣ.....</b>	<b>65</b>
8.1	Πληροφορίες Εφαρμογής για Ρηματικό Τύπο Εισόδου .....	65
8.2	Δείγματα Σελίδων της Εφαρμογής.....	67
<b>9.</b>	<b>ΟΔΗΓΙΕΣ ΕΓΚΑΤΑΣΤΑΣΗΣ ΤΗΣ ΕΦΑΡΜΟΓΗΣ.....</b>	<b>83</b>
<b>10.</b>	<b>ΣΥΜΠΕΡΑΣΜΑΤΑ .....</b>	<b>114</b>
	<b>ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ .....</b>	<b>115</b>
	<b>ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ .....</b>	<b>116</b>
	<b>ΑΝΑΦΟΡΕΣ .....</b>	<b>117</b>



## **ΠΡΟΛΟΓΟΣ**

Η παρούσα πτυχιακή εργασία διενεργήθηκε στο πλαίσιο των Προπτυχιακών Σπουδών μας στο Τμήμα Πληροφορικής και Τηλεπικοινωνιών του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών τον Μάρτιο του έτους 2016.

## 1. ΕΙΣΑΓΩΓΗ

Στο πλαίσιο της συγκεκριμένης πτυχιακής εργασίας πραγματοποιήθηκε η επέκταση του διαδικτυακού γλωσσικού εργαλείου του Εργαστηρίου Εκπαιδευτικής και Γλωσσικής Τεχνολογίας του ΕΚΠΑ που ονομάζεται 'VERBTAGGR++', το οποίο δέχεται ως είσοδο ένα μονολεκτικό ρηματικό τύπο της Νέας Ελληνικής και εμφανίζει πληροφορίες που αφορούν στη μορφοσυντακτική του αξία. Η αρχική πιλοτική εφαρμογή είχε υλοποιηθεί για τους ρηματικούς τύπους που λήγουν σε άτονο '-α' και '-ν' για τους οποίους και παρείχε αντίστοιχη μορφοσυντακτική αναγνώριση. Στην παρούσα εργασία η εφαρμογή ολοκληρώθηκε με την ενσωμάτωση όλων των υπόλοιπων ρηματικών δεδομένων. Επιπλέον, για κάθε τύπο εισόδου παρέχεται ένα σύνολο δεδομένων τα οποία χρειάζονται για την παραγωγή του λήμματος του τύπου, προκειμένου να είναι εύκολη μία μελλοντική επέκταση του εργαλείου.

Όσον αφορά στο όνομα της νέας έκδοσης του εργαλείου VERBTAGGR++, αυτό αποτελείται από τέσσερα συστατικά: 'VERB', 'TAG', 'GR' και '++'. Η λέξη 'VERB' σημαίνει 'ρήμα' στα Αγγλικά, ενώ το 'GR' προέρχεται από τη λέξη 'GREEK', που στα Αγγλικά είναι το επίθετο 'ελληνικός'. Το 'TAG' προέρχεται από τη λέξη 'TAGGING', που σαν όρος στην υπολογιστική γλωσσολογία χρησιμοποιείται για να δηλώσει τη διαδικασία αναγνώρισης και απόδοσης ετικέτας χαρακτηρισμού (tag) στις λέξεις ενός κειμένου, η οποία δηλώνει συγκεκριμένα χαρακτηριστικά σε διάφορα καθορισμένα εκ των προτέρων επίπεδα όπως φωνολογικό, μορφοσυντακτικό/γραμματικό, συντακτικό κ.λπ. Το '++' δηλώνει την ενσωμάτωση των υπολοίπων ρηματικών δεδομένων στην εφαρμογή καθώς και μία πρώτη επέκταση για την λειτουργικότητα της εφαρμογής ως προς την λημματοποίηση. Στην προκειμένη περίπτωση το όνομα 'VERBTAGGR++' παραπέμπει στην ολοκληρωμένη απόδοση μορφοσυντακτικής ετικέτας στους ρηματικούς τύπους της Νέας Ελληνικής.

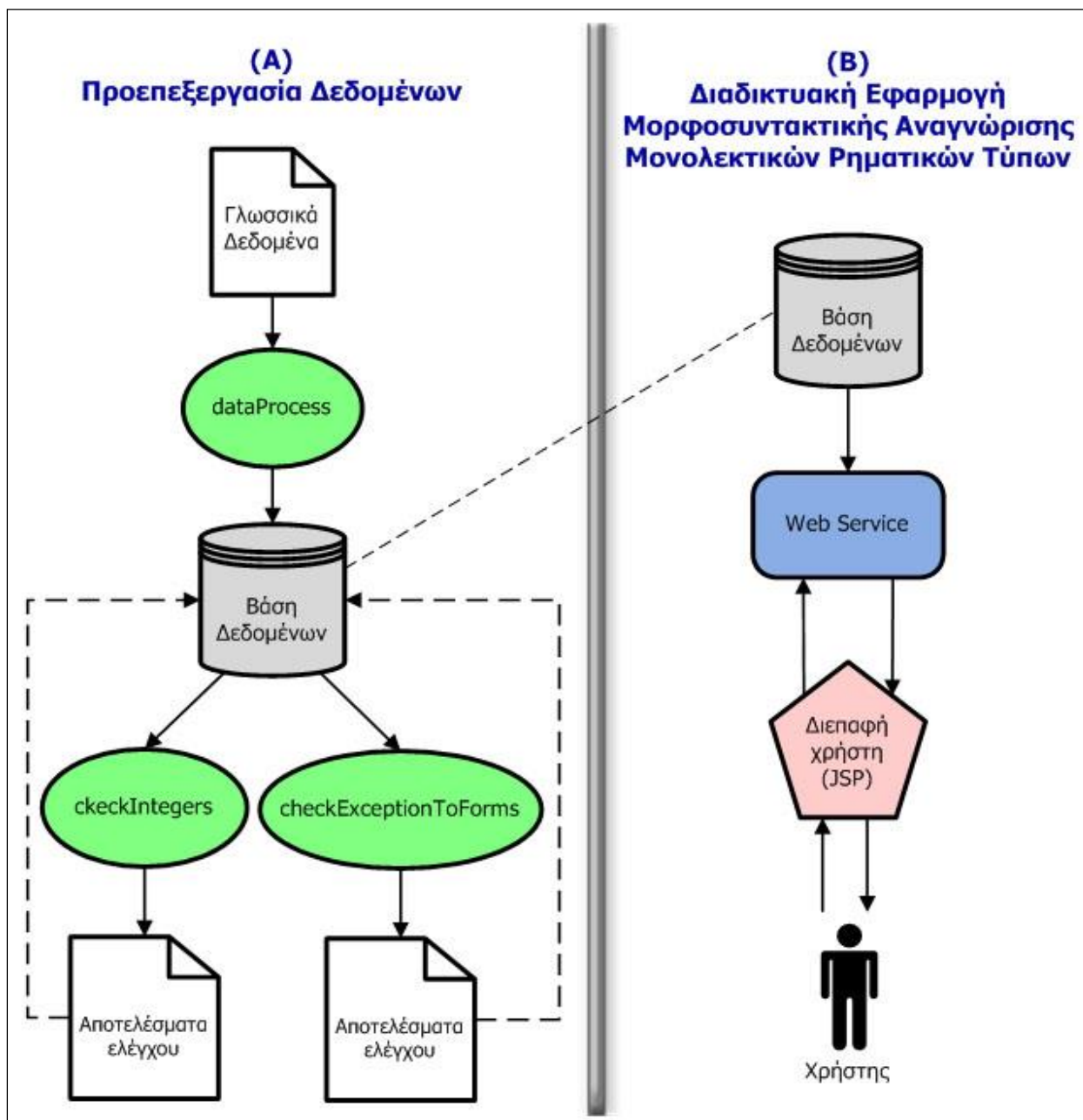
Τα γλωσσικά δεδομένα στα οποία βασίζεται το εργαλείο για να επιτελέσει τη μορφοσυντακτική αναγνώριση των ρηματικών τύπων και να παραγάγει τα στοιχεία που απαιτούνται για λημματοποίηση, προέρχονται από τη διδακτορική διατριβή της Πηνελόπης Λεμπέση [1]. Πιο συγκεκριμένα, η Λεμπέση διέθεσε ένα σύνολο αρχείων γλωσσικών δεδομένων με όμοια δομή, τα οποία περιείχαν πληροφορίες που η ίδια συνέλεξε στο πλαίσιο της δεκαετούς έρευνάς της και αφορούν στη μορφοσυντακτική αναγνώριση και τη λημματοποίηση των ρηματικών τύπων της Νέας Ελληνικής. Στη φάση της προεπεξεργασίας των γλωσσικών δεδομένων ελέγχθηκαν τα αρχεία αυτά ως προς την ορθότητα της δομής τους από λογισμικό που υλοποιήθηκε σε Java και στη συνέχεια διορθώθηκαν λάθη που εντοπίστηκαν και αναφέρθηκαν στη Λεμπέση. Έπειτα, με χρήση λογισμικού υλοποιημένου επίσης σε Java έγινε η επεξεργασία του περιεχομένου των αρχείων, προκειμένου να εισαχθεί στη βάση δεδομένων της εφαρμογής ένα σύνολο από δεδομένα τα οποία αντιστοιχούν στις πληροφορίες που περιέχονται στα αρχεία γλωσσικών δεδομένων. Μετά την εισαγωγή των δεδομένων στη βάση, αναπτύχθηκε μία διαδικτυακή εφαρμογή που επιτελεί μορφοσυντακτική αναγνώριση των μονολεκτικών ρηματικών τύπων της Νέας Ελληνικής. Η διεπαφή της εφαρμογής έχει υλοποιηθεί σε JSP και η λειτουργικότητά της βασίζεται σε ένα Web Service υλοποιημένο σε Java.

Η παρούσα εργασία είναι οργανωμένη σε 9 κεφάλαια. Στο επόμενο κεφάλαιο απεικονίζεται σχηματικά η αρχιτεκτονική του εργαλείου VERBTAGGR++, ενώ στο κεφάλαιο 3 παρουσιάζονται έννοιες και όροι που αφορούν στο ρηματικό σύστημα της Νέας Ελληνικής. Τα στοιχεία αυτά απαιτείται να είναι γνωστά στον αναγνώστη, διότι θα χρησιμοποιηθούν ευρέως στη συνέχεια. Στο κεφάλαιο 4 δίνονται γενικά στοιχεία για τη διδακτορική διατριβή της Λεμπέση [1], μία περιγραφή της δομής των αρχείων γλωσσικών δεδομένων που χρησιμοποιήθηκαν από το εργαλείο καθώς και του ελέγχου που πραγματοποιήθηκε στα νέα γλωσσικά δεδομένα που μας παραχωρήθηκαν. Στο κεφάλαιο

5 περιγράφεται η βάση δεδομένων του εργαλείου και παρατίθενται παραδείγματα με τα οποία φαίνεται πώς οι πληροφορίες των αρχείων γλωσσικών δεδομένων φυλάσσονται στη βάση. Με ανάλογο τρόπο στο κεφάλαιο 6, παρουσιάζεται ο τρόπος δημιουργίας της νέας βάσης δεδομένων, η οποία περιέχει τους κωδικούς λημματοποίησης και τους κανόνες που πρέπει να τηρηθούν ώστε να ολοκληρωθεί αργότερα η διαδικασία της λημματοποίησης. Ακολουθεί το κεφάλαιο 7 που αφορά σε θέματα υλοποίησης. Ειδικότερα, η πρώτη ενότητα του κεφαλαίου 7 παρουσιάζει το λογισμικό που αναπτύχθηκε και χρησιμοποιήθηκε στη φάση της προεπεξεργασίας των γλωσσικών δεδομένων, η δεύτερη ενότητα εξετάζει την τεχνολογία των Web Services που χρησιμοποιήθηκε στην υλοποίηση της διαδικτυακής εφαρμογής, η τρίτη ενότητα καταγράφει τη γενική μορφή του αλγόριθμου που ακολουθείται κατά τη διαδικασία μορφοσυντακτικής αναγνώρισης ενός μονολεκτικού ρηματικού τύπου της Νέας Ελληνικής, ενώ το κεφάλαιο ολοκληρώνεται με την ενότητα που περιγράφει το Web Service που υλοποιεί τον αλγόριθμο της προηγούμενης ενότητας. Στη συνέχεια, παρουσιάζεται στο κεφάλαιο 8 μία περιγραφή της διαδικτυακής εφαρμογής VERBTAGGR++ μαζί με εικόνες που απεικονίζουν τις σελίδες που εμφανίζονται, όταν δοθούν για αναγνώριση ενδεικτικές συμβολοσειρές, ενώ η εργασία ολοκληρώνεται με το κεφάλαιο 9, στο οποίο δίνονται αναλυτικές οδηγίες για την εγκατάσταση της διαδικτυακής εφαρμογής.

## 2. ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΕΡΓΑΛΕΙΟΥ VERBTAGGR++

Η αρχιτεκτονική του εργαλείου VERBTAGGR++ παρουσιάζεται στο παρόν κεφάλαιο προκειμένου ο αναγνώστης να έχει μία γενική εικόνα για τα τμήματα από τα οποία αυτό αποτελείται. Το ακόλουθο σχήμα απεικονίζει μία γενική δομή του εργαλείου:



Σχήμα 1: Αρχιτεκτονική εργαλείου VERBTAGGR++

Όπως φαίνεται και από το παραπάνω σχήμα, η εργασία που υλοποιήθηκε αποτελείται από δύο φάσεις:

### **(A) Φάση προεπεξεργασίας γλωσσικών δεδομένων:**

Στην πρώτη φάση έγινε η προεπεξεργασία των γλωσσικών δεδομένων στα οποία βασίστηκε το εργαλείο. Συγκεκριμένα, μία σειρά από αρχεία γλωσσικών δεδομένων δόθηκαν ως είσοδος στο αρχείο “dataProcess”, το οποίο έλεγξε τη δομή των γραμμών των αρχείων και εφόσον κανένα λάθος δεν εντοπίστηκε, δημιουργήθηκε ένα σύνολο από SQL επερωτήσεις, οι οποίες εισάγουν στη βάση δεδομένων του εργαλείου δεδομένα τα οποία αντιστοιχούν στις πληροφορίες του εκάστοτε αρχείου γλωσσικών δεδομένων. Στις περιπτώσεις που εμφανίστηκαν λάθη σε ορισμένες καταλήξεις ρηματικών τύπων, οι αντίστοιχες γραμμές των αρχείων στάλθηκαν στη Λεμπέση, προκειμένου να διορθωθούν και να γίνει σωστή εισαγωγή στη βάση. Στη συνέχεια, με χρήση των προγραμμάτων “checkIntegers” και “checkExceptionToForms” ελέγχθηκαν τα δεδομένα της βάσης και εντοπίστηκαν μη έγκυρες τιμές στα διάφορα πεδία των πινάκων. Τα λάθη αυτά αναφέρθηκαν στη Λεμπέση, από την οποία διατέθηκαν τα γλωσσικά δεδομένα και έπειτα ενημερώθηκε τελικά το περιεχόμενο της βάσης, ώστε να περιέχονται έγκυρα δεδομένα.

### **(B) Διαδικτυακή εφαρμογή μορφοσυντακτικής αναγνώρισης μονολεκτικών ρηματικών τύπων:**

Μετά τη δημιουργία εκ νέου του αρχείου “dataProcess” καθώς και τον έλεγχο και τη διόρθωση των μη έγκυρων τιμών της βάσης, χρησιμοποιήθηκε μία διαδικτυακή εφαρμογή η οποία δέχεται ως είσοδο ένα μονολεκτικό ρηματικό τύπο της Νέας Ελληνικής και επιστρέφει τα αποτελέσματα της μορφοσυντακτικής αναγνώρισής του. Ειδικότερα, ο χρήστης αλληλεπιδρά με μία διεπαφή υλοποιημένη σε JSP και εισάγει προς αναγνώριση μία συμβολοσειρά. Έπειτα, μέσω της διεπαφής καλείται ένα Web Service στο οποίο δίνεται ως είσοδος η προς αναγνώριση συμβολοσειρά και η γλώσσα στην οποία θα εκφραστούν τα αποτελέσματα της αναγνώρισης. Το Web Service αλληλεπιδρά με τη βάση κι επιστρέφει στη διεπαφή πληροφορίες που αφορούν στην αναγνώριση της συμβολοσειράς εισόδου. Οι πληροφορίες αυτές τελικά απεικονίζονται στο χρήστη. Αυτή η διαδικτυακή εφαρμογή υλοποιήθηκε από τη Σταυρούλα Ι. Κρουστάλλη στα πλαίσια της διπλωματικής της εργασίας το 2005.

### 3. ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΡΗΜΑΤΙΚΟΥ ΣΥΣΤΗΜΑΤΟΣ

Στο παρόν κεφάλαιο παρουσιάζονται βασικές έννοιες και στοιχεία για το ρηματικό σύστημα της Νέας Ελληνικής τα οποία προέρχονται από βιβλία που αφορούν στη γραμματική της Νέας Ελληνικής [2, 3, 4]. Τις πληροφορίες αυτές χρειάζεται να γνωρίζει ο αναγνώστης, διότι θα χρησιμοποιηθούν συχνά από εδώ και πέρα.

“Το σύστημα του ρήματος της Νέας Ελληνικής αποτελείται από το σύνολο των ρηματικών τύπων. Όπως και στα υπόλοιπα κλιτά συστήματα της γλώσσας όπως αυτό των ονομάτων ή των επιθέτων, κάθε ρηματικός τύπος απαρτίζεται από το ‘λεξικό μόρφημα’ που δηλώνει τη λεξική σημασία του ρήματος (π.χ. *γραφ-*, *παιζ-*, *τρεχ-*) και από τα ‘γραμματικά μορφήματα’, τα οποία δηλώνουν τα ‘παρεπόμενα’ του ρήματος όπως είναι αλλιώς γνωστές από την παραδοσιακή γραμματική οι γραμματικές πληροφορίες (γένος, αριθμός, πτώση, πρόσωπο, χρόνος, ποιόν ενέργειας, τροπικότητα κ.λπ). Οι ρηματικοί τύποι μπορεί να είναι μονολεκτικοί, π.χ. *τρώγαμε*, *φύγε* ή περιφραστικοί, π.χ. *έχουμε φάει*, *θα έχουμε τελειώσει*. Παραδοσιακά, για κάθε έναν τέτοιο συνδυασμό μορφημάτων που εκφράζουν οι ρηματικοί τύποι χρησιμοποιείται ο όρος «χρόνος» («χρόνοι του ρήματος»). Επειδή όμως ο χρόνος δεν είναι το μόνο στοιχείο που εκφράζουν αυτοί οι συνδυασμοί ρήματος και παρεπόμενων όπως αναφέρεται στη γραμματική των Κλαίρη – Μπαμπινιώτη [2], προτιμάμε και εμείς τον ουδέτερο όρο «ρηματικός τύπος», ο οποίος θα χρησιμοποιηθεί στο παρόν κείμενο.

Έτσι, “ένα ρήμα κλίνεται ως προς το πρόσωπο, τον αριθμό, το χρόνο, το ποιόν ενέργειας, τη φωνή και –σε κάποιο βαθμό– την έγκλιση” [3]. Στη συνέχεια δίνονται περισσότερες πληροφορίες για καθέναν από τους προαναφερθέντες όρους, οι οποίοι θα χρησιμοποιηθούν από εδώ και πέρα.

- **Πρόσωπο**

Το ‘πρώτο’ πρόσωπο δηλώνει τον ομιλητή, το ‘δεύτερο’ πρόσωπο δηλώνει το πρόσωπο / πράγμα στο οποίο απευθυνόμαστε και το ‘ τρίτο’ πρόσωπο δηλώνει την οντότητα για την οποία γίνεται λόγος.

- **Αριθμός**

Η γραμματική κατηγορία που αναφέρεται στο πλήθος των οντοτήτων που δηλώνονται από το ρήμα. Πρόκειται για τη διάκριση ανάμεσα σε ‘ενικό’ και ‘πληθυντικό’ στη Νέα Ελληνική, όπου ο ενικός ενός ρήματος υποδεικνύει ότι το υποκείμενο αποτελείται από μία μόνο οντότητα, ενώ ο πληθυντικός δηλώνει ότι το υποκείμενο αποτελείται από περισσότερες από μία οντότητες.

- **Χρόνος**

Βάσει των όσων αναφέρονται στη γραμματική των Holton – Mackridge – Φιλιππάκη [3], “σε γενικές γραμμές ο χρόνος στη λιγότερο μαρκαρισμένη οριστική έγκλιση και σαφέστερα στις κύριες προτάσεις είναι η ρηματική κατηγορία η οποία τοποθετεί την ενέργεια ή την κατάσταση που εκφράζει το ρήμα σε κάποια χρονική στιγμή: παρελθόν *έγραφα*, *έγραψα*, παρόν *γράφω*, μέλλον *θα γράφω*, *θα γράψω*, παρακείμενος *έχω γράψει*, υπερσυντέλικος *είχα γράψει*. Μερικοί από τους χρόνους σχηματίζονται ‘μονολεκτικά’, δηλ. με μεταβολή μέσα στον ίδιο το ρηματικό τύπο: ενεστώτας, απλός παρελθοντικός (ή αόριστος), παρατατικός. Αντίθετα άλλοι χρόνοι σχηματίζονται ‘περιφραστικά’ με τη χρήση του μορίου *θα*: συνοπτικός και μη συνοπτικός μέλλοντας (ή

αλλιώς απλός μέλλοντας και μέλλοντας διαρκείας), υποθετική (δομή που σχηματίζεται από το μόριο θα και τον παρατατικό), τετελεσμένη υποθετική (δομή που σχηματίζεται από το μόριο θα και τον υπερσυντέλικο), συντελεσμένος μέλλοντας. Οι πιο σύνθετοι χρόνοι, ο παρακείμενος, ο υπερσυντέλικος και ο συντελεσμένος μέλλοντας, χρησιμοποιούν το βοηθητικό ρήμα έχω”.

Στην ίδια γραμματική (των Holton – Mackridge – Φιλιππάκη [3]) γίνονται “δύο διαφορετικού είδους διακρίσεις χρόνου: (α) οι γενικές κατηγορίες παρελθόν και μη παρελθόν και (β) οι ειδικοί ρηματικοί τύποι ενεστώτας, παρατατικός, απλός παρελθοντικός, μέλλοντας, παρακείμενος κλπ”.

- **Ποιόν ενέργειας**

Όπως παρουσιάζεται στη γραμματική των Holton – Mackridge – Φιλιππάκη [3], “το ποιόν ενέργειας είναι μία γραμματική κατηγορία η οποία δηλώνει μία συγκεκριμένη διαφορά στον τρόπο με τον οποίο ο ομιλητής αντιλαμβάνεται και παρουσιάζει στον ακροατή την ενέργεια που εκφράζει το ρήμα. Μία ενέργεια μπορεί να παρουσιάζεται στην ολότητά της, ως ένα και μοναδικό γεγονός, ως μια επαναλαμβανόμενη συνήθεια, ως εξελισσόμενη ή τέλος ως ολοκληρωμένη στο παρελθόν αλλά με την ολοκλήρωσή της να έχει επιπτώσεις στην παρούσα κατάσταση των πραγμάτων. Μορφολογικά δηλαδή μέσα στον ίδιο το ρηματικό τύπο, διαφοροποιούνται δύο είδη ποιού ενέργειας: το ‘συνοπτικό’ και το ‘μη συνοπτικό’. Εξάλλου τα ρήματα είναι δυνατό να δηλώνουν και ένα τρίτο ποιόν ενέργειας, το ‘τετελεσμένο’. Αυτό εκφράζεται περιφραστικά με τη χρήση βοηθητικών ρημάτων”.

Συγκεκριμένα, “το μη συνοπτικό ποιόν ενέργειας παρουσιάζει την ενέργεια ή την κατάσταση που εκφράζει το ρήμα ως ένα μοναδικό αλλά συνεχιζόμενο γεγονός (εξελισσόμενο) ή ως επαναλαμβανόμενη συνήθεια. Το συνοπτικό ποιόν ενέργειας εκφράζει μία ενέργεια (λιγότερο μία κατάσταση) η οποία θεωρείται στην ολότητά της ως μοναδικό και ολοκληρωμένο γεγονός. Τέλος, το τετελεσμένο ποιόν ενέργειας περιγράφει μία ενέργεια ή κατάσταση η οποία έχει ολοκληρωθεί σε κάποια χρονική στιγμή, συνδυάζεται με την κατηγορία της χρονικής αναφοράς και σχηματίζει τρεις ρηματικούς χρόνους: παρακείμενο, υπερσυντέλικο και τετελεσμένο μέλλοντα” [3]. Για παράδειγμα, ο ενεστώτας *γράφω* είναι μη συνοπτικός, όπως και ο παρατατικός *έγραφα*: και οι δύο σχηματίζονται από το μη συνοπτικό θέμα *γραφ-*. Από την άλλη, ο αόριστος *έγραψα* είναι συνοπτικός ως προς το ποιόν ενέργειας και σχηματίζεται με το συνοπτικό θέμα *γραφ-*.

Εξετάζοντας την ορολογία και τα στοιχεία που αφορούν στο ποιόν ενέργειας βάσει του υλικού που παραθέτουν και άλλα βιβλία γραμματικής περιγραφής της γλώσσας, παρατηρούμε ότι στη γραμματική των Κλαίρη – Μπαμπινιώτη [2] χρησιμοποιούνται οι όροι ‘τέλειο’ και ‘ατελής’ αντί για το συνοπτικό και μη συνοπτικό των Holton – Mackridge – Φιλιππάκη [3]. Συγκεκριμένα, στη γραμματική των Κλαίρη – Μπαμπινιώτη αναφέρεται ότι “εκτός από την αντικειμενική χρονική σύσταση που δηλώνεται αυτόματα από τη σημασία του ρήματος, ο ομιλητής έχει τη δυνατότητα να δηλώνει και τον τρόπο με τον οποίο ο ίδιος αντιμετωπίζει το γεγονός ως προς τη χρονική του σύσταση, την οπτική γωνία από την οποία το βλέπει”. Δηλαδή ο ομιλητής μπορεί να δηλώσει αν βλέπει το γεγονός συνοπτικά, ως ενιαίο σύνολο ή αν το βλέπει στην εξέλιξή του, αν ενδιαφέρεται δηλαδή για την εσωτερική χρονική του σύνθεση, για τις φάσεις που το αποτελούν. Η διαφοροποίηση αυτή στη θεώρηση των γεγονότων εκφράζεται στην ελληνική γλώσσα από τη μορφή του ρήματος και μάλιστα από τα δύο βασικά θέματα του κάθε ρήματος, το θέμα του Ενεστώτα και το θέμα του Αορίστου. Για παράδειγμα:

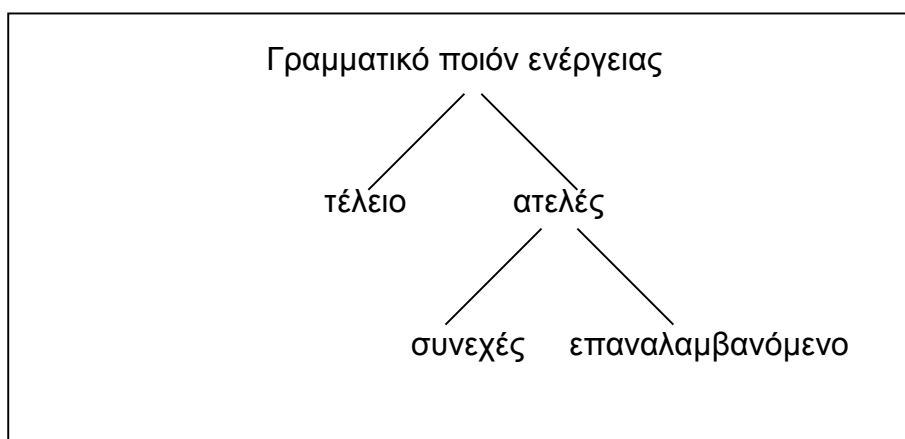
*Τον περασμένο μήνα ταξίδεψα στην Αφρική.*

*Τον περασμένο μήνα ταξίδεψα στην Αφρική.*

Η γραμματική κατηγορία που εκφράζει τον τρόπο θεώρησης των γεγονότων στη γραμματική των Κλαίρη – Μπαμπινιώτη [2] ονομάζεται ‘γραμματικό ποιόν ενέργειας’.

“Είναι όμως επίσης γνωστή με τους όρους ‘όψη’ ή ‘άποψη’ (aspect). Το γραμματικό ποιόν ενέργειας που εκφράζει τη συνοπτική θεώρηση ονομάζεται ‘τέλειο’, ενώ εκείνο που εκφράζει τη θεώρηση της ενέργειας ή της κατάστασης στην εξέλιξή της ονομάζεται ‘ατελής’. Το θέμα του Αορίστου εκφράζει το τέλειο και ο ομιλητής το χρησιμοποιεί όταν βλέπει το γεγονός στο σύνολό του, χωρίς να ενδιαφέρεται για τη διάρκειά του ή για τις φάσεις από τις οποίες αποτελείται. Αντίθετα το θέμα του Ενεστώτα εκφράζει το ατελής και ο ομιλητής το χρησιμοποιεί όταν βλέπει το γεγονός στην εξέλιξή του.” [2]

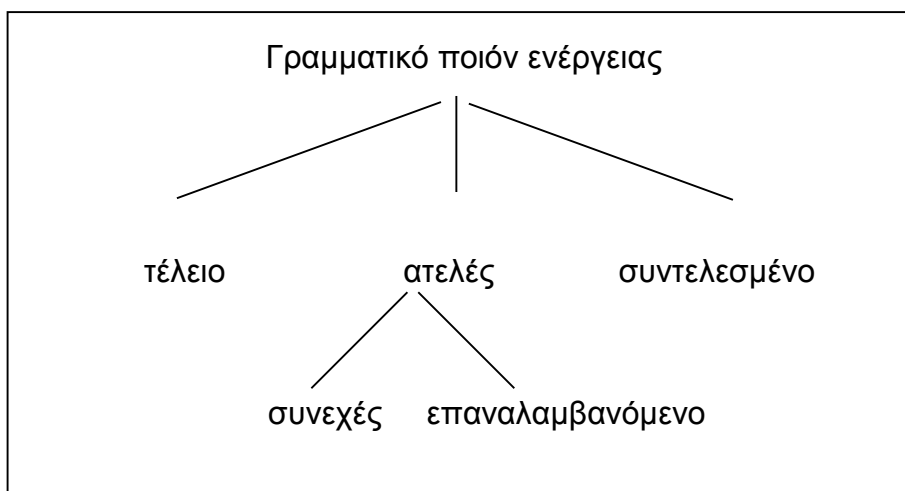
Συνεχίζοντας την παράθεση στοιχείων από τη γραμματική των Κλαίρη – Μπαμπινιώτη [2], αναφέρουμε ότι “το ατελής ποιόν ενέργειας έχει δύο δυνατές σημασίες: όταν το χρησιμοποιεί ο ομιλητής, παρουσιάζει το γεγονός είτε ως κάτι συνεχές, που βρίσκεται σε εξέλιξη, είτε ως κάτι που επαναλαμβάνεται με κανονικότητα”. Έτσι, προκύπτει το ακόλουθο σχήμα κατηγοριών του γραμματικού ποιού ενέργειας:



Σχήμα 2: Κατηγορίες ποιού ενέργειας (α)

Από το θέμα του Ενεστώτα *τρέχ-* σχηματίζονται, εκτός από τον ίδιο, ο Παρατατικός *έτρεχα* και ο ατελής Μέλλοντας *θα τρέχω* καθώς και η μετοχή σε *-όντας / ώντας*. Όλοι οι υπόλοιποι τύποι του ρήματος σχηματίζονται από το θέμα του Αορίστου. Οι λεγόμενοι συντελεσμένοι χρόνοι (Παρακείμενος, Υπερσυντέλικος, Συντελεσμένος Μέλλοντας) αποτελούν μια προβληματική κατηγορία: όπως δείχνει και η μορφή τους (σχηματίζονται από το θέμα του Αορίστου), είναι δυνατόν να θεωρηθεί ότι αποτελούν μια υποκατηγορία του τέλειου, δίνοντας μεγαλύτερη έμφαση στο συντελεσμένο του γεγονότος, τονίζοντας δηλαδή το ότι αυτό έχει ολοκληρωθεί. Είναι όμως δυνατόν επίσης να θεωρηθεί ότι η διαφορά των συντελεσμένων από τους άλλους χρόνους δεν εντοπίζεται στο ποιόν ενέργειας, αλλά στο χρόνο και ότι δηλώνουν το προτερόχρονο. Στη γραμματική των Κλαίρη – Μπαμπινιώτη [2] “επελέγη μία μέση οδός: η συντελικότητα αντιμετωπίστηκε ως μια ξεχωριστή κατηγορία και εντάχθηκε στο ποιόν ενέργειας με την παρατήρηση ότι έχει χαρακτηριστικά που τη διαφοροποιούν τόσο ως προς το ποιόν ενέργειας (τετελεσμένο) όσο και ως προς το χρόνο (προτερόχρονο)”. Το τελικό σχήμα λοιπόν διαμορφώνεται ως εξής:





Σχήμα 3: Κατηγορίες ποιού ενέργειας (β)

Επιπρόσθετα, πληροφορίες αντλήθηκαν και από τη γραμματική του P. Mackridg[e] [4]. Σε αυτήν χρησιμοποιείται ο όρος ‘τρόπος ενέργειας’ αντί για τον όρο ‘ποιόν ενέργειας’ και γίνεται η διάκριση σε ‘εξακολουθητικό’, ‘συνοπτικό’ και ‘συντελεσμένο’ τρόπο ενέργειας. Αναφέρεται βέβαια ότι “ο συντελεσμένος (ή συντελικός) τρόπος ενέργειας δεν έχει τόσο αποφασιστική σημασία όσο οι δύο άλλοι, ο εξακολουθητικός και ο συνοπτικός, ενώ υπάρχει και το γεγονός ότι ο συντελεσμένος τρόπος έχει πάντα περιφραστική εκφορά και η συνοχή της περίφρασης δεν έχει το «αδιαχώρητο» (π.χ. ανάμεσα στα δύο στοιχεία του περιφραστικού σχήματος είναι δυνατό να παρεμβάλλεται χρονικό επίρρημα και ολόκληρο λεκτικό σύνολο ακόμα, ονοματικό ή επιρρηματικό)”. Η σημασία του εξακολουθητικού και συνοπτικού τρόπου είναι αυτή που παρουσιάστηκε παραπάνω για το μη συνοπτικό (ή ατελή) και συνοπτικό (ή τέλειο) αντίστοιχα ποιόν ενέργειας. Όσον αφορά στο συντελεσμένο τρόπο, αναφέρεται ότι “διαφέρει από τους άλλους τρόπους, διότι ο χρόνος αναφοράς του είναι προγενέστερος από έναν άλλο χρόνο (είτε δηλαδή προγενέστερος από το χρόνο της λεκτικής εκφώνησης είτε προγενέστερος από το χρόνο ενέργειας ενός άλλου ρήματος του ίδιου λεκτικού περιβάλλοντος)”. Ο ιδιαίτερος χαρακτήρας του συντελεσμένου τρόπου είναι ότι αναφέρεται σε δύο διαβαθμίσεις του χρόνου και συνδυάζει τη μία με την άλλη (π.χ. παροντική με παρελθοντική ή παρελθοντική με πιο παρελθοντική ή μελλοντική με λιγότερο μελλοντική). Χρησιμεύει για να παριστάνουμε μία ενέργεια τελειωμένη (ως προς αυτό είναι όμοιος με το συνοπτικό τρόπο) και δε χρησιμοποιείται όπου υπάρχει χρονική σύμπτωση ρηματικών ενεργειών, ούτε όπου παρουσιάζεται μία ρηματική ενέργεια θεωρημένη διεξοδικά ή επαναληπτικά.

Ως παράδειγμα δίνονται στον παρακάτω πίνακα οι κύριοι μη απαρεμφατικοί τύποι ενεργητικής φωνής του ρήματος *αγαπώ* (στο πρώτο ενικό πρόσωπο). Κάθε τύπος στον πίνακα αυτό βρίσκεται σε ένα σημείο τομής χρόνου και ποιού (ή τρόπου) ενέργειας.

Πίνακας 1: Μη απαρεμφατικοί τύποι του ρήματος *αγαπώ*

		ΠΟΙΟΝ ΕΝΕΡΓΕΙΑΣ		
		Μη συνοπτικό	Συνοπτικό	Συντελεσμένο
ΧΡΟΝΟΣ	Μη παρελθοντικός	αγαπώ	αγαπήσω	έχω αγαπήσει

	Παρελθοντικός	αγαπούσα	αγάπησα	είχα αγαπήσει
--	---------------	----------	---------	---------------

Θα πρέπει να σημειωθεί ότι στη γραμματική του P. Mackridge δε χρησιμοποιούνται οι όροι ‘οριστική’ και ‘υποτακτική’ για να δηλωθούν μορφολογικές κατηγορίες, “γιατί δεν υπάρχει μορφολογική διαφοροποίηση τύπων μεταξύ τους (με βάση τα μορφολογικά χαρακτηριστικά του ίδιου ρήματος) αλλά διακρίνονται μόνο συντακτικά ανάλογα με το λεκτικό περιβάλλον”. Μάλιστα όπως αναφέρεται στην ίδια γραμματική, “μορφολογικά, η μόνη διαφορά που διακρίνει την οριστική και την υποτακτική ως εγκλίσεις της Νέας Ελληνικής είναι ότι στην οριστική αποκλείεται να γίνει χρήση του μη παρελθοντικού συνοπτικού, ενώ βέβαια στην υποτακτική έγκλιση η χρήση αυτού του τύπου δεν αποκλείεται. Συντακτικά εξάλλου το διακριτικό της υποτακτικής έγκλισης είναι ότι πριν απ’ αυτήν υπάρχει ένας ‘δείκτης υποτακτικής σύνδεσης’ (π.χ. *θα, να, ας, μη, ίσως, πριν*). Άρα η «υποτακτικότητα» δεν είναι ιδιότητα που ενυπάρχει στον ίδιο το ρηματικό τύπο αλλά είναι λειτουργία που επιτελείται από τους δείκτες υποτακτικής σύνταξης”.

Μία τελευταία παρατήρηση για τους συνοπτικούς μη παρελθοντικούς τύπους αφορά σε έναν ειδικό όρο που χρησιμοποιείται στη γραμματική των Holton – Mackridge – Φιλιππάκη [3]. Εκεί χρησιμοποιείται ο όρος ‘εξαρτημένος’ για τους ρηματικούς τύπους που σχηματίζονται “από το συνοπτικό θέμα με τις μη παρελθοντικές καταλήξεις. Ένας εναλλακτικός όρος όμως ο οποίος προτιμάται από πολλούς θεωρητικούς είναι ο «*συνοπτικός μη παρελθοντικός*» και αυτός είναι ο όρος που θα χρησιμοποιηθεί από εδώ και πέρα στο πλαίσιο της παρούσας πτυχιακής.

Θα πρέπει να σημειωθεί ότι ο συνοπτικός μη παρελθοντικός δεν απαντά ως ανεξάρτητος ρηματικός τύπος (δηλαδή δεν μπορεί να υπάρξει από μόνος του), γι’ αυτό και δεν μπορεί να εξαχθεί συμπέρασμα για την έγκλιση ή το χρόνο του. Βάσει στοιχείων από τη γραμματική των Holton – Mackridge – Φιλιππάκη [3], συναντάται με τα μόρια *θα, μη(ν), να ή ας*, μετά τους υποθετικούς ή χρονικούς συνδέσμους *αν, μόλις, όταν, αφού* κλπ, μετά τους συνδέσμους *πριν* και *ίσως*, μετά τις αοριστολογικές αναφορικές αντωνυμίες και επιρρήματα *όποιος, όποτε, όπου, όπως*. Ως παραδείγματα αναφέρονται οι τύποι *ίσως αγαπήσω, να είχα αγαπήσει, ας αγαπούσα, να αγαπήσω* κλπ.

- **Φωνή**

Τα ρήματα της Νέας Ελληνικής διαθέτουν δύο διαφορετικές ομάδες προσωπικών καταλήξεων οι οποίες σχετίζονται με τη γραμματική κατηγορία της φωνής: ‘ενεργητική’ και ‘(μεσο)παθητική’. Σε πολλές περιπτώσεις το ίδιο ρήμα απαντά και στις δύο φωνές, π.χ. *γράφω – γράφομαι*. Σε αυτές τις περιπτώσεις η τυπική χρήση της ενεργητικής φωνής είναι να δηλώνει ότι το υποκείμενο του ρήματος ενεργεί – δηλ. προκαλεί την ενέργεια που εκφράζεται από το ρήμα. Η χαρακτηριστική χρήση της αντίστοιχης παθητικής είναι να δηλώνει ότι το υποκείμενο δέχεται την ενέργεια – η οποία εκφράζεται από το ρήμα – και/ή επηρεάζεται από αυτή. Ωστόσο πολλά ρήματα απαντούν σε μία μόνο από τις δύο φωνές και σε μερικά από αυτά η αντιστοιχία ανάμεσα στις ενεργητικές και παθητικές καταλήξεις και στη σημασία της ενέργειας ή της αποδοχής της δεν ισχύει. Ως παράδειγμα αναφέρονται τα αποθετικά ρήματα τα οποία έχουν παθητική μορφολογία αλλά ενεργητική σημασία (π.χ. *έρχομαι, ξεκουράζομαι, γίνομαι* κλπ). [3]

- **Έγκλιση**

Στη γραμματική των Holton – Mackridge – Φιλιππάκη [3], η έγκλιση ορίζεται ως “το σύνολο των μορφολογικών και συντακτικών αντιθέσεων οι οποίες σχετίζονται με σημασιολογικές διαφοροποιήσεις που υποδηλώνουν τον τρόπο με τον οποίο ο ομιλητής επιθυμεί να παρουσιάσει την πληροφορία της πρότασης, είτε δηλ. ως δήλωση ενός γεγονότος (οριστική) είτε ως επιθυμία (υποτακτική) είτε ως διαταγή (προστακτική).”

- **Μετοχές**

Ολοκληρώνοντας τα βασικά χαρακτηριστικά των ρημάτων της Νέας Ελληνικής τα οποία απαιτείται να γνωρίζει ο αναγνώστης, κρίνεται απαραίτητο να γίνει ιδιαίτερη αναφορά στις μετοχές.

Οι άκλιτες μετοχές (ή ‘γερούνδια’ όπως ονομάζονται στη γραμματική των Holton – Mackridge – Φιλιππάκη [3]) χαρακτηρίζονται ως προς τη φωνή και το χρόνο στον οποίο ανήκουν (πληροφορίες για τη φωνή και το χρόνο αναφέρθηκαν παραπάνω). Πρόκειται ουσιαστικά για τις ενεργητικές μετοχές ενεστώτα οι οποίες σχηματίζονται με την προσθήκη του επιθήματος –οντας στο μη συνοπτικό θέμα των ρημάτων της πρώτης συζυγίας (γράφ-οντας, τρέχ-οντας) ή –ώντας στο θέμα των ρημάτων της δεύτερης συζυγίας (μιλ-ώντας). Επίσης, όπως αναφέρεται στη γραμματική αυτή [3], “υπάρχουν και μερικές περιθωριακές περιπτώσεις γερούνδιων παθητικών ρημάτων, όπως το *όντας* από το ρήμα *είμαι*, το *διηγώντας* από το ρήμα *διηγούμαι* και ίσως και κάποια άλλα”.

Από την άλλη οι κλιτές μετοχές κλίνονται ως προς το γένος, τον αριθμό και την πτώση, ενώ βέβαια επίσης χαρακτηρίζονται από τη φωνή και το χρόνο στον οποίο ανήκουν. Όσον αφορά στους προαναφερθέντες νέους όρους, η γραμματική των Holton – Mackridge – Φιλιππάκη [3] παραθέτει τους ακόλουθους ορισμούς:

**Αριθμός:** Η διάκριση ανάμεσα σε ‘ενικό’ και ‘πληθυντικό.’ Μία μετοχή σε ενικό δηλώνει μόνο μία οντότητα, ενώ η μετοχή σε πληθυντικό δηλώνει περισσότερες από μία οντότητες.

**Γένος:** Μία από τις τρεις γραμματικές κατηγορίες στην οποία ανήκει κάθε ουσιαστικό: ‘αρσενικό’, ‘θηλυκό’ και ‘ουδέτερο’. Όλες οι λέξεις που προσδιορίζουν ένα ουσιαστικό (μεταξύ των οποίων και οι μετοχές) πρέπει να συμφωνούν μαζί του ως προς το γένος.

**Πτώση:** Ένας από τους τύπους ενός ουσιαστικού (ή κάθε άλλου μέλους του ονομαστικού συστήματος), ο οποίος υποδεικνύει τη συντακτική λειτουργία του ουσιαστικού στην πρόταση: η ‘ονομαστική’ υποδεικνύει το υποκείμενο, η ‘αιτιατική’ το άμεσο αντικείμενο και η ‘γενική’ κατά κανόνα το έμμεσο αντικείμενο, το συμπλήρωμα ενός ουσιαστικού ή τον κτήτορα. Η ‘κλητική’ χρησιμοποιείται όταν απευθυνόμαστε σε κάποιον ή κάτι.

Παραδείγματα κλιτών μετοχών αποτελούν οι ακόλουθες: *τρεχούμενος*, *διατεταγμένος*, *προβλεπόμενος*, *ερωτώμενος* κλπ.

## 4. ΓΛΩΣΣΙΚΑ ΔΕΔΟΜΕΝΑ ΤΟΥ ΕΡΓΑΛΕΙΟΥ

### 4.1. Γενικές Πληροφορίες

Όπως αναφέρθηκε και στο εισαγωγικό κεφάλαιο, το εργαλείο VERBTAGGR++ που δημιουργήθηκε στο πλαίσιο της παρούσας διπλωματικής εργασίας, χρησιμοποιεί γλωσσικά δεδομένα που προέρχονται από τη διδακτορική διατριβή της Λεμπέση [1]. Σχετικές με τη διατριβή αυτή δημοσιεύσεις είναι οι [5, 6, 7]. Το αντικείμενο της διατριβής αφορά στη μορφοσυντακτική αναγνώριση και λημματοποίηση των ρηματικών τύπων της Νέας Ελληνικής και κάποια στοιχεία για αυτήν θα παρουσιαστούν στην παρούσα ενότητα.

Στο πλαίσιο της έρευνας της Λεμπέση δημιουργήθηκαν συνολικά 10 αρχεία γλωσσικών δεδομένων. Για την υλοποίηση της εφαρμογής της παρούσας πτυχιακής, χρησιμοποιήθηκαν δύο βασικά αρχεία από τη διδακτορική της έρευνα: 1. Το αρχείο των 151.534 στατιστικά εντοπισθεισών ρηματικών καταληκτικών γραφηματικών σειρών εκτός αυτών που λήγουν σε –α άτονο και σε –ν, καθώς αυτές είχαν χρησιμοποιηθεί στην προηγούμενη πιλοτική έκδοση του εργαλείου VERBTAGGR, και 2. Το αρχείο των κανόνων λημματοποίησης που αφορούν σε όλες τις ρηματικές καταληκτικές γραφηματικές σειρές της Νέας Ελληνικής. Για την πληρέστερη κατανόηση όμως της εργασίας μας θα αναφερθούμε στη συνέχεια και εν συντομία σε δύο επιπλέον αρχεία της Λεμπέση, τα οποία αποτελούν τη βάση της διδακτορικής έρευνάς της.

#### (1) Αρχείο 8.485 Ρηματικών Λημμάτων της Νέας Ελληνικής

Το αρχείο των ρηματικών λημμάτων της Λεμπέση αποτελείται από λεξικοποιημένα (και όχι απλά πιθανά) ρήματα, όπως αυτά έχουν καταγραφεί στα λεξικά:

- *Επίτομο Λεξικό της Νεοελληνικής Γλώσσας (Το Μικρό Λεξικό του Υπερλεξικού)* (1991), Αθήνα, Εκδόσεις Παγουλάτου
- Τεγόπουλος - Φυτράκης (1993), *Ελληνικό Λεξικό*, Αθήνα, Εκδόσεις Αρμονία
- Κριαράς Ε. (1995), *Νέο Ελληνικό Λεξικό*, Αθήνα, Εκδοτική Αθηνών
- Μπαμπινιώτης Γ. (1998), *Λεξικό της Νέας Ελληνικής Γλώσσας*, Αθήνα, Κέντρο Λεξικολογίας
- *Λεξικό της Κοινής Νεοελληνικής* (1998), Θεσσαλονίκη, Ινστιτούτο Νεοελληνικών Σπουδών

Στις περιπτώσεις αλλομορφίας, μόνο ένας τύπος έχει οριστεί ως λήμμα, ο οποίος και εμφανίζεται στην εφαρμογή, ενώ οι αλλομορφικοί αναγνωρίζονται ως κλιτοί τύποι του ρήματος εισόδου.

#### (2) Νέο Ρηματικό Κλιτικό Σύστημα 385 κλιτικών υποδειγμάτων της Νέας Ελληνικής

Η δημιουργία νέου κλιτικού συστήματος στο πλαίσιο της έρευνας της Λεμπέση κατέστη αναπόφευκτη λόγω της ανάγκης μονοσήμαντης ντετερμινιστικής αντιστοίχισης κάθε ρηματικού λήμματος προς ένα και μόνο κλιτικό υπόδειγμα, το οποίο να περιλαμβάνει όλους τους ελευθέρως εναλλασσόμενους τύπους. Οι πληροφορίες κλίσης που έχουν καταγραφεί έχουν ως αφετηρία τα κλιτικά συστήματα των Ιορδανίδου (1992) και Κυριακοπούλου (1994) και έχουν εμπλουτισθεί με όλες τις λεξικογραφικά καταχωρημένες στα προαναφερθέντα λεξικά ενδείξεις. Η ομαδοποίηση των λημμάτων ανά κλιτικό υπόδειγμα διευκολύνθηκε από την κατά αντίστροφη αλφαβητική σειρά ταξινόμηση των δεδομένων του "*Αντίστροφου Λεξικού της Νέας Ελληνικής*" (Κουρμούλης 1967). Βασικό κριτήριο διαχωρισμού αποτελεί η ελλειπτική συμπεριφορά των λημμάτων, δηλ. το γεγονός ότι συγκεκριμένα ρήματα δε σχηματίζουν όλους τους δυνατούς κλιτούς τύπους (π.χ αποθετικά ρήματα με μεσοπαθητικούς μόνο τύπους). Ας σημειωθεί ότι για αυτό το είδος ρημάτων έχουν χρησιμοποιηθεί οι αριθμοί από 305 έως και 388, ενώ οι αριθμοί από 302 έως και 304 δεν εμφανίζονται στην εφαρμογή, εφόσον παραμένουν εν εφεδρεία ενόψει ενδεχόμενων νέων κλιτικών υποδειγμάτων της γλώσσας φωνής ενεργητικής.

Προκειμένου να φανεί η συχνότητα εμφάνισης των νέων κλιτικών υποδειγμάτων, παρουσιάζεται ο ακόλουθος πίνακας, ο οποίος περιέχει τον αριθμό των ρημάτων της Νέας Ελληνικής που κλίνονται σύμφωνα με τα νέα ορισθέντα κλιτικά υποδείγματα:

Πίνακας 2: Συχνότητα εμφάνισης κλιτικών υποδειγμάτων

ΑΡΙΘΜΟΣ ΡΗΜΑΤΩΝ	ΚΛΙΤΙΚΟ ΥΠΟΔΕΙΓΜΑ
1095	047
1092	045
801	055
520	001
404	002
208	012
176	317
157	211, 288
134	109
125	169
122	208
114	015
110	205
102	244
95	041
80	049
67	344
62	313
56	171
54	170
47	118
46	322
44	122
44	182
42	149, 172
40	209

39	335
36	212
34	088, 129, 218
33	044
32	018
31	040, 048
30	017
29	125
28	062
25	110, 275, 337
24	380
23	016, 090, 193, 357
22	004, 022, 083, 206
21	096, 195, 226
20	039
19	154
18	120, 173, 236, 341
16	006, 099, 140, 267
15	021, 061, 073, 098, 107
14	014, 138, 324, 346
13	028, 034, 068, 097, 139
12	030, 051, 101, 185, 221
11	081, 103, 214, 217, 234, 280, 336, 349, 385
10	009, 019, 085, 137, 160, 199, 200, 257, 340, 365
9	013, 027, 050, 089, 094, 095, 116, 119, 132, 174, 281 287
8	020, 111, 121, 124, 128, 133, 144, 203, 207, 223, 305, 355, 382
7	010, 063, 067, 084, 146, 155, 215, 299, 316, 328, 331
6	003, 069, 086, 102, 145, 163, 164, 167, 180, 233, 237, 254, 348, 353
5	033, 057, 058, 065, 087, 092, 105, 106, 112, 148, 159, 161, 178, 179, 191, 194, 222, 228, 239, 242, 243, 255, 270, 326, 329

4	032, 056, 072, 143, 152, 158, 184, 189, 210, 225, 240, 262, 268, 279, 300, 315, 327, 333, 334, 342, 345, 351, 361, 379, 383
3	008, 011, 023, 029, 043, 060, 070, 075, 077, 100, 113, 115, 150, 176, 183, 188, 196, 201, 216, 220, 245, 252, 256, 269, 271, 276, 282, 286, 297, 298, 319, 343, 347, 350, 352, 370
2	005, 024, 026, 036, 038, 042, 052, 054, 059, 064, 066, 078, 093, 123, 131, 135, 147, 168, 175, 181, 186, 197, 224, 227, 230, 238, 249, 258, 265, 278, 285, 291, 295, 306, 314, 318, 320, 323, 330, 338, 339, 354, 358, 359, 360, 362, 363, 373, 378, 381, 384, 386, 388
1	007, 025, 031, 035, 037, 046, 053, 071, 074, 076, 079, 080, 082, 091, 104, 108, 114, 117, 126, 127, 130, 134, 136, 141, 142, 151, 153, 156, 157, 165, 162, 166, 177, 187, 190, 192, 198, 202, 204, 213, 219, 229, 231, 232, 235, 241, 246, 247, 248, 250, 251, 253, 259, 260, 261, 263, 264, 266, 272, 273, 274, 277, 283, 284, 289, 290, 292, 293, 294, 296, 301, 307, 308, 309, 310, 311, 312, 321, 325, 332, 356, 364, 366, 367, 368, 369, 371, 372, 374, 375, 376, 377, 387

### (3) 151.534 Στατιστικά Εντοπισθείσες Ρηματικές Καταληκτικές Γραφηματικές Σειρές

Η μορφοσυντακτική αναγνώριση των τύπων σύμφωνα με το μοντέλο της Λεμπέση δεν πραγματοποιείται με την κλασική μέθοδο κατάτμησης σε θέμα-κλιτικό επίθημα (κατάληξη) και σύγκρισης δεδομένων εισόδου με δεδομένα μορφολογικού λεξικού. Το νεωτεριστικό στοιχείο της προσέγγισης συνίσταται στη διεύρυνση των ορίων της αντίστροφης (άνευ μορφολογικού λεξικού) μεθόδου με λεπτομερέστερη διείσδυση στο εσωτερικό της λέξης ξεκινώντας από το τέλος αυτής, η οποία έχει ως στόχο τον εντοπισμό μη αμφίσημων καταληκτικών γραφηματικών σειρών μορφής ευρύτερης των παραδοσιακών αμφίσημων κλιτικών επιθημάτων.

Στο πλαίσιο αυτό χρησιμοποιήθηκε από τη Λεμπέση το πρόγραμμα στατιστικής ανάλυσης του T. Girard, το οποίο υλοποίησε ως μέλος της ομάδας ανάπτυξης SYSTRAN-EE/TELINDUS στο Λουξεμβούργο για την αυτόματη ενσωμάτωση της Τράπεζας Ορολογίας EURODICAUTOM στα λεξικά του συστήματος αυτόματης μετάφρασης SYSTRAN-EE. Λαμβάνοντας ως είσοδο 519.694 αυτομάτως παραχθέντες (βάσει του νέου κλιτικού συστήματος) ρηματικούς τύπους, το πρόγραμμα εντόπισε 151.534 καταληκτικές γραφηματικές σειρές, με κριτήριο διαχωρισμού το διαφορετικό κλιτικό υπόδειγμα των τύπων εισόδου. Τα αποτελέσματα αυτά ολοκληρώθηκαν με πληροφορίες μορφοσυντακτικού περιεχομένου και λημματοποίησης που προσέθεσε η Λεμπέση.

Συνεπώς, η πληροφορία μορφοσυντακτικού περιεχομένου που παρέχεται από την εφαρμογή αφορά πρώτιστα στην εκάστοτε εντοπισθείσα καταληκτική γραφηματική σειρά και κατ' επέκταση στον υπό αναγνώριση ρηματικό τύπο, γεγονός το οποίο δεν αποκλείει την αναγνώριση και μη υπαρκτών ρηματικών τύπων.

#### (4) Σύστημα Κανόνων Λημματοποίησης

Ο κωδικός κλιτικού υποδείγματος που παρέχεται από την εφαρμογή είναι σημαντικός και για την ενεργοποίηση των κατάλληλων κανόνων λημματοποίησης βάσει των οποίων παράγεται το λήμμα του ρηματικού τύπου εισόδου, εφόσον και οι κανόνες αυτοί έχουν ταξινομηθεί σε συνάρτηση με το νέο κλιτικό σύστημα.

##### 4.2. Δομή Αρχείων Γλωσσικών Δεδομένων

Για τις ανάγκες του εργαλείου VERBTAGGR++ διατέθηκε από τη Λεμπέση το αρχείο των καταληκτικών γραφηματικών σειρών, καταμημένο σε μικρότερα αρχεία με βάση τον καταληκτικό χαρακτήρα τους (εκτός του άτονου –α και –ν, όπως έχει ήδη αναφερθεί): -ε, -η, -ι, -ο, -ς, -υ, -ω και -'. Αυτά τα αρχεία κειμένου όμοιας δομής περιέχουν γλωσσικά δεδομένα χρήσιμα στην αυτόματη μορφοσυντακτική αναγνώριση και λημματοποίηση των μονολεκτικών ρηματικών τύπων της Νέας Ελληνικής και στην παρούσα ενότητα περιγράφεται η δομή και το είδος των πληροφοριών που περιέχουν.

Σε γενικές γραμμές κάθε γραμμή αρχείου αφορά κάποιους ρηματικούς τύπους οι οποίοι λήγουν στο ίδιο σύνολο χαρακτήρων. Αυτό το σύνολο χαρακτήρων ονομάζεται 'καταληκτική γραφηματική σειρά' και αφορά έναν ή περισσότερους χαρακτήρες με τους οποίους λήγουν οι ρηματικοί τύποι που δηλώνει η καταληκτική γραφηματική σειρά. Όσον αφορά στη μορφή των καταληκτικών γραφηματικών σειρών που εμφανίζονται στα αρχεία των γλωσσικών δεδομένων, ενδέχεται στην αρχή να εμφανίζεται ένας αστερίσκος ('\*') ή ένα αγγλικό ερωτηματικό ('?'). Ο αστερίσκος αντιστοιχεί σε έναν ή περισσότερους χαρακτήρες, οπότε η αντίστοιχη καταληκτική γραφηματική σειρά αφορά στους ρηματικούς τύπους που λήγουν στους χαρακτήρες που ακολουθούν τον αστερίσκο. Το αγγλικό ερωτηματικό αντιστοιχεί σε κανέναν, έναν ή περισσότερους χαρακτήρες, οπότε η αντίστοιχη καταληκτική γραφηματική σειρά δηλώνει το ρηματικό τύπο που αποτελείται ακριβώς από τους χαρακτήρες που ακολουθούν το ερωτηματικό καθώς και τους ρηματικούς τύπους που λήγουν σε αυτούς τους χαρακτήρες. Εάν μία καταληκτική γραφηματική σειρά δεν αρχίζει ούτε με αστερίσκο, ούτε με ερωτηματικό, τότε συνιστά έναν αυτούσιο ρηματικό τύπο της γλώσσας, αυτόν που αποτελείται από όλους τους χαρακτήρες της γραφηματικής σειράς.

Οι τύποι που λήγουν σε μία καταληκτική γραφηματική σειρά έχουν ένα ορισμένο μορφοσυντακτικό περιεχόμενο και ανήκουν σε κάποιο κλιτικό υπόδειγμα, ενώ το λήμμα τους παράγεται με χρήση ενός ορισμένου κωδικού λημματοποίησης. Το μορφοσυντακτικό περιεχόμενο αναπαρίσταται στα αρχεία με τη μορφή ενός κωδικού αριθμού, ενώ οι διάφορες καταληκτικές γραφηματικές σειρές αναπαρίστανται με λατινικούς χαρακτήρες. Η αντιστοιχία που ισχύει μεταξύ των ελληνικών και των λατινικών χαρακτήρων που εμφανίζονται στις καταληκτικές γραφηματικές σειρές έχει ως εξής:

A	→	A
B	→	V
Γ	→	G
Δ	→	D
E	→	E
Z	→	Z
H	→	I
Θ	→	C
I	→	J



Κ	→	Κ
Λ	→	Λ
Μ	→	Μ
Ν	→	Ν
Ξ	→	Χ
Ο	→	Ο
Π	→	Ρ
Ρ	→	Ρ
Σ	→	Σ
Τ	→	Τ
Υ	→	Υ
Φ	→	Φ
Χ	→	Χ
Ψ	→	Β
Ω	→	Ω
'	→	~
İ	→	J!
ÿ	→	Y!
ı	→	J!~
·ÿ	→	Y!~

Στη συνέχεια θα παρουσιαστούν ενδεικτικές γραμμές από τα αρχεία των γλωσσικών δεδομένων μέσα από τις οποίες θα γίνει εμφανής η ακριβής μορφή των αρχείων.

Ένα μεγάλο πλήθος γραμμών έχουν μορφή όπως αυτή που εμφανίζεται στο παράδειγμα 4.1:

➤ Παράδειγμα 4.1:

---

057 218,\*ABA 44% 39 2b otherwise -> (\*BA,211)

---

Οι πληροφορίες που παρέχονται από τη συγκεκριμένη γραμμή είναι οι ακόλουθες:

- Ο αριθμός 057 είναι τιμή μορφοσυντακτικού κωδικού.
- Ο αριθμός 218 είναι τιμή κλιτικού υποδείγματος.
- Οι χαρακτήρες \*ABA αποτελούν την καταληκτική γραφηματική σειρά στην οποία αφορούν οι πληροφορίες που περιέχει ολόκληρη η γραμμή.
- Το 44% εκφράζει τη σχετική συχνότητα, ενώ ο αριθμός 39 την απόλυτη συχνότητα των ρηματικών τύπων στους οποίους αφορά η γραμμή.
- Οι χαρακτήρες 2b αποτελούν κωδικό ληματοποίησης.

Επομένως, σύμφωνα με την παραπάνω γραμμή, από τους καταγεγραμμένους ρηματικούς τύπους που λήγουν σε *-αψα*, και οι 39 έχουν μορφοσυντακτικό περιεχόμενο που αναπαρίσταται από τον κωδικό 057, ανήκουν στο κλιτικό υπόδειγμα 218 και ο κωδικός που θα χρησιμοποιηθεί για τη λημματοποίηση είναι ο 2b. Αυτοί οι τύποι αποτελούν εξαίρεση των τύπων που λήγουν σε *-ψα*, οι οποίοι ανήκουν στο κλιτικό υπόδειγμα 211.

Σημειώνεται ότι η τιμή της σχετικής συχνότητας είναι δυνατόν να απουσιάζει, οπότε υπονοείται η τιμή 100%.

Σε περίπτωση που σε μία καταληκτική γραφηματική σειρά αντιστοιχούν περισσότεροι από ένας μορφοσυντακτικοί κωδικοί, οι γραμμές που εμφανίζονται είναι όπως αυτές του παραδείγματος 4.2:

➤ Παράδειγμα 4.2:

---

2msc 193,\*CA 14% 46 otherwise -> (\*A,47)

\*\*\*\*\*23/46=042:\*8 + 23/46=080:8

---

Σύμφωνα με τις παραπάνω γραμμές, υπάρχουν 46 καταγεγραμμένοι ρηματικοί τύποι που λήγουν σε *-θα* και ανήκουν στο κλιτικό υπόδειγμα 193. Αυτοί οι τύποι αποτελούν εξαίρεση των τύπων που λήγουν σε *-α*, οι οποίοι ανήκουν στο κλιτικό υπόδειγμα 47. Από αυτούς τους 46 οι 23 έχουν μορφοσυντακτικό περιεχόμενο που αναπαρίσταται από τον κωδικό 042 και το λήμμα τους παράγεται με χρήση του κωδικού \*8, ενώ οι υπόλοιποι 23 έχουν μορφοσυντακτικό περιεχόμενο που αναπαρίσταται από τον κωδικό 080 και κωδικό λημματοποίησης ίσο με 8.

Εάν αντιθέτως σε μία καταληκτική γραφηματική σειρά αντιστοιχούν περισσότεροι από ένας μορφοσυντακτικοί κωδικοί, αλλά όλοι οι αντίστοιχοι ρηματικοί τύποι έχουν τον ίδιο κωδικό λημματοποίησης, τότε ο κωδικός λημματοποίησης εμφανίζεται στη βασική γραμμή που αρχίζει με τους χαρακτήρες 'Xmsc' (όπου X ακέραιος αριθμός) και όχι στην ακόλουθη γραμμή που αρχίζει με αστερίσκους, όπως συμβαίνει στην παραπάνω περίπτωση. Ένα παράδειγμα αυτής της περίπτωσης αποτελεί το 4.3:

➤ Παράδειγμα 4.3:

---

2msc 122,\*Y~NAN 33% 43 4b otherwise -> (\*NAN,55)

\*\*\*\*\* 7/43=086+ 36/43=062

---

Σύμφωνα με αυτές τις γραμμές, υπάρχουν 43 καταγεγραμμένοι ρηματικοί τύποι που λήγουν σε *-ύναν* και ανήκουν στο κλιτικό υπόδειγμα 122. Όλοι αυτοί οι τύποι έχουν κωδικό λημματοποίησης τον 4b και αποτελούν εξαίρεση των τύπων που λήγουν σε *-ναν*, οι οποίοι ανήκουν στο κλιτικό υπόδειγμα 55. Από αυτούς τους 43 οι 7 έχουν μορφοσυντακτικό περιεχόμενο που αναπαρίσταται από τον κωδικό 086, ενώ οι υπόλοιποι 36 έχουν μορφοσυντακτικό περιεχόμενο που αναπαρίσταται από τον κωδικό 062.

Επιπλέον, στις περιπτώσεις γραμμών με περισσότερους από έναν μορφοσυντακτικούς κωδικούς, είναι δυνατόν στη γραμμή με τους αστερίσκους να αναφέρονται εντός παρενθέσεων ένας ή περισσότεροι συγκεκριμένοι ρηματικοί τύποι διαχωριζόμενοι με το χαρακτήρα '/', τους οποίους αφορά ο συγκεκριμένος μορφοσυντακτικός κωδικός. Μία τέτοια περίπτωση εμφανίζεται στις γραμμές του παραδείγματος 4.4:

➤ Παράδειγμα 4.4:

---

2msc 122,\*TYNA 57% 8 3b otherwise -> (\*A,47)

\*\*\* 2/8=081 (PLA~TYNA/LE~PTYNA) + 6/8=057

---

Βάσει των παραπάνω πληροφοριών υπάρχουν καταγεγραμμένοι 8 ρηματικοί τύποι που λήγουν σε *-τυνα*, οι οποίοι ανήκουν στο κλιτικό υπόδειγμα 122, έχουν κωδικό λημματοποίησης τον 3b και αποτελούν εξαίρεση των τύπων που λήγουν σε *-α*, οι οποίοι ανήκουν στο κλιτικό υπόδειγμα 47. Από αυτούς τους 8 οι 2 είναι οι ρηματικοί τύποι 'πλάτυνα' και 'λέπτυνα', των οποίων η μορφοσυντακτική αξία αναπαρίσταται από τον κωδικό 081, ενώ η μορφοσυντακτική αξία των υπόλοιπων 6 αναπαρίσταται από τον κωδικό 057.

Μία ξεχωριστή περίπτωση γραμμών στις οποίες αντιστοιχούν περισσότεροι από ένας μορφοσυντακτικοί κωδικοί είναι οι γενικότερες γραμμές των αρχείων οι οποίες αφορούν σε ρηματικούς τύπους που δεν αποτελούν εξαίρεση άλλων τύπων και ως εκ τούτου εμφανίζουν την ένδειξη 'no otherwise'. Στις γραμμές αυτές λόγω του μεγάλου αριθμού των τύπων που καλύπτονται δεν αναφέρεται πληροφορία για τη συχνότητα εμφάνισης του κάθε μορφοσυντακτικού κωδικού χωριστά. Ως παράδειγμα αναφέρονται οι γραμμές του παραδείγματος 4.5 οι οποίες αφορούν στην καταληκτική γραφηματική σειρά *-α* και εμφανίζονται στην αρχή του αρχείου που περιέχει πληροφορίες για τους ρηματικούς τύπους που λήγουν σε *-α* άτονο:

➤ Παράδειγμα 4.5:

---

7msc 47,\*A 19% 7921 no otherwise

\*\*\*010:5a,020:5b,033:1a,039:6a,040:6a,057:2b,064:5b

---

Οι πληροφορίες που παρέχουν οι παραπάνω γραμμές είναι ότι από τους καταγεγραμμένους ρηματικούς τύπους που λήγουν σε *-α* οι 7.921 ανήκουν στο κλιτικό υπόδειγμα 47. Η μορφοσυντακτική αξία αυτών των τύπων αναπαρίσταται συνολικά από τους κωδικούς 010, 020, 033, 039, 040, 057, 064, ενώ οι αντίστοιχοι κωδικοί λημματοποίησης είναι οι 5a, 5b, 1a, 6a, 6a, 2b, 5b. Ωστόσο, όπως είναι προφανές λόγω του εκτεταμένου αριθμού των περιπτώσεων ιδίου κλιτικού υποδείγματος που καλύπτονται (εφόσον το κλιτικό υπόδειγμα 47 είναι το συχνότερο όλων), δεν παρέχεται πληροφορία για το ακριβές πλήθος των τύπων στο οποίο αφορά ο κάθε μορφοσυντακτικός κωδικός. Το κενό αυτό μπορεί να συμπληρωθεί μόνο κατόπιν περαιτέρω στατιστικής ανάλυσης τύπων ιδίου κλιτικού υποδείγματος αλλά διαφορετικής μορφοσυντακτικής αξίας.

Μία ακόμη περίπτωση γραμμών αποτελούν αυτές που ξεκινούν με τους χαρακτήρες '(+'. Μία τέτοια γραμμή εκφράζει ότι για την καταληκτική γραφηματική σειρά της συγκεκριμένης γραμμής ισχύουν επιπλέον οι πληροφορίες που παρέχονται από την αμέσως προηγούμενη βασική γραμμή που δεν ξεκινά με τους χαρακτήρες '(+'. Ένα παράδειγμα της περίπτωσης αυτής είναι το 4.6:

➤ Παράδειγμα 4.6:

---

064 72,\*RASTA~CIKA 50% 2 5b otherwise -> (\*ASTA~CIKA,189)  
(+064 261,PARASTA~CIKA 1 8b otherwise -> (\*RASTA~CIKA,72)MA)

---

Οι παραπάνω γραμμές, εκτός των βασικών πληροφοριών για τους τύπους που λήγουν σε *-ραστάθηκα*, εκφράζουν ότι ο τύπος *παραστάθηκα* εμφανίζεται σε δύο κλιτικά υποδείγματα δύο διαφορετικών ρηματικών λημμάτων με την ίδια γραμματική αξία, διαφορετική όμως σημασιολογική. Με άλλα λόγια πρόκειται για δύο ομόγραφους κλιτούς τύπους που προέρχονται από διαφορετικά ρήματα. Συγκεκριμένα:

- Ο ένας τύπος *παραστάθηκα* ανήκει στο κλιτικό υπόδειγμα 72, έχει μορφοσυντακτικό κωδικό 064, κωδικό λημματοποίησης 5b και αποτελεί εξαίρεση των τύπων που λήγουν σε *-αστάθηκα*.
- Ο άλλος τύπος *παραστάθηκα* ανήκει στο κλιτικό υπόδειγμα 261, έχει μορφοσυντακτικό κωδικό 064, κωδικό λημματοποίησης 8b και αποτελεί εξαίρεση των τύπων που λήγουν σε *-ραστάθηκα*.

Πληροφοριακά αναφέρεται ότι τα δύο λήμματα κατά την κλίση των οποίων παράγεται ο τύπος *παραστάθηκα* είναι το *παραστέκω* (από το κλιτικό υπόδειγμα 72) και το *παριστάνω* / *παριστώ* (από το κλιτικό υπόδειγμα 261).

Σημείωση:

Όπως φαίνεται από το παραπάνω παράδειγμα, είναι δυνατόν μία γραμμή να περιέχει στο τέλος αυτής την ένδειξη 'MA' ή 'MM'. Η ένδειξη 'MA' (*manually added*) σημαίνει ότι η συγκεκριμένη γραμμή προστέθηκε από την ίδια τη Λεμπέση και όχι από το πρόγραμμα στατιστικής ανάλυσης του T. Girard, όπως και όλες οι γραμμές που περιέχουν σχόλια και εισάγονται με αστερίσκους, ενώ η ένδειξη 'MM' (*manually modified*) σημαίνει ότι η Λεμπέση πραγματοποίησε κάποια αλλαγή στη γραμμή που παράχθηκε από το πρόγραμμα στατιστικής ανάλυσης του T. Girard.

Άλλο ένα παράδειγμα γραμμής που ξεκινά με τους χαρακτήρες '(+' είναι το 4.7:

➤ Παράδειγμα 4.7:

---

057 165,CA~RREBA 1 3 otherwise -> (\*RREBA,211)  
(+057 208,CA~RREBA 1 2b otherwise -> (\*RREBA,211)MA)

---

Βάσει των γραμμών αυτών ο τύπος *θάρρεψα* επίσης παράγεται κατά την κλίση δύο διαφορετικών λημμάτων και έχουν την ίδια γραμματική αλλά διαφορετική σημασιολογική

αξία. Έτσι, ο ένας τύπος *θάρρεψα* ανήκει στο κλιτικό υπόδειγμα 165, που περιλαμβάνει το μοναδικό λήμμα *θαρρώ*, έχει μορφοσυντακτικό κωδικό 057, κωδικό λημματοποίησης 3 και αποτελεί εξαίρεση των τύπων που λήγουν σε *-ρρεψα*, ενώ ο άλλος ομόγραφος τύπος *θάρρεψα* ανήκει στο ευρύτερα εμφανιζόμενο κλιτικό υπόδειγμα 208 (που καλύπτει 123 ρηματικά λήμματα μεταξύ των οποίων και το ρήμα *θαρρεύω*), έχει μορφοσυντακτικό κωδικό 057, κωδικό λημματοποίησης 2b και αποτελεί εξαίρεση επίσης των τύπων που λήγουν σε *-ρρεψα*.

Επιπλέον ενδέχεται να εμφανιστεί γραμμή που αρχίζει με τους χαρακτήρες ‘+’ η οποία περιέχει διαφορετικό μορφοσυντακτικό κωδικό από αυτόν της βασικής της γραμμής που προηγείται (παράδειγμα 4.8):

➤ Παράδειγμα 4.8:

---

081 173,\*ERA 50% 20 1 otherwise -> (\*RA,169)  
(+057 147,E~FERA 1 1b otherwise -> (\*ERA,173)MA)

---

Για τον τύπο *έφεραισχούν* οι πληροφορίες από την πρώτη (βασική) γραμμή:

μορφοσυντακτικός κωδικός → 081

κλιτικό υπόδειγμα → 173

κωδικός λημματοποίησης → 1

(λήμμα → *φέρω*)

εξαίρεση των τύπων που λήγουν σε → *-ρα*

αλλά και οι πληροφορίες από τη δεύτερη γραμμή:

μορφοσυντακτικός κωδικός → 057

κλιτικό υπόδειγμα → 147

κωδικός λημματοποίησης → 1b

(λήμμα → *φέρνω*)

εξαίρεση των τύπων που λήγουν σε → *-ερα*.

Επίσης είναι δυνατόν να υπάρξει συνδυασμός των γραμμών που ξεκινούν με χαρακτήρες τύπου ‘Xmsc’ (όπου X ακέραιος αριθμός) με γραμμές που ξεκινούν με ‘+’ όπως συμβαίνει στο παράδειγμα 4.9:

➤ Παράδειγμα 4.9:

---

2msc 90,\*LA 17% 52 otherwise -> (\*A,47)

\*\*\*\*\*26/52=057:1b + 26/52=033:1a

(+057 50,E~VALA 1 2b otherwise -> (\*LA,90)MA)

---

Οι πληροφορίες που αντλούνται από το παράδειγμα 4.9 είναι ότι έχουν καταγραφεί 52 ρηματικοί τύποι που λήγουν σε *-λα*, οι οποίοι ανήκουν στο κλιτικό υπόδειγμα 90 και αποτελούν εξαίρεση των τύπων που λήγουν σε *-α*. Από αυτούς οι 26 έχουν μορφοσυντακτικό κωδικό 057 και κωδικό λημματοποίησης 1b, ενώ οι υπόλοιποι 26 έχουν μορφοσυντακτικό κωδικό 033 και κωδικό λημματοποίησης 1a. Όσον αφορά ειδικότερα στο ρηματικό τύπο *έβαλα*, ισχύουν κατ' αρχάς οι πληροφορίες που αντλούνται από την τελευταία γραμμή, δηλαδή:

μορφοσυντακτικός κωδικός → 057

κλιτικό υπόδειγμα → 50

κωδικός λημματοποίησης → 2b

(λήμμα → *βάλλω*)

εξαίρεση των τύπων που λήγουν σε → *-λα*.

Επίσης, εφόσον η γραμμή του *έβαλα* ξεκινά με '(+', γίνεται αντιληπτό ότι υπάρχει καταγεγραμμένος άλλος ένας ρηματικός τύπος *έβαλα*, για τον οποίο πληροφορίες θα αντληθούν από τις προηγούμενες γραμμές. Δεδομένου ότι στην προηγούμενη βασική γραμμή αντιστοιχούν περισσότεροι από ένας μορφοσυντακτικοί κωδικοί, για τον τύπο *έβαλα* θα ληφθούν υπόψη οι πληροφορίες που αφορούν στο μορφοσυντακτικό κωδικό 057, διότι αυτός είναι ο κωδικός που επίσης εμφανίζεται στη γραμμή που ξεκινά με '(+' και αφορά στο *έβαλα* Επομένως, για τον τύπο *έβαλα* ισχύουν επιπλέον οι πληροφορίες:

μορφοσυντακτικός κωδικός → 057

κλιτικό υπόδειγμα → 90

κωδικός λημματοποίησης → 1b

εξαίρεση των τύπων που λήγουν σε → *-α*.

Τέλος, αναφέρεται ότι ενδέχεται να υπάρχουν στη σειρά περισσότερες από μια γραμμές που όλες ξεκινούν με τους χαρακτήρες '(+', γεγονός το οποίο σημαίνει ότι για όλες τις καταληκτικές γραφηματικές σειρές που εμφανίζονται σε αυτές τις γραμμές, ισχύουν επιπλέον πληροφορίες που αντλούνται από τη βασική γραμμή που υπάρχει πριν από τις γραμμές που ξεκινούν με '(+':

➤ Παράδειγμα 4.10:

---

064 188,\*ASTA~ΗΤΙΚΑ 60% 3 6 otherwise -> (\*STA~ΗΤΙΚΑ,48)

(+064 48,VASTA~ΗΤΙΚΑ 1 5b otherwise -> (\*ASTA~ΗΤΙΚΑ,188)MA)

(+064 48,ANAVASTA~ΗΤΙΚΑ 1 5b otherwise -> (\*ASTA~ΗΤΙΚΑ,188)MA)

---

Για τον τύπο *βαστάχτηκα* ισχύουν πληροφορίες που αντλούνται από τη 2<sup>η</sup> γραμμή:

μορφοσυντακτικός κωδικός → 064

κλιτικό υπόδειγμα → 48

κωδικός λημματοποίησης → 5b

(λήμμα → *βαστάζω*)εξαίρεση των τύπων που λήγουν σε → *-αστάχτηκα*

καθώς και οι πληροφορίες που αντλούνται από τη βασική γραμμή:

μορφοσυντακτικός κωδικός → 064

κλιτικό υπόδειγμα → 188

κωδικός λημματοποίησης → 6

(λήμμα → *βαστώ*)

εξαίρεση των τύπων που λήγουν σε → *-στάχτηκα*.

Όμοια για τον τύπο *αναβαστάχτηκα* ισχύουν πληροφορίες που αντλούνται από την 3<sup>η</sup> γραμμή:

μορφοσυντακτικός κωδικός → 064

κλιτικό υπόδειγμα → 48

(λήμμα → *αναβαστάζω*)

κωδικός λημματοποίησης → 5b

εξαίρεση των τύπων που λήγουν σε → *-αστάχτηκα*

καθώς και οι πληροφορίες που αντλούνται από τη βασική γραμμή:

μορφοσυντακτικός κωδικός → 064

κλιτικό υπόδειγμα → 188

κωδικός λημματοποίησης → 6

(λήμμα → *αναβαστώ*)

εξαίρεση των τύπων που λήγουν σε → *-στάχτηκα*.

### 4.3 Έλεγχος των ήδη υπάρχοντων Γλωσσικών Δεδομένων

Στο πλαίσιο της παρούσας πτυχιακής τα αρχεία .sql, τα οποία είχαν υλοποιηθεί από την προηγούμενη πιλοτική διπλωματική εργασία της Κρουστάλλη, και περιείχαν τα δεδομένα με καταλήξεις σε '-α' άτονο και '-ν', ελέγχθηκαν, διορθώθηκαν και ξαναπεράστηκαν στη βάση δεδομένων.

Εν συνεχεία, με τη βοήθεια του προγράμματος dataProcess, το οποίο δημιουργήσαμε εκ νέου, ελέγξαμε όλα τα αρχεία που περιέχουν τα γλωσσικά/στατιστικά δεδομένα των καταληκτικών γραφηματικών σειρών που δεν λήγουν σε '-α' άτονο και '-ν' και σε συνεργασία με την κα Π. Λεμπέση έγιναν μεμονωμένες διορθώσεις. Τέλος, δημιουργήθηκαν επιπλέον αρχεία .sql, τα οποία περιέχουν επερωτήσεις (queries) για την εισαγωγή των υπολοίπων αρχείων στη βάση δεδομένων.

## 5. ΠΕΡΙΓΡΑΦΗ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ ΑΝΑΓΝΩΡΙΣΗΣ ΡΗΜΑΤΙΚΩΝ ΤΥΠΩΝ

Τα δεδομένα που προήλθαν από τα αρχεία γλωσσικών δεδομένων των οποίων η δομή παρουσιάστηκε στο προηγούμενο κεφάλαιο (βλ. §4.2.), αποθηκεύονται σε μία βάση δεδομένων MySQL, προκειμένου να χρησιμοποιηθούν από τη διαδικτυακή εφαρμογή VERBTAGGR++. Η βάση ονομάζεται verbDB και αποτελείται από τρεις πίνακες, οι οποίοι περιγράφονται στις επόμενες ενότητες.

### 5.1. Πίνακας ms\_codeExplain

Περιλαμβάνει πληροφορίες σχετικές με την ερμηνεία των διαφόρων μορφοσυντακτικών κωδικών, τιμές των οποίων περιέχονται στο πεδίο 'ms\_code' του πίνακα verb\_info, ο οποίος περιγράφεται παρακάτω. Συγκεκριμένα, κάθε γραμμή αυτού του πίνακα εκφράζει και μία διαφορετική ερμηνεία ενός μορφοσυντακτικού κωδικού. Τα πεδία από τα οποία αποτελείται είναι τα εξής:

- **primaryKey**: Περιέχει έναν αύξοντα αριθμό με τιμή διαφορετική για κάθε γραμμή του πίνακα. Χρησιμοποιείται ως πρωτεύον κλειδί.
- **ms\_code**: Περιέχει συμβολοσειρά που εκφράζει ένα μορφοσυντακτικό κωδικό. Οι τιμές του κυμαίνονται από '000' μέχρι '102'. Σημειώνεται ότι η γραμμή εκφράζει μία ερμηνεία αυτού του κωδικού.
- **person\_partcplInfo**: Εάν ο μορφοσυντακτικός κωδικός της γραμμής αφορά σε ρήμα, τότε το συγκεκριμένο πεδίο περιέχει πληροφορίες για το πρόσωπο και τον αριθμό του ρήματος, ενώ αν ο κωδικός αφορά σε μετοχή, το πεδίο περιέχει πληροφορίες για το γένος, την πτώση και τον αριθμό της μετοχής. Συγκεκριμένα:
  - Αν η ερμηνεία αφορά σε ρήμα, το πεδίο περιέχει μία από τις συμβολοσειρές 1S, 2S, 3S, 1P, 2P, 3P, οι οποίες αντίστοιχα εκφράζουν τις πληροφορίες 1<sup>ου</sup> ενικού, 2<sup>ου</sup> ενικού, 3<sup>ου</sup> ενικού, 1<sup>ου</sup> πληθυντικού, 2<sup>ου</sup> πληθυντικού και 3<sup>ου</sup> πληθυντικού προσώπου.
  - Αν η ερμηνεία αφορά κλιτή μετοχή, το πεδίο περιέχει μία συμβολοσειρά τριών χαρακτήρων εκ των οποίων ο πρώτος εκφράζει το γένος της μετοχής ('M' για το αρσενικό γένος, 'F' για το θηλυκό, 'N' για το ουδέτερο), ο δεύτερος εκφράζει την πτώση της μετοχής ('N' για την ονομαστική πτώση, 'G' για τη γενική, 'A' για την αιτιατική), ενώ ο τρίτος εκφράζει τον αριθμό ('S' για τον ενικό, 'P' για τον πληθυντικό).
  - Αν η ερμηνεία αφορά σε άκλιτη μετοχή, το πεδίο περιέχει τη συμβολοσειρά 'invariable'.
- **tense**: Πρόκειται για πληροφορία που αφορά στο χρόνο. Οι δυνατές τιμές είναι οι ακόλουθες συμβολοσειρές:
  - present: για τον ενεστώτα
  - past: για τον αόριστο
  - imperfect: για τον παρατατικό
  - perfective: για ρηματικούς τύπους συνοπτικού μη παρελθοντικού χρόνου
  - perfective perfect formant: για ρηματικούς τύπους συνοπτικού μη παρελθοντικού χρόνου που μπορούν να χρησιμοποιηθούν στο σχηματισμό των συντελικών χρόνων.



Υπενθυμίζεται ότι, όπως αναφέρθηκε στο κεφάλαιο 3, οι συνοπτικοί μη παρελθοντικοί τύποι δεν απαντούν ως ανεξάρτητοι ρηματικοί τύποι, γι' αυτό και δεν μπορεί να εξαχθεί συμπέρασμα για την έγκλιση ή το χρόνο τους. Για το λόγο αυτό, οι γραμμές του πίνακα `ms_codeExplain` που αφορούν σε μορφοσυντακτικούς κωδικούς οι οποίοι αντιστοιχούν σε συνοπτικούς μη παρελθοντικούς τύπους, περιέχουν στο πεδίο `tense` μία κατάλληλη ένδειξη και όχι ένα συγκεκριμένο χρόνο.

- **voice:** Αφορά σε πληροφορία για τη φωνή. Έτσι οι δυνατές συμβολοσειρές που μπορεί να περιέχονται στο συγκεκριμένο πεδίο είναι 'active' για την ενεργητική φωνή ή 'passive' για τη μεσοπαθητική.
- **mood:** Αφορά σε πληροφορία για την έγκλιση. Οι δυνατές τιμές είναι 'indicative' για την οριστική, 'imperative' για την προστακτική ή NULL, όταν η γραμμή αφορά σε μετοχή ή συνοπτικό μη παρελθοντικό τύπο. (Όπως αναφέρθηκε παραπάνω, οι συνοπτικοί μη παρελθοντικοί τύποι δεν εκφράζουν μία συγκεκριμένη έγκλιση.)
- **type:** Λαμβάνει ως τιμή τη συμβολοσειρά 'verb', στην περίπτωση που η γραμμή αφορά σε ρήμα ή τη συμβολοσειρά 'participle', αν η γραμμή αφορά σε μετοχή.

Σημειώνεται ότι τα δεδομένα που περιέχονται στον πίνακα `ms_codeExplain` προέρχονται από υλικό που διέθεσε η Λεμπέση και οι SQL εντολές που εισάγουν αυτά τα δεδομένα περιέχονται στο αρχείο `SYNODEΥΤΙΚΟ_YΛΙΚΟ\DB_files\createDB.sql` του `cd`. Μερικές ενδεικτικές γραμμές του πίνακα αυτού παρουσιάζονται στον ακόλουθο πίνακα:

Πίνακας 3: Παραδείγματα γραμμών του πίνακα `ms_codeExplain`

primaryKey	ms_code	person_partcplInfo	tense	voice	mood	type
1	000	invariable	present	active	NULL	participle
5	003	MAS	present	passive	NULL	participle
6	003	NNS	present	passive	NULL	participle
7	003	NAS	present	passive	NULL	participle
88	064	1S	past	passive	indicative	verb

## 5.2. Πίνακας `verb_info`

Περιέχει τα κυρίως δεδομένα της εφαρμογής, τα οποία περιλαμβάνουν πληροφορίες για τις διάφορες καταληκτικές γραφηματικές σειρές των ρηματικών τύπων, όπως αυτές προκύπτουν από τα αρχεία γλωσσικών δεδομένων. Κάθε γραμμή του πίνακα αφορά σε ένα μορφοσυντακτικό κωδικό μίας καταληκτικής γραφηματικής σειράς. Τα πεδία που ο πίνακας περιέχει είναι τα εξής:

- **final\_grapheme\_seq:** Περιέχει μία συμβολοσειρά που αναπαριστά την καταληκτική γραφηματική σειρά στην οποία αφορούν οι πληροφορίες που περιέχει ολόκληρη η γραμμή.

- **model**: Περιέχει μία ή περισσότερες τιμές κλιτικών υποδειγμάτων χωρισμένες μεταξύ τους με παύλα. Η τιμή κάθε κλιτικού υποδείγματος κυμαίνεται από 1 έως 388 εκτός των 302, 303, 304.
- **ms\_code**: Περιέχει μία συμβολοσειρά που αναπαριστά μορφοσυντακτικό κωδικό. Η τιμή αυτή είναι κάποια από τις συμβολοσειρές που περιέχονται στο πεδίο `ms_code` του πίνακα `ms_codeExplain` με έναν αστερίσκο (\*) πιθανόν να προηγείται. Ο αστερίσκος εκφράζει πληροφορία που είναι χρήσιμη στη διαδικασία της λημματοποίησης.
- **relativeFreq**: Περιέχει μία συμβολοσειρά που λήγει με το χαρακτήρα '%' και εκφράζει τιμή σχετικής συχνότητας.
- **absoluteFreq**:
  - Εάν η γραμμή αφορά σε καταληκτική γραφηματική σειρά η οποία δηλώνει περισσότερους από έναν μορφοσυντακτικούς κωδικούς (το οποίο γίνεται αντιληπτό από το ότι ο πίνακας `verb_info` περιέχει περισσότερες από μία γραμμές για τη συγκεκριμένη γραφηματική σειρά), το πεδίο αυτό περιέχει το πλήθος των ρηματικών τύπων που λήγουν στη συγκεκριμένη γραφηματική σειρά και χαρακτηρίζονται από το μορφοσυντακτικό κωδικό της γραμμής σε σχέση με το συνολικό πλήθος των τύπων που λήγουν στη συγκεκριμένη γραφηματική σειρά. Συγκεκριμένα στην περίπτωση αυτή περιέχεται συμβολοσειρά της μορφής 'x/y' όπου x, y ακέραιοι αριθμοί εκφράζοντας έτσι ότι οι x από τους y ρηματικούς τύπους που λήγουν στη γραφηματική σειρά του πεδίου `final_grapheme_seq` καλύπτονται από το μορφοσυντακτικό κωδικό του πεδίου `ms_code`.
  - Εάν η γραμμή αφορά σε καταληκτική γραφηματική σειρά η οποία δηλώνει ένα μόνο μορφοσυντακτικό κωδικό (το οποίο γίνεται αντιληπτό από το ότι ο πίνακας `verb_info` περιέχει ακριβώς μία γραμμή για τη συγκεκριμένη γραφηματική σειρά), το πεδίο περιέχει μία ή περισσότερες τιμές απόλυτων συχνοτήτων χωρισμένες μεταξύ τους με παύλα, οι οποίες κατά σειρά αντιστοιχούν στα κλιτικά υποδείγματα που περιέχονται στο πεδίο `model`.
- **lemmaCode**: Περιέχει έναν ή περισσότερους κωδικούς λημματοποίησης χωρισμένους μεταξύ τους με παύλα. Στην περίπτωση που περιέχονται περισσότεροι από ένας κωδικοί, οι τιμές τους αντιστοιχούν κατά σειρά στα κλιτικά υποδείγματα που περιέχονται στο πεδίο `model` και στις απόλυτες συχνότητες που περιέχονται στο πεδίο `absoluteFreq`. Σημειώνεται ότι ο κωδικός λημματοποίησης είναι απαραίτητος για την παράγωγη του λήμματος των κλιτών ρηματικών τύπων κι έτσι δε χρησιμοποιείται από την παρούσα έκδοση της εφαρμογής VERBTAGGR++, η οποία επιτελεί μόνο τη μορφοσυντακτική αναγνώριση των μονολεκτικών ρηματικών τύπων. Ωστόσο, φυλάσσεται ως πληροφορία στη βάση της εφαρμογής, έτσι ώστε εύκολα μελλοντικά να δημιουργηθεί μία νέα έκδοση της εφαρμογής, η οποία θα υποστηρίξει και τη λημματοποίηση των ρηματικών τύπων.
- **exceptionToForms**: Περιέχει μία ή περισσότερες καταληκτικές γραφηματικές σειρές χωρισμένες μεταξύ τους με παύλα, εξαίρεση των οποίων αποτελεί η καταληκτική γραφηματική σειρά του πεδίου `final_grapheme_seq`, εφόσον διαφέρει στην τιμή του κλιτικού υποδείγματος και αποτελείται από τουλάχιστον ένα λιγότερο γραφηματικό χαρακτήρα στα αριστερά αυτής. Οι γραφηματικές σειρές που περιέχονται στο συγκεκριμένο πεδίο αντιστοιχούν κατά σειρά στα κλιτικά υποδείγματα του πεδίου `model`, στις απόλυτες συχνότητες του πεδίου `absoluteFreq` και στους κωδικούς λημματοποίησης του πεδίου `lemmaCode`. Σε περίπτωση που η γραφηματική σειρά του πεδίου `final_grapheme_seq` δεν αποτελεί εξαίρεση άλλης γραφηματικής σειράς, το πεδίο περιέχει τη συμβολοσειρά 'no'.

### Σημείωση:

Για να γίνει καλύτερα κατανοητή η αντιστοιχία των τιμών που περιέχονται στα πεδία `model`, `absoluteFreq`, `lemmaCode` και `exceptionToForms`, οι οποίες διαχωρίζονται μεταξύ τους με μία παύλα ('-'), αναφέρεται το ακόλουθο παράδειγμα: Αν το πεδίο `model` περιέχει συμβολοσειρά της μορφής 'mod<sub>1</sub> – mod<sub>2</sub>' (mod<sub>1</sub>, mod<sub>2</sub> ακέραιοι αριθμοί), το πεδίο `absoluteFreq` θα περιέχει συμβολοσειρά της μορφής 'fr<sub>1</sub>-fr<sub>2</sub>' (fr<sub>1</sub>, fr<sub>2</sub>: ακέραιοι αριθμοί), το πεδίο `lemmaCode` συμβολοσειρά της μορφής 'lem<sub>1</sub>-lem<sub>2</sub>' (lem<sub>1</sub>, lem<sub>2</sub>: συμβολοσειρές που εκφράζουν κωδικό λημματοποίησης) και το πεδίο `exceptionToForms` συμβολοσειρά της μορφής 'seq<sub>1</sub>- seq<sub>2</sub>' (seq<sub>1</sub>, seq<sub>2</sub>: συμβολοσειρές που εκφράζουν καταληκτικές γραφηματικές σειρές). Αυτό σημαίνει ότι fr<sub>1</sub> από τους τύπους που λήγουν στη γραφηματική σειρά του πεδίου `final_grapheme_seq` ανήκουν στο κλιτικό υπόδειγμα mod<sub>1</sub>, έχουν κωδικό λημματοποίησης lem<sub>1</sub> και αποτελούν εξαίρεση των τύπων που λήγουν στη γραφηματική σειρά seq<sub>1</sub>, ενώ οι υπόλοιποι fr<sub>2</sub> ανήκουν στο κλιτικό υπόδειγμα mod<sub>2</sub>, έχουν κωδικό λημματοποίησης lem<sub>2</sub> και αποτελούν εξαίρεση των τύπων που λήγουν στη γραφηματική σειρά seq<sub>2</sub>. Όλοι όμως οι τύποι (που το πλήθος τους είναι το ίδιο με το άθροισμα των fr<sub>1</sub> και fr<sub>2</sub>) χαρακτηρίζονται από το μορφοσυντακτικό κωδικό που περιέχεται στο πεδίο `ms_code` της συγκεκριμένης γραμμής του πίνακα `verb_info`.

- **extraMeaning:** Περιέχει μία συμβολοσειρά που εκφράζει καταληκτική γραφηματική σειρά ή τη συμβολοσειρά 'no'. Η σημασία αυτού του πεδίου, στην περίπτωση που περιέχει συμβολοσειρά διαφορετική της 'no', είναι ότι για τους τύπους που λήγουν στη γραφηματική σειρά που περιέχεται στο πεδίο 'final\_grapheme\_seq' ισχύει επιπλέον η μορφοσυντακτική πληροφορία που αφορά στην καταληκτική γραφηματική σειρά του εν λόγω πεδίου. Αυτό αφορά σε γραμμές από τα αρχεία γλωσσικών δεδομένων οι οποίες αρχίζουν με τους χαρακτήρες '(+'. Εάν κάτι τέτοιο δε συμβαίνει, το πεδίο περιέχει τη συμβολοσειρά 'no'.

Το πρωτεύον κλειδί του πίνακα `verb_info` είναι ο συνδυασμός των πεδίων `final_grapheme_seq` και `ms_code`, γεγονός το οποίο σημαίνει ότι κάθε γραμμή του πίνακα αφορά σε ένα διαφορετικό μορφοσυντακτικό κωδικό μίας καταληκτικής γραφηματικής σειράς.

Θα πρέπει να αναφερθεί ότι οι SQL ερωτήσεις που εισάγουν δεδομένα στον πίνακα `verb_info` παράγονται με χρήση ενός προγράμματος Java, το οποίο δέχεται ως είσοδο ένα αρχείο γλωσσικών δεδομένων που έχει τη δομή που παρουσιάστηκε στην §4.2 και ως έξοδο παράγει ένα αρχείο με τις SQL ερωτήσεις που αντιστοιχούν στο περιεχόμενο του αρχείου εισόδου. (Πληροφορίες για το πρόγραμμα αυτό υπάρχουν στην §7.1.1.) Στο πλαίσιο αυτής της εργασίας, έχουν παραχθεί τα εν λόγω αρχεία που περιέχουν τις SQL ερωτήσεις που εισάγουν δεδομένα για τους όλους τους ρηματικούς τύπους. Τα αρχεία αυτά περιλαμβάνονται στον κατάλογο `DB_files`.

Προκειμένου να γίνει αντιληπτή η αντιστοιχία μεταξύ των δεδομένων των αρχείων γλωσσικών δεδομένων και των δεδομένων του πίνακα 'verb\_info', κρίνεται σκόπιμο να αναφερθούν μερικά ενδεικτικά παραδείγματα. Έτσι, στα παραδείγματα που ακολουθούν, παρατίθενται μερικές γραμμές από τα αρχεία γλωσσικών δεδομένων οι οποίες αναλύθηκαν στα παραδείγματα του κεφαλαίου 4 καθώς και οι αντίστοιχες γραμμές του πίνακα `verb_info`:

### ➤ Παράδειγμα 5.1:

Οι πληροφορίες που αφορούν στην παρακάτω γραμμή και που παρουσιάστηκε στο παράδειγμα 4.1:

---

057 218,\*ABA 44% 39 2b otherwise -> (\*BA,211)

---

αποθηκεύονται στον πίνακα 'verb\_info' ως εξής:

Πίνακας 4: Γραμμή του πίνακα verb\_info για τη γραφηματική σειρά '\*αψα'

final_grapheme_seq	model	Ms_code	relativeFreq	absoluteFreq	lemmaCode	exceptionToForms	extraMeaning
*ABA	218	057	44%	39	2b	*BA	no

➤ Παράδειγμα 5.2:

Οι πληροφορίες που αφορούν στις γραμμές του παραδείγματος 4.2:

---

2msc 193,\*CA 14% 46 otherwise -> (\*A,47)

\*\*\*\*\*23/46=042:\*8 + 23/46=080:8

---

αποθηκεύονται στον πίνακα 'verb\_info' ως εξής:

Πίνακας 5: Γραμμές του πίνακα verb\_info για τη γραφηματική σειρά '\*θα'

final_grapheme_seq	model	ms_code	relativeFreq	absoluteFreq	lemmaCode	exceptionToForms	extraMeaning
*CA	193	042	14%	23/46	*8	*A	no
*CA	193	080	14%	23/46	8	*A	no

➤ Παράδειγμα 5.3:

Οι πληροφορίες που αφορούν στις γραμμές του παραδείγματος 4.7:

---

057 165,CA~RREBA 1 3 otherwise -> (\*RREBA,211)

(+057 208,CA~RREBA 1 2b otherwise -> (\*RREBA,211)MA)

---

αποθηκεύονται στον πίνακα 'verb\_info' ως εξής:

Πίνακας 6: Γραμμή του πίνακα verb\_info για τη γραφηματική σειρά 'θάρρεψα'

final_grapheme_seq	model	ms_code	relativeFreq	absoluteFreq	lemmaCode	exceptionToForms	extraMeaning
CA~RREBA	165-208	057	100%	1-1	3-2b	*RREBA-*RREBA	no

➤ Παράδειγμα 5.4:

Οι πληροφορίες που αφορούν στις γραμμές που παρουσιάστηκαν στο παράδειγμα 4.9:

2msc 90,\*LA 17% 52 otherwise -> (\*A,47)

\*\*\*\*\*26/52=057:1b + 26/52=033:1a

(+057 50,E~VALA 1 2b otherwise -> (\*LA,90)MA)

αποθηκεύονται στον πίνακα 'verb\_info' ως εξής:

Πίνακας 7: Γραμμές του πίνακα verb\_info για τις γραφηματικές σειρές '\*λα' και 'έβαλα'

final_grapheme_seq	model	ms_code	relativeFreq	absoluteFreq	lemmaCode	exceptionToForms	extraMeaning
*LA	90	057	17%	26/52	1b	*A	no
*LA	90	033	17%	26/52	1a	*A	no
E~VALA	50	057	100%	1	2b	*LA	*LA

### 5.3. Πίνακας equalsMscodes

Ο πίνακας αυτός περιέχει ζεύγη μορφοσυντακτικών κωδικών των οποίων το περιεχόμενο μοιάζει αρκετά (π.χ. διαφέρουν μόνο στο χρόνο ή την έγκλιση). Τα ζεύγη αυτά έχουν καθοριστεί από τη Λεμπέση και είναι χρήσιμα κατά τη διαδικασία της αναγνώρισης, όταν χρησιμοποιούνται δεδομένα από γραμμές των αρχείων γλωσσικών δεδομένων που αρχίζουν με τους χαρακτήρες '+'.  
Αποτελείται από δύο πεδία με ονόματα 'msc1' και 'msc2', καθένα από τα οποία περιέχει μία συμβολοσειρά που εκφράζει ένα μορφοσυντακτικό κωδικό.

## 6. ΠΕΡΙΓΡΑΦΗ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ ΛΗΜΜΑΤΟΠΟΙΗΣΗΣ ΡΗΜΑΤΙΚΩΝ ΤΥΠΩΝ

Δύο είναι τα αρχεία που μας παραχωρήθηκαν από τη Λεμπέση για τις ανάγκες της λημματοποίησης όλων των ρηματικών τύπων της Νέας Ελληνικής. Το πρώτο περιέχει τους μορφολογικούς κανόνες που απαιτούνται για την επιστροφή στο λήμμα κάθε κλιτού ρηματικού τύπου και καλείται από τον κωδικό λημματοποίησης που εμφανίζεται σε κάθε γραμμή καταληκτικής γραφηματικής σειράς, όπως έχει ήδη περιγραφεί, σε συνδυασμό με τον αριθμό κλιτικού υποδείγματος που αναγνωρίζει η εφαρμογή. Έτσι είναι και αυτό δομημένο σε 382 κλιτικά υποδείγματα, από το 1 έως το 385 εκτός των αριθμών 302, 303 και 304, οι οποίοι έχουν διατηρηθεί κενοί για μελλοντική κάλυψη ρηματικής κλίσης ενεργητικής φωνής, εάν χρειαστεί.

Λόγω του περιορισμένου χρόνου ολοκλήρωσης της παρούσας πτυχιακής εργασίας δεν δημιουργήθηκε ο σχετικός αλγόριθμος λημματοποίησης της εφαρμογής. Ενδεικτικά όμως παρουσιάζουμε ένα μικρό απόσπασμα από το αρχείο και δη τους κανόνες που απαιτούνται για τη λημματοποίηση των ρηματικών τύπων τριών σύνθετων ρημάτων (διάγω, κατάγω και περιάγω) που κλίνονται σύμφωνα με το σχετικά απλό κλιτικό υπόδειγμα 023

023	0		DJA'GW	21	
	1	PY	W	2	PERJI'GA 33,77
	2	PY	W	2	PERJA'GEJ 23,34,38
	3	PY	W	2	KATA'GEJS/I'GAGA 22,25,26,36,37,45,57,93
	4	PY	W	2	DJA'GOYME/OYNE 24,26,47,58,62
	5		W	2	PERJAGA'GETE 46,49,50,60,61
	6		W	2	PERJAGA'GOYME 48,50

Ο κωδικός λημματοποίησης εμφανίζεται στην αρχή της κάθε γραμμής (0, 1, 2, 3, 4, 5, 6) και αντιστοιχεί στον αριθμό των τελικών χαρακτήρων του κάθε ρηματικού τύπου που πρέπει να διαγραφεί, με τους χαρακτήρες PY (Prefix: Yes) καλείται το 2<sup>ο</sup> περιφερειακό αρχείο εσωτερικής αύξησης, το οποίο θα παρουσιαστεί στη συνέχεια, καθορίζεται ο αριθμός της συλλαβής του παραχθέντος λήμματος που θα φέρει τον τόνο, 2<sup>η</sup> από το τέλος της λέξης στην προκειμένη περίπτωση, και παρατίθενται πληροφοριακά αλλά και για πρόσθετο έλεγχο στο τέλος κάθε γραμμής οι κωδικοί μορφοσυντακτικού περιεχομένου για τους οποίους πρέπει να ενεργοποιηθεί ο κάθε κανόνας.

Το 2<sup>ο</sup> παραχωρηθέν από τη Λεμπέση αρχείο, το οποίο ονομάζεται AUGMENTREVERSEfinal.doc και το οποίο ενεργοποιείται μόνο για τους κανόνες που περιέχουν την ένδειξη PY (Prefix:Yes), ελέγχει την παρουσία συγκεκριμένων προθημάτων και επιτελεί τις κατάλληλες αλλαγές, προκειμένου να απαληφθεί η εσωτερική αύξηση σύνθετων παρελθοντικών ρηματικών τύπων.

Πιο συγκεκριμένα στο αρχείο αυτό ορίζεται πόσοι και ποιοι χαρακτήρες θα αφαιρεθούν ή θα αντικατασταθούν από το πρόθημα ή τη ρίζα του ρηματικού τύπου, ώστε να προκύψει το λήμμα από το οποίο έχει προέλθει.

➤ Παράδειγμα 6.1:

```
CODE "A" : REMPLACER LE DERNIER CARACTERE DU PREFIXE  
          PAR 'A'
```

Μετατροπή του τελευταίου χαρακτήρα του προθέματος σε 'Α'

Στην παραπάνω περίπτωση, αν ο κωδικός της αλλαγής εντολής είναι «Α», αφού γίνει διαχωρισμός προθήματος και ρίζας, ο **τελευταίος** χαρακτήρας του προθήματος θα αντικατασταθεί με «Α».

➤ Παράδειγμα 6.2:

```
CODE "P" : ENLEVER PREMIER CARACTERE DE LA RACINE
```

Αφαίρεση του πρώτου χαρακτήρα της ρίζας

Στην περίπτωση αυτή αν ο κωδικός της εντολής αλλαγής είναι «Ρ», αφού γίνει διαχωρισμός προθήματος και ρίζας, ο **πρώτος** χαρακτήρας της ρίζας θα αφαιρεθεί.

Τονίζεται ότι όταν γίνονται αλλαγές στο πρόθημα, αυτές γίνονται στους τελευταίους χαρακτήρες του, ενώ αντίθετα στην μοναδική περίπτωση που μετατρέπεται η ρίζα, η αλλαγή γίνεται στον πρώτο χαρακτήρα της.

Έτσι, για τις ανάγκες της λημματοποίησης δημιουργήθηκε μία νέα βάση δεδομένων lemmaDB, η οποία περιλαμβάνει έναν πίνακα με όνομα 'lemma'.

## 6.1. Πίνακας lemma

Ο πίνακας αυτός περιλαμβάνει εγγραφές, τις οποίες εισαγάγαμε, όπου σύμφωνα με τον εκάστοτε κωδικό εσωτερικής αλλαγής ορίζεται τί θα τροποποιηθεί.

Συγκεκριμένα, κάθε γραμμή αυτού του πίνακα εκφράζει και μία διαφορετική ερμηνεία ενός κωδικού αλλαγής. Τα πεδία από τα οποία αποτελείται είναι τα εξής:

- **primaryKey:** Περιέχει έναν αύξοντα αριθμό, αριθμητικού (Integer) τύπου δεδομένων, με τιμή διαφορετική για κάθε γραμμή του πίνακα. Χρησιμοποιείται ως πρωτεύον κλειδί.
- **lemmaCode:** Περιέχει έναν κωδικό αλφαριθμητικού (Varchar) τύπου δεδομένων, από Α έως R, όπως ορίζεται στο αρχείο AUGMENTREVERSEfinal.doc.
- **modifyPrefix:** Πρόκειται για ένα πεδίο λογικού (Boolean) τύπου δεδομένων, το οποίο έχει τιμή TRUE στην περίπτωση που πρέπει να τροποποιηθεί το πρόθημα, ενώ FALSE σε αντίθετη περίπτωση.
- **modifyRoot:** Πρόκειται για ένα πεδίο λογικού (Boolean) τύπου δεδομένων, το οποίο έχει τιμή TRUE στην περίπτωση που πρέπει να τροποποιηθεί η ρίζα, ενώ FALSE σε αντίθετη περίπτωση.
- **numOfCharactersToBeReplaced:** Το πεδίο αυτό είναι αριθμητικού (Integer) τύπου δεδομένων, το οποίο παρέχει πληροφορία για τον αριθμό των χαρακτήρων που πρέπει να αντικατασταθούν είτε από το πρόθημα είτε από τη ρίζα ανάλογα με τις τιμές των ανωτέρω δύο μεταβλητών.
- **Replacewith:** Αλφαριθμητικός (Varchar) τύπος δεδομένων που περιλαμβάνει με τί θα αντικατασταθούν οι χαρακτήρες που θα τροποποιηθούν είτε από το πρόθημα είτε από τη ρίζα. Στην περίπτωση που η τιμή είναι NULL, γίνεται αφαίρεση χαρακτήρων χωρίς αντικατάσταση.

Για την παραπάνω βάση έχει δημιουργηθεί το αρχείο createLemmaDB.sql, στο οποίο περιλαμβάνονται οι εντολές (queries) για την δημιουργία της καθώς και του πίνακα lemma, όπως περιγράφηκε ανωτέρω. Επιπλέον έχουν συμπεριληφθεί οι κατάλληλες εντολές για την εισαγωγή των απαιτούμενων δεδομένων σύμφωνα με τα οποία θα γίνει η λημματοποίηση των ρηματικών τύπων.

➤ Παράδειγμα 6.3:

```
INSERT INTO lemma (lemmaCode, modifyPrefix, modifyRoot, numOfCharactersToBeReplaced, replacewith) VALUES ("A", TRUE, FALSE, 1, "A");
```

```
INSERT INTO lemma (lemmaCode, modifyPrefix, modifyRoot, numOfCharactersToBeReplaced, replacewith) VALUES ("P", FALSE, TRUE, 1, NULL);
```

Συνεπώς, προκειμένου να ολοκληρωθεί η λημματοποίηση των ρηματικών τύπων της εφαρμογής, θα πρέπει να δημιουργηθεί αλγόριθμος που θα ενσωματώσει αφ'ενός τους μορφολογικούς κανόνες λημματοποίησης της Λεμπέση που προαναφέραμε, ο οποίος θα ενεργοποιείται βάσει του κωδικού λημματοποίησης των καταληκτικών γραφηματικών σειρών αναγνώρισης και αφ'ετέρου τη νέα βάση δεδομένων lemmaDB, η οποία απαλείφει την εσωτερική αύξηση, όπου αυτή εμφανίζεται.



## 7. ΥΛΟΠΟΙΗΣΗ ΕΡΓΑΛΕΙΟΥ VERBTAGGR++

Το παρόν κεφάλαιο ασχολείται με θέματα που αφορούν στην υλοποίηση του εργαλείου VERBTAGGR++. Οι τεχνολογίες που χρησιμοποιήθηκαν είναι οι εξής: Java (J2SE 1.4.2), JSP, Web Services και MySQL.

Συγκεκριμένα:

- Η προεπεξεργασία των γλωσσικών δεδομένων πραγματοποιήθηκε με προγράμματα υλοποιημένα σε γλώσσα προγραμματισμού Java.
- Η λειτουργία της αναγνώρισης ενός ρηματικού τύπου παρέχεται από ένα Web Service υλοποιημένο επίσης σε Java, το οποίο αναπτύχθηκε με χρήση του εργαλείου Java Web Services Developer Pack 1.5.
- Η διαδικτυακή διεπαφή του χρήστη έχει υλοποιηθεί σε JSP.
- Τα γλωσσικά δεδομένα που χρησιμοποιούνται από τη διαδικτυακή εφαρμογή αποθηκεύονται σε μία βάση δεδομένων MySQL 4.1.
- Τέλος, ως web container χρησιμοποιήθηκε ο Tomcat 5.0.

Επομένως, το περιεχόμενο του παρόντος κεφαλαίου περιλαμβάνει τις εξής ενότητες: Η §7.1 περιγράφει το λογισμικό που υλοποιήθηκε για την προεπεξεργασία και τον έλεγχο της ορθότητας των αρχείων των γλωσσικών δεδομένων καθώς επίσης και την αυτοματοποιημένη εισαγωγή στη βάση δεδομένων που προήλθαν από αυτά τα αρχεία. Στην §7.2 αναφέρονται τα στοιχεία για την τεχνολογία των Web Services. Η §7.3 περιλαμβάνει μία γενική μορφή του βασικού αλγόριθμου που ακολουθήσαμε για την υλοποίηση του Web Service που χρησιμοποιείται από τη διαδικτυακή εφαρμογή VERBTAGGR++. Τέλος, στην §7.4 περιγράφεται αυτό το Web Service που υλοποιεί τον αλγόριθμο της §7.3 και πραγματοποιεί τη λειτουργία της μορφοσυντακτικής αναγνώρισης όλων των μονολεκτικών ρηματικών τύπων της Νέας Ελληνικής.

### 7.1. Λογισμικό Προεπεξεργασίας Γλωσσικών Δεδομένων

Κατά την πρώτη φάση της προεπεξεργασίας των γλωσσικών δεδομένων υλοποιήθηκαν τρία προγράμματα σε γλώσσα Java, τα οποία επεξεργάζονται και ελέγχουν την ορθότητα των αρχείων γλωσσικών δεδομένων έχοντας τη δομή που παρουσιάστηκε στην §4.2. Το πρώτο πρόγραμμα, με όνομα 'dataProcess', λαμβάνει ως είσοδο ένα αρχείο γλωσσικών δεδομένων και παράγει ως έξοδο ένα αρχείο κειμένου, το οποίο περιέχει ένα πλήθος επερωτήσεων που πρέπει να εκτελεστούν, ώστε να εισαχθούν στη βάση τα δεδομένα που αντιστοιχούν στις πληροφορίες που περιέχει το αρχείο εισόδου. Το δεύτερο πρόγραμμα, με όνομα 'checkIntegers', ελέγχει την ορθότητα των τιμών που περιέχονται στα πεδία 'absoluteFreq', 'relativeFreq', 'ms\_code' και 'model' του πίνακα 'verb\_info' της βάσης. Το τρίτο πρόγραμμα έχει όνομα 'checkExceptionToForms' και ελέγχει την ορθότητα του περιεχομένου του πεδίου 'exceptionToForms' του πίνακα 'verb\_info' της βάσης. Σημειώνεται ότι τα προγράμματα περιέχονται στο cd που παραδόθηκε μαζί με το παρόν έγγραφο εντός του καταλόγου SYNODEYTIKO\_YLIKO \ linguistic\_data\_software\ και για να χρησιμοποιηθούν, θα πρέπει να έχει προηγηθεί η εγκατάσταση της πλατφόρμας της Java J2SDK (για οδηγίες εγκατάστασης βλ. κεφάλαιο 8) και έπειτα μέσω μίας γραμμής εντολών να δοθεί η εντολή:

*java filename*

όπου filename: το όνομα του αρχείου που ο χρήστης επιθυμεί να χρησιμοποιήσει, δηλαδή ένα εκ των dataProcess, checkIntegers, checkExceptionToForms.

Περισσότερες πληροφορίες για τη λειτουργικότητα των προγραμμάτων δίνονται παρακάτω.

### 7.1.1. Πρόγραμμα dataProcess

Το πρόγραμμα 'dataProcess' δέχεται ως είσοδο το όνομα ενός αρχείου γλωσσικών δεδομένων (το όνομα μπορεί να είναι πλήρες ή σχετικό με τον κατάλογο που βρίσκεται το πρόγραμμα). Για την ορθότερη λειτουργία του προγράμματος, εάν το όνομα δεν αντιστοιχεί σε υπαρκτό αρχείο, το πρόγραμμα ζητά από το χρήστη ένα νέο όνομα και η διαδικασία αυτή συνεχίζεται, μέχρι να εισαχθεί το όνομα ενός υπαρκτού αρχείου. Στη συνέχεια το πρόγραμμα διαβάζει κάθε γραμμή του αρχείου και εφόσον η γραμμή έχει την αναμενόμενη μορφή (σύμφωνα με όσα αναφέρθηκαν παραπάνω στην §4.2), σχηματίζει μία κατάλληλη SQL επερώτηση, η οποία αφού εκτελεστεί, θα εισάγει στη βάση δεδομένων του εργαλείου τα δεδομένα που αντιστοιχούν στη γραμμή που διαβάστηκε. Η συμβολοσειρά που περιέχει την SQL επερώτηση γράφεται σε ένα αρχείο εξόδου, το οποίο αποθηκεύεται σε ένα φάκελο με όνομα "output". Εάν κατά την ανάγνωση των γραμμών του αρχείου εισόδου εντοπιστεί κάποιο πρόβλημα, για παράδειγμα αν κάποια γραμμή δεν έχει την αναμενόμενη μορφή, τυπώνεται μήνυμα σχετικό με το πρόβλημα που εντοπίστηκε και διακόπτεται η λειτουργία του προγράμματος.

Εφόσον κανένα πρόβλημα δεν εντοπιστεί και πραγματοποιηθεί η ανάγνωση όλων των γραμμών του αρχείου εισόδου, ενημερώνεται ο χρήστης για την επιτυχή επεξεργασία του αρχείου εισόδου καθώς και για το όνομα του αρχείου εξόδου που δημιουργήθηκε.

Κατά συνέπεια, το πρόγραμμα 'dataProcess' είναι υπεύθυνο για την αυτοματοποιημένη επεξεργασία των αρχείων γλωσσικών δεδομένων και την εισαγωγή των αντίστοιχων δεδομένων στη βάση. Όπως ήδη αναφέρθηκε, στο πλαίσιο της παρούσας πτυχιακής εργασίας διατέθηκαν και υπέστησαν επεξεργασία τα αρχεία γλωσσικών δεδομένων που αφορούν σε όλους τους ρηματικούς τύπους. Επιπλέον, άλλη μία χρησιμότητα του προγράμματος 'dataProcess' αφορά στον εντοπισμό ασυνεπειών στη δομή των αρχείων εισόδου, οι οποίες αναφέρθηκαν και διορθώθηκαν από τη Λεμπέση.

### 7.1.2. Πρόγραμμα checkIntegers

Το πρόγραμμα 'checkIntegers' (καθώς και το 'checkExceptionToForms') χρησιμοποιείται, αφού γίνει η εισαγωγή δεδομένων στη βάση, προκειμένου να ελεγχθεί η ορθότητα των τιμών που περιέχουν κάποια από τα πεδία του πίνακα 'verb\_info'. Συγκεκριμένα, το εν λόγω πρόγραμμα πραγματοποιεί τους ακόλουθους ελέγχους:

- Ελέγχει αν κάθε συμβολοσειρά που περιέχεται στο πεδίο 'absoluteFreq' του πίνακα 'verb\_info':
  - είτε εκφράζει έναν ακέραιο αριθμό
  - είτε έχει τη μορφή 'x/y' όπου x, y ακέραιοι αριθμοί
  - είτε έχει τη μορφή 'x<sub>1</sub>- x<sub>2</sub>- ...- x<sub>n</sub>' όπου x<sub>i</sub> ακέραιος αριθμός, 1 ≤ i ≤ n.
- Ελέγχει αν κάθε συμβολοσειρά που περιέχεται στο πεδίο 'relativeFreq' του πίνακα 'verb\_info' έχει τη μορφή 'x%', όπου x ακέραιος αριθμός.
- Ελέγχει αν κάθε συμβολοσειρά που περιέχεται στο πεδίο 'ms\_code' του πίνακα 'verb\_info' έχει τη μορφή 'x' ή '\*x', όπου x ένας ακέραιος με τιμή όχι μεγαλύτερη από 102.
- Ελέγχει αν κάθε συμβολοσειρά που περιέχεται στο πεδίο 'model' του πίνακα 'verb\_info':

- είτε εκφράζει έναν ακέραιο αριθμό με τιμή από 1 έως 388 εκτός των 302, 303, 304
- είτε έχει τη μορφή ' $x_1 - x_2 - \dots - x_n$ ' όπου  $x_i$  ακέραιος αριθμός με τιμή από 1 έως 388 εκτός των 302, 303, 304,  $1 \leq i \leq n$ .

(Πληροφορίες για τη σημασία του περιεχομένου των πεδίων του πίνακα 'verb\_info' δόθηκαν παραπάνω, στο κεφάλαιο 5.)

Το πρόγραμμα δημιουργεί ένα αρχείο εξόδου με όνομα 'checkIntegers\_output.txt', στο οποίο περιέχονται ενδεικτικά μηνύματα για τα λάθη που εντοπίστηκαν. Δηλαδή, για κάθε μη αποδεκτή τιμή που εντοπίστηκε σε ένα πεδίο καταγράφεται ένα μήνυμα, που περιέχει τη συγκεκριμένη τιμή καθώς και την τιμή του πεδίου 'final\_grapheme\_seq' της αντίστοιχης γραμμής, ώστε εύκολα να ανευρεθεί η γραμμή που περιέχει το λάθος.

### 7.1.3. Πρόγραμμα checkExceptionToForms

Το πρόγραμμα 'checkExceptionToForms' ελέγχει την ορθότητα των τιμών που περιέχονται στο πεδίο 'exceptionToForms' του πίνακα 'verb\_info'. Συγκεκριμένα, ελέγχεται αν κάθε συμβολοσειρά που περιέχεται στο πεδίο 'exceptionToForms':

- είτε αποτελείται από τους χαρακτήρες 'no'
- είτε αποτελεί μία καταληκτική γραφηματική σειρά, δηλαδή μία από τις τιμές του πεδίου 'final\_grapheme\_seq' του ίδιου πίνακα
- είτε έχει τη μορφή ' $x_1 - x_2 - \dots - x_n$ ', όπου κάθε  $x_i$  είναι μία από τις τιμές του πεδίου 'final\_grapheme\_seq' του ίδιου πίνακα.

Ως έξοδος παράγονται τρία αρχεία καθένα από τα οποία, σε περίπτωση που εντοπιστούν λάθη, περιέχει γραμμές της εξής μορφής:

*"final\_grapheme\_sequence" -> "invalid\_exceptionToForms"*

όπου *"invalid\_exceptionToForms"* μη αποδεκτή τιμή του πεδίου 'exceptionToForms' μίας γραμμής του πίνακα 'verb\_info' και *"final\_grapheme\_seq"* η τιμή του πεδίου 'final\_grapheme\_seq' της ίδιας γραμμής. Επομένως, όλα τα αρχεία εξόδου που παράγονται αφορούν σε γραμμές του πίνακα 'verb\_info' που περιέχουν μη αποδεκτή τιμή στο πεδίο 'exceptionToForms' και συγκεκριμένα πρόκειται για τα ακόλουθα αρχεία:

- *checkExceptionToForms\_changes.txt*: Το αρχείο αυτό αφορά σε τιμές του πεδίου 'exceptionToForms' που αρχίζουν με αστερίσκο (\*) οι οποίες γίνονται αποδεκτές, αν ο αστερίσκος αντικατασταθεί από αγγλικό ερωτηματικό (?). Δηλαδή, η συμβολοσειρά που προκύπτει, αν γίνει η προαναφερθείσα αντικατάσταση, αντιστοιχεί σε μία από τις τιμές του πεδίου 'final\_grapheme\_seq' του πίνακα 'verb\_info'.
- *checkExceptionToForms\_not\_changes.txt*: Το αρχείο αυτό αφορά σε τιμές του πεδίου 'exceptionToForms' που αρχίζουν με αστερίσκο, οι οποίες όμως δε γίνονται αποδεκτές, αν ο αστερίσκος αντικατασταθεί από αγγλικό ερωτηματικό.
- *checkExceptionToForms\_not\_changes2.txt*: Το αρχείο αυτό αφορά σε μη αποδεκτές τιμές του πεδίου 'exceptionToForms' που δεν αρχίζουν με αστερίσκο.

Θα πρέπει να σημειωθεί ότι τα προγράμματα 'checkIntegers' και 'checkExceptionToForms' ήταν ιδιαίτερα χρήσιμα, διότι με αυτά εντοπίστηκαν λάθη στα αρχεία γλωσσικών δεδομένων, τα οποία δεν ήταν δυνατό να εντοπιστούν με το

πρόγραμμα 'dataProcess', διότι οι γραμμές που περιείχαν τα λάθη είχαν κατά τα άλλα σωστή δομή.

## 7.2. Τεχνολογία των Web Services

Μετά τη φάση της προεπεξεργασίας των γλωσσικών δεδομένων αναβαθμίστηκε η διαδικτυακή εφαρμογή *VerbTagGr*, που επιτελεί μορφοσυντακτική αναγνώριση των μονολεκτικών ρηματικών τύπων της Νέας Ελληνικής. Αυτή η εφαρμογή βασίζεται στην τεχνολογία των Web Services και για το λόγο αυτό κρίνεται σκόπιμο να παρουσιαστούν κάποια στοιχεία για τα Web Services, που προέρχονται από την τρέχουσα βιβλιογραφία. Κατά συνέπεια, αναφέρονται βασικοί ορισμοί των Web Services, οι φορείς από τους οποίους αποτελείται το μοντέλο τους, τα πιο σημαντικά πρότυπα που ορίζουν την υποδομή τους, ο κύκλος ζωής ενός Web Service και η ενότητα ολοκληρώνεται με μία σύνοψη και συμπεράσματα.

### 7.2.1. Ορισμός των Web Services

Μερικοί από τους ορισμούς που έχουν δοθεί για τα Web Services είναι οι ακόλουθοι:

➤ Ορισμός από το W3C [8]:

Ένα Web Service είναι ένα σύστημα λογισμικού που προσδιορίζεται από ένα URI και του οποίου η δημόσια διεπαφή (public interface) καθορίζεται και περιγράφεται με χρήση XML. Ο ορισμός του μπορεί να αναζητηθεί από άλλα συστήματα λογισμικού, τα οποία μπορούν τότε να αλληλεπιδράσουν με το Web Service με τρόπο που υπαγορεύεται από τον ορισμό του, χρησιμοποιώντας μηνύματα βασισμένα στην XML, τα οποία διέπονται από πρωτόκολλα του Internet.

➤ Ορισμός από την IBM [9]:

Ένα Web Service είναι μία διεπαφή που περιγράφει ένα σύνολο από λειτουργίες δικτυακά προσπελάσιμες μέσω μηνυμάτων βασισμένων στην XML, το οποίο εκπληρώνει μία συγκεκριμένη εργασία ή ένα σύνολο εργασιών. Περιγράφεται με χρήση τυπικής XML, η οποία παρέχει όλες τις απαραίτητες λεπτομέρειες για την αλληλεπίδραση με το Web Service, περιλαμβάνοντας τη δομή των μηνυμάτων, τα πρωτόκολλα μεταφοράς και την τοποθεσία (location).

➤ Ορισμός από τη SUN [10]:

Τα Web Services είναι συστατικά λογισμικού (software components) που μπορούν αυτόματα να αναζητηθούν, να συνδυαστούν και να επαναχρησιμοποιηθούν, ώστε να παρέχουν λύσεις στα προβλήματα και τις αιτήσεις των χρηστών.

➤ Ορισμός από τη Microsoft [11]:

Ένα Web Service είναι μία μονάδα λογικής εφαρμογής που παρέχει δεδομένα και υπηρεσίες σε άλλες εφαρμογές. Οι εφαρμογές προσπελαίνουν τα Web Services μέσω ευρέως διαδεδομένων web πρωτοκόλλων και δομών δεδομένων όπως είναι το HTTP, η XML, το SOAP, χωρίς να απαιτείται γνώση για τον τρόπο υλοποίησης των Web Services.

### 7.2.2. Μοντέλο των Web Services

Το μοντέλο των Web Services περιλαμβάνει τους ακόλουθους τρεις ρόλους [12]:

➤ **Παροχέας υπηρεσίας (Service provider)**

Ο παροχέας υπηρεσίας είναι ο φορέας που παρέχει ως Web Services εφαρμογές λογισμικού, οι οποίες καλύπτουν συγκεκριμένες ανάγκες. Οι παροχείς υπηρεσιών είναι υπεύθυνοι για την ανάπτυξη, την περιγραφή, τη δημοσίευση, την ενημέρωση και την ανάκληση των Web Services που παρέχουν μέσω του Internet. Από αρχιτεκτονική άποψη, πρόκειται ουσιαστικά για την πλατφόρμα που κρατά την υλοποίηση του Web Service.

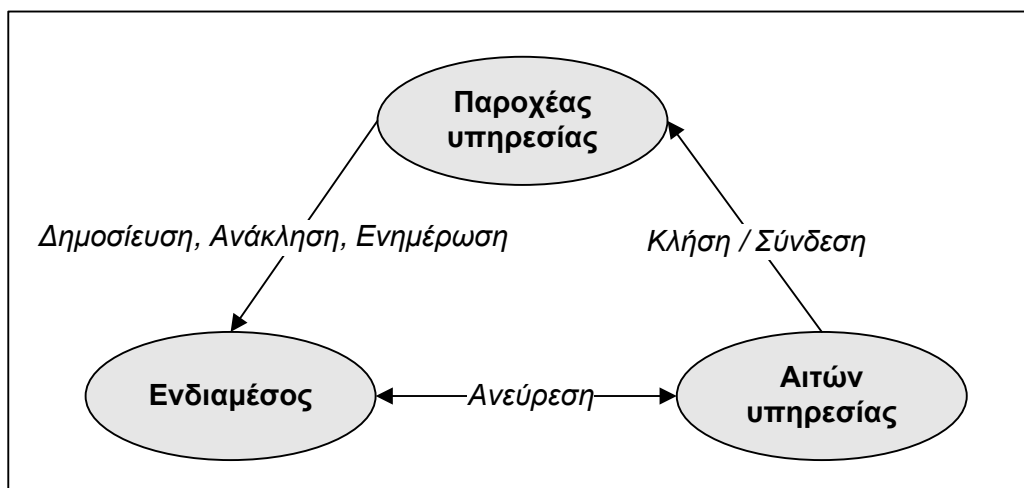
➤ **Αιτών υπηρεσίας (Service requester)**

Ο αιτών υπηρεσίας είναι ο φορέας που έχει κάποια ανάγκη, η οποία μπορεί να καλυφθεί από μία υπηρεσία διαθέσιμη στο Internet. Είναι υπεύθυνος για την αναζήτηση της περιγραφής ενός Web Service που παρέχεται από έναν ενδιάμεσο καθώς και για τη χρήση της περιγραφής αυτής, προκειμένου να γίνει κλήση (invoke) και σύνδεση (bind) του Web Service που παρέχεται από έναν παροχέα. Από αρχιτεκτονική άποψη, πρόκειται για την εφαρμογή που αναζητά και καλεί μία υπηρεσία. Θα μπορούσε δηλαδή να είναι ένας άνθρωπος – χρήστης που έχει πρόσβαση σε μία υπηρεσία μέσω ενός browser ή ένα πρόγραμμα εφαρμογής ή ακόμη κι ένα άλλο Web Service.

➤ **Ενδιάμεσος (Service broker)**

Ο ενδιάμεσος είναι ο φορέας που παρέχει μία αποθήκη με περιγραφές των Web Services που καταχώρησαν παροχείς υπηρεσιών, ώστε να διευκολυνθούν οι ενδιαφερόμενοι κατά την αναζήτηση υπηρεσιών. Από τις περιγραφές των Web Services, οι ενδιαφερόμενοι λαμβάνουν πληροφορίες σχετικά με την κλήση των Web Services. Ως παράδειγμα των ενδιάμεσων αναφέρονται οι UDDI αποθήκες. [14, 15]

Η αλληλεπίδραση μεταξύ των τριών φορέων του μοντέλου των Web Services φαίνεται στο σχήμα που ακολουθεί (σχήμα 4):



Σχήμα 4: Μοντέλο των Web Services

Δεδομένου ότι οι φορείς του μοντέλου των Web Services επικοινωνούν μεταξύ τους, προφανώς είναι απαραίτητο να χρησιμοποιούν πρότυπες τεχνολογίες για την περιγραφή των υπηρεσιών, τη μεταξύ τους επικοινωνία και τη δόμηση των δεδομένων. Η χρήση πρότυπων τεχνολογιών επιτρέπει την υλοποίηση των Web Services με τρόπο

ανεξάρτητο πλατφόρμας και προγραμματιστικής γλώσσας. Στην επόμενη ενότητα παρουσιάζονται πρότυπες τεχνολογίες που πολύ συχνά χρησιμοποιούνται στο χώρο των Web Services.

### 7.2.3. Πρότυπα των Web Services

Αρκετά πρότυπα που ορίζουν την υποδομή των Web Services καθώς και ένας μεγάλος αριθμός από εργαλεία ανάπτυξης που υποστηρίζουν τα Web Services έχουν σήμερα αναπτυχθεί. Μία από τις διαθέσιμες και αρκετά διαδεδομένες επιλογές είναι τα πρότυπα WSDL, UDDI και SOAP, τα οποία συνοπτικά περιγράφονται στη συνέχεια. Ο λόγος που επιλέχθηκε να παρουσιαστεί αυτό το σύνολο προτύπων είναι επειδή σε αυτά βασίζεται η τεχνολογία που χρησιμοποιήθηκε κατά την ανάπτυξη του Web Service που υλοποιήθηκε στα πλαίσια της παρούσας πτυχιακής εργασίας.

#### ➤ WSDL (Web Services Description Language) [12]

Η WSDL, όπως δηλώνει και το όνομά της, είναι μία γλώσσα περιγραφής της προγραμματιστικής διεπαφής ενός Web Service. Πρόκειται για μία XML γραμματική που προσδιορίζει διάφορες ιδιότητες του Web Service, όπως το *τι* κάνει, *πού* βρίσκεται και *πώς* καλείται. Έτσι, τα WSDL έγγραφα βασίζονται στην XML και περιγράφουν τις δυνατότητες που προσφέρουν τα Web Services καθώς και τα πρωτόκολλα και τις δομές που χρησιμοποιούνται από αυτά.

#### ➤ SOAP (Simple Object Access Protocol) [12, 13]

Το SOAP είναι ένα πρότυπο χρήσιμο για την αποστολή μηνυμάτων και την πραγματοποίηση απομακρυσμένων κλήσεων (remote procedure calls) μέσω του Internet. Αυτό διαχειρίζεται τις επικοινωνίες μεταξύ των συστατικών (components) και είναι ανεξάρτητο από την προγραμματιστική γλώσσα, το μοντέλο αντικειμένων, το λειτουργικό σύστημα και την πλατφόρμα στην οποία μία εφαρμογή τρέχει. Χρησιμοποιεί το HTTP ως πρωτόκολλο μεταφοράς και XML σύνταξη για την κωδικοποίηση των δεδομένων και την αποστολή εντολών κλήσης εφαρμογών που εκτελούνται σε απομακρυσμένους servers. (Σημειώνεται ότι παρόλ' αυτά και άλλα πρωτόκολλα μεταφοράς μπορούν να χρησιμοποιηθούν, όπως το FTP, SMTP κ.ά.)

#### ➤ UDDI (Universal Description, Discovery, Integration) [12, 13]

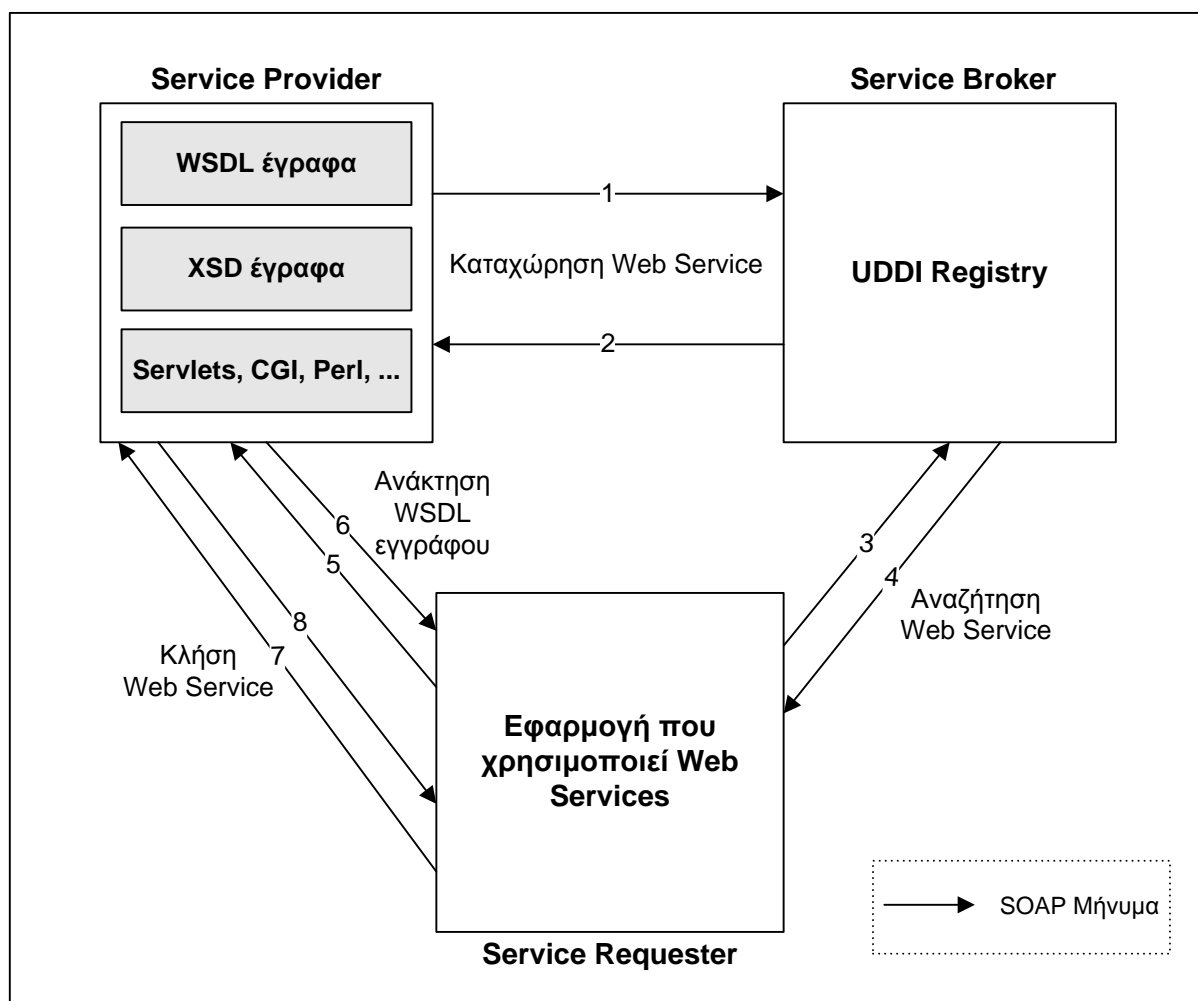
Το πρότυπο UDDI παρέχει ένα μέσο για τη δημοσίευση πληροφοριών που αφορούν σε Web Services. Επερωτήσεις που πραγματοποιούνται από εφαρμογές σε μία UDDI registry που βασίζεται στην XML επιστρέφουν πληροφορίες για νέα Web Services, για την κατάσταση συγκεκριμένων υπηρεσιών και τη διαθεσιμότητα συμβατών υπηρεσιών, ενώ επιτρέπουν να ελεγχθεί αν κάποιος συγκεκριμένος οργανισμός ή εταιρεία παρέχει ένα συγκεκριμένο Web Service, να αναζητηθούν εταιρείες που παρέχουν κάποιο συγκεκριμένο τύπο Web Services και να ανακτηθούν οι τεχνικές λεπτομέρειες που απαιτούνται για την αλληλεπίδραση με ένα Web Service.

Οι UDDI προδιαγραφές αποτελούνται από ένα XML σχήμα για SOAP μηνύματα (διότι τα UDDI μηνύματα μεταφέρονται με χρήση του SOAP πρωτοκόλλου) καθώς και περιγραφές για APIs των UDDI. Τα APIs των UDDI περιέχουν μηνύματα για την αλληλεπίδραση με UDDI registries.

Έτσι, το UDDI παρέχει έναν παγκόσμιο, κατακεντρωμένο κατάλογο βασισμένο στην XML, στον οποίο δημοσιεύονται περιγραφές για Web Services, ενώ μπορούν αυτόματα να αναζητηθούν και να ενοποιηθούν διαθέσιμες υπηρεσίες.

### ➤ Αλληλεπίδραση των Προτύπων WSDL, UDDI και SOAP

Μετά την περιγραφή των προτύπων WSDL, UDDI και SOAP, κρίνεται σκόπιμο να παρουσιαστεί πώς αυτά αλληλεπιδρούν από άποψη χρήσης, ώστε να φανεί η διαδικασία που ακολουθείται στο περιβάλλον των Web Services. Σχηματικά, η επικοινωνία μεταξύ των προτύπων γίνεται ως εξής (σχήμα 5):



Σχήμα 5: Συσχέτιση προτύπων WSDL, UDDI και SOAP

Όπως φαίνεται από το σχήμα 5, σε έναν τυπικό κύκλο επεξεργασίας ενός Web Service μπορούν να υπάρξουν τα παρακάτω βήματα:

- Ο παροχέας υπηρεσίας καταχωρεί σε μία UDDI Registry μία περιγραφή για ένα Web Service μέσω της ανταλλαγής UDDI SOAP μηνυμάτων με τον ενδιαμέσο. Επίσης ο παροχέας μπορεί να ενημερώσει ή και να διαγράψει καταχωρήσεις από τη UDDI Registry. Όλες αυτές οι ενέργειες μπορούν να πραγματοποιηθούν μέσω μιας Web διεπαφής ή ειδικών εργαλείων (βέλη 1, 2).

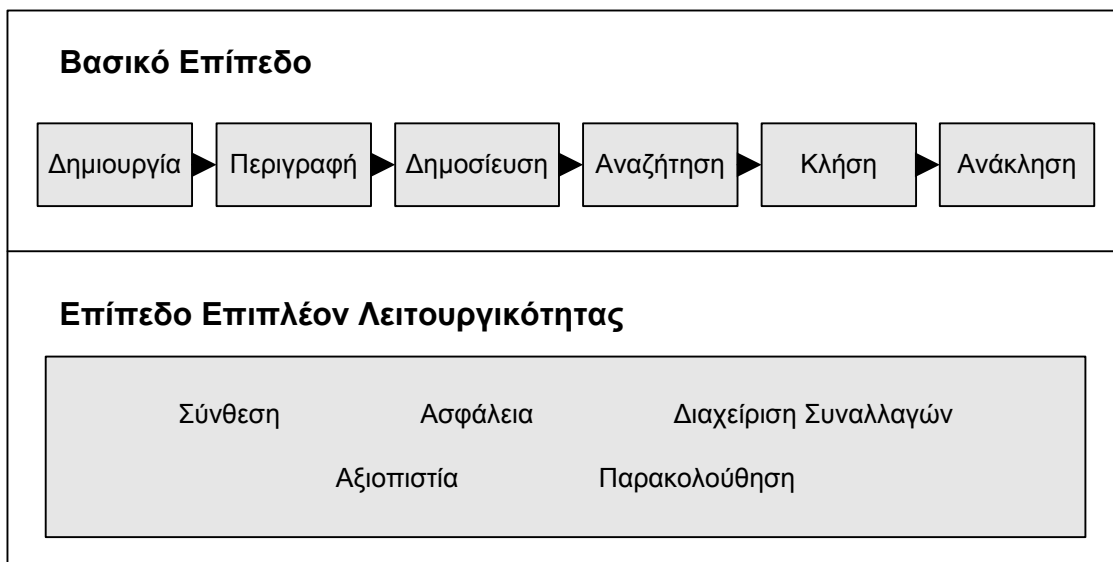
Ο αιτών υπηρεσίας αναζητεί πληροφορίες για Web Services στη UDDI Registry. Η αναζήτηση μπορεί να γίνει είτε κατά το χρόνο ανάπτυξης (development time) είτε κατά το χρόνο εκτέλεσης (run time).

Κατά το χρόνο ανάπτυξης, μπορεί να πραγματοποιηθεί αναζήτηση κατάλληλων Web Services στη UDDI Registry (βέλη 3, 4) και οι περιγραφές που θα επιστραφούν να χρησιμοποιηθούν για την ανάκτηση του αντίστοιχου WSDL εγγράφου από τον παροχέα (βέλη 5, 6). Εφόσον μελετηθούν οι πληροφορίες του WSDL εγγράφου, δημιουργούνται client proxies μέσω των οποίων η εφαρμογή έχει πρόσβαση στο Web Service (βέλη 7, 8).

Κατά το χρόνο εκτέλεσης, η εφαρμογή δε χρειάζεται να χρησιμοποιήσει τη UDDI Registry, αν το Web Service είναι διαρκώς διαθέσιμο και πάντα το ίδιο (όπως συμβαίνει με την εφαρμογή που αναπτύχθηκε στην παρούσα πτυχιακή). Ωστόσο, αν το Web Service κάποια στιγμή σταματήσει να είναι διαθέσιμο, η εφαρμογή μπορεί με κατάλληλες επερωτήσεις στη UDDI Registry να διαπιστώσει αν το Web Service έχει απλά μετακινηθεί σε άλλη τοποθεσία. Επιπλέον, είναι δυνατόν η επιλογή του κατάλληλου Web Service να γίνει κατά το χρόνο εκτέλεσης (πχ μία εφαρμογή μπορεί να αναζητεί κατά το χρόνο εκτέλεσης το Web Service που παρέχει τα πιο πρόσφατα οικονομικά νέα).

#### 7.2.4. Κύκλος Ζωής των Web Services

Ο κύκλος ζωής ενός Web Service περιλαμβάνει ένα πλήθος από δραστηριότητες (σχήμα 6):



Σχήμα 6: Κύκλος ζωής ενός Web Service

Προφανώς, άλλες ενέργειες εκτελούνται από τον αιτούντα υπηρεσίας και άλλες από τον ενδιάμεσο ή τον παροχέα.

Η πρώτη ενέργεια στον κύκλο ζωής ενός Web Service είναι η δημιουργία του, η οποία ενδέχεται να βασίζεται σε υπάρχοντα Web Services. Η δεύτερη ενέργεια είναι η περιγραφή του, ενώ ακολουθεί η δημοσίευσή του σε μία Registry. Ένα Web Service που έχει δημοσιευτεί είναι δυνατόν κατά τη διάρκεια της ζωής του να χρειαστεί να ενημερωθεί



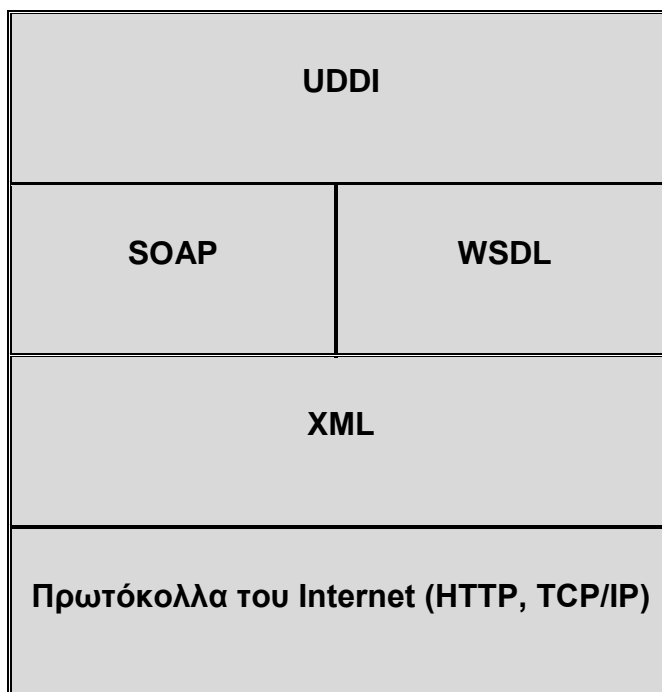
ή να ανακληθεί. Η αναζήτηση μπορεί να διευκολυνθεί από έναν ενδιάμεσο ο οποίος αναμένεται να υποστηρίζει την ανάλυση των απαιτήσεων και την περιγραφή των αναγκών των χρηστών, να συγκρίνει τις ανάγκες των χρηστών με τα υπάρχοντα Web Services καθώς και να στηρίζει τη σύνθεση των Web Services. Τελικά, ένα Web Service μπορεί να ανακληθεί, αν δεν είναι πια διαθέσιμο ή αναγκαίο. Επίσης, κατά τη διάρκεια του κύκλου ζωής είναι δυνατόν να υπάρξουν και επιπλέον ενέργειες που σχετίζονται με την ασφάλεια, την αξιοπιστία, την παρακολούθηση και τη διαχείριση των συναλλαγών των Web Services.

Κατά συνέπεια, οι ενέργειες που εκτελούνται κατά τη διάρκεια του κύκλου ζωής ενός Web Service μπορούν να διαχωριστούν σε δύο επίπεδα:

- Στο ‘βασικό επίπεδο’ περιλαμβάνονται οι ενέργειες της δημιουργίας, της περιγραφής, της δημοσίευσης, της αναζήτησης, της κλήσης και της ανάκλησης. Αυτές οι ενέργειες πρέπει να υποστηρίζονται από κάθε περιβάλλον που χρησιμοποιεί Web Services.
- Στο ‘επίπεδο επιπλέον λειτουργικότητας’ συγκαταλέγονται ενέργειες που προσφέρουν επιπλέον λειτουργικότητα σε ένα περιβάλλον χρήσης Web Services και συνεισφέρουν στην αύξηση της αποδοτικότητας. Ως τέτοιες ενέργειες αναφέρονται η σύνθεση, η ασφάλεια, η αξιοπιστία, η παρακολούθηση και η διαχείριση των συναλλαγών που εμπλέκουν Web Services.

#### 7.2.5. Σύνοψη και Συμπεράσματα

Βάσει των όσων ειπώθηκαν στις προηγούμενες ενότητες του κεφαλαίου, συμπεραίνεται ότι η χρήση των προτύπων WSDL, UDDI και SOAP βοηθούν, ώστε τα προβλήματα της ενοποίησης και της διαλειτουργικότητας των εφαρμογών να απλοποιηθούν σε επίπεδα [12] (σχήμα 7):



Σχήμα 7: Πρότυπα WSDL, UDDI και SOAP από όψη επιπέδων

Η XML παρέχει μία ανεξαρτήτου πλατφόρμας προσέγγιση για την κωδικοποίηση και τη δόμηση των δεδομένων. Το πρότυπο SOAP, το οποίο βασίζεται στην XML, καθορίζει έναν απλό τρόπο για την ανταλλαγή πληροφορίας εκτός των ορίων των συστημάτων και με τρόπο που δεν εξαρτάται από τις προγραμματιστικές γλώσσες και τα λειτουργικά συστήματα τα οποία χρησιμοποιούν τα επικοινωνούντα μέρη. Το πρότυπο WSDL είναι μία XML γραμματική που καθορίζει ιδιότητες των Web Services, ενώ οι UDDI προδιαγραφές συνιστούν ένα υψηλότερο επίπεδο που επιτρέπει στους ενδιαφερόμενους να πραγματοποιήσουν επερωτήσεις για Web Services καθώς και να περιγράψουν τα δικά τους Web Services.

Η εφαρμογή που δημιουργήθηκε στο πλαίσιο της διπλωματικής εργασίας της Κρουστάλλη και αναβαθμίστηκε στο πλαίσιο της παρούσας πτυχιακής εργασίας βασίζεται σε ένα Web Service, το οποίο δέχεται ως είσοδο ένα ρηματικό τύπο και μία ένδειξη για τη γλώσσα στην οποία θα επιστραφούν τα αποτελέσματα και επιστρέφει πληροφορίες που αφορούν στη μορφοσυντακτική του αναγνώριση (για περισσότερες πληροφορίες βλ. §7.4). Η απόφαση να στηριχτεί η βασική λειτουργικότητα της εφαρμογής σε ένα Web Service δικαιολογείται ως ακολούθως:

- Εφόσον το Web Service γίνει διαθέσιμο στο internet μέσω ενός web container, μπορεί οποιοσδήποτε χρήστης ή άλλη εφαρμογή να καλέσει μία δημόσια μέθοδο του Web Service με τα κατάλληλα ορίσματα κι έτσι να λάβει πληροφορίες που αφορούν στη μορφοσυντακτική αξία και τη λημματοποίηση του τύπου εισόδου. Κατά συνέπεια, η εργασία που έγινε στο πλαίσιο της παρούσας πτυχιακής δεν περιορίζεται στην υλοποίηση της εφαρμογής VERBTAGGR++, αλλά είναι δυνατόν να αποδειχθεί χρήσιμη και σε άλλες εφαρμογές που θα καλέσουν το Web Service μέσω του internet, χωρίς να ενδιαφέρει ούτε ο τρόπος υλοποίησης του Web Service ούτε η δομή της βάσης και η διαδικασία που επιτελείται, προκειμένου να αναγνωριστεί ο ρηματικός τύπος εισόδου.
- Η εφαρμογή γίνεται εύκολα επεκτάσιμη, ώστε να είναι δυνατόν μελλοντικά να δημιουργηθεί νέα έκδοση, η οποία εκτός της μορφοσυντακτικής αναγνώρισης θα υποστηρίζει και τη λημματοποίηση των ρηματικών τύπων της Νέας Ελληνικής. Αυτό δικαιολογείται, διότι μπορεί μελλοντικά να υλοποιηθεί ένα άλλο Web Service, το οποίο θα επιτελεί τη λημματοποίηση ενός ρηματικού τύπου και θα καλείται από την εφαρμογή VERBTAGGR++.

Για τους παραπάνω λόγους αποφασίστηκε να υλοποιηθεί ένα Web Service το οποίο θα επιτελεί τη λειτουργία της μορφοσυντακτικής αναγνώρισης ενός μονολεκτικού ρηματικού τύπου, ενώ στη συνέχεια αναπτύχθηκε η διαδικτυακή εφαρμογή, η οποία καλεί αυτό το Web Service.

### **7.3. Αλγόριθμος Μορφοσυντακτικής Αναγνώρισης Μονολεκτικών Ρηματικών Τύπων της Νέας Ελληνικής**

Το Web Service που υλοποιήθηκε επιτελεί μορφοσυντακτική αναγνώριση των μονολεκτικών ρηματικών τύπων της Νέας Ελληνικής και η γενική μορφή του αλγόριθμου που εφαρμόζει αποτελεί το αντικείμενο της παρούσας ενότητας. Η βασική ιδέα του αλγόριθμου σε μορφή ψευδοκώδικα εμφανίζεται στο σχήμα 8, όπου γίνονται εμφανή τα βήματα που ακολουθούνται από τη συνάρτηση `recognGrVerbForm`, η οποία δέχεται ως είσοδο μία συμβολοσειρά που αναπαριστά ένα ρηματικό τύπο της Νέας Ελληνικής (`verb`) και επιστρέφει πληροφορίες σχετικές με τη μορφοσυντακτική του αναγνώριση. Οι συμβολισμοί που χρησιμοποιούνται είναι οι ακόλουθοι:

- `length_of_verb`:

Το πλήθος των χαρακτήρων από τους οποίους αποτελείται η είσοδος `verb`.

▪ `substring(verb, i)`:

Το τμήμα των χαρακτήρων της συμβολοσειράς `verb` που ξεκινά από την  $i$ -οστή θέση και εκτείνεται μέχρι τον τελευταίο χαρακτήρα.

▪ `string(str1 + str2)`:

Η συμβολοσειρά που αποτελείται από τη συνένωση των συμβολοσειρών `str1` και `str2`.

▪ `str is_a_final_grapheme_sequence`:

Η παράσταση είναι αληθής, αν η συμβολοσειρά `str` είναι αποδεκτή καταληκτική γραφηματική σειρά. Υπενθυμίζεται ότι οι καταληκτικές γραφηματικές σειρές στις οποίες βασίζεται η αναγνώριση των ρηματικών τύπων έχουν προκύψει στατιστικά από τη διδακτορική διατριβή της Λεμπέση και φυλάσσονται στο πεδίο `final_grapheme_seq` του πίνακα `verb_info` της βάσης.

▪ `info(str)`:

Πληροφορίες που αφορούν στο μορφοσυντακτικό περιεχόμενο και στη λημματοποίηση των τύπων που λήγουν στην καταληκτική γραφηματική σειρά `str`.

**function recognGrVerbForm (verb)**

```
1. if (verb is_a_final_grapheme_sequence)
2.     return info(verb);
3. else {
4.     matchNotFound = true;
5.     i = 1;
6.     while (matchNotFound && i ≤ length_of_verb) {
7.         subVerb = substring(verb, i);
8.         if(string('*' + subVerb) is_a_final_grapheme_sequence) {
9.             matchNotFound = false;
10.            return info(subVerb);
11.        }
12.        else if(string('?' + subVerb)
13.            is_a_final_grapheme_sequence) {
14.            matchNotFound = false;
15.            return info(subVerb);
16.        }
17.        i = i+1;
18.    } //end of while
19.    if (matchNotFound)
20.        return ('verb_not_recognized');
```

## Σχήμα 8: Αλγόριθμος μορφοσυντακτικής αναγνώρισης μονολεκτικού ρηματικού τύπου

Με τον παραπάνω αλγόριθμο επιχειρείται να βρεθεί μία καταληκτική γραφηματική σειρά βάσει της οποίας γίνεται η αναγνώριση της συμβολοσειράς εισόδου. Αυτή η γραφηματική σειρά καλύπτει όσο το δυνατόν περισσότερους χαρακτήρες από τη συμβολοσειρά εισόδου ξεκινώντας από αριστερά προς τα δεξιά. Συγκεκριμένα, εξετάζεται αν αυτούσια η συμβολοσειρά εισόδου αποτελεί καταληκτική γραφηματική σειρά (γραμμή 1), οπότε επιστρέφονται πληροφορίες που αφορούν στο συγκεκριμένο τύπο και ο αλγόριθμος ολοκληρώνεται (γραμμή 2). Διαφορετικά, γίνονται αρχικοποιήσεις μεταβλητών (γραμμές 4 και 5) και ακολουθεί ένας βρόχος επανάληψης (γραμμές 6 έως 17), ο οποίος εκτελείται, όσο δεν έχει βρεθεί ταίριασμα με καταληκτική γραφηματική σειρά και δεν έχει γίνει σάρωση ολόκληρης της συμβολοσειράς εισόδου. Εντός του βρόχου επανάληψης η μεταβλητή *subVerb* λαμβάνει ως τιμή το τμήμα των χαρακτήρων της συμβολοσειράς εισόδου που ξεκινά από την *i*-οστή θέση και εκτείνεται μέχρι το τέλος αυτής (γραμμή 7). Σε κάθε επανάληψη η τιμή του *i* αυξάνεται κατά ένα (γραμμή 16) ξεκινώντας από την τιμή 1. Κατά συνέπεια η μεταβλητή *subVerb* την πρώτη φορά έχει ως τιμή ολόκληρη τη συμβολοσειρά εισόδου και σε κάθε επανάληψη λαμβάνει ως νέα τιμή τη συμβολοσειρά που προκύπτει, αν από την υπάρχουσα τιμή της αφαιρεθεί ο πρώτος από αριστερά χαρακτήρας. Εφόσον η μεταβλητή *subVerb* λάβει τη νέα τιμή της, εξετάζονται κατά σειρά οι ακόλουθες δύο συνθήκες:

- Πρώτα ελέγχεται αν η συμβολοσειρά που αποτελείται από τη συνένωση ενός αστερίσκου με το περιεχόμενο της *subVerb* αποτελεί καταληκτική γραφηματική σειρά και αν αυτό συμβαίνει, επιστρέφονται πληροφορίες που αφορούν στη συγκεκριμένη γραφηματική σειρά (γραμμές 8 – 11).

Αν ο προηγούμενος έλεγχος αποτύχει, εξετάζεται αν η συμβολοσειρά που αποτελείται από τη συνένωση ενός αγγλικού ερωτηματικού με το περιεχόμενο της *subVerb* αποτελεί καταληκτική γραφηματική σειρά και αν αυτό συμβαίνει, επιστρέφονται πληροφορίες για αυτήν τη γραφηματική σειρά (γραμμές 12 – 15).

Αν τελικά η μεταβλητή *subVerb* πάρει ως τιμή τον τελευταίο χαρακτήρα της συμβολοσειράς εισόδου (δηλ. όταν το *i* πάρει ως τιμή το `length_of_verb`) και πάλι δεν προκύψει μία καταληκτική γραφηματική σειρά καταχωρημένη στην εφαρμογή, τότε αυτό σημαίνει πως η συμβολοσειρά εισόδου δεν αναγνωρίστηκε ως αποδεκτός ρηματικός τύπος, οπότε εκτελούνται οι γραμμές 18 – 19.

Για παράδειγμα, αν ως είσοδος δοθεί η συμβολοσειρά 'κυκλοφορούν', τότε κατά την εκτέλεση του παραπάνω αλγόριθμου, θα γίνουν τα εξής:

- Πρώτα θα εξεταστεί αν αυτούσια η συμβολοσειρά 'κυκλοφορούν' αποτελεί καταληκτική γραφηματική σειρά καταχωρημένη στην εφαρμογή.
- Ο παραπάνω έλεγχος αποτυγχάνει, γι' αυτό το *i* θα πάρει την τιμή 1 και θα εκτελεστεί ο βρόχος του `while`. Έτσι, θα γίνει ο ίδιος έλεγχος για τη συμβολοσειρά '\*κυκλοφορούν'.
- Ο έλεγχος για το '\*κυκλοφορούν' επίσης θα αποτύχει, γι' αυτό ο έλεγχος θα γίνει για τη συμβολοσειρά '?κυκλοφορούν'.
- Επειδή ούτε το '?κυκλοφορούν' αποτελεί αποδεκτή καταληκτική γραφηματική σειρά, το *i* θα πάρει την τιμή 2 και θα εκτελεστεί δεύτερη επανάληψη του βρόχου `while` κατά την οποία θα ελεγχθούν ανεπιτυχώς αν οι συμβολοσειρές '\*υκλοφορούν' και '?υκλοφορούν' αποτελούν καταληκτικές γραφηματικές σειρές.

- Τελικά, κατά την τρίτη επανάληψη θα ελεγχθεί η συμβολοσειρά *\*κλοφορούν*, για την οποία θα διαπιστωθεί ότι αποτελεί αποδεκτή καταληκτική γραφηματική σειρά, οπότε θα επιστραφούν οι πληροφορίες που αφορούν στη γραφηματική σειρά *\*κλοφορούν*.

Εάν ως είσοδος δινόταν η συμβολοσειρά *‘ξανακυκλοφορούν*’, τότε ομοίως, θα επιστρεφόταν οι πληροφορίες που αφορούν στην καταληκτική γραφηματική σειρά *‘\*ακυκλοφορούν*’.

Στο σημείο αυτό θα πρέπει να αναφερθεί πως υπάρχουν περιπτώσεις κατά τις οποίες δε γίνεται απόλυτα σωστή η αναγνώριση της συμβολοσειράς εισόδου και οι αδυναμίες αυτές προκύπτουν από την έρευνα της Λεμπέση. Όπως παρουσιάστηκε στο κεφάλαιο 4, από την εργασία της Λεμπέση παράχθηκε ένα σύνολο καταληκτικών γραφηματικών σειρών των οποίων το κριτήριο διαχωρισμού είναι το διαφορετικό κλιτικό υπόδειγμα των ρηματικών τύπων που εξετάστηκαν. Κατά συνέπεια, υπάρχουν περιπτώσεις ρηματικών τύπων οι οποίοι ανήκουν στο ίδιο κλιτικό υπόδειγμα και αναγνωρίζονται βάσει της ίδιας καταληκτικής γραφηματικής σειράς αλλά έχουν διαφορετικό μορφοσυντακτικό περιεχόμενο. Για τους τύπους αυτούς, ως πιθανό μορφοσυντακτικό περιεχόμενο δίνονται οι πληροφορίες που αφορούν σε όλους τους τύπους που δηλώνονται από τη συγκεκριμένη καταληκτική γραφηματική σειρά. Επειδή προφανώς δεν αφορούν όλες αυτές οι πληροφορίες τον εκάστοτε δοθέντα ρηματικό τύπο, το αποτέλεσμα είναι σε τέτοιες περιπτώσεις να δοθούν εκτός των σωστών μορφοσυντακτικών πληροφοριών και κάποιες επιπλέον λανθασμένες. Ως παράδειγμα αναφέρονται τα ρήματα *έφυγα* και *έφευγα*, τα οποία ανήκουν στο κλιτικό υπόδειγμα 30 και αναγνωρίζονται βάσει της καταληκτικής γραφηματικής σειράς *\*υγα*. Ωστόσο, η μορφοσυντακτική αξία των τύπων διαφέρει στην πληροφορία του χρόνου – αόριστος για τον τύπο *έφυγα*, παρατατικός για τον τύπο *έφευγα*. Έτσι, όταν εισαχθεί προς αναγνώριση κάποιος από τους δύο τύπους, στην έξοδο θα δοθούν και οι δύο πληροφορίες για το χρόνο, ενώ έγκυρη είναι κάθε φορά η μία από τις δύο. Στην §7.2 απεικονίζεται η έξοδος της εφαρμογής για τον τύπο *έφυγα*. Όπως έχει ήδη προαναφερθεί, μόνο περαιτέρω στατιστική ανάλυση εις βάθος τύπων του ίδιου κλιτικού υποδείγματος θα μπορέσει να εντοπίσει διάφορες καταληκτικές γραφηματικές σειρές διαφορετικού μορφοσυντακτικού περιεχομένου όπως *\*ευγα* (παρατατικός) ή *\*φυγα* (αόριστος) για το συγκεκριμένο παράδειγμα.

Μία ακόμη αδυναμία που θα πρέπει να σημειωθεί προκύπτει από το ότι η πληροφορία μορφοσυντακτικού περιεχομένου που παρέχεται αφορά στην εκάστοτε εντοπισθείσα καταληκτική γραφηματική σειρά και κατ' επέκταση στον υπό αναγνώριση ρηματικό τύπο. Το γεγονός αυτό δεν αποκλείει την αναγνώριση και μη υπαρκτών ρηματικών τύπων. Για παράδειγμα, η ύπαρξη της καταληκτικής γραφηματικής σειράς *\*ακυκλοφορούν* έχει ως αποτέλεσμα να αναγνωρίζονται όλες οι συμβολοσειρές που αρχίζουν με οποιοδήποτε σύνολο χαρακτήρων και λήγουν στους χαρακτήρες *‘ακυκλοφορούν*’. Η αδυναμία αυτή είναι εγγενής της αντίστροφης στατιστικής προσέγγισης και δε μπορεί να αντιμετωπιστεί.

#### 7.4. Περιγραφή Web Service Μορφοσυντακτικής Αναγνώρισης Μονολεκτικών Ρηματικών Τύπων της Νέας Ελληνικής

Το Web Service που δημιουργήθηκε κατά την ανάπτυξη της διαδικτυακής εφαρμογής μορφοσυντακτικής αναγνώρισης των μονολεκτικών ρηματικών τύπων της Νέας Ελληνικής υλοποιήθηκε σε Java και περιλαμβάνει μία δημόσια μέθοδο με όνομα **recognGrVerbForm**.

Η μέθοδος αυτή δέχεται ως ορίσματα δύο συμβολοσειρές: το πρώτο αναπαριστά έναν ελληνικό ρηματικό τύπο τον οποίο η μέθοδος θα προσπαθήσει να αναγνωρίσει μορφοσυντακτικά, ενώ το δεύτερο εκφράζει τη γλώσσα στην οποία θα επιστραφούν τα αποτελέσματα της αναγνώρισης (*‘gr’* για Ελληνικά, *‘en’* για Αγγλικά). Η μέθοδος

επιστρέφει μία δομή HashMap με πληροφορίες που αφορούν στα αποτελέσματα της αναγνώρισης.

(Σημείωση:

Μία δομή HashMap στη γλώσσα Java περιέχει ένα πλήθος από στοιχεία καθένα από τα οποία αποτελείται από ένα ζεύγος ‘όνομα κλειδιού’ – ‘τιμή’ εκφράζοντας ότι ένα στοιχείο συγκεκριμένου ονόματος περιέχει την αντίστοιχη τιμή.)

Συγκεκριμένα, στη δομή που επιστρέφεται περιέχεται ένα στοιχείο με όνομα ‘mode’, το οποίο υποδεικνύει αν η αναγνώριση πραγματοποιήθηκε με επιτυχία ή υπήρξε πρόβλημα. Οι δυνατές τιμές αυτού του στοιχείου είναι οι ακόλουθες:

- Η συμβολοσειρά ‘-1’, εάν κατά τη διαδικασία αναγνώρισης προκλήθηκε μία εξαίρεση (excerption). Στην περίπτωση αυτή, ένα άλλο στοιχείο της δομής με όνομα ‘exMsg’ έχει ως τιμή μία περιγραφή του λάθους που προκλήθηκε.
- Η συμβολοσειρά ‘NOT\_GREEK\_VERB\_FORM’, εάν το πρώτο όρισμα δεν αποτελείται αποκλειστικά από ελληνικούς χαρακτήρες.
- Η συμβολοσειρά ‘WRONG\_LEMINFO\_SIZE’, στην περίπτωση που η αναγνώριση του ρηματικού τύπου εισόδου πραγματοποιήθηκε με χρήση μίας γραμμής του πίνακα ‘verb\_info’ στην οποία το πλήθος των κλιτικών υποδειγμάτων του πεδίου model, το πλήθος των κωδικών ληματοποίησης του πεδίου lemmaCode, το πλήθος των απόλυτων συχνοτήτων του πεδίου ‘absoluteFreq’ και το πλήθος των καταληκτικών γραφηματικών σειρών του πεδίου ‘exceptionToForms’ δεν είναι όλα μεταξύ τους ίσα.

(Σημείωση: Τα διάφορα κλιτικά υποδείγματα, οι κωδικοί ληματοποίησης, οι απόλυτες συχνότητες και οι καταληκτικές γραφηματικές σειρές που περιέχονται στα τέσσερα προαναφερθέντα πεδία, διαχωρίζονται μεταξύ τους με μία παύλα ‘-’. Για περισσότερες πληροφορίες βλ. §5.2).

- Τη συμβολοσειρά ‘EQUALITY\_NOT\_FOUND’, στην περίπτωση που η αναγνώριση του ρηματικού τύπου εισόδου πραγματοποιήθηκε με χρήση μίας γραμμής του πίνακα ‘verb\_info’ στην οποία το πεδίο ‘extraMeaning’ περιέχει την τιμή μίας καταληκτικής γραφηματικής σειράς στην οποία αντιστοιχούν περισσότεροι από ένας μορφοσυντακτικοί κωδικοί και κανένας από αυτούς δεν είναι ίσος με το μορφοσυντακτικό κωδικό της γραμμής βάσει της οποίας έγινε η αναγνώριση.
- Τη συμβολοσειρά ‘NOT\_VALID\_MSCODE’, στην περίπτωση που κατά την αναγνώριση του ρηματικού τύπου εισόδου εντοπίστηκε στο πεδίο ‘ms\_code’ του πίνακα ‘verb\_info’ μη αποδεκτή τιμή μορφοσυντακτικού κωδικού, δηλαδή τιμή η οποία δεν αντιστοιχεί σε καμία από τις τιμές που περιέχονται στο πεδίο ‘ms\_code’ του πίνακα ‘ms\_codeExplain’.
- Τη συμβολοσειρά ‘WRONG\_EXCEPTION\_TO\_FORMS’, στην περίπτωση που η αναγνώριση του ρηματικού τύπου εισόδου πραγματοποιήθηκε με χρήση μίας γραμμής του πίνακα ‘verb\_info’ η οποία στο πεδίο ‘exceptionToForms’ περιέχει μη αποδεκτή τιμή καταληκτικής γραφηματικής σειράς, δηλαδή τιμή η οποία δεν αντιστοιχεί σε καμία από τις τιμές του πεδίου ‘final\_grapheme\_seq’ του πίνακα ‘verb\_info’.
- Τη συμβολοσειρά ‘WRONG\_EXTRA\_MEANING’, στην περίπτωση που η αναγνώριση του ρηματικού τύπου εισόδου πραγματοποιήθηκε με χρήση μίας γραμμής του πίνακα ‘verb\_info’ η οποία στο πεδίο ‘extraMeaning’ περιέχει μη αποδεκτή τιμή καταληκτικής γραφηματικής σειράς, δηλαδή τιμή η οποία δεν αντιστοιχεί σε καμία από τις τιμές του πεδίου ‘final\_grapheme\_seq’ του πίνακα ‘verb\_info’.

- Τη συμβολοσειρά '1', αν η αναγνώριση πραγματοποιήθηκε με επιτυχία, δηλαδή χωρίς να προκύψει κάποιο από τα προβλήματα που αναφέρθηκαν παραπάνω. Στην περίπτωση αυτή η δομή HashMap που επιστρέφει η μέθοδος περιέχει και ένα πλήθος από άλλα στοιχεία:
  - Ένα στοιχείο με όνομα 'verbFormCategory' και τιμή μία από τις ακόλουθες συμβολοσειρές:
    - Τη συμβολοσειρά '1' στην περίπτωση που η αναγνώριση πραγματοποιήθηκε βάσει μίας τουλάχιστον γραμμής του πίνακα 'verb\_info' η οποία περιέχει ως τιμή στο πεδίο 'final\_grapheme\_seq' αυτούσιο το ρηματικό τύπο εισόδου.
    - Τη συμβολοσειρά '2' στην περίπτωση που η αναγνώριση πραγματοποιήθηκε βάσει μίας τουλάχιστον γραμμής του πίνακα 'verb\_info' η οποία περιέχει ως τιμή στο πεδίο 'final\_grapheme\_seq' μία συμβολοσειρά που αποτελείται από το χαρακτήρα '?' και όλους τους χαρακτήρες του ρηματικού τύπου εισόδου.
    - Τη συμβολοσειρά '3' στην περίπτωση που η αναγνώριση πραγματοποιήθηκε βάσει μίας τουλάχιστον γραμμής του πίνακα 'verb\_info' η οποία περιέχει ως τιμή στο πεδίο 'final\_grapheme\_seq' μία συμβολοσειρά που αποτελείται από το χαρακτήρα '\*' και όλους τους χαρακτήρες του ρηματικού τύπου εισόδου. Αυτό σημαίνει ότι η συμβολοσειρά που δόθηκε ως όρισμα δεν υφίσταται ως ρηματικός τύπος της Νέας Ελληνικής, αλλά συναντάται ως καταληκτική γραφηματική σειρά της γλώσσας.
    - Τη συμβολοσειρά '4' στην περίπτωση που η αναγνώριση πραγματοποιήθηκε βάσει μίας τουλάχιστον γραμμής του πίνακα 'verb\_info' η οποία περιέχει ως τιμή στο πεδίο 'final\_grapheme\_seq' μία συμβολοσειρά που αποτελείται από το χαρακτήρα '\*' και ένα μέρος των χαρακτήρων του ρηματικού τύπου εισόδου.
    - Τη συμβολοσειρά '5' στην περίπτωση που η αναγνώριση πραγματοποιήθηκε βάσει μίας τουλάχιστον γραμμής του πίνακα 'verb\_info' η οποία περιέχει ως τιμή στο πεδίο 'final\_grapheme\_seq' ένα string που αποτελείται από το χαρακτήρα '?' και ένα μέρος των χαρακτήρων του ρηματικού τύπου εισόδου.
    - Τη συμβολοσειρά '6' στην περίπτωση που η συμβολοσειρά του πρώτου ορίσματος δεν αναγνωρίστηκε ως κάποιος αποδεκτός ρηματικός τύπος της Νέας Ελληνικής. Τότε η επιστρεφόμενη δομή περιέχει μόνο το στοιχείο 'mode' και το 'verbForm' που περιγράφεται στη συνέχεια.
  - Ένα στοιχείο με όνομα 'verbForm' και τιμή τη συμβολοσειρά εισόδου με πεζούς ελληνικούς χαρακτήρες.
  - Ένα στοιχείο με όνομα 'final\_grapheme\_seq' και τιμή την καταληκτική γραφηματική σειρά βάσει της οποίας έγινε η αναγνώριση. Αυτή η συμβολοσειρά ενδέχεται να αποτελείται από όλους τους χαρακτήρες του ρηματικού τύπου εισόδου, γεγονός το οποίο συμβαίνει, όταν το στοιχείο 'verbFormCategory' έχει την τιμή '1'.

- Ένα στοιχείο με όνομα *'totalAbsFreq'* και τιμή το συνολικό πλήθος των ρηματικών τύπων που λήγουν στη γραφηματική σειρά στην οποία βασίστηκε η αναγνώριση.
- Ένα στοιχείο με όνομα *'lemmaInfo'* το οποίο αφορά σε πληροφορίες χρήσιμες για τη λημματοποίηση του δοθέντος τύπου. Έτσι, η τιμή του στοιχείου με όνομα *'lemmaInfo'* είναι μία δομή *ArrayList*, η οποία περιέχει τόσα στοιχεία όσα και τα κλιτικά υποδείγματα στα οποία ανήκουν οι τύποι που λήγουν στη γραφηματική σειρά στην οποία βασίστηκε η αναγνώριση. Συγκεκριμένα, κάθε στοιχείο της *ArrayList* είναι μία δομή *HashMap* με τα εξής στοιχεία:
  - Ένα στοιχείο με όνομα *'model'* και τιμή το κλιτικό υπόδειγμα που αφορά η συγκεκριμένη δομή *HashMap*.
  - Ένα στοιχείο με όνομα *'absFreq'* και τιμή την απόλυτη συχνότητα που αφορά το συγκεκριμένο κλιτικό υπόδειγμα, δηλαδή το πλήθος των ρηματικών τύπων που λήγουν στη γραφηματική σειρά στην οποία βασίστηκε η αναγνώριση και ανήκουν στο συγκεκριμένο κλιτικό υπόδειγμα.
  - Ένα στοιχείο με όνομα *'exceptionToForms'* και τιμή της μορφής *'xxx (yyy)'* όπου *'xxx'* η καταληκτική γραφηματική σειρά της οποίας εξαίρεση αποτελούν οι τύποι που λήγουν στη γραφηματική σειρά στην οποία βασίστηκε η αναγνώριση και ανήκουν στο κλιτικό υπόδειγμα που εκφράζει η συγκεκριμένη δομή *HashMap* και *'yyy'* η τιμή του κλιτικού υποδείγματος στο οποίο ανήκει η γραφηματική σειρά *'xxx'*. Σημειώνεται ότι στην περίπτωση που δεν υφίσταται τέτοια εξαίρεση, το στοιχείο αυτό έχει ως τιμή μία παύλα ('-').
  - Ένα στοιχείο με όνομα *'lemmaData'* και τιμή μία δομή *ArrayList* της οποίας κάθε στοιχείο είναι μία δομή *HashMap* με περιεχόμενο τα ακόλουθα δύο στοιχεία:
    - Ένα στοιχείο με όνομα *'msc'* και τιμή μία συμβολοσειρά που εκφράζει ένα μορφοσυντακτικό κωδικό.
    - Ένα στοιχείο με όνομα *'lemmaCode'* και τιμή μία συμβολοσειρά που εκφράζει έναν κωδικό λημματοποίησης.Σημειώνεται ότι τα δεδομένα του στοιχείου *'lemmaData'* δεν εμφανίζονται στην εφαρμογή, ωστόσο επιστρέφονται μέσω του web service, διότι χρειάζονται στη διαδικασία της λημματοποίησης που μελλοντικά θα υποστηρίξει η εφαρμογή.
- Ένα στοιχείο με όνομα *'mscInfo'* το οποίο περιέχει πληροφορίες που αφορούν στη μορφοσυντακτική αξία της καταληκτικής γραφηματικής σειράς στην οποία βασίστηκε η αναγνώριση. Συγκεκριμένα, το στοιχείο αυτό περιέχει μία δομή *ArrayList*, κάθε στοιχείο της οποίας είναι μία δομή *HashMap* η οποία εκφράζει ένα διαφορετικό μορφοσυντακτικό κωδικό κι έτσι περιέχει τα εξής στοιχεία:
  - Ένα στοιχείο με όνομα *'ms\_code'* και τιμή το μορφοσυντακτικό κωδικό στον οποίο αφορά η συγκεκριμένη δομή *HashMap*.
  - Ένα στοιχείο με όνομα *'occurFreq'* και τιμή τη συχνότητα εμφάνισης που αφορά στο συγκεκριμένο μορφοσυντακτικό κωδικό. Ειδικότερα, αυτή η συχνότητα εμφάνισης εκφράζει το πλήθος των ρηματικών τύπων που λήγουν στη γραφηματική σειρά στην οποία βασίστηκε η αναγνώριση και των οποίων η μορφοσυντακτική αξία εκφράζεται από τον κωδικό που αναπαριστά η δομή

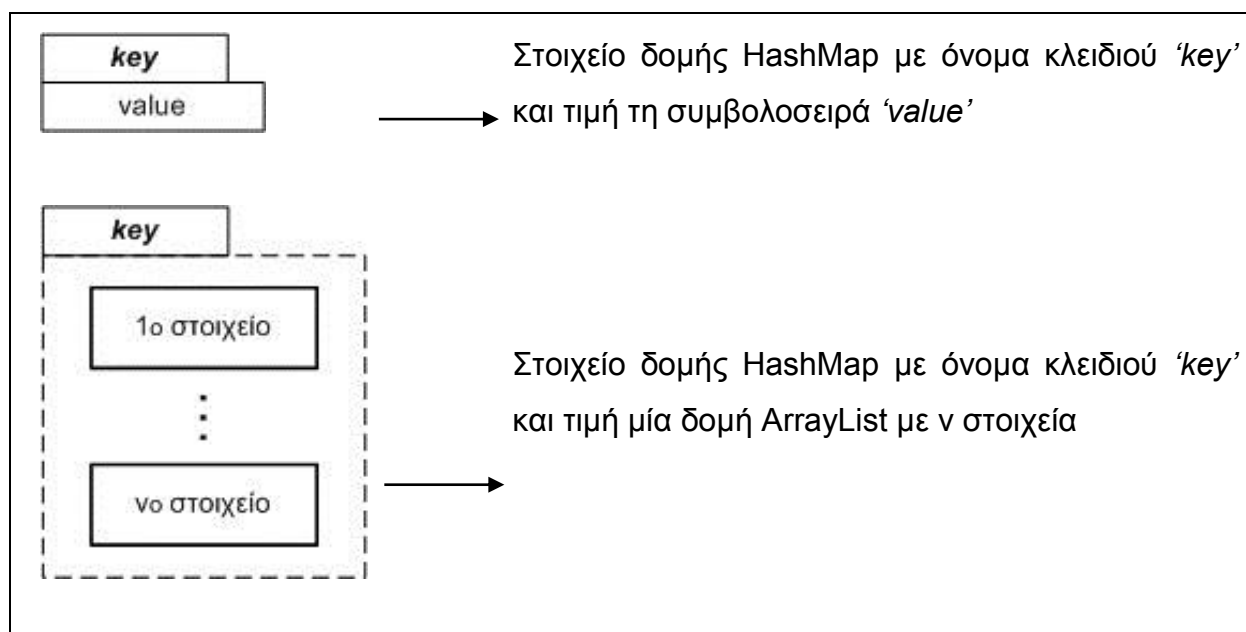


HashMap σε σχέση με το συνολικό πλήθος των τύπων που λήγουν στη γραφηματική σειρά που βασίστηκε η αναγνώριση.

- Ένα στοιχείο με όνομα *'MSC\_explain'* και τιμή μία δομή ArrayList, η οποία περιέχει τόσα στοιχεία όσες και οι ερμηνείες που αφορούν στο μορφοσυντακτικό κωδικό της συγκεκριμένης δομής HashMap. Σημειώνεται ότι η γλώσσα στην οποία είναι εκφρασμένες οι πληροφορίες που περιέχονται στη δομή ArrayList εξαρτώνται από το δεύτερο όρισμα της μεθόδου: αν το δεύτερο όρισμα είναι η συμβολοσειρά *'gr'*, οι πληροφορίες είναι εκφρασμένες στα Ελληνικά, διαφορετικά είναι εκφρασμένες στα Αγγλικά. Ειδικότερα, κάθε στοιχείο αυτής της δομής ArrayList είναι μία δομή HashMap η οποία περιέχει τα ακόλουθα στοιχεία:
  - Ένα στοιχείο με όνομα *'type'* και τιμή τη συμβολοσειρά *'verb'*, αν η ερμηνεία αφορά σε ρήμα της Νέας Ελληνικής ή τη συμβολοσειρά *'invariable participle'*, αν η ερμηνεία αφορά σε άκλιτη μετοχή της Νέας Ελληνικής ή τη συμβολοσειρά *'variable participle'*, αν η ερμηνεία αφορά σε κλιτή μετοχή της Νέας Ελληνικής.
  - Αν η ερμηνεία αφορά σε ρήμα της Νέας Ελληνικής, στη δομή περιέχονται επιπλέον τα εξής στοιχεία:
    - Ένα στοιχείο με όνομα *'person'* και τιμή που εκφράζει το πρόσωπο του ρήματος, δηλαδή μία από τις συμβολοσειρές *'1o'*, *'2o'*, *'3o'* (στα Ελληνικά) ή *'1st'*, *'2nd'*, *'3rd'* (στα Αγγλικά).
    - Ένα στοιχείο με όνομα *'number'* και τιμή που εκφράζει τον αριθμό του ρήματος, δηλαδή μία από τις συμβολοσειρές *'ΕΝΙΚΟΣ'*, *'ΠΛΗΘΥΝΤΙΚΟΣ'* (στα Ελληνικά) ή *'SINGULAR'*, *'PLURAL'* (στα Αγγλικά).
    - Ένα στοιχείο με όνομα *'tense'* και τιμή που εκφράζει το χρόνο, δηλαδή μία από τις συμβολοσειρές *'ΕΝΕΣΤΩΤΑΣ'*, *'ΑΟΡΙΣΤΟΣ'*, *'ΠΑΡΑΤΑΤΙΚΟΣ'*, *'ΣΥΝΟΠΤΙΚΟΣ ΜΗ ΠΑΡΕΛΘΟΝΤΙΚΟΣ'* (στα Ελληνικά) ή *'PRESENT'*, *'PAST'*, *'IMPERFECT'*, *'PERFECTIVE NON-PAST'* (στα Αγγλικά).
    - Ένα στοιχείο με όνομα *'voice'* και τιμή που εκφράζει τη φωνή, δηλαδή μία από τις συμβολοσειρές *'ΕΝΕΡΓΗΤΙΚΗ'*, *'ΜΕΣΟΠΑΘΗΤΙΚΗ'* (στα Ελληνικά) ή *'ACTIVE'*, *'MEDIOPASSIVE'* (στα Αγγλικά).
    - Ένα στοιχείο με όνομα *'mood'* και τιμή που εκφράζει την έγκλιση, δηλαδή μία από τις τιμές *'ΟΡΙΣΤΙΚΗ'*, *'ΠΡΟΣΤΑΚΤΙΚΗ'* (στα Ελληνικά) ή *'INDICATIVE'*, *'IMPERATIVE'* (στα Αγγλικά). Μπορεί επίσης να περιέχει τη συμβολοσειρά *'-'*, γεγονός το οποίο συμβαίνει στην περίπτωση συνοπτικών μη παρελθοντικών τύπων, διότι όπως αναφέρθηκε στο κεφάλαιο 3, οι τύποι αυτοί δεν εκφράζουν μία συγκεκριμένη έγκλιση.
  - Αν η ερμηνεία αφορά σε κλιτή μετοχή της Νέας Ελληνικής, στη δομή περιέχονται τα εξής στοιχεία:
    - Ένα στοιχείο με όνομα *'gender'* και τιμή που εκφράζει το γένος της μετοχής, δηλαδή μία από τις συμβολοσειρές *'ΑΡΣΕΝΙΚΟ'*, *'ΘΗΛΥΚΟ'*,

- ‘ΟΥΔΕΤΕΡΟ’ (στα Ελληνικά) ή ‘MASCULINE’, ‘FEMININE’, ‘NEUTER’ (στα Αγγλικά).
- Ένα στοιχείο με όνομα ‘case’ και τιμή που εκφράζει την πτώση της μετοχής, δηλαδή μία από τις συμβολοσειρές ‘ΟΝΟΜΑΣΤΙΚΗ’, ‘ΓΕΝΙΚΗ’, ‘ΑΙΤΙΑΤΙΚΗ’ (στα Ελληνικά) ή ‘NOMINATIVE’, ‘GENITIVE’, ‘ACCUSATIVE’ (στα Αγγλικά).
  - Ένα στοιχείο με όνομα ‘number’ και τιμή που εκφράζει τον αριθμό της μετοχής, δηλαδή μία από τις συμβολοσειρές ‘ΕΝΙΚΟΣ’, ‘ΠΛΗΘΥΝΤΙΚΟΣ’ (στα Ελληνικά) ή ‘SINGULAR’, ‘PLURAL’ (στα Αγγλικά).
  - Ένα στοιχείο με όνομα ‘tense’ και τιμή που εκφράζει το χρόνο, δηλαδή μία από τις συμβολοσειρές ‘ΕΝΕΣΤΩΤΑΣ’, ‘ΑΟΡΙΣΤΟΣ’ (στα Ελληνικά) ή ‘PRESENT’, ‘PAST’ (στα Αγγλικά).
  - Ένα στοιχείο με όνομα ‘voice’ και τιμή που εκφράζει τη φωνή, δηλαδή μία από τις συμβολοσειρές ‘ΕΝΕΡΓΗΤΙΚΗ’, ‘ΜΕΣΟΠΑΘΗΤΙΚΗ’ (στα Ελληνικά) ή ‘ACTIVE’, ‘MEDIOPASSIVE’ (στα Αγγλικά).
- Αν η ερμηνεία αφορά σε άκλιτη μετοχή της Νέας Ελληνικής, στη δομή περιέχονται μόνο οι δύο τελευταίες πληροφορίες που αναφέρθηκαν στην περίπτωση των κλιτών μετοχών, δηλαδή ο χρόνος και η φωνή της μετοχής.

Στη συνέχεια παρουσιάζονται σχηματικά οι δομές HashMap που επιστρέφονται από τη μέθοδο ‘recognGrVerbForm’, όταν κληθεί με παραμέτρους συγκεκριμένες συμβολοσειρές, ώστε παραστατικά να φανεί η επιστρεφόμενη δομή για ενδεικτικές εισόδους. Στα σχήματα που ακολουθούν χρησιμοποιούνται οι εξής συμβολισμοί:



Σχήμα 9: Χρησιμοποιούμενοι συμβολισμοί

### ➤ **Παράδειγμα 7.1: Επιστρεφόμενη δομή για τον τύπο παράγουν**

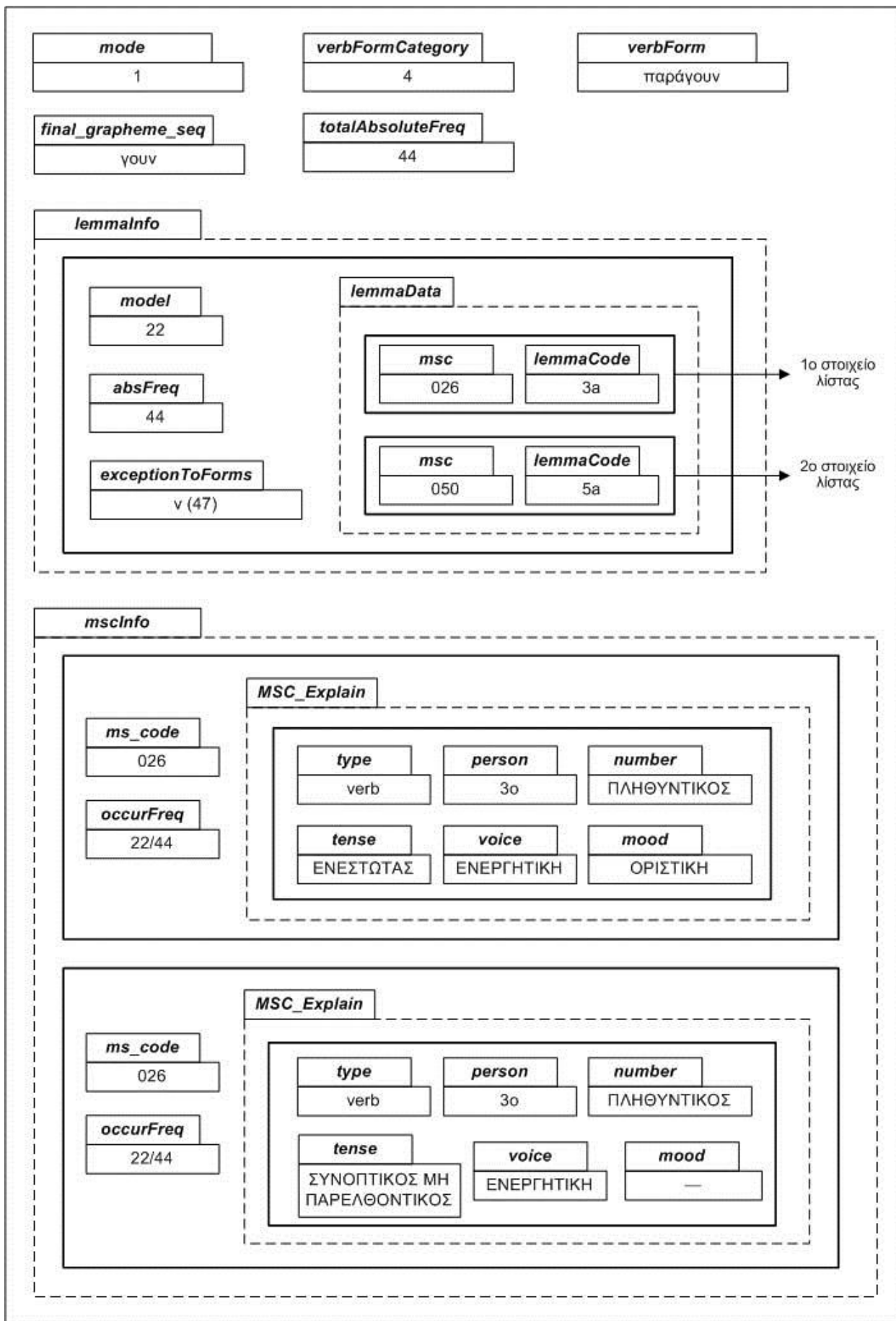
Στην περίπτωση που επιθυμούμε να γίνει η αναγνώριση του τύπου παράγουν στα Ελληνικά, θα πρέπει να κληθεί η μέθοδος ‘recognGrVerbForm’ με παραμέτρους τις

συμβολοσειρές ‘παράγουν’ και ‘gr’. Σχετικά με τον τύπο *παράγουν*, εκτελώντας τον αλγόριθμο της §7.3, διαπιστώνουμε ότι αναγνωρίζεται βάσει της καταληκτικής γραφηματικής σειράς ‘-γουν’. Οι γραμμές του πίνακα ‘verb\_info’ της βάσης που αφορούν στην καταληκτική γραφηματική σειρά ‘-γουν’ έχουν ως εξής (για τη δομή του πίνακα ‘verb\_info’ βλ. §6.2):

Πίνακας 8: Γραμμές του πίνακα verb\_info για τη γραφηματική σειρά ‘-γουν’

final_grapheme_seq	model	ms_code	relativeFreq	absoluteFreq	lemmaCode	exceptionToForms	extraMeaning
*GOYN	22	026	24%	22/44	3a	*N	no
*GOYN	22	050	24%	22/44	5a	*N	no

Κατά συνέπεια, οι πληροφορίες που προκύπτουν από τη βάση είναι ότι υπάρχουν 44 ρηματικοί τύποι οι οποίοι λήγουν σε *-γουν*, εκ των οποίων οι 22 έχουν μορφοσυντακτικό κωδικό ‘026’ και κωδικό λημματοποίησης ‘3a’, ενώ οι υπόλοιποι 22 έχουν μορφοσυντακτικό κωδικό ‘050’ και κωδικό λημματοποίησης ‘5a’. Όλοι αυτοί οι 44 τύποι ανήκουν στο κλιτικό υπόδειγμα 22 και αποτελούν εξαίρεση του συνόλου των ρηματικών τύπων που λήγουν σε *-v*. Επίσης, από τη γραμμή του πίνακα ‘verb\_info’ που αφορά στη γραφηματική σειρά ‘\*N’ προκύπτει ότι οι τύποι που λήγουν σε *-v* ανήκουν στο κλιτικό υπόδειγμα 47, ενώ από γραμμές του πίνακα ‘ms\_codeExplain’ προκύπτει η ερμηνεία των μορφοσυντακτικών κωδικών 026 και 050, η οποία φαίνεται στο σχήμα που ακολουθεί. Σημειώνεται ότι το στοιχείο ‘verbFormCategory’ της δομής HashMap που θα επιστραφεί θα έχει την τιμή ‘4’, διότι ο τύπος αναγνωρίστηκε με χρήση καταληκτικής γραφηματικής σειράς που αρχίζει με αστερίσκο και καλύπτει ένα μέρος των χαρακτήρων του τύπου εισόδου (για περισσότερες πληροφορίες για το στοιχείο ‘verbFormCategory’ βλ. παραπάνω στην παρούσα ενότητα). Έτσι, σχηματικά, η δομή HashMap που επιστρέφεται από τη μέθοδο, εφόσον δεν προκληθεί καμία εξαίρεση κατά την αλληλεπίδραση με τη βάση, έχει ως εξής:



Σχήμα 10: Επιστρεφόμενη δομή HashMap για τον τύπο παράγουν

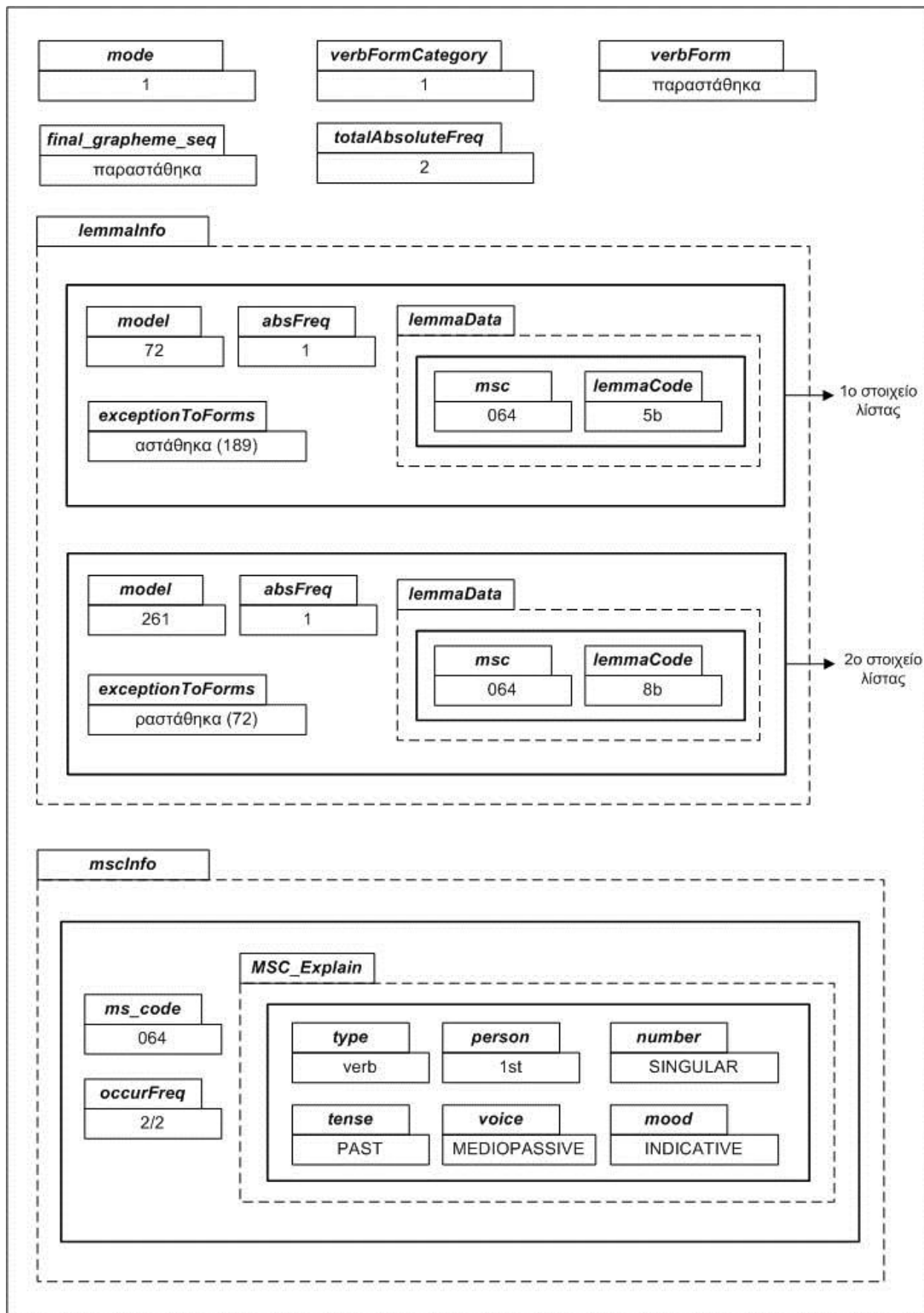
➤ **Παράδειγμα 7.2: Επιστρεφόμενη δομή για τον τύπο παραστάθηκα**

Για να γίνει η αναγνώριση του ρηματικού τύπου *παραστάθηκα* στα Αγγλικά, θα πρέπει να κληθεί η μέθοδος ‘recognGrVerbForm’ με παραμέτρους τις συμβολοσειρές ‘παραστάθηκα’ και ‘en’. Μετά την εκτέλεση του αλγόριθμου της §7.3 προκύπτει ότι ο τύπος *παραστάθηκα* περιέχεται αυτούσιος ως καταληκτική γραφηματική σειρά αποθηκευμένη στη βάση, κατά συνέπεια το στοιχείο ‘verbFormCategory’ της δομής HashMap θα έχει την τιμή ‘1’. Οι γραμμές του πίνακα ‘verb\_info’ οι οποίες θα χρησιμοποιηθούν για την αναγνώριση του *παραστάθηκα* είναι οι εξής:

Πίνακας 9: Γραμμές του πίνακα verb\_info για τις γραφηματικές σειρές ‘-ραστάθηκα’ και ‘παραστάθηκα’

final_grapheme_seq	model	ms_code	relativeFreq	absoluteFreq	lemmaCode	exceptionToForms	extraMeaning
* <i>RASTA~CIKA</i>	72	064	50%	2	5b	* <i>ASTA~CIKA</i>	no
<i>PARASTA~CIK</i> A	261	064	100%	1	8b	* <i>RASTA~CIK</i> A	* <i>RASTA~CIK</i> A

Από τα παραπάνω προκύπτει ότι ο τύπος *παραστάθηκα* παράγεται κατά την κλίση δύο λημμάτων (εφόσον στο πεδίο ‘extraMeaning’ περιέχεται μία καταληκτική γραφηματική σειρά). Κατά συνέπεια η απόλυτη συχνότητα του τύπου ισούται με 2. Ο ένας τύπος *παραστάθηκα* ανήκει στο κλιτικό υπόδειγμα 72, έχει μορφοσυντακτικό κωδικό 064, κωδικό λημματοποίησης 5b και αποτελεί εξαίρεση των τύπων σε ‘-αστάθηκα’ (οι οποίοι ανήκουν στο κλιτικό υπόδειγμα 189 όπως προκύπτει από άλλη γραμμή του πίνακα ‘verb\_info’), ενώ ο άλλος τύπος ανήκει στο κλιτικό υπόδειγμα 261, έχει μορφοσυντακτικό κωδικό 064, κωδικό λημματοποίησης 8b και αποτελεί εξαίρεση των τύπων που λήγουν σε ‘-ραστάθηκα’ (οι οποίοι, όπως φαίνεται και από το σχήμα, ανήκουν στο κλιτικό υπόδειγμα 72). Τελικά, η δομή HashMap που επιστρέφεται, εφόσον δεν παρουσιαστούν προβλήματα κατά την επικοινωνία με τη βάση, περιέχει τις τιμές που φαίνονται στο σχήμα που ακολουθεί:



Σχήμα 11: Επιστρεφόμενη δομή HashMap για τον τύπο 'παραστάθηκα'

➤ **Παράδειγμα 7.3: Επιστρεφόμενη δομή για τον τύπο αναφύμενα**

Η αναγνώριση της μετοχής αναφύμενα διαπιστώνεται ότι γίνεται με χρήση της καταληκτικής γραφηματικής σειράς ‘φύμενα’. Η σχετική γραμμή του πίνακα ‘verb\_info’ είναι η ακόλουθη:

Πίνακας 10: Γραμμή του πίνακα verb\_info για τη γραφηματική σειρά ‘-φύμενα’

final_grapheme_seq	model	ms_code	relativeFreq	absoluteFreq	lemmaCode	exceptionToForms	extraMeaning
?FYO~MENA	313	010	100%	4	5a	*YO~MENA	no

Από τη γραμμή του πίνακα ‘verb\_info’ που αφορά στη γραφηματική σειρά ‘-φύμενα’ προκύπτει ότι η συγκεκριμένη σειρά ανήκει στο κλιτικό υπόδειγμα 205, ενώ σύμφωνα με τα δεδομένα του πίνακα ‘ms\_codeExplain’ ο μορφοσυντακτικός κωδικός ‘010’ αναπαριστά τις παρακάτω πληροφορίες:

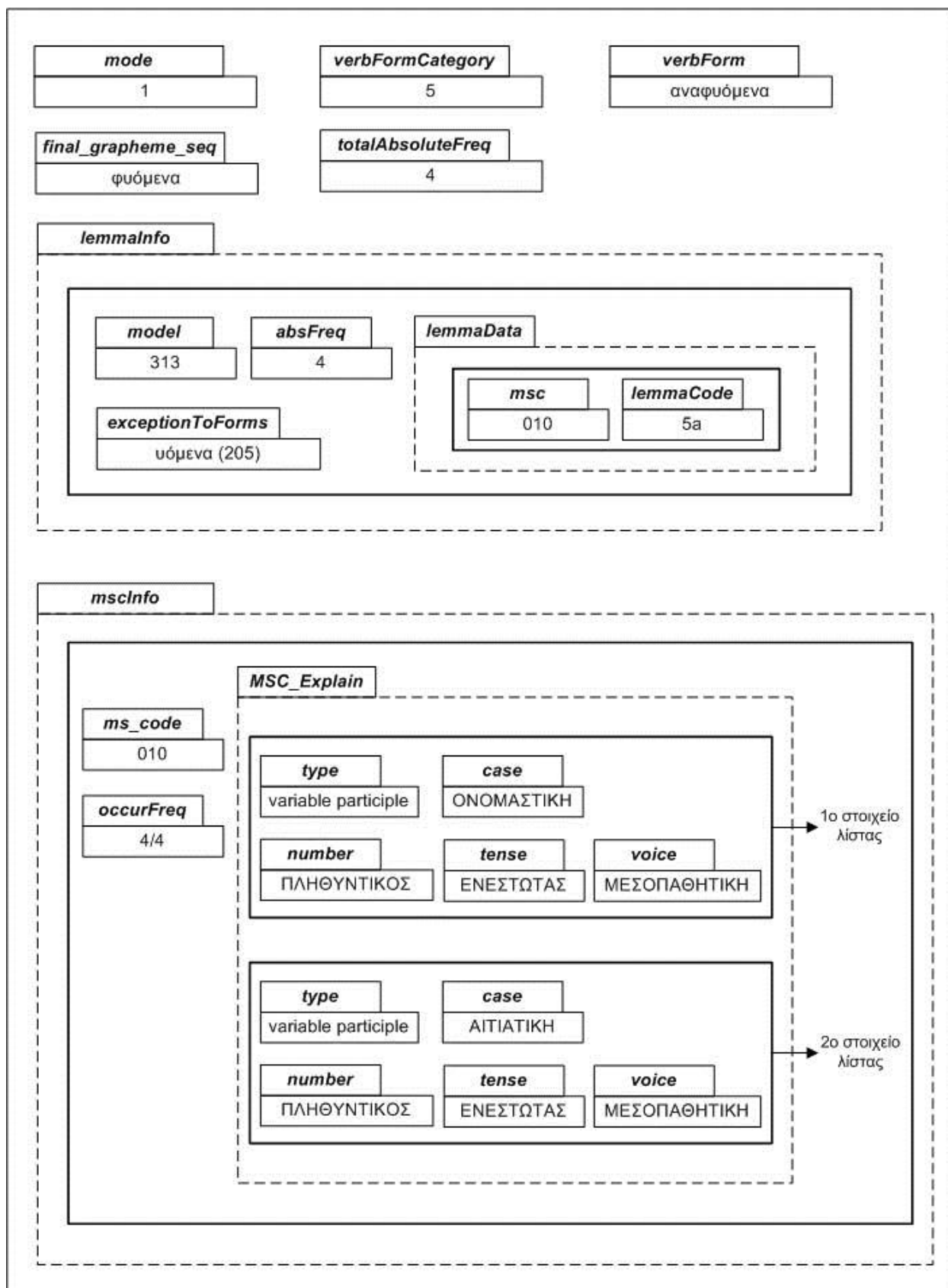
▪ Κλιτή μετοχή

Γένος: Ουδέτερο  
Αριθμός: Πληθυντικός  
Πτώση: Ονομαστική  
Χρόνος: Ενεστώτας  
Φωνή: Μεσοπαθητική

▪ Κλιτή μετοχή

Γένος: Ουδέτερο  
Αριθμός: Πληθυντικός  
Πτώση: Αιτιατική  
Χρόνος: Ενεστώτας  
Φωνή: Μεσοπαθητική

Κατά συνέπεια, η δομή HashMap που θα επιστρέψει η μέθοδος ‘recognGrVerbForm’, όταν κληθεί με παραμέτρους τις συμβολοσειρές ‘αναφύμενα’ και ‘gr’ και σε περίπτωση που δεν υπάρξει πρόβλημα κατά την επικοινωνία με τη βάση, θα περιέχει τις τιμές που φαίνονται στο σχήμα που ακολουθεί:



Σχήμα 12: Επιστρεφόμενη δομή HashMap για τον τύπο 'αναφύομενα'



## 8. ΠΕΡΙΓΡΑΦΗ ΔΙΑΔΙΚΤΥΑΚΗΣ ΕΦΑΡΜΟΓΗΣ

Σκοπός αυτού του κεφαλαίου είναι να περιγραφεί η γενικότερη μορφή της διαδικτυακής εφαρμογής που επιτελεί μορφοσυντακτική αναγνώριση των μονολεκτικών ρηματικών τύπων της Νέας Ελληνικής, να ερμηνευτούν οι πληροφορίες που εμφανίζονται, όταν ένας χρήστης εισάγει προς αναγνώριση κάποιο ρηματικό τύπο και να παρουσιαστούν ενδεικτικά δείγματα της σελίδας που περιέχει αυτές τις πληροφορίες.

Θα πρέπει να σημειωθεί ότι η εφαρμογή είναι διαθέσιμη στην ελληνική και την αγγλική γλώσσα – σε κάθε σελίδα ο χρήστης έχει τη δυνατότητα επιλογής γλώσσας πιέζοντας ένα εικονίδιο με τη σημαία της αντίστοιχης γλώσσας.

Ένα τελευταίο σημείο που θα πρέπει να αναφερθεί είναι ότι για να μπορέσει ένας χρήστης να χρησιμοποιήσει την εφαρμογή – εφόσον αυτή έχει εγκατασταθεί σύμφωνα με τις οδηγίες του κεφαλαίου 9 – θα πρέπει μέσω ενός browser να καλέσει την εφαρμογή ως εξής:

*“http://server\_name:port/verbttaggr”*

όπου server\_name το όνομα του μηχανήματος που έχει εγκατασταθεί ο Tomcat (web container) και port ο αριθμός της θύρας που ο Tomcat δέχεται αιτήσεις.

### 8.1. Πληροφορίες Εφαρμογής για Ρηματικό Τύπο Εισόδου

Όπως θα φανεί και από τις εικόνες της §8.2, όταν ο χρήστης εισάγει μία συμβολοσειρά η οποία θα αναγνωριστεί ως αποδεκτός ρηματικός τύπος, παρουσιάζονται οι εξής πληροφορίες:

#### 1) Μορφοσυντακτικό περιεχόμενο

Πρόκειται για πληροφορίες που αφορούν στη μορφοσυντακτική αξία των ρηματικών τύπων που λήγουν στη γραφηματική σειρά βάσει της οποίας πραγματοποιήθηκε η αναγνώριση (σχετικά με τις γραφηματικές σειρές βλ. κεφάλαιο 4). Συγκεκριμένα, ανάλογα με το ρηματικό τύπο εισόδου, η εφαρμογή εμφανίζει τα ακόλουθα στοιχεία:

- Αν ο τύπος αναγνωρίστηκε ως ρήμα, παρουσιάζονται:
  - Πρόσωπο
  - Αριθμός
  - Έγκλιση
  - Χρόνος
  - Φωνή
  - Συχνότητα εμφάνισης
- Σημειώνεται ότι στην περίπτωση που ο τύπος εισόδου είναι τύπος εξαρτημένος συνοπτικού ποιού ενέργειας και μη παρελθοντικού χρόνου (όπως είναι οι τύποι *γράφω, πληρωθώ, διαβάσει* κ.λπ.), ως χρόνος αναφέρεται η ένδειξη ΣΥΝΟΠΤΙΚΟΣ ΜΗ ΠΑΡΕΛΘΟΝΤΙΚΟΣ, ενώ δεν καταγράφεται πληροφορία για την έγκλιση, εφόσον ο συνοπτικός μη παρελθοντικός τύπος δεν απαντά ως ανεξάρτητος ρηματικός τύπος και ως εκ τούτου δεν μπορεί να εξαχθεί συμπέρασμα για την έγκλιση και το χρόνο του (βλ. κεφάλαιο 3).

- Αν ο τύπος αναγνωρίστηκε ως μετοχή κλιτή, παρουσιάζονται:
  - Γένος
  - Αριθμός
  - Πτώση
  - Χρόνος
  - Φωνή
  - Συχνότητα εμφάνισης
  
- Αν ο τύπος αναγνωρίστηκε ως μετοχή άκλιτη, παρουσιάζονται:
  - Χρόνος
  - Φωνή
  - Συχνότητα εμφάνισης

Όσον αφορά στη συχνότητα εμφάνισης του μορφοσυντακτικού περιεχομένου, αυτή αφορά στο πλήθος των καταγεγραμμένων ρηματικών τύπων που λήγουν στη γραφηματική σειρά βάσει της οποίας έγινε η αναγνώριση και οι οποίοι εκφράζουν το συγκεκριμένο μορφοσυντακτικό περιεχόμενο σε σχέση με το συνολικό πλήθος των τύπων που λήγουν στη γραφηματική σειρά αναγνώρισης. Επιπλέον, θα πρέπει να σημειωθεί ότι σε πολλές περιπτώσεις η μορφοσυντακτική αμφισημία αίρεται μόνο με περαιτέρω συντακτική ανάλυση. Ως εκ τούτου, η εφαρμογή είναι δυνατόν να παραγάγει πλέον του ενός μορφοσυντακτικά περιεχόμενα ως έξοδο για ένα ρηματικό τύπο. Συνεπώς, για κάθε μορφοσυντακτικό περιεχόμενο η συχνότητα εμφάνισης παρατίθεται ως ένδειξη για την πιθανότητα ισχύος του αντίστοιχου περιεχομένου. Σε περίπτωση που δεν υπάρχουν στοιχεία για τη συχνότητα εμφάνισης, όταν π.χ. η αναγνώριση βασίζεται μόνο στον τελικό γραφηματικό χαρακτήρα της λέξης, ο οποίος συνήθως εκφράζει πληθώρα μορφοσυντακτικών περιεχομένων του ίδιου κλιτικού υποδείγματος, στο πεδίο αυτό εμφανίζεται προς το παρόν μία παύλα, εφόσον απαιτείται περαιτέρω στατιστική ανάλυση των τύπων ίδιου κλιτικού υποδείγματος.

## **2) Κλιτικό υπόδειγμα, Λήμμα, Απόλυτη Συχνότητα και Εξαίρεση Τύπων Που Λήγουν Σε**

Για τον εκάστοτε ρηματικό τύπο εισόδου παρέχεται ένας αριθμητικός κωδικός, ο οποίος δηλώνει το κλιτικό υπόδειγμα σύμφωνα με το οποίο κλίνεται το ρήμα. Στις ολιγάριθμες σχετικά περιπτώσεις στις οποίες ο τύπος ανήκει σε περισσότερα του ενός κλιτικά υποδείγματα, καταγράφονται όλοι οι αντίστοιχοι κωδικοί από το νέο κλιτικό σύστημα των 382 υποδειγμάτων (σχετικά με το κλιτικό σύστημα βλ. κεφάλαιο 4). Για κάθε κλιτικό υπόδειγμα έχει προβλεφθεί αντίστοιχο πεδίο εμφάνισης του λήμματος, ενώ επιπλέον καταγράφεται και μία τιμή απόλυτης συχνότητας, η οποία δηλώνει το πλήθος των ρηματικών τύπων που λήγουν στη γραφηματική σειρά αναγνώρισης και ανήκουν στο συγκεκριμένο κλιτικό υπόδειγμα. Αν η αναγνώριση βασίστηκε σε όλους τους γραφηματικούς χαρακτήρες του τύπου εισόδου, η απόλυτη συχνότητα ισούται με ένα (1) και το σχετικό πεδίο φέρει τον τίτλο "Απόλυτη Συχνότητα του Τύπου". Το πεδίο αυτό δηλαδή φέρει τον τίτλο "Απόλυτη Συχνότητα του Τύπου", όταν η γραφηματική σειρά αναγνώρισης, και κατ' επέκταση ο τύπος εισόδου, αποτελεί αυτόνομο υπαρκτό τύπο της γλώσσας. Αντιθέτως, αν η αναγνώριση έγινε βάσει τμήματος μόνο του συνόλου των

χαρακτήρων του τύπου εισόδου, η απόλυτη συχνότητα ενδέχεται να είναι αριθμός μεγαλύτερος της μονάδας. Στις περιπτώσεις αυτές, στο πεδίο απόλυτης συχνότητας, το οποίο φέρει τον τίτλο "Απόλυτη Συχνότητα Τύπων που λήγουν σε -xxx", δηλώνεται το πλήθος των ρηματικών τύπων που λήγουν στη γραφηματική σειρά αναγνώρισης και ανήκουν στο κλιτικό υπόδειγμα που αναγράφεται στην αντίστοιχη γραμμή. Τέλος, στο πεδίο που φέρει τον τίτλο "Εξαίρεση Τύπων Που Λήγουν Σε" καταγράφεται η καταληκτική γραφηματική σειρά εξαίρεση της οποίας αποτελεί η καταληκτική γραφηματική σειρά αναγνώρισης. Είναι υπερσύνολο της καταληκτικής γραφηματικής σειράς βάσει της οποίας πραγματοποιείται η αναγνώριση, εφόσον καλύπτει αυτήν καθώς και άλλες, και διαφοροποιείται από αυτήν ως προς τον κωδικό κλιτικού υποδείγματος και έναν τουλάχιστον γραφηματικό χαρακτήρα λιγότερο στα αριστερά αυτής. Ο διαφορετικός κωδικός κλιτικού υποδείγματος που δηλώνει αυτή η λιγότερων χαρακτήρων καταληκτική γραφηματική σειρά αναγράφεται εντός παρενθέσεων. Στις ολιγάριθμες σχετικά περιπτώσεις στις οποίες ο τύπος ανήκει σε περισσότερα του ενός κλιτικά υποδείγματα, καταγράφονται όλες οι καταληκτικές γραφηματικές σειρές των οποίων αποτελεί εξαίρεση η καταληκτική γραφηματική σειρά αναγνώρισης μαζί με τους αντίστοιχους κωδικούς κλιτικού υποδείγματος που αυτές δηλώνουν.

### **3) Σύνθετοι Καταγραφέντες Τύποι Που Σχηματίζονται Από Τον Απλό Τύπο xxx / 2ο Συνθετικό Σύνθετων Καταγραφέντων Τύπων**

Το πεδίο αυτό ενεργοποιείται και περιέχει πληροφορία, μόνο όταν από τα στατιστικά δεδομένα της διδακτορικής διατριβής [1] προκύπτει ότι η καταληκτική γραφηματική σειρά αναγνώρισης μπορεί να εμφανιστεί είτε ως αυτόνομος απλός τύπος είτε ως 2ο συνθετικό ρηματικών λεξημάτων του ίδιου κλιτικού υποδείγματος με τον απλό. Ειδικότερα, υπάρχουν οι εξής δύο περιπτώσεις:

- Όταν η καταληκτική γραφηματική σειρά αναγνώρισης είναι όμοια με έναν απλό τύπο εισόδου, το πεδίο φέρει τον τίτλο "2ο Συνθετικό Σύνθετων Καταγραφέντων Τύπων" και ως τιμή περιέχει το πλήθος των σύνθετων τύπων που έχουν καταγραφεί να περιέχουν τον απλό τύπο εισόδου ως 2ο συνθετικό ανήκοντας στο ίδιο κλιτικό υπόδειγμα με αυτόν.
- Όταν η καταληκτική γραφηματική σειρά αναγνώρισης αποτελεί 2ο συνθετικό ενός σύνθετου τύπου εισόδου του ίδιου κλιτικού υποδείγματος με τον αντίστοιχο απλό, το πεδίο φέρει τον τίτλο "Σύνθετοι Καταγραφέντες Τύποι Που Σχηματίζονται Από Τον Απλό Τύπο xxx", όπου 'xxx' η καταληκτική γραφηματική σειρά αναγνώρισης. Ως τιμή περιέχει το πλήθος των σύνθετων τύπων που έχουν καταγραφεί να περιέχουν τον απλό τύπο ως 2ο συνθετικό ανήκοντας στο ίδιο κλιτικό υπόδειγμα με αυτόν.


Σε όλες τις υπόλοιπες περιπτώσεις, δηλαδή όταν το κλιτικό υπόδειγμα σύνθετων τύπων είναι διαφορετικό από το κλιτικό υπόδειγμα των αντίστοιχων απλών, το πεδίο αυτό δεν εμφανίζεται, γεγονός το οποίο δεν πρέπει να παρερμηνευθεί ως πλήρης απουσία σύνθετων τύπων σχηματιζόμενων από έναν ενδεχομένως απλό τύπο εισόδου. Η μη ενεργοποίησή του θα πρέπει συνεπώς να ερμηνευθεί ως ένδειξη διαφορετικού κλιτικού υποδείγματος μεταξύ ενός τύπου απλού και των σύνθετων που μπορούν να σχηματιστούν από αυτόν.

## **8.2. Δείγματα Σελίδων της Εφαρμογής**

Στην παρούσα ενότητα παρουσιάζονται μία σειρά από εικόνες που απεικονίζουν ενδεικτικές σελίδες που εμφανίζονται από την εφαρμογή, όταν δοθούν προς αναγνώριση διάφορες συμβολοσειρές. Πάνω από κάθε εικόνα αναγράφεται η συμβολοσειρά που δίνεται κάθε φορά ως είσοδος μαζί με κάποιες απαραίτητες διευκρινίσεις και επεξηγήσεις. Υπενθυμίζεται ότι έχει προβλεφθεί χώρος στον οποίο μελλοντικά θα αναγράφεται το λήμμα του τύπου εισόδου, ενώ προς το παρόν εμφανίζεται μία παύλα ('-'). Τέλος, σημειώνεται ότι στις εικόνες που ακολουθούν τα αποτελέσματα της αναγνώρισης είναι στα Ελληνικά, ωστόσο οι πληροφορίες είναι διαθέσιμες και στα Αγγλικά και εμφανίζονται, όταν ο χρήστης πιέσει το εικονίδιο της αγγλικής σημαίας.

### 1) Συμβολοσειρά εισόδου: 'παραστάθηκα'

Η γραφηματική σειρά βάσει της οποίας γίνεται η αναγνώριση του τύπου *παραστάθηκα* αποτελείται από όλους τους χαρακτήρες της συμβολοσειράς και σύμφωνα με τα δεδομένα της εφαρμογής πρόκειται για τύπο που εμφανίζεται σε δύο κλιτικά υποδείγματα δύο διαφορετικών ρηματικών λημμάτων, οι οποίοι έχουν την ίδια γραμματική αλλά διαφορετική σημασιολογική αξία. Υπάρχουν δηλαδή δύο ομόγραφοι κλιτοί τύποι εκ των οποίων ο ένας ανήκει στο κλιτικό υπόδειγμα 72 και αποτελεί εξαίρεση των τύπων που λήγουν σε *-αστάθηκα* (οι οποίοι ανήκουν στο κλιτικό υπόδειγμα 189), ενώ ο άλλος τύπος ανήκει στο κλιτικό υπόδειγμα 261 και αποτελεί εξαίρεση των τύπων που λήγουν σε *-ραστάθηκα* (οι οποίοι ανήκουν στο κλιτικό υπόδειγμα 72). Όλες αυτές οι πληροφορίες απεικονίζονται από την εφαρμογή ως εξής:





Μορφοσυντακτική Αναγνώριση / Λημματοποίηση των Μονολεκτικών Ρηματικών Τύπων της Νέας Ελληνικής

Αρχική σελίδα

Αναγνώριση ρήματος

Επικοινωνία

Βοήθεια

 Ελληνικά  
 English

**Συμβολοσειρά προς αναγνώριση: "παραστάθηκα"**

ΜΟΡΦΟΣΥΝΤΑΚΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ	
ΡΗΜΑ	
ΠΡΟΣΩΠΟ	1ο
ΑΡΙΘΜΟΣ	ΕΝΙΚΟΣ
ΕΓΚΛΙΣΗ	ΟΡΙΣΤΙΚΗ
ΧΡΟΝΟΣ	ΑΟΡΙΣΤΟΣ
ΦΩΝΗ	ΜΕΣΟΠΑΘΗΤΙΚΗ
ΣΥΧΝΟΤΗΤΑ ΕΜΦΑΝΙΣΗΣ	2 / 2

ΚΛΙΤΙΚΟ ΥΠΟΔΕΙΓΜΑ	ΛΗΜΜΑ	ΑΠΟΛΥΤΗ ΣΥΧΝΟΤΗΤΑ ΤΟΥ ΤΥΠΟΥ	ΕΞΑΙΡΕΣΗ ΤΥΠΩΝ ΠΟΥ ΛΗΓΟΥΝ ΣΕ
72	-	1	-αστάθηκα (189)
261	-	1	-ραστάθηκα (72)

Αναγνώριση Νέου Ρήματος


Εικόνα 1: Αναγνώριση της συμβολοσειράς ‘παραστάθηκα’

## 2) Συμβολοσειρά εισόδου: ‘πακετάρουν’

Ο τύπος *πακετάρουν* αναγνωρίζεται βάσει της καταληκτικής γραφηματικής σειράς ‘-ακετάρουν’, η οποία αναγράφεται στον τίτλο του πεδίου της απόλυτης συχνότητας. Όπως φαίνεται από την εικόνα που ακολουθεί, έχει καταγραφεί ένας τύπος ο οποίος λήγει στους συγκεκριμένους χαρακτήρες και αυτός ο τύπος έχει διπλή μορφοσυντακτική αξία: ενδέχεται να είναι ρήμα 3<sup>ου</sup> πληθυντικού προσώπου, έγκλισης οριστικής, χρόνου ενεστώτα, φωνής ενεργητικής ή ρήμα 3<sup>ου</sup> πληθυντικού προσώπου, χρόνου συνοπτικού μη παρελθοντικού και φωνής ενεργητικής. Σε τέτοιες περιπτώσεις η εφαρμογή εμφανίζει όλες τις εκδοχές με τον τρόπο που παρουσιάζεται στην εικόνα 2.

Σημειώνεται ότι η μορφοσυντακτική αμφισημία που αφορά τον τύπο μπορεί να αρθεί μόνο με περαιτέρω συντακτική ανάλυση, όπως φαίνεται και από τα παραδείγματα που ακολουθούν:

Στην πρόταση 'Αυτή τη στιγμή πακετάρουν τα πράγματα' η μορφοσυντακτική αξία του τύπου 'πακετάρουν' εκφράζεται από την πρώτη εκδοχή (χρόνος ενεστώτας), ενώ στην πρόταση 'Τους συνέστησε να πακετάρουν προσεκτικά τα πράγματα' η μορφοσυντακτική αξία του 'πακετάρουν' εκφράζεται από τη δεύτερη εκδοχή (χρόνος συνοπτικός μη παρελθοντικός).


Μορφοσυντακτική Αναγνώριση / Δημματοποίηση των Μονολεκτικών Ρηματικών Τύπων της Νέας Ελληνικής



---

Αρχική σελίδα

Αναγνώριση ρήματος

Επικοινωνία

Βοήθεια

 Ελληνικά  
 English

Συμβολοσειρά προς αναγνώριση: "πακετάρουν"

ΜΟΡΦΟΣΥΝΤΑΚΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ		
ΡΗΜΑ		
ΠΡΟΣΩΠΟ	3ο	3ο
ΑΡΙΘΜΟΣ	ΠΛΗΘΥΝΤΙΚΟΣ	ΠΛΗΘΥΝΤΙΚΟΣ
ΕΓΚΛΙΣΗ	ΟΡΙΣΤΙΚΗ	-
ΧΡΟΝΟΣ	ΕΝΕΣΤΩΤΑΣ	ΣΥΝΟΠΤΙΚΟΣ ΜΗ ΠΑΡΕΛΘΟΝΤΙΚΟΣ
ΦΩΝΗ	ΕΝΕΡΓΗΤΙΚΗ	ΕΝΕΡΓΗΤΙΚΗ
ΣΥΧΝΟΤΗΤΑ ΕΜΦΑΝΙΣΗΣ	1 / 1	

ΚΛΙΤΙΚΟ ΥΠΟΔΕΙΓΜΑ	ΛΗΜΜΑ	ΑΠΟΛΥΤΗ ΣΥΧΝΟΤΗΤΑ ΤΥΠΩΝ ΠΟΥ ΛΗΓΟΥΝ ΣΕ -ακετάρουν	ΕΞΑΙΡΕΣΗ ΤΥΠΩΝ ΠΟΥ ΛΗΓΟΥΝ ΣΕ
171	-	1	-ουν (169)


  

Αναγνώριση Νέου Ρήματος

Εικόνα 2: Αναγνώριση της συμβολοσειράς 'πακετάρουν'


### 3) Συμβολοσειρά εισόδου: 'δέχθηκαν'


Ο ρηματικός τύπος *δέχθηκαν* αναγνωρίζεται βάσει της γραφηματικής σειράς '– δέχθηκαν', η οποία μπορεί να εμφανιστεί είτε ως σύνθετος είτε ως αυτόνομος απλός τύπος. Έτσι, εκτός των υπόλοιπων πληροφοριών, εμφανίζεται ένα πεδίο με τίτλο «2ο ΣΥΝΘΕΤΙΚΟ ΣΥΝΘΕΤΩΝ ΚΑΤΑΓΡΑΦΕΝΤΩΝ ΤΥΠΩΝ» και τιμή 6 με την οποία δηλώνεται ότι ο τύπος *δέχθηκαν* αποτελεί το δεύτερο συνθετικό 6 σύνθετων καταγραφέντων τύπων.





Μορφοσυντακτική Αναγνώριση / Λημματοποίηση των  
Μονολεκτικών Ρηματικών Τύπων της Νέας Ελληνικής


---


  
Αρχική σελίδα

  
Αναγνώριση ρήματος

  
Επικοινωνία

  
Βοήθεια

  
Ελληνικά

  
English


---

**Συμβολοσειρά προς αναγνώριση: "δέχθηκαν"**

ΜΟΡΦΟΣΥΝΤΑΚΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ	
ΡΗΜΑ	
ΠΡΟΣΩΠΟ	3ο
ΑΡΙΘΜΟΣ	ΠΛΗΘΥΝΤΙΚΟΣ
ΕΓΚΛΙΣΗ	ΟΡΙΣΤΙΚΗ
ΧΡΟΝΟΣ	ΑΟΡΙΣΤΟΣ
ΦΩΝΗ	ΜΕΣΟΠΑΘΗΤΙΚΗ
ΣΥΧΝΟΤΗΤΑ ΕΜΦΑΝΙΣΗΣ	7 / 7

ΚΛΙΤΙΚΟ ΥΠΟΔΕΙΓΜΑ	ΛΗΜΜΑ	ΑΠΟΛΥΤΗ ΣΥΧΝΟΤΗΤΑ ΤΥΠΩΝ ΠΟΥ ΛΗΓΟΥΝ ΣΕ -δέχθηκαν	ΕΞΑΙΡΕΣΗ ΤΥΠΩΝ ΠΟΥ ΛΗΓΟΥΝ ΣΕ
340	-	7	-έχθηκαν (27)


2ο ΣΥΝΘΕΤΙΚΟ ΣΥΝΘΕΤΩΝ ΚΑΤΑΓΡΑΦΕΝΤΩΝ ΤΥΠΩΝ
6


Αναγνώριση Νέου Ρήματος


Εικόνα 3: Αναγνώριση της συμβολοσειρά 'δέχθηκαν'


#### 4) Συμβολοσειρά εισόδου: 'αποδέχθηκαν'


Ο τύπος *αποδέχθηκαν* αναγνωρίζεται επίσης βάσει της καταληκτικής γραφηματικής σειράς '-δέχθηκαν', όπως συνέβη και με τον απλό τύπο *δέχθηκαν*. Τώρα όμως ως είσοδος δόθηκε ένας από τους σύνθετους τύπους, οπότε εκτός των άλλων πληροφοριών εμφανίζεται ένα πεδίο με τίτλο «ΣΥΝΘΕΤΟΙ ΚΑΤΑΓΡΑΦΕΝΤΕΣ ΤΥΠΟΙ ΠΟΥ ΣΧΗΜΑΤΙΖΟΝΤΑΙ ΑΠΟ ΤΟΝ ΑΠΛΟ ΤΥΠΟ *δέχθηκαν*» και τιμή 6 με την οποία δηλώνεται ότι υπάρχουν 6 σύνθετοι τύποι με δεύτερο συνθετικό τον τύπο *δέχθηκαν*.






Μορφοσυντακτική Αναγνώριση / Λημματοποίηση των Μονολεκτικών Ρηματικών Τύπων της Νέας Ελληνικής

  
Αρχική σελίδα

  
Αναγνώριση ρήματος

  
Επικοινωνία

  
Βοήθεια

 Ελληνικά  
 English

**Συμβολοσειρά προς αναγνώριση: "αποδέχθηκαν"**

ΜΟΡΦΟΣΥΝΤΑΚΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ	
ΡΗΜΑ	
ΠΡΟΣΩΠΟ	3ο
ΑΡΙΘΜΟΣ	ΠΛΗΘΥΝΤΙΚΟΣ
ΕΓΚΛΙΣΗ	ΟΡΙΣΤΙΚΗ
ΧΡΟΝΟΣ	ΑΟΡΙΣΤΟΣ
ΦΩΝΗ	ΜΕΣΟΠΑΘΗΤΙΚΗ
ΣΥΧΝΟΤΗΤΑ ΕΜΦΑΝΙΣΗΣ	7 / 7


ΚΛΙΤΙΚΟ ΥΠΟΔΕΙΓΜΑ	ΛΗΜΜΑ	ΑΠΟΛΥΤΗ ΣΥΧΝΟΤΗΤΑ ΤΥΠΩΝ ΠΟΥ ΛΗΓΟΥΝ ΣΕ -δέχθηκαν	ΕΞΑΙΡΕΣΗ ΤΥΠΩΝ ΠΟΥ ΛΗΓΟΥΝ ΣΕ
340	-	7	-έχθηκαν (27)

ΣΥΝΘΕΤΟΙ ΚΑΤΑΓΡΑΦΕΝΤΕΣ ΤΥΠΟΙ ΠΟΥ ΣΧΗΜΑΤΙΖΟΝΤΑΙ ΑΠΟ ΤΟΝ ΑΠΛΟ ΤΥΠΟ δέχθηκαν

6

 Αναγνώριση Νέου Ρήματος

Εικόνα 4: Αναγνώριση της συμβολοσειράς ‘αποδέχθηκαν’

### 5) Συμβολοσειρά εισόδου: ‘έφυγα’


Η καταληκτική γραφηματική σειρά ‘-υγα’ χρησιμοποιείται για την αναγνώριση του τύπου *έφυγα*. Όμως, και ο τύπος *έφευγα* αναγνωρίζεται βάσει της ίδιας καταληκτικής γραφηματικής σειράς, διότι και οι δύο τύποι ανήκουν στο ίδιο κλιτικό υπόδειγμα (30) και λήγουν στους ίδιους χαρακτήρες. Όπως αναφέρθηκε στην §7.3, μία αδυναμία της εφαρμογής προκύπτει από το γεγονός ότι το σύνολο των καταληκτικών γραφηματικών σειρών που παράχθηκαν από την έρευνα της Λεμπέση έχουν ως κριτήριο διαχωρισμού το διαφορετικό κλιτικό υπόδειγμα των ρηματικών τύπων που εξετάστηκαν. Το γεγονός αυτό έχει ως αποτέλεσμα την ύπαρξη ρηματικών τύπων οι οποίοι, ανήκουν μεν στο ίδιο κλιτικό υπόδειγμα και αναγνωρίζονται βάσει της ίδιας καταληκτικής γραφηματικής σειράς, αλλά έχουν διαφορετικό μορφοσυντακτικό περιεχόμενο, όπως συμβαίνει με τους τύπους *έφυγα* και *έφευγα*. Σε τέτοιες περιπτώσεις ως πιθανό μορφοσυντακτικό περιεχόμενο δίνονται οι πληροφορίες που αφορούν σε όλους τους τύπους που εκφράζονται από τη συγκεκριμένη καταληκτική γραφηματική σειρά, οπότε προφανώς κάποιες ερμηνείες είναι

Ζέρβα Ι. – Κωλέτση Β. – Μωραγιάννη Ν.


72





λανθασμένες. Έτσι, η εφαρμογή εμφανίζει ότι ο τύπος 'έφυγα' είναι εξίσου πιθανό να είναι ρήμα χρόνου παρατατικού ή αορίστου, ενώ σωστό είναι μόνο το δεύτερο. Συγκεκριμένα, σύμφωνα με τις πληροφορίες της παρακάτω εικόνας, η μορφοσυντακτική αξία των 19 από τους 38 τύπους που λήγουν σε *-υγα* είναι αυτή που καταγράφεται στον πρώτο πίνακα μορφοσυντακτικού περιεχομένου, ενώ η αξία των υπόλοιπων 19 τύπων που λήγουν σε *-υγα* καταγράφεται στο δεύτερο πίνακα. Όμοια για τον τύπο *έφευγα* η εφαρμογή εμφανίζει το ίδιο μορφοσυντακτικό περιεχόμενο, ενώ η σωστή πληροφορία για το χρόνο είναι ο παρατατικός. Ευτυχώς η αμφισημία αυτού του είδους είναι ιδιαίτερα περιορισμένη, καθώς χαρακτηρίζει μόνο 11 ρήματα ενεργητικής φωνής που ακολουθούν το κλιτικό υπόδειγμα 30 για την παραγωγή των κλιτών τύπων τους, όλα σύνθετα του απλού ρήματος *φεύγω* καθώς και το 1 και μοναδικό σύνθετο ρήμα *αποφεύγω*, το οποίο εμφανίζει και μεσοπαθητικούς τύπους και ως εκ τούτου κλίνεται σύμφωνα με το κλιτικό υπόδειγμα 31. Ο χρήστης όμως οφείλει να είναι ενήμερος για τον συγκεκριμένο περιορισμό της προσέγγισης.






Μορφοσυντακτική Αναγνώριση / Λημματοποίηση των  
Μονολεκτικών Ρηματικών Τύπων της Νέας Ελληνικής

  
Αρχική σελίδα

  
Αναγνώριση ρήματος

  
Επικοινωνία

  
Βοήθεια

 Ελληνικά  
 English


**Συμβολοσειρά προς αναγνώριση: "έφυγα"**

ΜΟΡΦΟΣΥΝΤΑΚΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ	
ΡΗΜΑ	
ΠΡΟΣΩΠΟ	1ο
ΑΡΙΘΜΟΣ	ΕΝΙΚΟΣ
ΕΓΚΛΙΣΗ	ΟΡΙΣΤΙΚΗ
ΧΡΟΝΟΣ	ΠΑΡΑΤΑΤΙΚΟΣ
ΦΩΝΗ	ΕΝΕΡΓΗΤΙΚΗ
ΣΥΧΝΟΤΗΤΑ ΕΜΦΑΝΙΣΗΣ	19/38

**ή**

ΜΟΡΦΟΣΥΝΤΑΚΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ	
ΡΗΜΑ	
ΠΡΟΣΩΠΟ	1ο
ΑΡΙΘΜΟΣ	ΕΝΙΚΟΣ
ΕΓΚΛΙΣΗ	ΟΡΙΣΤΙΚΗ
ΧΡΟΝΟΣ	ΑΟΡΙΣΤΟΣ
ΦΩΝΗ	ΕΝΕΡΓΗΤΙΚΗ
ΣΥΧΝΟΤΗΤΑ ΕΜΦΑΝΙΣΗΣ	19/38

ΚΛΙΤΙΚΟ ΥΠΟΔΕΙΓΜΑ	ΛΗΜΜΑ	ΑΠΟΛΥΤΗ ΣΥΧΝΟΤΗΤΑ ΤΥΠΩΝ ΠΟΥ ΛΗΓΟΥΝ ΣΕ -υγα	ΕΞΑΙΡΕΣΗ ΤΥΠΩΝ ΠΟΥ ΛΗΓΟΥΝ ΣΕ
30	-	38	-γα (12)

 Αναγνώριση Νέου Ρήματος

Εικόνα 5: Αναγνώριση συμβολοσειράς 'έφυγα'

### 6) Συμβολοσειρά εισόδου: 'καλούμενον'

Ο τύπος *καλούμενον* είναι μία κλιτή μετοχή με τριπλή μορφοσυντακτική αξία, εφόσον το γένος και η πτώση της μετοχής προσδιορίζονται με βεβαιότητα μόνο μετά από συντακτική ανάλυση. Οι μορφοσυντακτικές πληροφορίες που γενικά αναγράφονται για τις κλιτές μετοχές φαίνονται στην παρακάτω εικόνα:



Μορφοσυντακτική Αναγνώριση / Λημματοποίηση των Μονολεκτικών Ρηματικών Τύπων της Νέας Ελληνικής

Αρχική σελίδα

Αναγνώριση ρήματος

Επικοινωνία

Βοήθεια

  
Ελληνικά
   
English

**Συμβολοσειρά προς αναγνώριση: "καλούμενον"**

ΜΟΡΦΟΣΥΝΤΑΚΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ			
ΜΕΤΟΧΗ ΚΛΙΤΗ			
ΓΕΝΟΣ	ΑΡΣΕΝΙΚΟ	ΟΥΔΕΤΕΡΟ	ΟΥΔΕΤΕΡΟ
ΑΡΙΘΜΟΣ	ΕΝΙΚΟΣ	ΕΝΙΚΟΣ	ΕΝΙΚΟΣ
ΠΤΩΣΗ	ΑΠΠΙΑΤΙΚΗ	ΟΝΟΜΑΣΤΙΚΗ	ΑΠΠΙΑΤΙΚΗ
ΧΡΟΝΟΣ	ΕΝΕΣΤΩΤΑΣ	ΕΝΕΣΤΩΤΑΣ	ΕΝΕΣΤΩΤΑΣ
ΦΩΝΗ	ΜΕΣΟΠΑΘΗΤΙΚΗ	ΜΕΣΟΠΑΘΗΤΙΚΗ	ΜΕΣΟΠΑΘΗΤΙΚΗ
ΣΥΧΝΟΤΗΤΑ ΕΜΦΑΝΙΣΗΣ	1 / 1		


ΚΛΙΤΙΚΟ ΥΠΟΔΕΙΓΜΑ	ΛΗΜΜΑ	ΑΠΟΛΥΤΗ ΣΥΧΝΟΤΗΤΑ ΤΟΥ ΤΥΠΟΥ	ΕΞΑΙΡΕΣΗ ΤΥΠΩΝ ΠΟΥ ΛΗΓΟΥΝ ΣΕ
80	-	1	-αλούμενον (81)

Αναγνώριση Νέου Ρήματος

Εικόνα 6: Αναγνώριση συμβολοσειράς 'καλούμενον'


### 7) Συμβολοσειρά εισόδου: 'ρέμαγα'


Όταν δοθεί προς αναγνώριση μία συμβολοσειρά η οποία αποτελεί μία από τις καταληκτικές γραφηματικές σειρές, χωρίς ωστόσο να απαντά ως αυτοτελής ρηματικός τύπος, καταγράφεται το γεγονός αυτό και από κάτω παρουσιάζονται οι πληροφορίες που αφορούν στη συγκεκριμένη καταληκτική γραφηματική σειρά. Μία τέτοια περίπτωση αποτελεί η συμβολοσειρά 'ρέμαγα' για την οποία η εφαρμογή εμφανίζει τη σελίδα της εικόνας που ακολουθεί:





Μορφοσυντακτική Αναγνώριση / Δημματοποίηση των Μονολεκτικών Ρηματικών Τύπων της Νέας Ελληνικής


---


  
Αρχική σελίδα

  
Αναγνώριση ρήματος

  
Επικοινωνία

  
Βοήθεια

  
Ελληνικά


  
English

**Συμβολοσειρά προς αναγνώριση: "ρέμαγα"**

- ◆ ΜΗ ΥΠΑΡΚΤΟΣ ΡΗΜΑΤΙΚΟΣ ΤΥΠΟΣ
- ◆ ΑΝΑΓΝΩΡΙΣΘΕΙΣΑ ΚΑΤΑΛΗΚΤΙΚΗ ΓΡΑΦΗΜΑΤΙΚΗ ΣΕΙΡΑ

ΜΟΡΦΟΣΥΝΤΑΚΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ	
ΡΗΜΑ	
ΠΡΟΣΩΠΟ	1ο
ΑΡΙΘΜΟΣ	ΕΝΙΚΟΣ
ΕΓΚΛΙΣΗ	ΟΡΙΣΤΙΚΗ
ΧΡΟΝΟΣ	ΠΑΡΑΤΑΤΙΚΟΣ
ΦΩΝΗ	ΕΝΕΡΓΗΤΙΚΗ
ΣΥΧΝΟΤΗΤΑ ΕΜΦΑΝΙΣΗΣ	2 / 2

ΚΛΙΤΙΚΟ ΥΠΟΔΕΙΓΜΑ	ΛΗΜΜΑ	ΑΠΟΛΥΤΗ ΣΥΧΝΟΤΗΤΑ ΤΥΠΩΝ ΠΟΥ ΛΗΓΟΥΝ ΣΕ -ρέμαγα	ΕΞΑΙΡΕΣΗ ΤΥΠΩΝ ΠΟΥ ΛΗΓΟΥΝ ΣΕ
84	-	2	-έμαγα (15)

 Αναγνώριση Νέου Ρήματος

Εικόνα 7: Αναγνώριση συμβολοσειράς 'ρέμαγα'

### 8) Συμβολοσειρά εισόδου: αβγδ

Η παρακάτω σελίδα εμφανίζεται σε περίπτωση που η συμβολοσειρά εισόδου δεν αναγνωριστεί ως αποδεκτός ρηματικός τύπος.

VERBTAGGR Μορφοσυντακτική Αναγνώριση / Δηματοποίηση των Μονολεκτικών Ρηματικών Τύπων της Νέας Ελληνικής

Αρχική σελίδα Αναγνώριση ρήματος Επικοινωνία Βοήθεια

Ελληνικά English

Συμβολοσειρά προς αναγνώριση: "αβγδ"


◆ ΜΗ ΑΝΑΓΝΩΡΙΣΘΕΙΣΑ ΣΥΜΒΟΛΟΣΕΙΡΑ  
◆ ΜΗ ΥΠΑΡΚΤΟΣ ΡΗΜΑΤΙΚΟΣ ΤΥΠΟΣ

Αναγνώριση Νέου Ρήματος

Εικόνα 8: Αναγνώριση συμβολοσειράς 'αβγδ'


### 9) Συμβολοσειρά εισόδου: 'προελέχθη'

Η γραφηματική σειρά βάσει της οποίας γίνεται η αναγνώριση του τύπου *προελέχθη* αποτελείται από όλους τους χαρακτήρες της συμβολοσειράς και σύμφωνα με τα δεδομένα της εφαρμογής πρόκειται για τύπο που εμφανίζεται σε ένα κλιτικό υπόδειγμα, δηλαδή στο κλιτικό υπόδειγμα 26. Αποτελεί εξαίρεση των τύπων που λήγουν σε *-θη* (οι οποίοι ανήκουν στο κλιτικό υπόδειγμα 22). Όλες αυτές οι πληροφορίες απεικονίζονται από την εφαρμογή ως εξής:



Μορφοσυντακτική Αναγνώριση / Λημματοποίηση των Μονολεκτικών Ρηματικών Τύπων της Νέας Ελληνικής

---

  
 Αρχική σελίδα

  
 Αναγνώριση ρήματος

  
 Επικοινωνία

  
 Βοήθεια


  
 Ελληνικά English

Συμβολοσειρά προς αναγνώριση: "προελέχθη"

ΜΟΡΦΟΣΥΝΤΑΚΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ	
ΡΗΜΑ	
ΠΡΟΣΩΠΟ	3ο
ΑΡΙΘΜΟΣ	ΕΝΙΚΟΣ
ΕΓΚΛΙΣΗ	ΟΡΙΣΤΙΚΗ
ΧΡΟΝΟΣ	ΑΟΡΙΣΤΟΣ
ΦΩΝΗ	ΜΕΣΟΠΑΘΗΤΙΚΗ
ΣΥΧΝΟΤΗΤΑ ΕΜΦΑΝΙΣΗΣ	1 / 1

ΚΛΙΤΙΚΟ ΥΠΟΔΕΙΓΜΑ	ΛΗΜΜΑ	ΑΠΟΛΥΤΗ ΣΥΧΝΟΤΗΤΑ ΤΟΥ ΤΥΠΟΥ	ΕΞΑΙΡΕΣΗ ΤΥΠΩΝ ΠΟΥ ΛΗΓΟΥΝ ΣΕ
26	-	1	-θη (22)

 Αναγνώριση Νέου Ρήματος

Εικόνα 9: Αναγνώριση συμβολοσειράς ‘προελέχθη’

### 10) Συμβολοσειρά εισόδου: ‘κάνε’

Η γραφηματική σειρά βάσει της οποίας γίνεται η αναγνώριση του τύπου *κάνε* αποτελείται από όλους τους χαρακτήρες της συμβολοσειράς και σύμφωνα με τα δεδομένα της εφαρμογής πρόκειται για τύπο που εμφανίζεται σε ένα κλιτικό υπόδειγμα, δηλαδή στο κλιτικό υπόδειγμα 111. Αποτελεί εξαίρεση των τύπων που λήγουν σε *-κάνε* (οι οποίοι ανήκουν στο κλιτικό υπόδειγμα 15). Όλες αυτές οι πληροφορίες απεικονίζονται από την εφαρμογή ως εξής:

*VERBTAGGR++* Μορφοσυντακτική Αναγνώριση / Λημματοποίηση των Μονολεκτικών Ρηματικών Τύπων της Νέας Ελληνικής


 Αρχική σελίδα	 Αναγνώριση ρήματος	 Επικοινωνία	 Βοήθεια
---	--	---	---



**Συμβολοσειρά προς αναγνώριση: "κάνε"**

ΜΟΡΦΟΣΥΝΤΑΚΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ		
ΡΗΜΑ		
ΠΡΟΣΩΠΟ	2ο	2ο
ΑΡΙΘΜΟΣ	ΕΝΙΚΟΣ	ΕΝΙΚΟΣ
ΕΓΚΛΙΣΗ	ΠΡΟΣΤΑΚΤΙΚΗ	ΠΡΟΣΤΑΚΤΙΚΗ
ΧΡΟΝΟΣ	ΕΝΕΣΤΩΤΑΣ	ΑΟΡΙΣΤΟΣ
ΦΩΝΗ	ΕΝΕΡΓΗΤΙΚΗ	ΕΝΕΡΓΗΤΙΚΗ
ΣΥΧΝΟΤΗΤΑ ΕΜΦΑΝΙΣΗΣ	1 / 1	

ΚΛΙΤΙΚΟ ΥΠΟΔΕΙΓΜΑ	ΛΗΜΜΑ	ΑΠΟΛΥΤΗ ΣΥΧΝΟΤΗΤΑ ΤΟΥ ΤΥΠΟΥ	ΕΞΑΙΡΕΣΗ ΤΥΠΩΝ ΠΟΥ ΛΗΓΟΥΝ ΣΕ
111	-	1	-κάνε (15)

 Αναγνώριση Νέου Ρήματος

Εικόνα 10: Αναγνώριση συμβολοσειράς 'κάνε'

## 11) Συμβολοσειρά εισόδου: 'εκλάβαμε'

Η γραφηματική σειρά βάσει της οποίας γίνεται η αναγνώριση του τύπου *εκλάβαμε* αποτελείται από όλους τους χαρακτήρες της συμβολοσειράς και σύμφωνα με τα δεδομένα της εφαρμογής πρόκειται για τύπο που εμφανίζεται σε ένα κλιτικό υποδείγμα, δηλαδή στο κλιτικό υπόδειγμα 107. Αποτελεί εξαίρεση των τύπων που λήγουν σε *-βαμε* (οι οποίοι ανήκουν στο κλιτικό υπόδειγμα 4). Όλες αυτές οι πληροφορίες απεικονίζονται από την εφαρμογή ως εξής:


Μορφοσυντακτική Αναγνώριση / Ληματοποίηση των Μονολεκτικών Ρηματικών Τύπων της Νέας Ελληνικής

 Αρχική σελίδα

 Αναγνώριση ρήματος

 Επικοινωνία

 Βοήθεια



  
 Ελληνικά English

**Συμβολοσειρά προς αναγνώριση: "εκλάβαμε"**

ΜΟΡΦΟΣΥΝΤΑΚΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ	
ΡΗΜΑ	
ΠΡΟΣΩΠΟ	1ο
ΑΡΙΘΜΟΣ	ΠΛΗΘΥΝΤΙΚΟΣ
ΕΓΚΛΙΣΗ	ΟΡΙΣΤΙΚΗ
ΧΡΟΝΟΣ	ΑΟΡΙΣΤΟΣ
ΦΩΝΗ	ΕΝΕΡΓΗΤΙΚΗ
ΣΥΧΝΟΤΗΤΑ ΕΜΦΑΝΙΣΗΣ	15 / 15

ΚΛΙΤΙΚΟ ΥΠΟΔΕΙΓΜΑ	ΛΗΜΜΑ	ΑΠΟΛΥΤΗ ΣΥΧΝΟΤΗΤΑ ΤΥΠΩΝ ΠΟΥ ΛΗΓΟΥΝ ΣΕ -άβαμε	ΕΞΑΙΡΕΣΗ ΤΥΠΩΝ ΠΟΥ ΛΗΓΟΥΝ ΣΕ
107	-	15	-βαμε (4)


 Αναγνώριση Νέου Ρήματος

Εικόνα 11: Αναγνώριση συμβολοσειράς 'εκλάβαμε'




## 12) Συμβολοσειρά εισόδου: 'κάψω'


Ο ρηματικός τύπος *κάψω* αναγνωρίζεται βάσει της γραφηματικής σειράς '-κάψω', η οποία μπορεί να εμφανιστεί είτε ως σύνθετος είτε ως αυτόνομος απλός τύπος. Έτσι, εκτός των υπόλοιπων πληροφοριών, εμφανίζεται ένα πεδίο με τίτλο «2ο ΣΥΝΘΕΤΙΚΟ ΣΥΝΘΕΤΩΝ ΚΑΤΑΓΡΑΦΕΝΤΩΝ ΤΥΠΩΝ» και τιμή 6 με την οποία δηλώνεται ότι ο τύπος *κάψω* αποτελεί το δεύτερο συνθετικό 6 σύνθετων καταγραφέντων τύπων.





Μορφοσυντακτική Αναγνώριση / Λημματοποίηση των Μονολεκτικών Ρηματικών Τύπων της Νέας Ελληνικής




Ελληνικό English

  
Αρχική σελίδα

  
Αναγνώριση ρήματος

  
Επικοινωνία

  
Βοήθεια

**Συμβολοσειρά προς αναγνώριση: "κάψω"**

ΜΟΡΦΟΣΥΝΤΑΚΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ	
ΡΗΜΑ	
ΠΡΟΣΩΠΟ	1ο
ΑΡΙΘΜΟΣ	ΕΝΙΚΟΣ
ΕΓΚΛΙΣΗ	-
ΧΡΟΝΟΣ	ΣΥΝΟΠΤΙΚΟΣ ΜΗ ΠΑΡΕΛΘΟΝΤΙΚΟΣ
ΦΩΝΗ	ΕΝΕΡΓΗΤΙΚΗ
ΣΥΧΝΟΤΗΤΑ ΕΜΦΑΝΙΣΗΣ	7 / 7


  

ΚΛΙΤΙΚΟ ΥΠΟΔΕΙΓΜΑ	ΛΗΜΜΑ	ΑΠΟΛΥΤΗ ΣΥΧΝΟΤΗΤΑ ΤΥΠΩΝ ΠΟΥ ΛΗΓΟΥΝ ΣΕ -κάψω	ΕΞΑΙΡΕΣΗ ΤΥΠΩΝ ΠΟΥ ΛΗΓΟΥΝ ΣΕ
63	-	7	-άψω (218)

**2ο ΣΥΝΘΕΤΙΚΟ ΣΥΝΘΕΤΩΝ ΚΑΤΑΓΡΑΦΕΝΤΩΝ ΤΥΠΩΝ**

6

 Αναγνώριση Νέου Ρήματος

Εικόνα 12: Αναγνώριση συμβολοσειράς 'κάψω'

### 13) Συμβολοσειρά εισόδου: 'ράψε'

Η γραφηματική σειρά βάσει της οποίας γίνεται η αναγνώριση του τύπου *ράψε* αποτελείται από όλους τους χαρακτήρες της συμβολοσειράς και σύμφωνα με τα δεδομένα της εφαρμογής πρόκειται για τύπο που εμφανίζεται σε ένα κλιτικό υπόδειγμα, δηλαδή στο κλιτικό υπόδειγμα 4. Αποτελεί εξαίρεση των τύπων που λήγουν σε *-άψε* (οι οποίοι ανήκουν στο κλιτικό υπόδειγμα 242). Όλες αυτές οι πληροφορίες απεικονίζονται από την εφαρμογή στην αγγλική γλώσσα, όπως έχει επιλεγθεί, ως εξής:

The screenshot shows the VERBTAGGR++ web application interface. At the top, the title reads "VERBTAGGR++ Morphosyntactic Recognition / Lemmatization of One-word Modern Greek Verb Forms". Below the title are navigation buttons: Home, Recognize Verb, Contact, and Help. To the right are the Greek and English flags with the text "Ελληνικά English".

The main content area displays the string to recognize: "ράψε". Below this is a table titled "MORPHOSYNTACTIC CONTENT" with the following data:

MORPHOSYNTACTIC CONTENT	
VERB	
PERSON	2nd
NUMBER	SINGULAR
MOOD	IMPERATIVE
TENSE	PAST
VOICE	ACTIVE
OCCURRENCE FREQUENCY	1 / 1

Below the morphosyntactic content table is another table with the following data:

CONJUGATION MODEL	LEMMA	ABSOLUTE FREQUENCY OF THE FORM	EXCEPTION TO FORMS ENDING IN
4	-	1	-άψε (242)

At the bottom of the interface is a button labeled "Recognize New Verb".

Εικόνα 13: Αναγνώριση συμβολοσειράς 'ράψε'

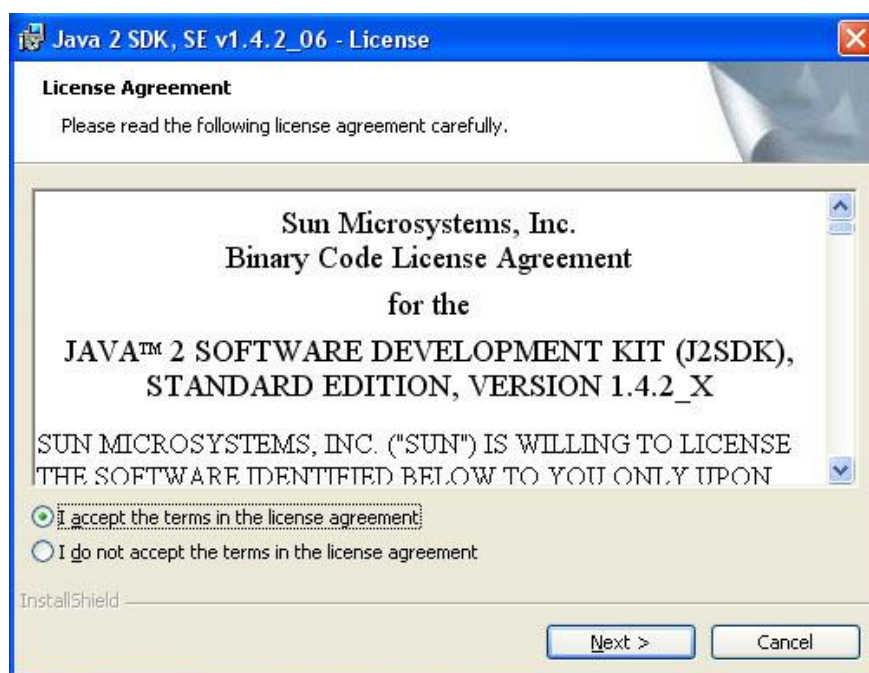
## 9. ΟΔΗΓΙΕΣ ΕΓΚΑΤΑΣΤΑΣΗΣ ΕΦΑΡΜΟΓΗΣ

Σε αυτό το κεφάλαιο περιέχονται αναλυτικές οδηγίες με κείμενο και εικόνες που περιγράφουν τις ενέργειες που πρέπει να εκτελεστούν ώστε να εγκατασταθεί η διαδικτυακή εφαρμογή VERBTAGGR++. Όλο το απαραίτητο λογισμικό περιλαμβάνεται στο cd που παραδόθηκε στο πλαίσιο της παρούσας πτυχιακής εργασίας. Σημειώνεται ότι η εφαρμογή έχει δοκιμαστεί στα λειτουργικά συστήματα Microsoft Windows XP Professional, Microsoft Windows 2000, Windows 8 και Windows 8.1.

Για να γίνει λοιπόν η εγκατάσταση της εφαρμογής, θα πρέπει να γίνουν τα εξής βήματα:

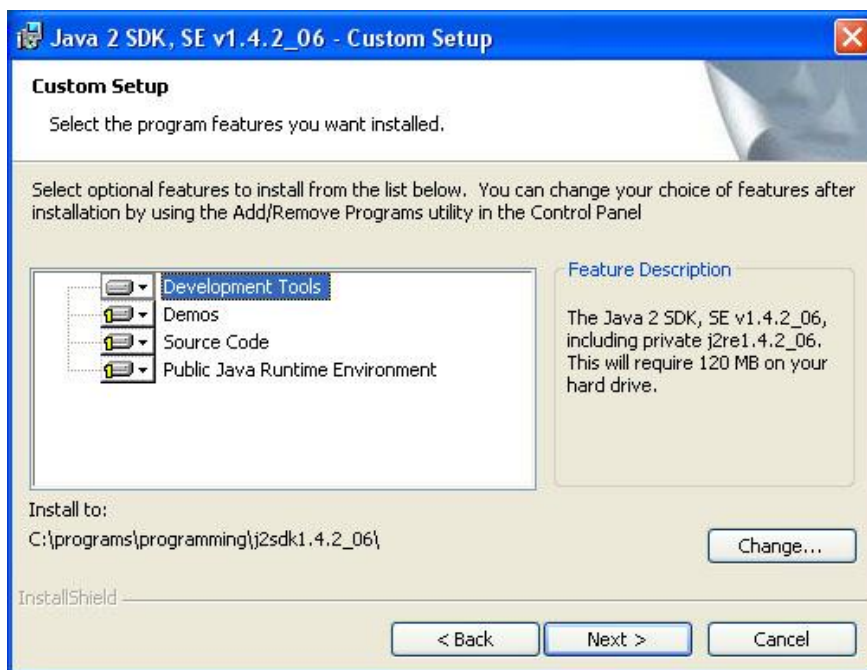
### 1) Εγκατάσταση της πλατφόρμας Java™ 2 SDK, Standard Edition έκδοση 1.4.2 (J2SE 1.4.2\_06 SDK)

Εκτελείτε το αρχείο software\_to\_install \ J2SEv 1.4.2\_06 SDK \ j2sdk-1\_4\_2\_06-windows-i586-p.exe που υπάρχει στο cd, οπότε σας παρουσιάζονται τα ακόλουθα παράθυρα και εκτελείτε τις οδηγίες που καταγράφονται κάτω από κάθε εικόνα:



Εικόνα 9: Εγκατάσταση J2SDK – βήμα 1

Επιλέγεται *'I accept the terms in the license agreement'* και πιέζετε το κουμπί *'Next'*.



Εικόνα 10: Εγκατάσταση J2SDK – βήμα 2

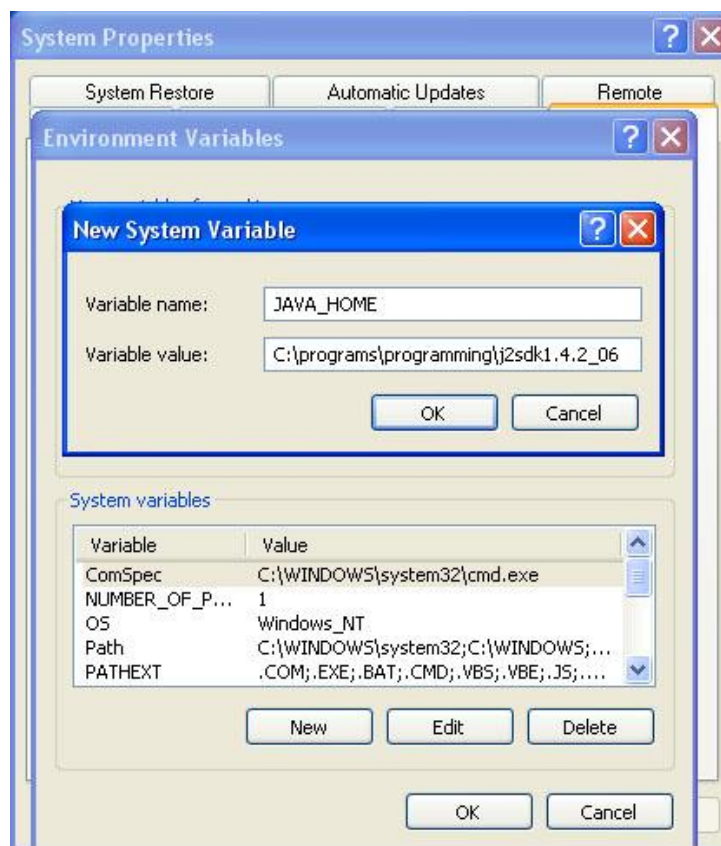
Πιέζοντας το κουμπί *'Change'*, σας δίνεται η δυνατότητα να αλλάξετε τον κατάλογο στον οποίο θα γίνει η εγκατάσταση της πλατφόρμας. Στο παράδειγμα της εικόνας έχει επιλεγεί ο κατάλογος `C:\ programs \ programming \ j2sdk1.4.2_06\`. Στη συνέχεια πιέζετε το κουμπί *'Next'*.



Εικόνα 11: Εγκατάσταση J2SDK – βήμα 3

Πιέζετε το κουμπί *'Finish'* ώστε να ολοκληρωθεί η εγκατάσταση της πλατφόρμας.

Στη συνέχεια θέτετε μία μεταβλητή περιβάλλοντος με όνομα JAVA\_HOME και τιμή το μονοπάτι εγκατάστασης του J2SDK (πχ για το μονοπάτι που φαίνεται στις προηγούμενες εικόνες, η JAVA\_HOME θα πρέπει να πάρει την τιμή C:\programs\programming\j2sdk1.4.2\_06). Οι μεταβλητές περιβάλλοντος στα Windows XP ορίζονται από το Control Panel → System → Advanced → Environment Variables → New (System Variable) όπου στο Variable name δίνετε το όνομα της μεταβλητής και στο Variable value την τιμή της όπως φαίνεται στην παρακάτω εικόνα:



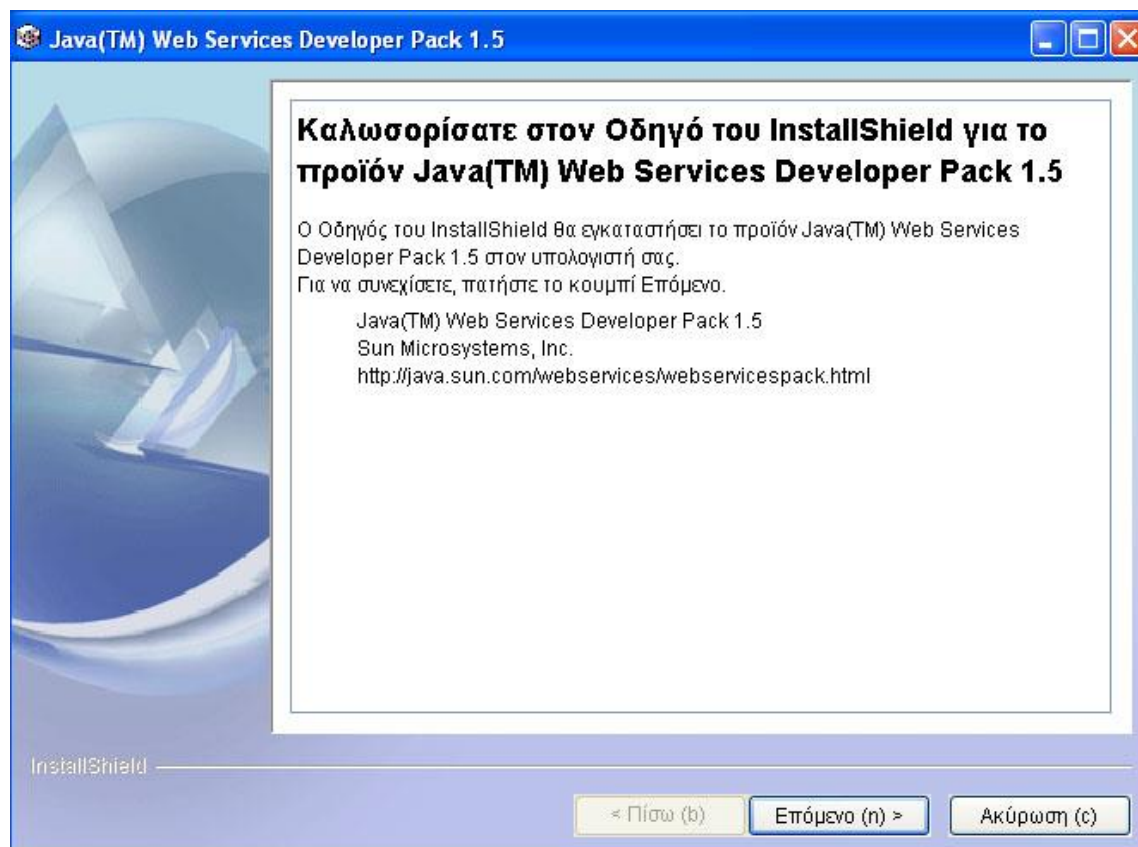
Εικόνα 12: Δημιουργία μεταβλητής περιβάλλοντος JAVA\_HOME

## 2) Εγκατάσταση του web container Tomcat

Για την εγκατάσταση του Tomcat θα πρέπει απλά να αποσυμπιέσετε το αρχείο software\_to\_install \ Tomcat by Sun \ tomcat50-jwsdp.zip του cd σε κάποιο σημείο στο δίσκο, έστω στο C:\ programs \ programming οπότε πλέον υπάρχει ένας κατάλογος με όνομα tomcat50-jwsdp εντός του C:\ programs \ programming.

## 3) Εγκατάσταση του Java Web Services Developer Pack έκδοσης 1.5 (JWSDP 1.5)

Για να γίνει η εγκατάσταση του συγκεκριμένου εργαλείου, εκτελείτε το αρχείο software\_to\_install \ JWSDP 1.5 \ jwsdp-1\_5-windows-i586.exe του cd, οπότε παρουσιάζονται τα παράθυρα των παρακάτω εικόνων στα οποία πραγματοποιείτε τις ενέργειες που αναφέρονται κάτω από κάθε εικόνα:



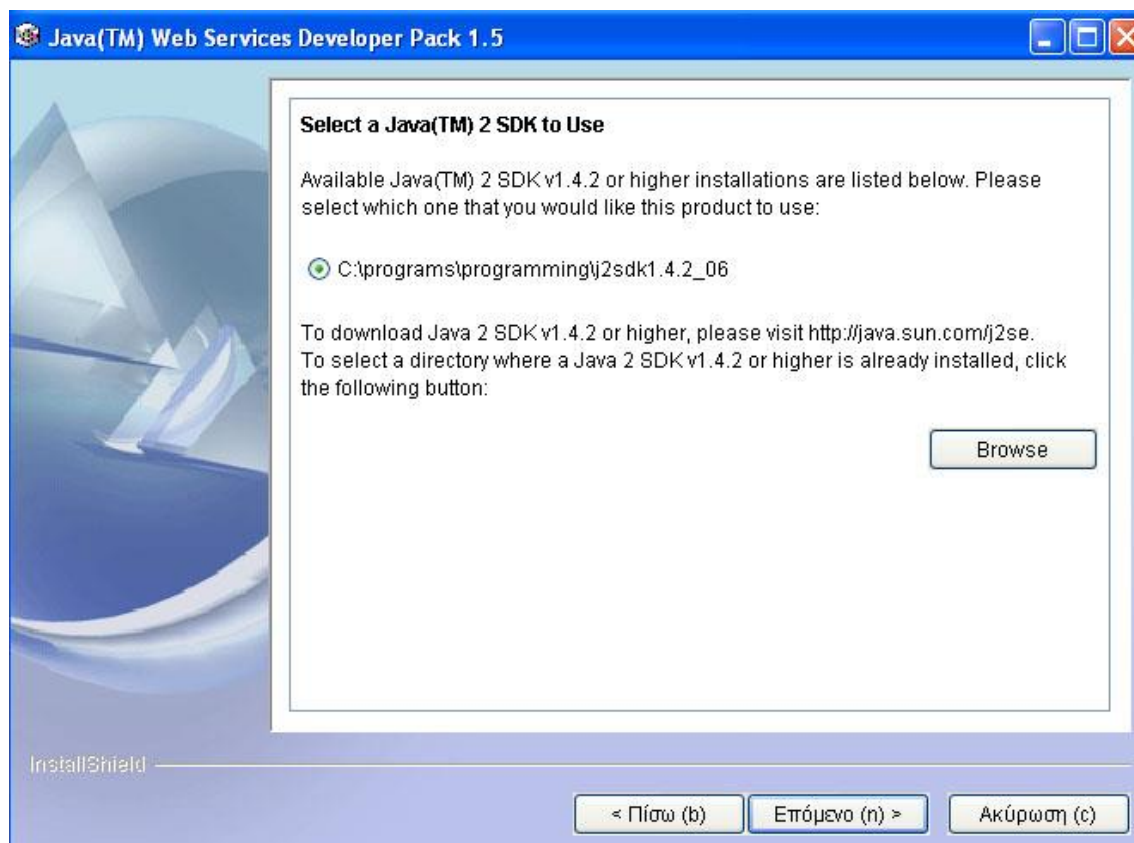
Εικόνα 13: Εγκατάσταση JWSDP – βήμα 1

Πιέζετε το κουμπί 'Επόμενο'.



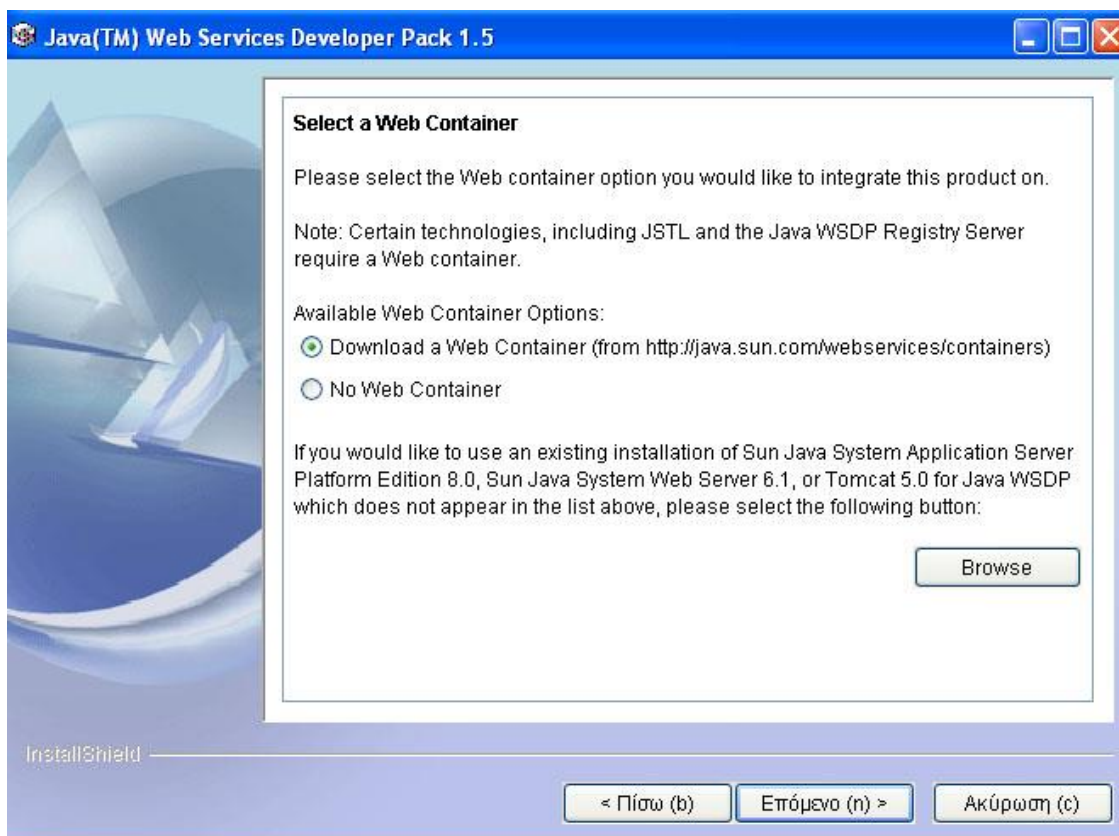
Εικόνα 14: Εγκατάσταση JWSDP – βήμα 2

Επιλέγετε 'Approve' και πιέζετε το κουμπί 'Επόμενο'



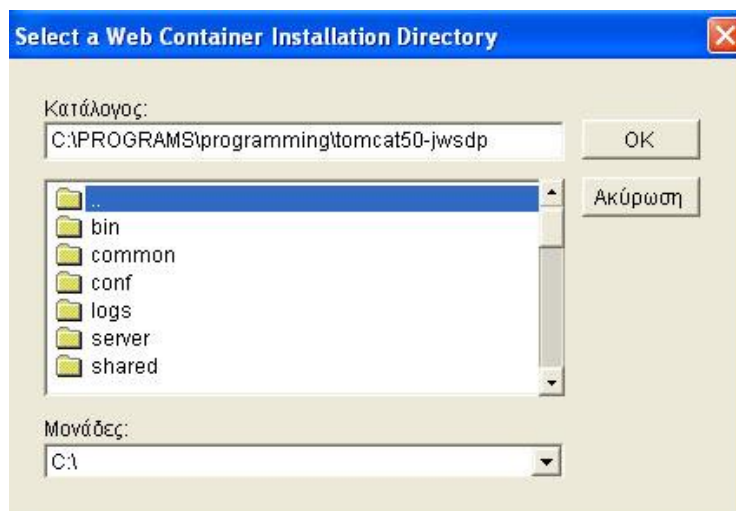
Εικόνα 15: Εγκατάσταση JWSDP – βήμα 3

Πιέζετε το κουμπί 'Επόμενο'.



Εικόνα 16: Εγκατάσταση JWSDP – βήμα 4

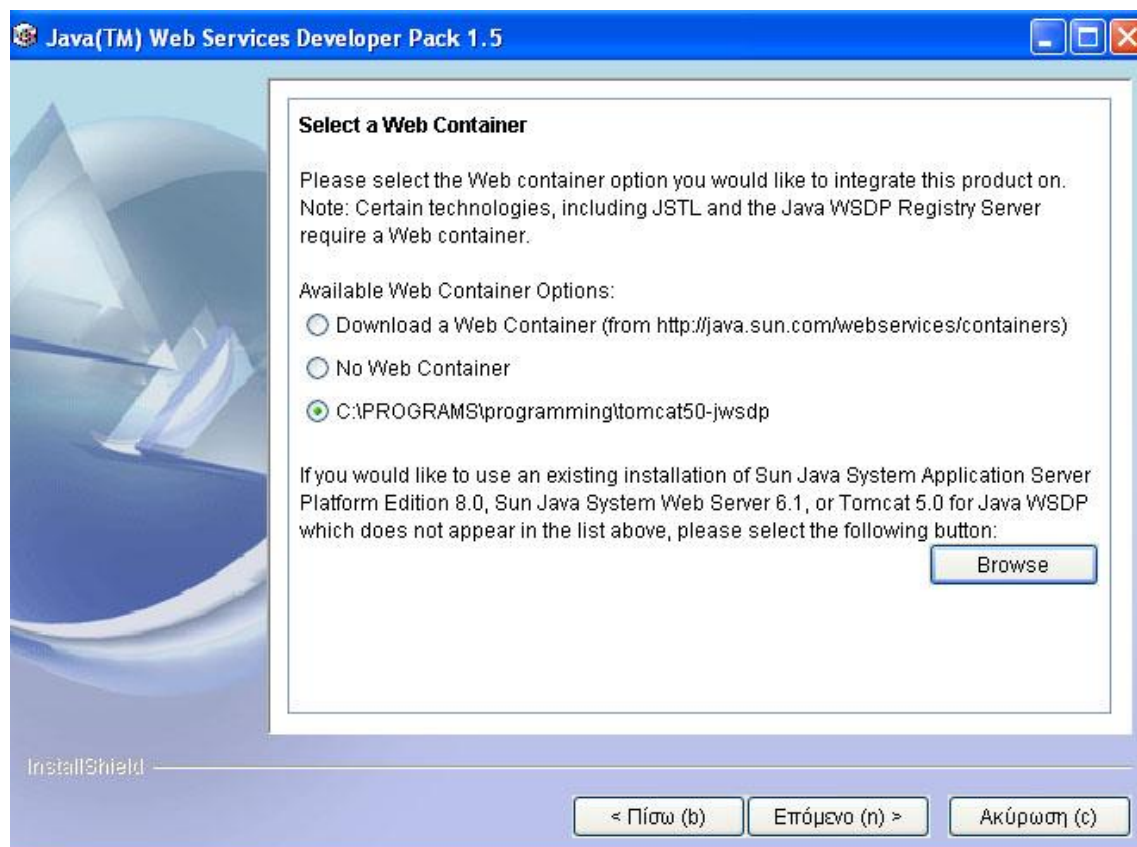
Πιέζετε το κουμπί 'Browse', οπότε σας παρουσιάζεται το ακόλουθο παράθυρο:



Εικόνα 17: Εγκατάσταση JWSDP – βήμα 5

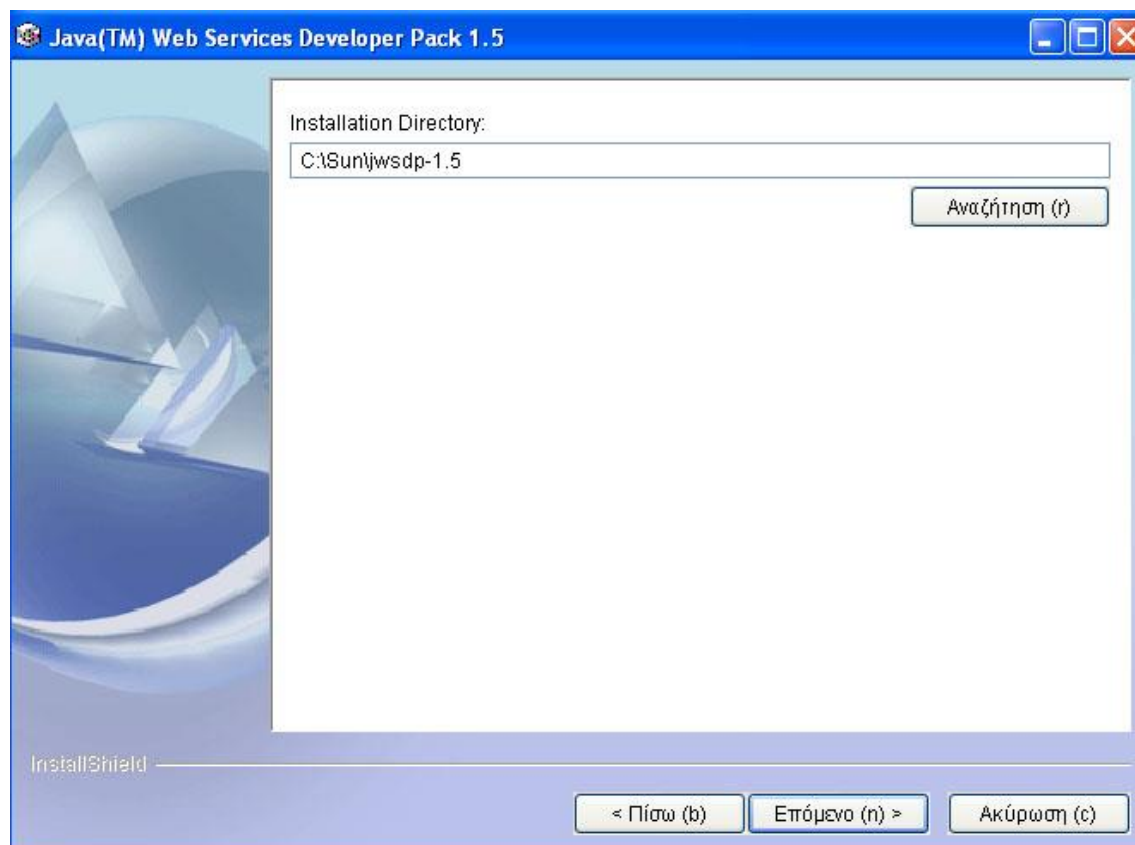
Στο παράθυρο που φαίνεται στην παραπάνω εικόνα, επιλέγετε το μονοπάτι στο οποίο έχει γίνει η εγκατάσταση του Tomcat (στο παράδειγμα των εικόνων είναι το C:\programs \ programming \ tomcat50-jwsdp) και πιέζετε το κουμπί 'OK'.





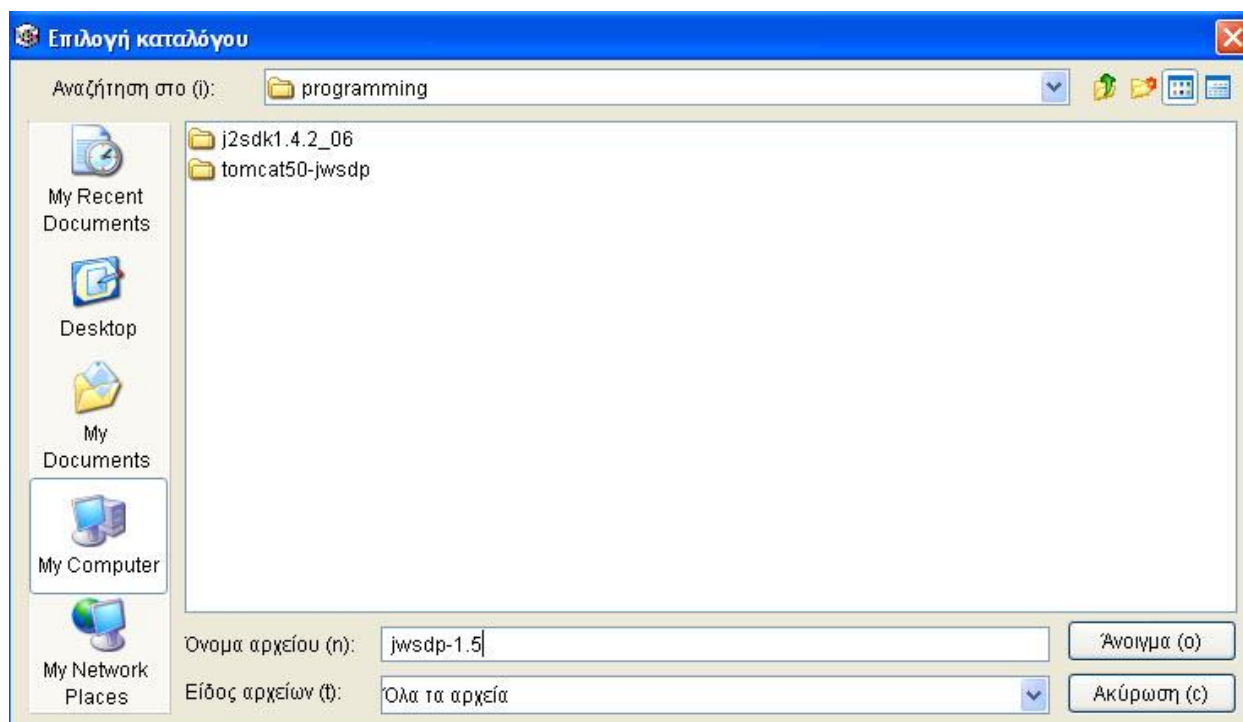
Εικόνα 18: Εγκατάσταση JWSDP – βήμα 6

Στο παραπάνω παράθυρο εμφανίζεται πλέον το μονοπάτι εγκατάστασης του Tomcat (εδώ το C:\ programs \ programming \ tomcat50-jwsdp) το οποίο επιλέγετε και στη συνέχεια πιέζετε το κουμπί 'Επόμενο'.



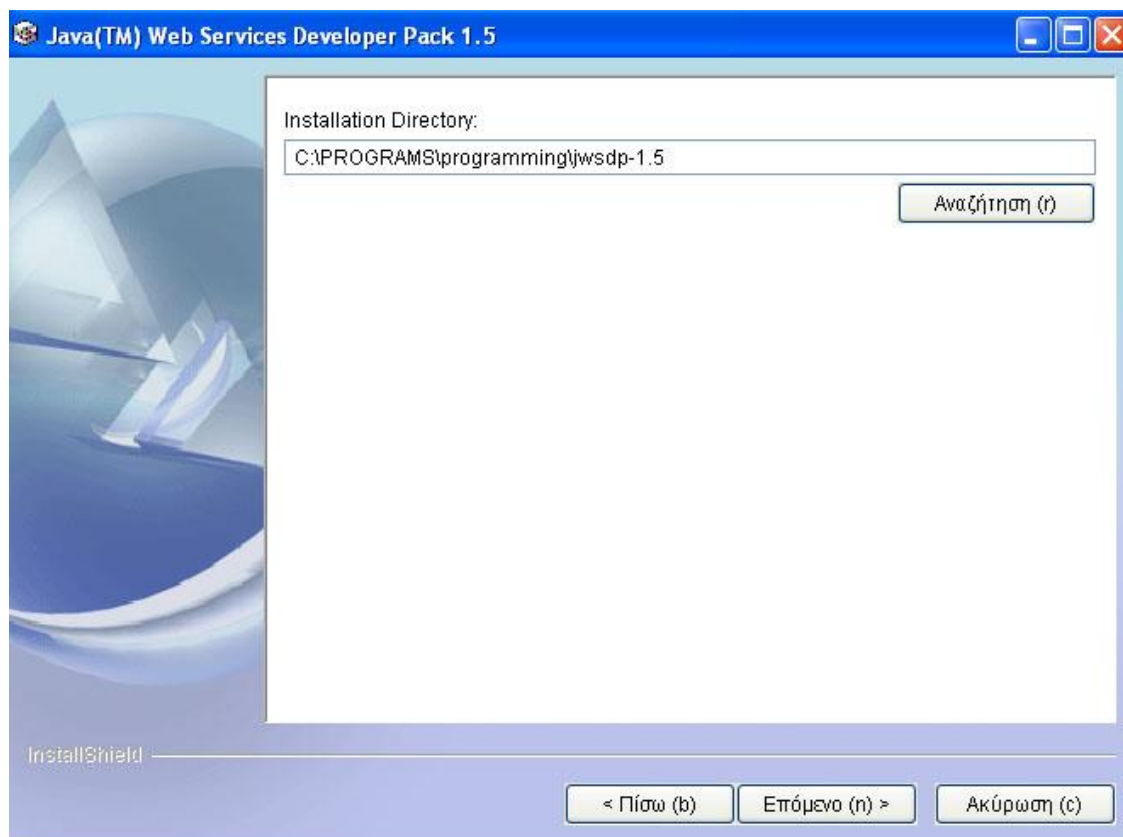
Εικόνα 19: Εγκατάσταση JWSDP – βήμα 7

Πιέζετε το κουμπί 'Αναζήτηση,' ώστε να επιλέξετε το μονοπάτι εγκατάστασης.



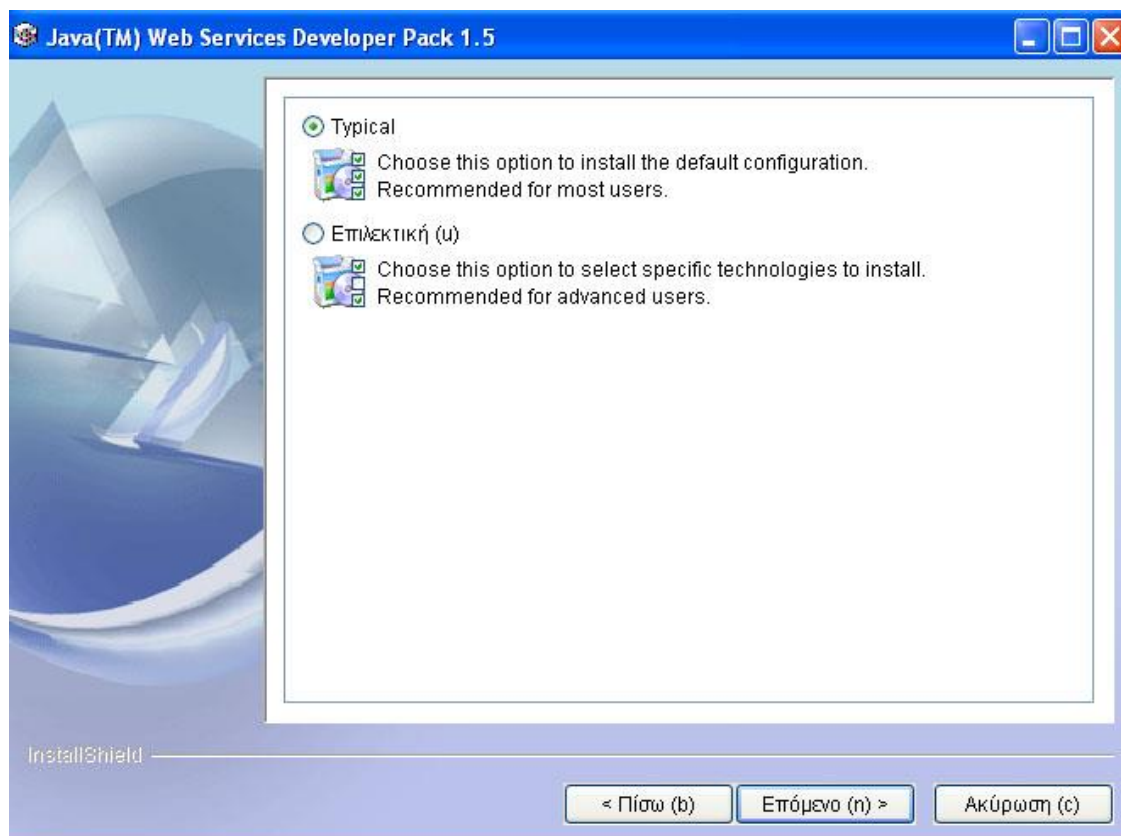
Εικόνα 20: Εγκατάσταση JWSDP – βήμα 8

Στο παραπάνω παράθυρο επιλέγετε το μονοπάτι στο οποίο επιθυμείτε να γίνει η εγκατάσταση του εργαλείου JWSDP 1.5.



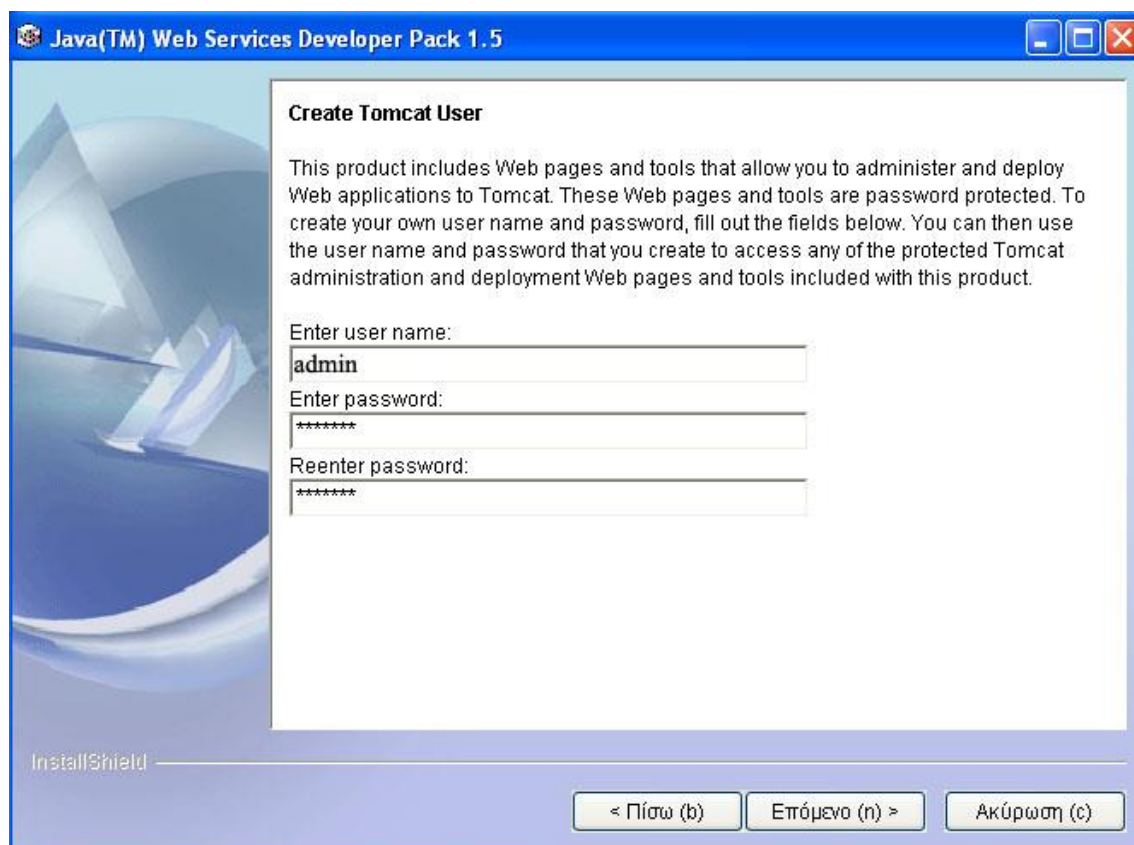
Εικόνα 21: Εγκατάσταση JWSDP – βήμα 9

Πιέζετε το κουμπί 'Επόμενο'.



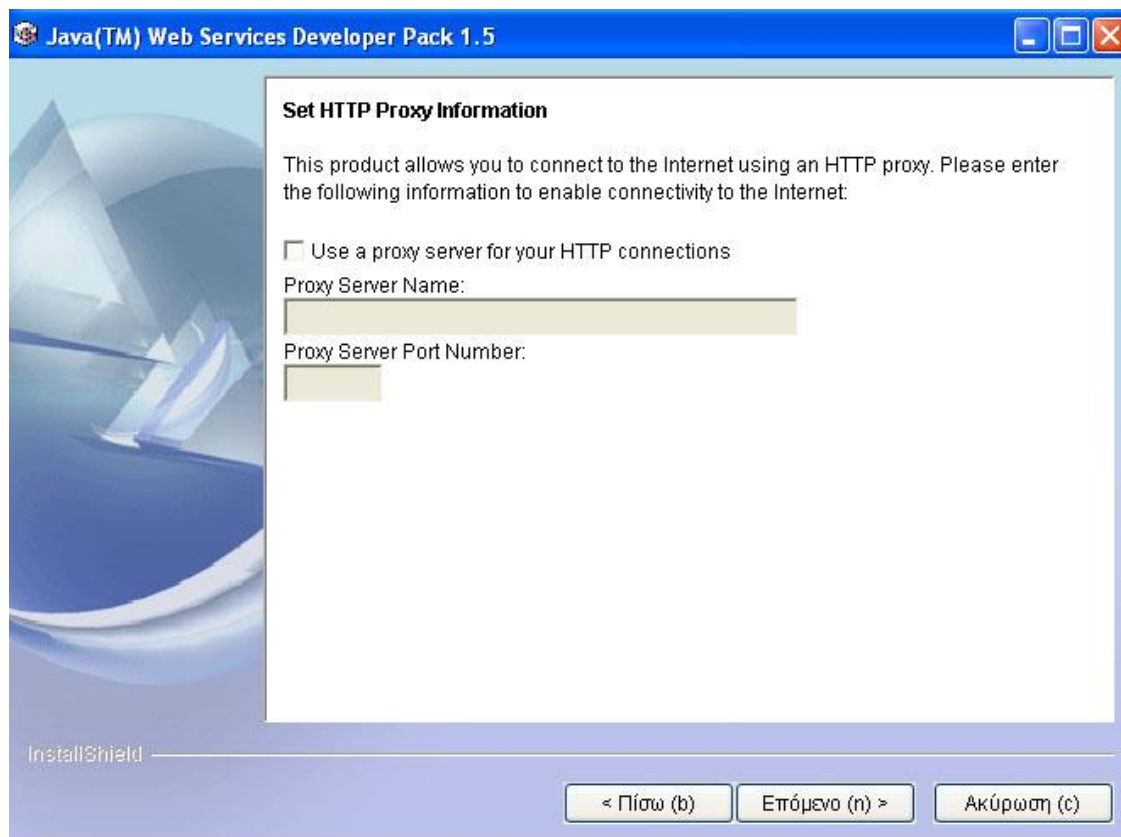
Εικόνα 22: Εγκατάσταση JWSDP – βήμα 10

Επιλέγετε *'Typical'* και πιέζετε το κουμπί *'Επόμενο'*.



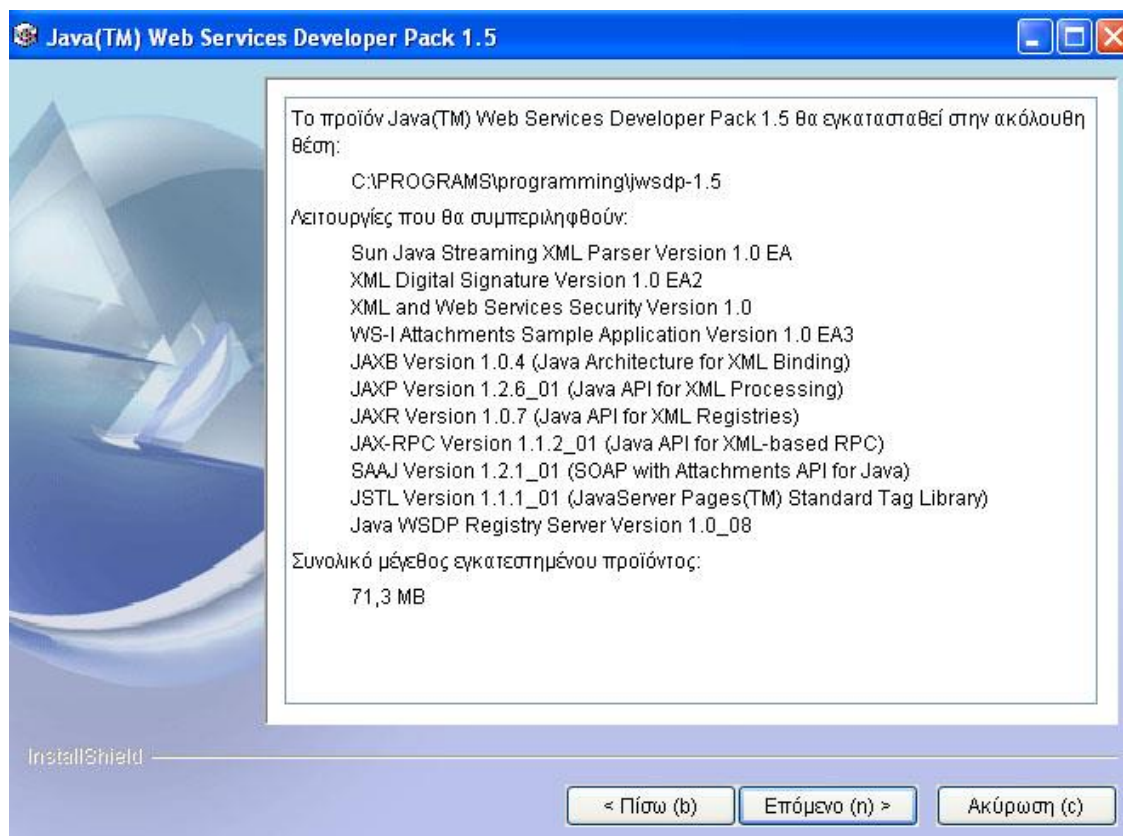
Εικόνα 23: Εγκατάσταση JWSDP – βήμα 11

Εισάγετε ένα όνομα χρήστη της επιλογής σας και δύο φορές τον ίδιο κωδικό πρόσβασης και πιέζετε το κουμπί *‘Επόμενο’*.



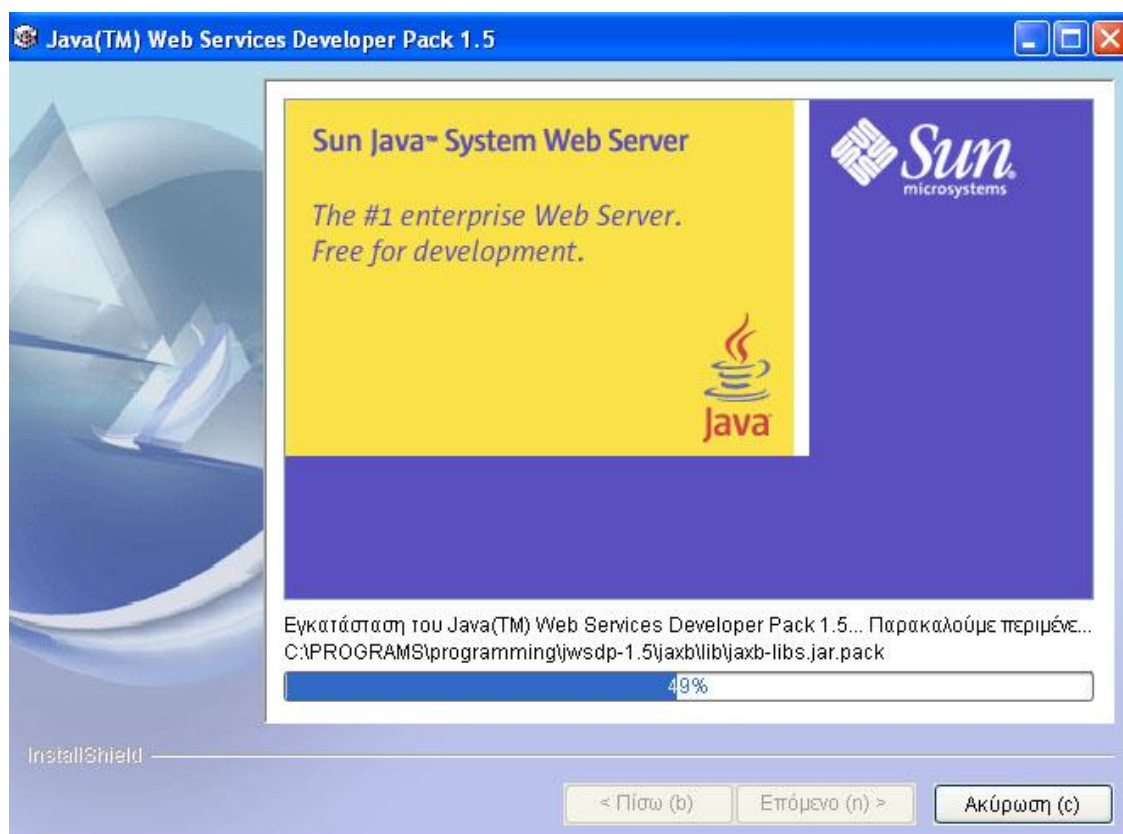
Εικόνα 24 :Εγκατάσταση JWSDP – βήμα 12

Πιέζετε το κουμπί *‘Επόμενο’*.



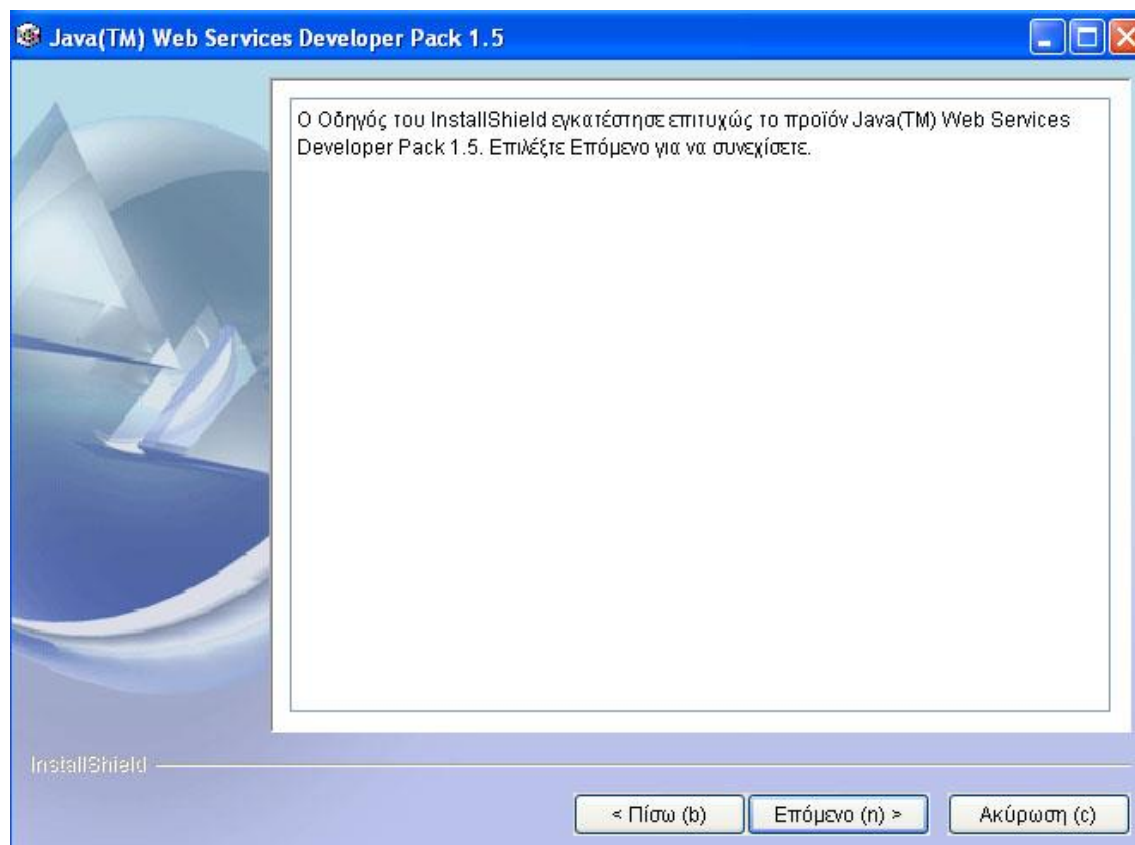
Εικόνα 25: Εγκατάσταση JWSDP – βήμα 13

Πιέζετε το κουμπί 'Επόμενο'.



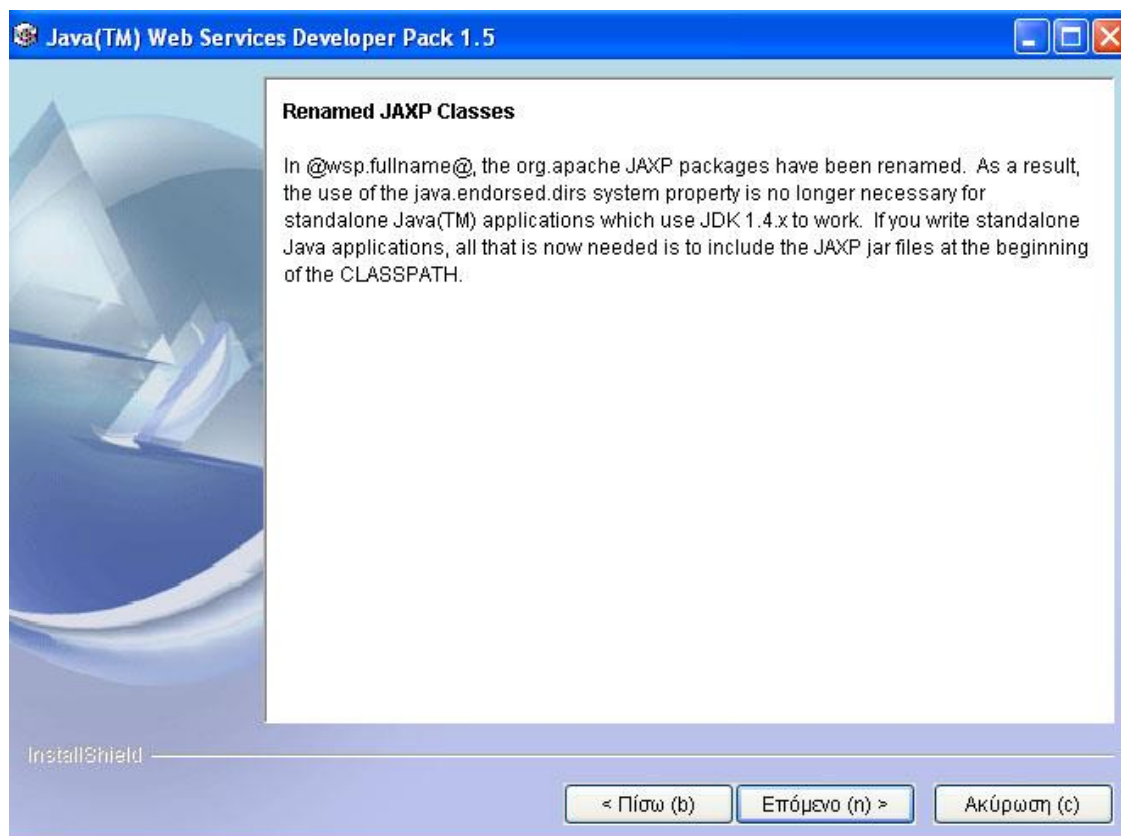
Εικόνα 26: Εγκατάσταση JWSDP – βήμα 14

Περιμένετε μέχρι να πραγματοποιηθεί η εγκατάσταση.



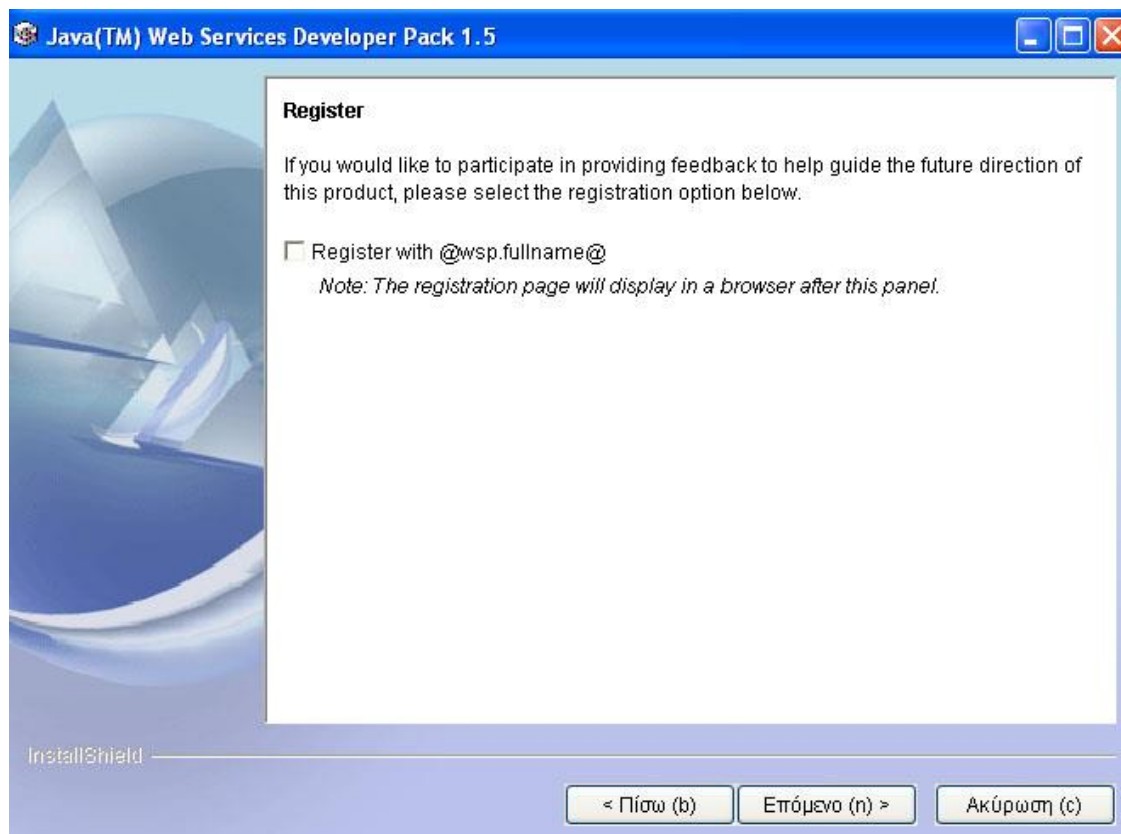
Εικόνα 27: Εγκατάσταση JWSDP – βήμα 15

Πιέζετε το κουμπί 'Επόμενο'.



Εικόνα 28: Εγκατάσταση JWSDP – βήμα 16

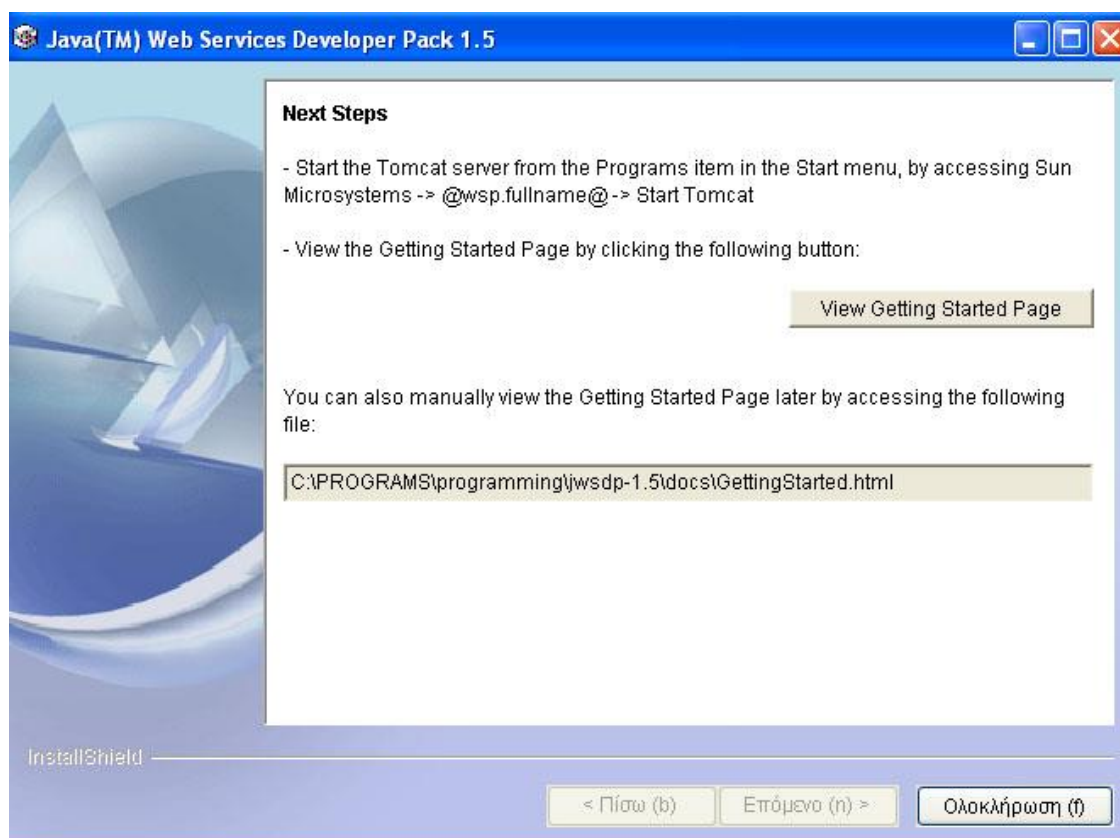
Πιέζετε το κουμπί 'Επόμενο'.



Εικόνα 29: Εγκατάσταση JWSDP – βήμα 17



Πιέζετε το κουμπί 'Επόμενο'.



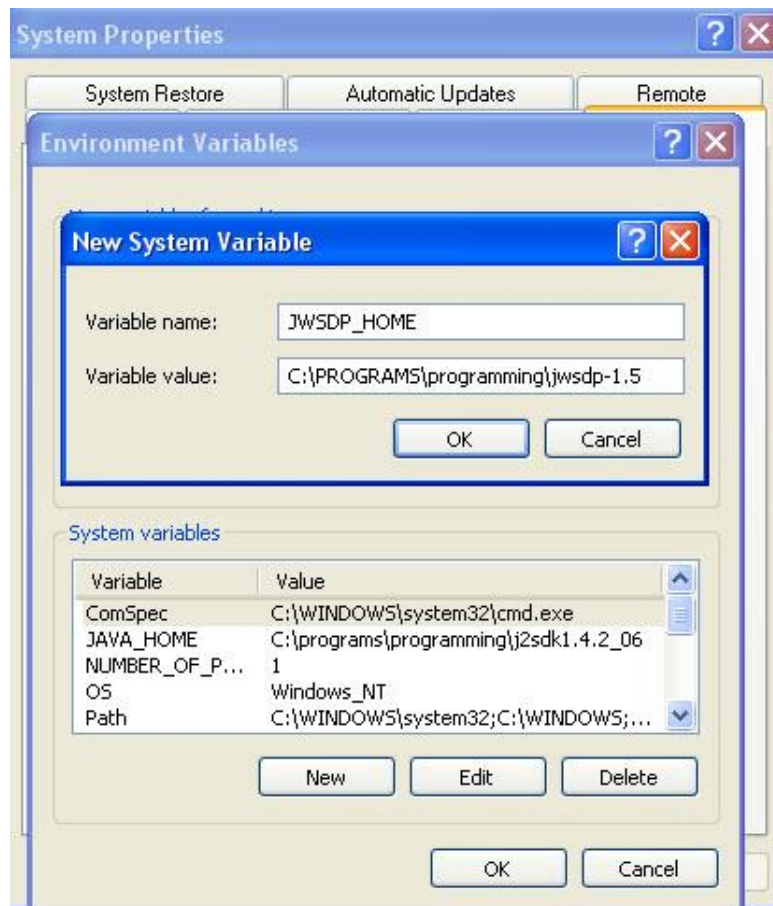
Εικόνα 30: Εγκατάσταση JWSDP – βήμα 18

Πιέζετε το κουμπί 'Ολοκλήρωση' ώστε να ολοκληρωθεί η εγκατάσταση.

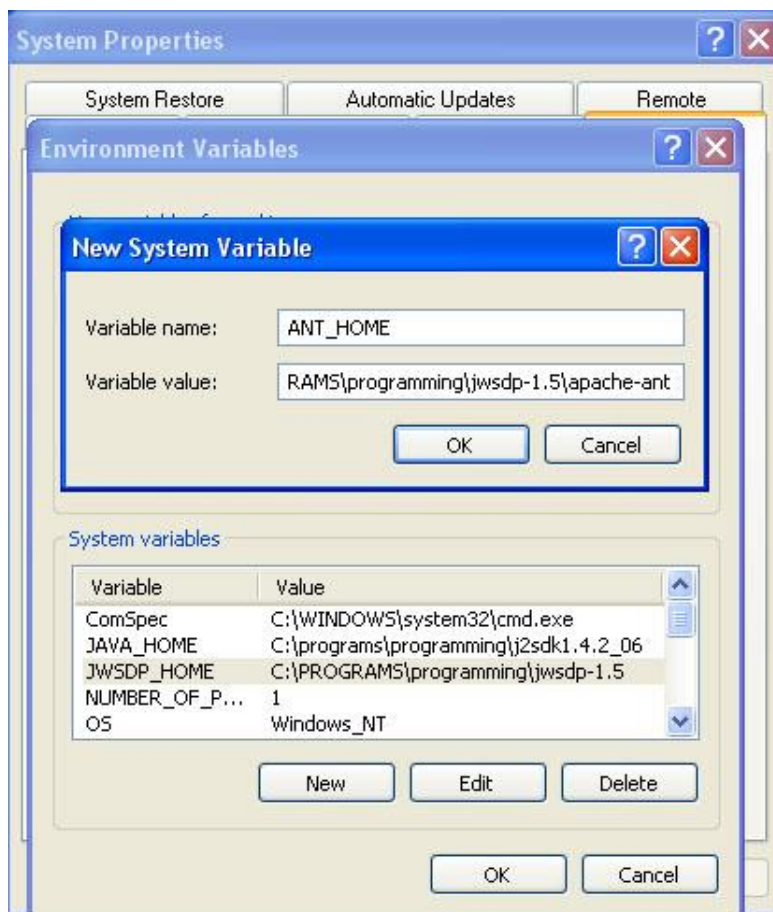
Στη συνέχεια θέτετε τις εξής μεταβλητές περιβάλλοντος:

- Μία μεταβλητή με όνομα JWSDP\_HOME και τιμή το μονοπάτι εγκατάστασης του JWSDP (πχ για το παράδειγμα των παραπάνω εικόνων, η τιμή που πρέπει να αποδοθεί είναι η C:\programs\programming\jwsdp-1.5).
- Μία μεταβλητή με όνομα ANT\_HOME και τιμή το μονοπάτι του καταλόγου apache\_ant που βρίσκεται εντός του καταλόγου εγκατάστασης του JWSDP (πχ για το παράδειγμα των παραπάνω εικόνων, η τιμή που πρέπει να αποδοθεί είναι η C:\programs\programming\jwsdp-1.5\apache\_ant).

Η δημιουργία των μεταβλητών αυτών φαίνεται στις επόμενες δύο εικόνες:



Εικόνα 31: Δημιουργία μεταβλητής περιβάλλοντος JWSDP\_HOME



Εικόνα 32: Δημιουργία μεταβλητής περιβάλλοντος ANT\_HOME

#### 4) Τροποποίηση της τιμής της μεταβλητής περιβάλλοντος PATH

Στη συνέχεια θα πρέπει να ενημερώσετε την τιμή της μεταβλητής περιβάλλοντος PATH. Στα Windows XP αυτό γίνεται μέσω του Control Panel → System → Advanced → Environment Variables → Επιλογή της system variable PATH → Edit. Στο τέλος της υπάρχουσας τιμής της PATH, θα πρέπει να προστεθεί το ακόλουθο όπου στη θέση του 'J2SDK\_install\_path' και του 'JWSDP\_install\_path' τοποθετείτε αντίστοιχα το μονοπάτι εγκατάστασης της J2SDK πλατφόρμας και του JWSDP εργαλείου:

---

```
;J2SDK_install_path\bin;JWSDP_install_path\jaxrpc\bin;JWSDP_install_path\apache-ant\bin
```

---

Για το παράδειγμα των εικόνων, θα πρέπει να προστεθεί το εξής:

```
;C:\PROGRAMS\programming\j2sdk1.4.2_06\bin;C:\PROGRAMS\programming\jwsdp-1.5\jaxrpc\bin;C:\PROGRAMS\programming\jwsdp-1.5\ apache-ant\bin
```

#### 5) Εγκατάσταση του συστήματος διαχείρισης βάσεων δεδομένων MySQL

Εκτελείτε το αρχείο software\_to\_install \ MySQL \ mysql-4.1.7-essential-win.msi του cd, οπότε παρουσιάζονται τα παράθυρα των παρακάτω εικόνων στα οποία πραγματοποιείτε τις ενέργειες που αναφέρονται κάτω από κάθε εικόνα:



Εικόνα 33: Εγκατάσταση MySQL – βήμα 1

Πιέζετε το κουμπί 'Next'.



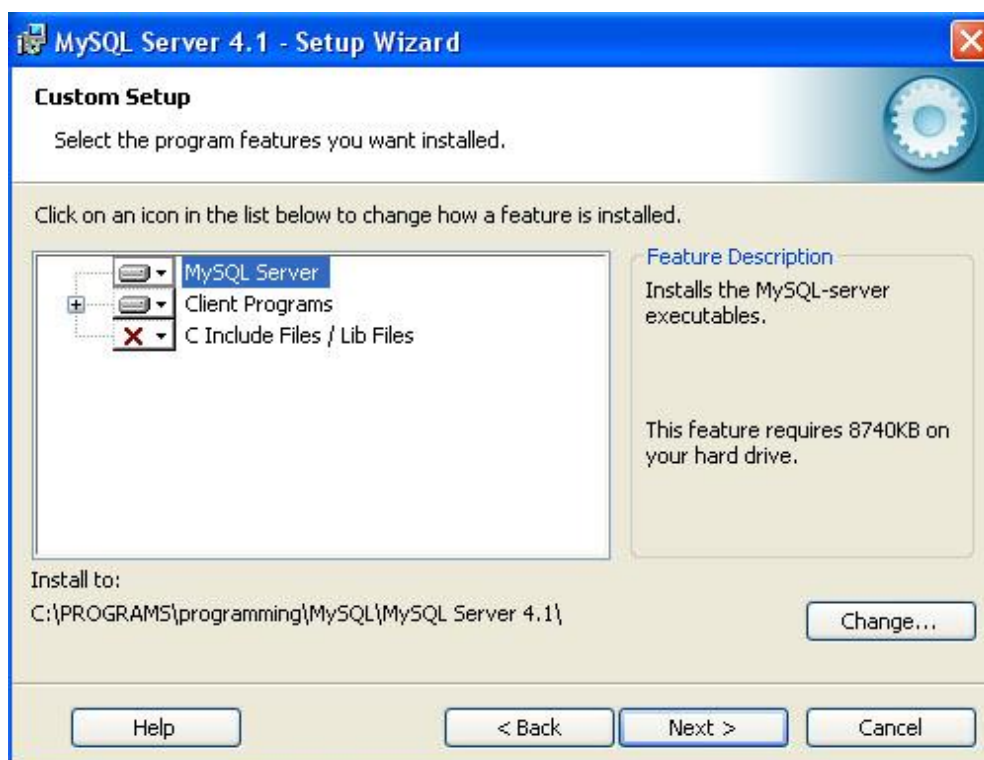
Εικόνα 34: Εγκατάσταση MySQL – βήμα 2

Επιλέγετε 'Custom' και πιέζετε το κουμπί 'Next'.



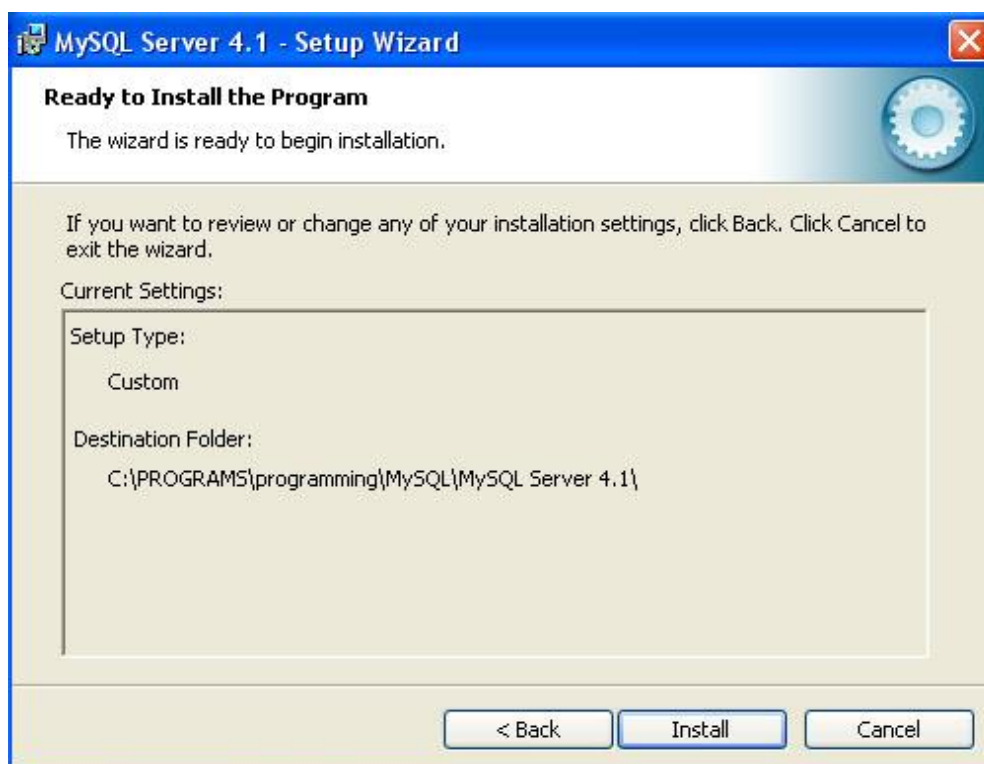
Εικόνα 35: Εγκατάσταση MySQL – βήμα 3

Πιέζοντας το κουμπί 'Change', σας δίνεται η δυνατότητα να αλλάξετε τον κατάλογο στον οποίο θα γίνει η εγκατάσταση της MySQL. Στο παράδειγμα της εικόνας επιλέχθηκε ο κατάλογος C:\ programs \ programming \ MySQL \ MySQL Server 4.1\, οπότε εμφανίστηκε το παράθυρο της επόμενης εικόνας.



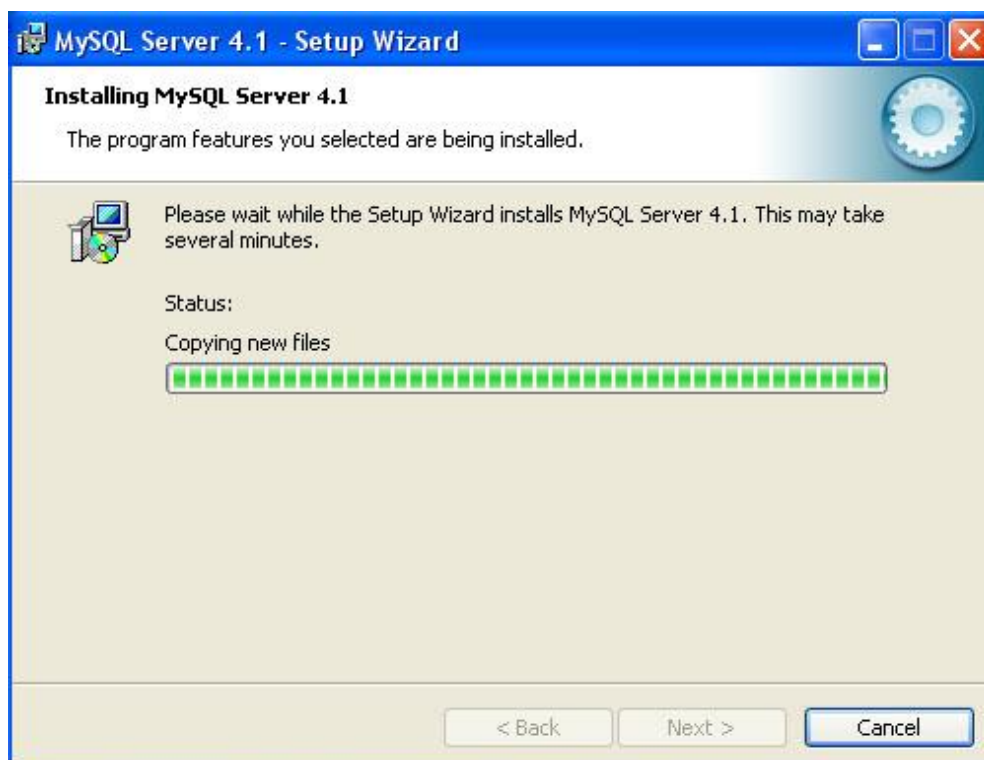
Εικόνα 36: Εγκατάσταση MySQL – βήμα 4

Πιέζετε το κουμπί 'Next'.



Εικόνα 37: Εγκατάσταση MySQL – βήμα 5

Πιέζετε το κουμπί 'Install'.



Εικόνα 38: Εγκατάσταση MySQL – βήμα 6

Περιμένετε μέχρι να πραγματοποιηθεί η εγκατάσταση.



Εικόνα 39: Εγκατάσταση MySQL – βήμα 7

Επιλέγετε 'Skip Sign-Up' και πιέζετε το κουμπί 'Next'.



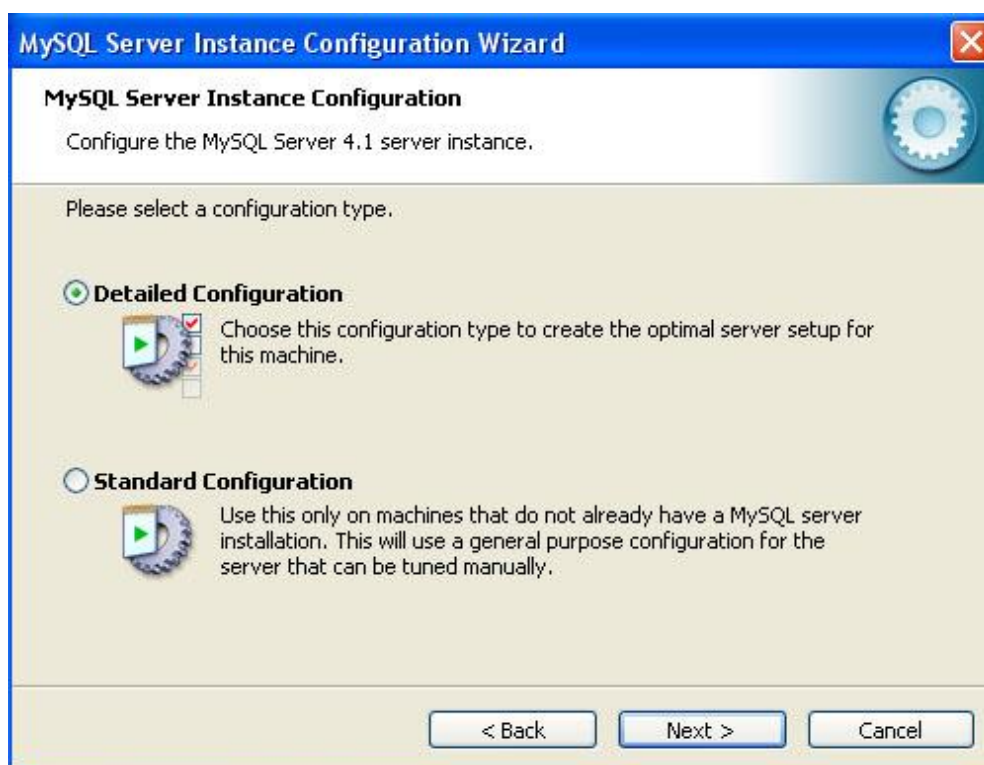
Εικόνα 40: Εγκατάσταση MySQL – βήμα 8

Επιλέγετε την πρόταση *'Configure the MySQL Server now'* και πιέζετε το κουμπί *'Finish'*.



Εικόνα 41: Εγκατάσταση MySQL – βήμα 9

Πιέζετε το κουμπί *'Next'*.



Εικόνα 42: Εγκατάσταση MySQL – βήμα 10

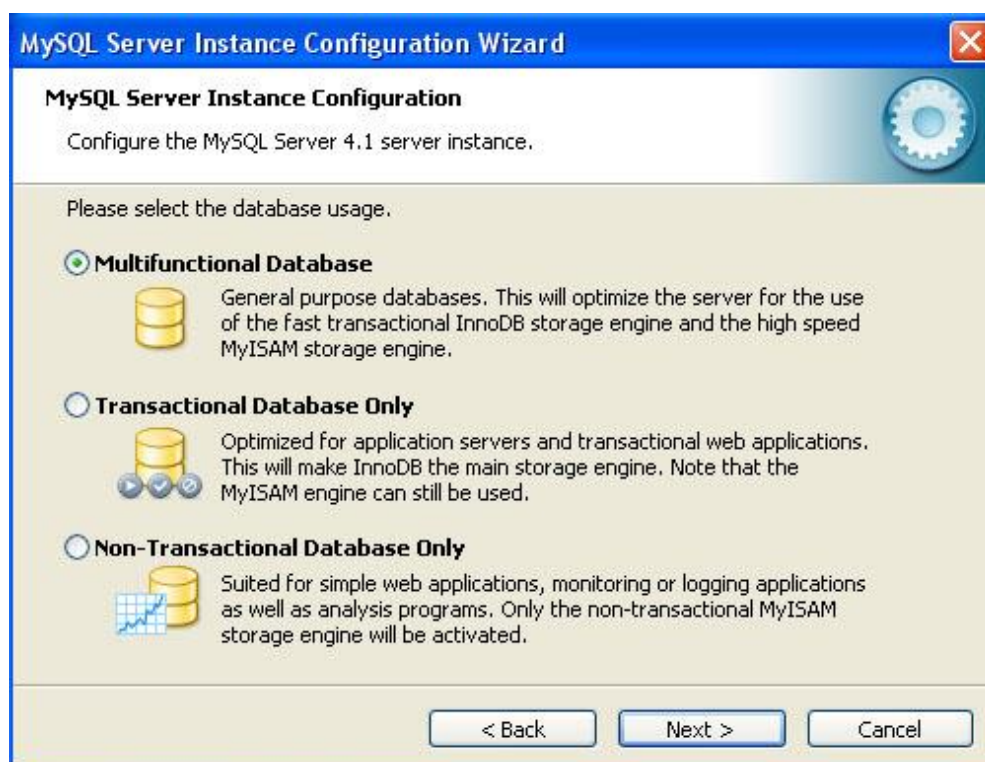


Επιλέγεται *'Detailed Configuration'* και πιέζεται το κουμπί *'Next'*.



Εικόνα 43: Εγκατάσταση MySQL – βήμα 11

Επιλέγεται *'Developer Machine'* και πιέζεται το κουμπί *'Next'*.



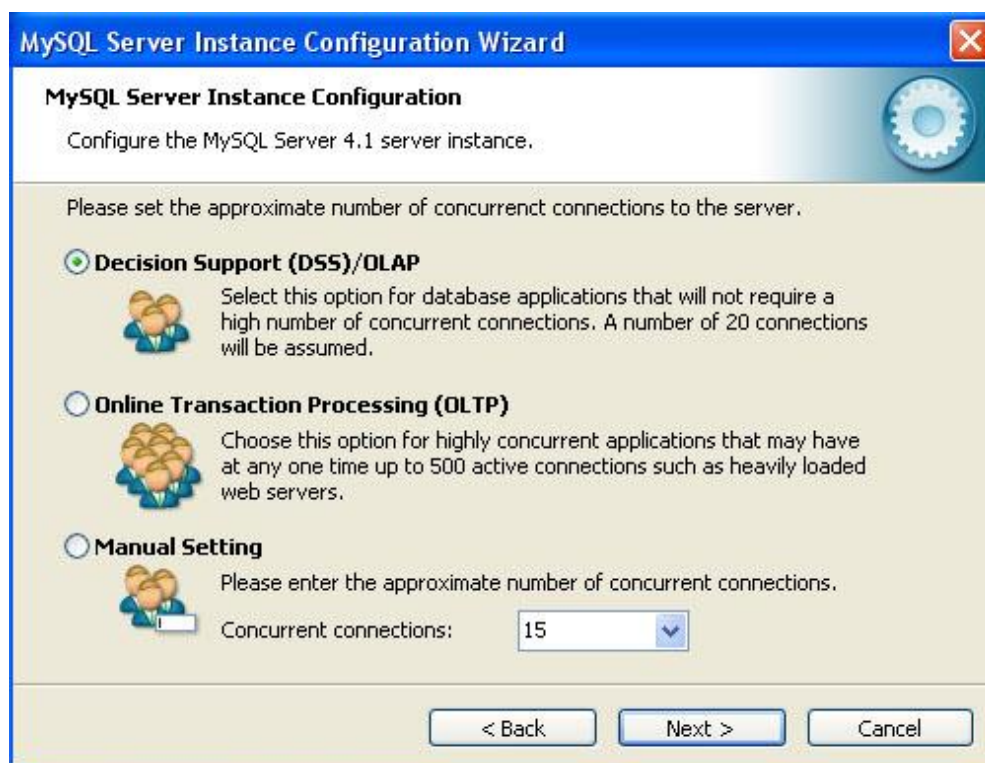
Εικόνα 44: Εγκατάσταση MySQL – βήμα 12

Επιλέγεται *'Multifunctional Database'* και πιέζεται το κουμπί *'Next'*.



Εικόνα 45: Εγκατάσταση MySQL – βήμα 13

Πιέζεται το κουμπί *'Next'*.



Εικόνα 46: Εγκατάσταση MySQL – βήμα 14

Επιλέγετε *'Decision Support (DSS)/OLAP'* και πιέζετε το κουμπί *'Next'*.



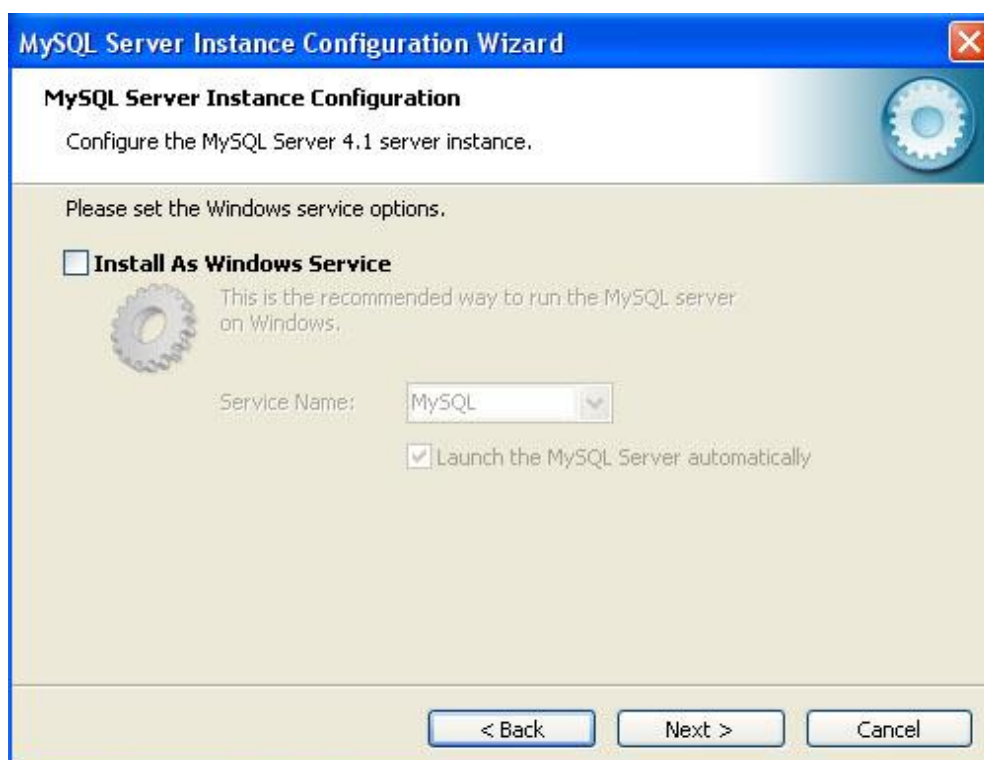
Εικόνα 47: Εγκατάσταση MySQL – βήμα 15

Δεν επιλέγετε *'Enable TCP/IP Networking'* εφόσον επιθυμείτε να χρησιμοποιήσετε τη MySQL τοπικά στον υπολογιστή σας και πιέζετε το κουμπί *'Next'*.



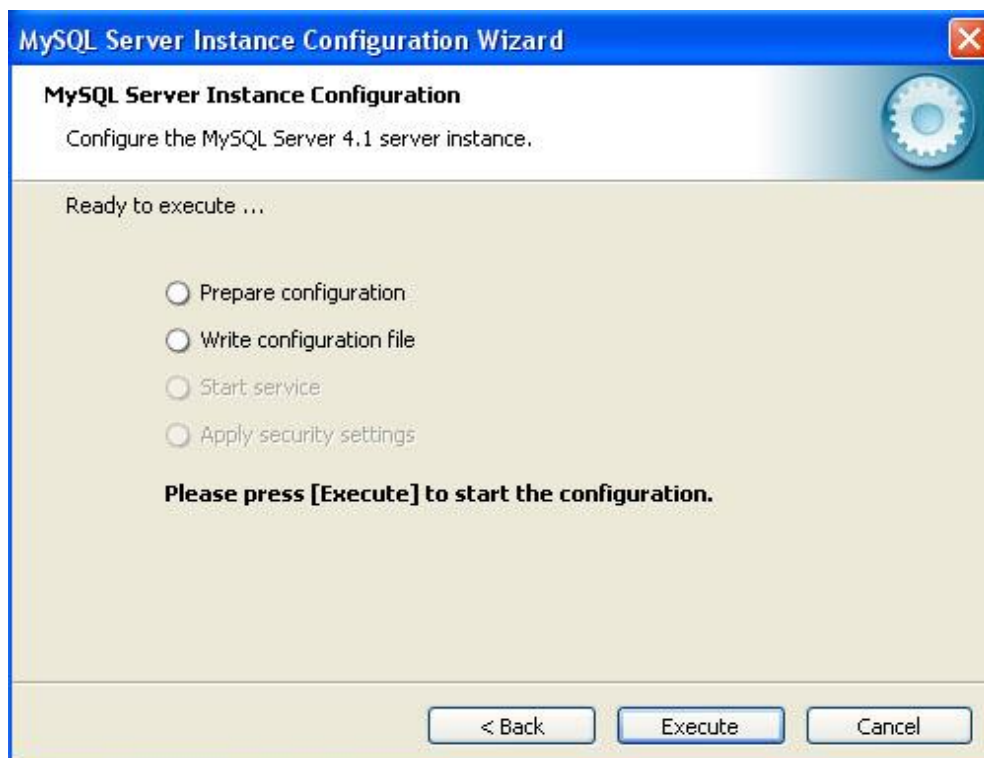
Εικόνα 48: Εγκατάσταση MySQL – βήμα 16

Επιλέγεται *'Manual Selected Default Character Set / Collation'* και στο πεδίο του *'Character Set'* δίνεται *'greek'*. Στη συνέχεια πιέζετε το κουμπί *'Next'*.



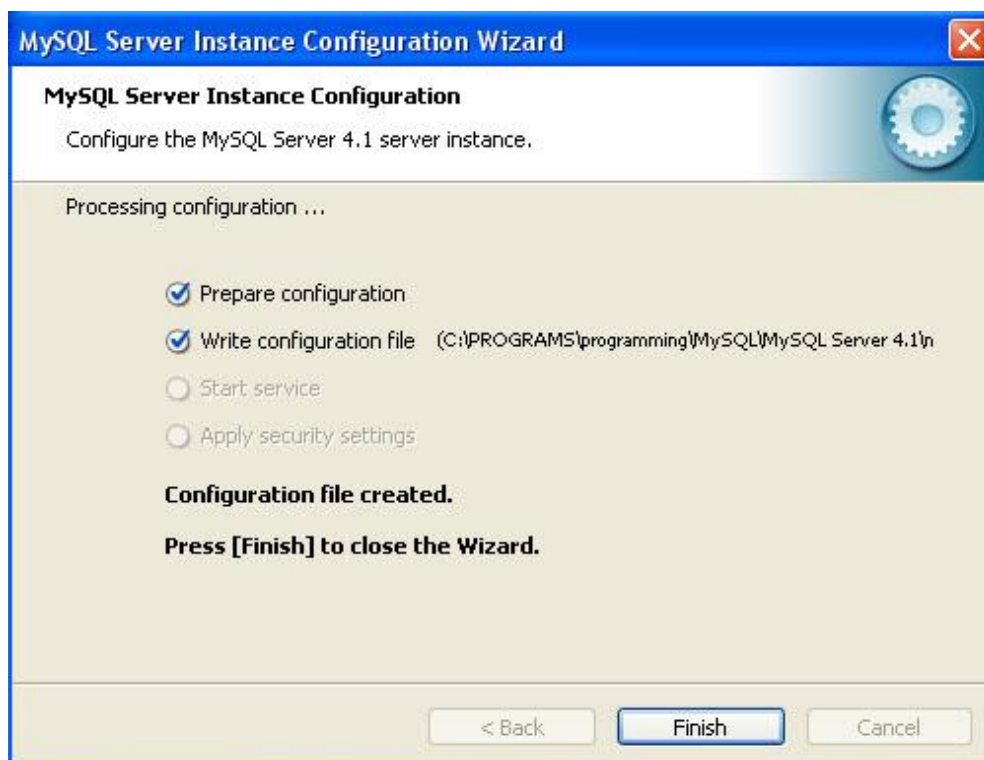
Εικόνα 49: Εγκατάσταση MySQL – βήμα 17

Δεν επιλέγεται *'Install As Windows Service'* και πιέζετε το κουμπί *'Next'*.



Εικόνα 50: Εγκατάσταση MySQL – βήμα 18

Πιέζετε το κουμπί *'Execute'*.



Εικόνα 51: Εγκατάσταση MySQL – βήμα 19

Πιέζετε το κουμπί *'Finish'*, ώστε να ολοκληρωθεί η εγκατάσταση.

Μετά την εγκατάσταση της MySQL, καλό είναι να θέσετε ένα νέο κωδικό πρόσβασης για το root, διότι διαφορετικά ο root θα συνδέεται με κενό κωδικό πρόσβασης, γεγονός το οποίο δεν είναι ασφαλές. Για να γίνει αυτό, εκτελείτε τα ακόλουθα:

- Εκκινείτε το server της MySQL:

Ανοίγετε ένα παράθυρο DOS Command Prompt και πληκτρολογείτε:

```
MySQL_install_path\bin\mysqld
```

όπου 'MySQL\_install\_path' ο κατάλογος εγκατάστασης της MySQL.

- Συνδέεστε με το server της MySQL μέσω ενός client:

Ανοίγετε ένα νέο παράθυρο DOS Command Prompt και πληκτρολογείτε:

```
MySQL_install_path\bin\mysql -u root
```

όπου 'MySQL\_install\_path' ο κατάλογος εγκατάστασης της MySQL. Στη συνέχεια δίνετε τις παρακάτω εντολές, με τις οποίες τίθεται για παράδειγμα το 'root\_pass' ως κωδικός πρόσβασης του root, ενώ επίσης διαγράφεται ο χρήστης της MySQL που έχει κενό όνομα χρήστη και κενό κωδικό πρόσβασης, το οποίο καλό είναι να γίνει επίσης για λόγους ασφάλειας, ώστε να μην μπορεί κανείς να συνδεθεί με το server της MySQL με κενό όνομα χρήστη και κενό κωδικό πρόσβασης:

---

```
mysql> use mysql;
```

```
mysql> UPDATE user SET password = PASSWORD('root_pass') WHERE user = 'root';
```

```
mysql> DELETE FROM user WHERE user = '';
```

---

- Τερματίζετε τον client και το server της MySQL:

Στο παράθυρο του προηγούμενου βήματος πληκτρολογείτε 'exit', στη συνέχεια δίνετε την εντολή 'MySQL\_install\_path\bin\mysqladmin -u root -p shutdown',

όπου 'MySQL\_install\_path' ο κατάλογος εγκατάστασης της MySQL και εισάγετε τον κωδικό πρόσβασης του root που θέσατε στο προηγούμενο βήμα.

## 6) Εγκατάσταση του JDBC driver για τη MySQL

Αποσυμπιέζετε το αρχείο software\_to\_install \ MySQL JDBC driver \ mysql-connector-java-3.0.16-ga.zip του cd και αντιγράφετε το αρχείο mysql-connector-java-3.0.16-ga \ mysql-connector-java-3.0.16-ga-bin.jar στους ακόλουθους καταλόγους, όπου 'J2SDK\_install\_path' είναι ο κατάλογος εγκατάστασης της πλατφόρμας της Java και 'Tomcat\_install\_path' ο κατάλογος εγκατάστασης του Tomcat:

- 'J2SDK\_install\_path' \ jre \ lib \ ext\
- 'Tomcat\_install\_path' \ common \ lib\

## 7) Δημιουργία βάσης δεδομένων και εισαγωγή στοιχείων σε αυτήν

- Αντιγράφετε τον κατάλογο SYNODEYTIKO\_YLIKO \ DB\_files\ του cd, έστω εντός του C:\.

- Εκκινείτε το server της MySQL: Ανοίγετε ένα παράθυρο DOS Command Prompt και πληκτρολογείτε:

```
MySQL_install_path\bin\mysqld
```

όπου 'MySQL\_install\_path' ο κατάλογος εγκατάστασης της MySQL.

- Συνδέεστε με το server της MySQL μέσω ενός client: Ανοίγετε ένα νέο παράθυρο DOS Command Prompt και πληκτρολογείτε:

```
MySQL_install_path\bin\mysql -u root -p
```

όπου 'MySQL\_install\_path' ο κατάλογος εγκατάστασης της MySQL και εισάγετε τον κωδικό πρόσβασης του root που είχατε θέσει στο βήμα 5 (στο παράδειγμα των οδηγιών ήταν 'root\_pass').

- Δίνετε τις παρακάτω εντολές, με τις οποίες θα εκτελεστούν οι SQL επερωτήσεις όλων των αρχείων που περιέχονται στον φάκελο DB\_files :

---

```
mysql> source C:/DB_files/createDB.sql;  
mysql> source C:/DB_files/data_A.sql;  
mysql> source C:/DB_files/data_AN.sql;  
mysql> source C:/DB_files/data_restN.sql;  
mysql> source C:/DB_files/stat~edited.sql;
```

---

Με την εκτέλεση των επερωτήσεων που περιέχονται στο αρχείο createDB.sql, πραγματοποιούνται τα εξής:

- Δημιουργείται μία βάση δεδομένων με όνομα 'verbDB'.
- Δημιουργούνται οι πίνακες 'ms\_codeExplain', 'verb\_info' και 'equalsMscodes' εντός της βάσης 'verbDB'.
- Εισάγονται δεδομένα στον πίνακα 'ms\_codeExplain'.
- Δημιουργείται ένας χρήστης της MySQL με όνομα χρήστη 'verbDB\_user', ο οποίος έχει δικαιώματα προσπέλασης μόνο στη βάση δεδομένων 'verbDB'.

Οι επερωτήσεις οι οποίες περιέχονται στα αρχεία \*.sql (εκτός του createDB.sql) εισάγουν στον πίνακα 'verb\_info' δεδομένα που αφορούν σε όλους πλέον τους ρηματικούς τύπους.

Η πρώτη εντολή που εκτελεί τις επερωτήσεις του createDB.sql είναι κρίσιμο να κληθεί, καθώς τα επόμενα αρχεία εισάγουν στοιχεία στη βάση η οποία θα πρέπει να έχει ήδη δημιουργηθεί.

Τέλος, για τη δημιουργία της βάσης lemmaDB θα πρέπει να κληθεί η εντολή

```
mysql> source C:/lemmaDB_file/createLemmaDB.sql;
```

- Τερματίζετε τον client και το server της MySQL: Στο παράθυρο του προηγούμενου βήματος πληκτρολογείτε 'exit', στη συνέχεια δίνετε την εντολή 'MySQL\_install\_path\bin\mysqladmin -u root -p shutdown', όπου

'MySQL\_install\_path' ο κατάλογος εγκατάστασης της MySQL και εισάγετε τον κωδικό πρόσβασης του root που είχατε θέσει στο βήμα 5.

## 8) Αντιγραφή αρχείων του web site

Αντιγράφετε τον κατάλογο verbttaggr του cd εντός του ακόλουθου καταλόγου (όπου 'Tomcat\_install\_path' ο κατάλογος εγκατάστασης του Tomcat):

```
Tomcat_install_path \ webapps \
```

## 9) Εκκίνηση του Tomcat και του server της MySQL

### ▪ Εκκίνηση του Tomcat:

Ανοίγετε ένα παράθυρο DOS Command Prompt και πληκτρολογείτε:

```
Tomcat_install_path\bin\catalina.bat start
```

όπου 'Tomcat\_install\_path' ο κατάλογος εγκατάστασης του Tomcat.

(Εναλλακτικά, η εκκίνηση μπορεί να γίνει μέσω του start menu:

Επιλέγετε start → All Programs → ... → Java(TM) Web Services Developer Pack 1.5 → Start Tomcat)

### ▪ Εκκίνηση του server της MySQL:

Ανοίγετε ένα παράθυρο DOS Command Prompt και πληκτρολογείτε:

```
MySQL_install_path\bin\mysqld
```

όπου 'MySQL\_install\_path' ο κατάλογος εγκατάστασης της MySQL.

## 10) Ανάπτυξη του Web Service

Για να γίνει η ανάπτυξη του Web Service το οποίο χρησιμοποιεί η εφαρμογή VERBTAGGR, εκτελείτε τα ακόλουθα:

- Καλείτε μέσω οποιουδήποτε browser τη σελίδα:

```
"http://localhost:8080/manager/html"
```

- Στο παράθυρο που εμφανίζεται εισάγετε το όνομα χρήστη και τον κωδικό πρόσβασης που είχατε θέσει κατά την εγκατάσταση του εργαλείου JWSDP (βλ. εικόνα 23)
- Στην ενότητα 'WAR file to deploy' της σελίδας που παρουσιάζεται, πιέζετε το κουμπί 'Browse' και μέσω του παραθύρου που εμφανίζεται επιλέγετε το αρχείο web\_service \ war\_file\_to\_deploy\greekVerbRecWS.war που βρίσκεται στο cd.
- Πιέζετε το κουμπί 'Deploy' της ενότητας 'WAR file to deploy'.

Μετά την εκτέλεση των προαναφερθεισών οδηγιών, μπορείτε να χρησιμοποιήσετε την εφαρμογή VERBTAGGR++ από τον υπολογιστή στον οποίο έγινε η εγκατάσταση με χρήση ενός browser μέσω του οποίου θα καλέσετε τη σελίδα:

<http://localhost:8080/verbttaggr>

## 11) Τερματισμός του Tomcat και του server της MySQL



Μετά τη χρήση της εφαρμογής, τερματίζετε τον Tomcat και το server της MySQL:

- Σε ένα παράθυρο DOS Command Prompt πληκτρολογείτε:

*Tomcat\_install\_path*\bin\catalina.bat stop

όπου '*Tomcat\_install\_path*' ο κατάλογος εγκατάστασης του Tomcat.

(Εναλλακτικά, μπορείτε να επιλέξετε start → All Programs → ... → Java(TM) Web Services Developer Pack 1.5 → Stop Tomcat)

- Σε ένα παράθυρο DOS Command Prompt πληκτρολογείτε:

*MySQL\_install\_path*\bin\mysqladmin -u root -p shutdown

όπου '*MySQL\_install\_path*' ο κατάλογος εγκατάστασης της MySQL και εισάγετε τον κωδικό πρόσβασης του root που είχατε θέσει στο βήμα 5.

## 10. ΣΥΜΠΕΡΑΣΜΑΤΑ

Συμπερασματικά, στο πλαίσιο της παρούσας πτυχιακής εργασίας, μας δώθηκε η διπλωματική εργασία της Σταυρούλας Κρουστάλλη, “Διαδικτυακό Εργαλείο Στατιστικής Μορφοσυντακτικής Αναγνώρισης των Μονολεκτικών Ρηματικών Τύπων της Νέας Ελληνικής «*VERBTAGGR*»” [16], η οποία εκπονήθηκε το 2005 στο Τμήμα Πληροφορικής και Τηλεπικοινωνιών και η διδακτορική διατριβή της Πηνελόπης Λεμπέση “Στατιστική Μορφοσυντακτική Αναγνώριση και Λημματοποίηση του Ρηματικού Συνόλου της Νέας Ελληνικής” [1], που πραγματοποιήθηκε στο Πανεπιστήμιο Marc Bloch του Στρασβούργου το 2005, οι οποίες συνέβαλαν στη δημιουργία μίας πιλοτικής διαδικτυακής εφαρμογής «*VERBTAGGR*». Αρχικά είχαν συμπεριληφθεί μόνο τα δεδομένα που αφορούσαν στους τύπους που λήγουν σε άτονο ‘-α’ και ‘-ν’. Στο πλαίσιο της παρούσας εργασίας, αναβαθμίστηκε το εργαλείο σε «*VERBTAGGR++*», συμπεριλήφθηκαν και πλέον υποστηρίζονται όλοι οι μονολεκτικοί ρηματικοί τύποι της Νέας Ελληνικής. Επιπλέον, πραγματοποιήθηκαν κάποια πρώτα βήματα με σκοπό η παραπάνω εφαρμογή να υποστηρίζει στο μέλλον και τη λημματοποίηση των νεοελληνικών ρηματικών τύπων.

## ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ

<b>Ξενόγλωσσος όρος</b>	<b>Ελληνικός Όρος</b>
Morphosyntactic	Μορφοσυντακτικός
Computational Linguistics	Υπολογιστική Γλωσσολογία
Lemmatization	Λημματοποίηση
Web Service	Υπηρεσίες Διαδικτύου
Invoke	Κλήση
Bind	Σύνδεση

## ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ

VerbTagGr++	Verbal Tagging Greek ++
URL	Uniform Resource Locator
XML	Extensible Markup Language
IBM	International Business Machines
HTTP	Hypertext Transfer Protocol
UDDI	Universal Description, Discovery and Integration
SOAP	Simple Object Access Protocol
WSDL	Web Services Description Language
FTP	File Transfer Protocol
SMTP	Simple Mail Transfer Protocol
TCP/IP	Transmission Control Protocol/ Internet Protocol
W3C	World Wide Web Consortium
ΕΚΠΑ	Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

## ΑΝΑΦΟΡΕΣ

1. Πηνελόπη. Λεμπέση, “Στατιστική Μορφοσυντακτική Αναγνώριση και Λημματοποίηση του Ρηματικού Συνόλου της Νέας Ελληνικής”, Διδακτορική Διατριβή, Πανεπιστήμιο Marc Bloch, Στρασβούργο, 2005.
2. Χ. Κλαίρης, Γ. Μπαμπινιώτης, “Γραμματική της Νέας Ελληνικής, Δομολειτουργική – Επικοινωνιακή, ΙΙ. Το ρήμα”, Εκδόσεις Ελληνικά Γράμματα, 1999, pp. 2, 75-78.
3. D. Holton, P. Mackridge, E. Φιλιππάκη – Warburton, “Γραμματική της Ελληνικής Γλώσσας”, Εκδόσεις Πατάκη, 1999, pp. 111-114, 212-224, 234-235, 477-489.
4. P. Mackridge , “Η Νεοελληνική Γλώσσα”, Εκδόσεις Πατάκη, 1990, pp. 189-173.
5. P. Lembessi, "Statistical Retrograde Morphosyntactic Disambiguation and Linear Lemmatization of the Modern Greek Verbal Class", Selected Papers of the 16th International Symposium on Theoretical and Applied Linguistics, 2005, pp. 104-118.
6. Π. Λεμπέση, "Άνευ Λεξικού Αυτόματη Υπολογιστική Μορφοσυντακτική Ανάλυση και Λημματοποίηση του Ρηματικού Συνόλου της Νέας Ελληνικής", Πρακτικά 21ης Ετήσιας Συνάντησης Εργασίας του Τομέα Γλωσσολογίας της Φιλοσοφικής Σχολής του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, 2001, pp. 383-395.
7. P. Lembessi, "A Retrograde Morphemic Analysis of the Modern Greek Verbal Category and its Computational Implementation", Proceedings of the 2nd International Conference on Greek Linguistics, 1997, pp. 359-368.
8. D. Austin, A. Barbir, C. Ferris and S. Garg, “Web Services Architecture Requirements”, World Wide Web Consortium (W3C) Working Group note, February 2004; <http://www.w3.org/TR/wsa-reqs/>.
9. IBM Web Services Architecture team, “Web Services architecture overview”, September 2000; <http://www-128.ibm.com/developerworks/web/library/w-ovr/>.
10. Ed Ort, “Service-Oriented Architecture and Web Services: Concepts, Technologies and Tools”, April 2005; <http://java.sun.com/developer/technicalArticles/WebServices/soa2/index.html/>.
11. Microsoft Corporation, “Web Services and Other Distributed Technologies Developer Center”, 2005; <http://msdn.microsoft.com/webservices/>.
12. A. Tsalgatidou and T. Pilioura, “An Overview of Standards and Related Technology in Web Services”, International Journal of Distributed and Parallel Databases, Sept. 2002, vol. 12, pp. 135-162.
13. Attachmate Corporation, “Taking advantage of Web services”, November 2004; [http://download.attachmate.com/downloads/acrobat/english/02\\_0034\\_WebServices\\_Whitepaper.pdf](http://download.attachmate.com/downloads/acrobat/english/02_0034_WebServices_Whitepaper.pdf).
14. IBM Corporation , “IBM UDDI Registry”, 2005; <http://www-3.ibm.com/services/uddi/>.
15. Microsoft Corporation, “Microsoft UDDI Business Registry”, 2003; <http://uddi.microsoft.com/>.
16. Σταυρούλα Ι. Κρουστάλλη, Κρουστάλλη “Διαδικτυακό Εργαλείο Στατιστικής Μορφοσυντακτικής Αναγνώρισης των Μονολεκτικών Ρηματικών Τύπων της Νέας Ελληνικής «VERBTAGGR»”, Διπλωματική Εργασία, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, Σχολή Θετικών Επιστημών Τμήμα Πληροφορικής και Τηλεπικοινωνιών, Πρόγραμμα Μεταπτυχιακών Σπουδών, Αθήνα 2005.