Μεταπτυχιακή Διατριβή

Γεωργουλάκης Ευστράτιος
A.M 2008019003

Τμήμα Μηχανικών Παραγωγής και Διοίκησης
Τομέας Συστημάτων Παραγωγής

Αυτόματος Σημασιολογικός Εντοπισμός Γεγονότων σε Ακολουθίες Βίντεο βάσει Συμφραζομένων (Context)

Επιβλέπων
Δουλάμης Αναστάσιος

Χανιά, Δεκέμβριος 2011

Abstract

Η αυτοματοποίηση της διαδικασίας ανάλυσης ακολουθιών βίντεο κρίνεται πλέον ως ανάγκη επιτακτική δεδομένης της πληθώρας διαθέσιμου πολυμεσικού υλικού, αλλά και του φρενήρη ρυθμού παραγωγής νέου. Η εργασία αυτή επικεντρώνεται στην ανάπτυξη μιας τέτοιου είδους εφαρμογής στη γλώσσα προγραμματισμού Java, συνδυάζοντας δοκιμασμένες τεχνικές και εργαλεία όπως τα WordNet, Stanford POS-Tagger και JMF. Με τον τρόπο αυτό επιτυγχάνεται εξαγωγή των ιδιαίτερων χαρακτηριστικών του εκάστοτε βίντεο, βάσει της πολυεπίπεδης ανάλυσής του. Το αποτέλεσμα είναι η δημιουργία μιας επεκτάσιμης πλατφόρμας, η οποία στην τρέχουσα έκδοση επιτυγχάνει μια 2 σταδίων εκτίμηση της θέσης των γεγονότων που περιγράφονται στο υπό ανάλυση βίντεο.

Πρόλογος

Η ραγδαία ανάπτυξη της τεχνολογίας συνεπάγεται και την ραγδαία συσσώρευση ανεπεξέργαστου πληροφοριακού υλικού. Η ροή των πληροφοριών είναι καταιγιστική, με αποτέλεσμα την επιδερμική αφομοίωση και την αδυναμία χειρισμού και ειδικότερα ευρετηριασμού πολυμεσικού υλικού (multimedia), όπως αρχεία βίντεο, δίχως τη χρήση κατάλληλων εφαρμογών ανάλυσης και διαχείρισης.

Πολλές είναι οι τεχνολογίες που έχουν αναπτυχθεί με στόχο την επεξεργασία πολυμεσικού υλικού. Διάφορες τεχνικές ανάλυσης ήχου και εικόνας έχουν επινοηθεί με στόχο την απομόνωση ιδιαίτερων χαρακτηριστικών και οι οποίες μεμονωμένα έχουν χρησιμοποιηθεί σε πληθώρα εφαρμογών. Ένα εργαλείο συνδυασμού τέτοιων τεχνικών για την ανάλυση πολυμεσικού υλικού και συγκεκριμένα αρχείων βίντεο θα μπορούσε να δώσει ακόμη καλύτερα αποτελέσματα.

Τα αποτελέσματα μια τέτοιας ανάλυσης μπορούν να βρουν πλήθος εφαρμογών με μια από τις σημαντικότερες να αφορά στην ταξινόμηση και εύκολη αναζήτηση συγκεκριμένων πληροφοριών σε μεγάλο όγκο δεδομένων. Ταυτόχρονα γίνεται εφικτή η απομόνωση μέρους ενός βίντεο με τα επιθυμητά χαρακτηριστικά διευκολύνοντας ακόμη περισσότερο την εκάστοτε αναζήτηση. Οι μη αυτοματοποιημένες τεχνικές αναζήτησης επικεντρώνονται κυρίως σε περιγραφή η οποία προσδίδεται «χειροκίνητα» για ένα βίντεο μέσω λέξεων «κλειδιών» ή άλλων χαρακτηριστικών. Αυτό απαιτεί μεγάλη, συχνά απαγορευτικά, προσπάθεια από το χρήστη, ενώ αναγκαστικά περιορίζεται στα

υποκειμενικά χαρακτηριστικά που ο καθένας αναγνωρίζει και καταγράφει.

Σκοπός της εργασίας αυτής είναι η ανάπτυξη μιας εφαρμογής, η οποία θα αποτελέσει μια επεκτάσιμη πλατφόρμα για την επεξεργασία πολυμεσικού υλικού, κάνοντας χρήση συνδυασμού τεχνικών. Συγκεκριμένα, στην τρέχουσα έκδοση της εφαρμογής αυτής, ενορχηστρώνεται ο σημασιολογικός εντοπισμός γεγονότων ενός βίντεο, με την ανάπτυξη λειτουργιών ανάλυσης των πλαισίων βίντεό του αλλά και της αντίστοιχης λεκτικής του περιγραφής. Για την επίτευξη του παραπάνω στόχου, η εφαρμογή ενσωματώνει και χρησιμοποιεί συνδυαστικά ευρέως διαδεδομένα εργαλεία ανάλυσης πολυμεσικού υλικού (WordNet, Stanford POS-Tagger, JMF). Ταυτόχρονα αναπτύχθηκαν τεχνικές για την βελτιστοποίηση των επιδόσεων του συστήματος τόσο όσον αφορά την ταχύτητα απόκρισης αλλά και την ακρίβεια των αποτελεσμάτων.

Το ειδικά σχεδιασμένο γραφικό περιβάλλον παρέχει στον χρήστη εύκολη πρόσβαση στην λειτουργικότητα του συστήματος. Διάφορες ρυθμίσεις και αλλαγές είναι εφικτές κατά τη διάρκεια εκτέλεσης και δίνεται η δυνατότητα στο χρήστη να βλέπει άμεσα την εξέλιξη και το αποτέλεσμα ανάλυσης σύμφωνα με τις επιλογές του.

Πρωτοτυπία της παρούσας εργασίας είναι η χρήση της γλώσσας προγραμματισμού Java για την ανάπτυξη τόσο της λειτουργικότητας όσο και του γραφικού περιβάλλοντος διεπαφής χρήστη. Η χρήση οντοκεντρικής γλώσσας προγραμματισμού οδήγησε στην μοντελοποίηση των λειτουργιών με τρόπο άρτια δομημένο και εύκολα αναγνωρίσιμο. Για την σχεδίαση του κάθε επίπεδου του συστήματος τηρήθηκαν αυστηρά όροι αρθρωτής (modular) αρχιτεκτονικής, ενώ η υλοποίηση έγινε με γνώμονα τη μελλοντική επέκταση, είτε με την αντικατάσταση υπαρχόντων δομών, είτε με την ενσωμάτωση εντελώς νέων. Δεδομένου ότι η εφαρμογή «κληρονομεί» όλα τα χαρακτηριστικά της γλώσσας Java όσον αφορά την φορητότητα και τη συμβατότητα με κάθε πλατφόρμα, καθίσταται δυνατή η λειτουργία της σε πλήθος υπολογιστικών συστημάτων και με κάθε λογής λειτουργικό σύστημα.

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή

Η ανάπτυξη των τεχνολογιών αποθήκευσης και διάδοσης πληροφοριών στις μέρες μας, έχει επιφέρει μια ολοένα αυξανόμενη συσσώρευση «ακατέργαστου» πολυμεσικού υλικού. Με τον όρο ακατέργαστο εννοούμε υλικό το οποίο εμπεριέχει πληροφορίες οι οποίες δεν είναι εύκολα προσβάσιμες ή και αναγνωρίσιμες, όταν γίνεται η αναζήτηση και ανάκτησή τους. Τέτοιου είδους πολυμεσικό υλικό κυρίως αφορά σε ακολουθίες βίντεο τα οποία έχουν το χαρακτηριστικό της πληθώρας διαθέσιμων πληροφοριών σε σύντομο χρονικό διάστημα οι οποίες όμως δύσκολα ταξινομούνται, ευρετηριάζονται και ανακτώνται.

Η λύση στο πρόβλημα αυτό είναι η αντιπροσώπευση του εκάστοτε βίντεο από μια περιγραφή. Η περιγραφή αυτή θα πρέπει να εμπεριέχει τα κύρια χαρακτηριστικά του με τρόπο τέτοιο ώστε να είναι σύντομη αλλά και ταυτόχρονα πλήρης. Με τον τρόπο αυτό οποιαδήποτε αναζήτηση σε βίντεο ανάγεται σε αναζήτηση στην αντίστοιχή του περιγραφή.

Η παραδοσιακή, μη αυτοματοποιημένη μέθοδος περιγραφής των περιεχομένων ενός βίντεο αφήνει το χρήστη να προσδώσει ο ίδιος τα χαρακτηριστικά του βίντεο που παράγει. Μέθοδος η οποία αφενός απαιτεί πολύ μεγάλο κόστος σε χρόνο και προσπάθεια, αφετέρου η αποδοτικότητά της εξαρτάται από τον εκάστοτε χρήστη.

Για τους παραπάνω λόγους έχουν αναπτυχθεί τεχνικές και εργαλεία της αυτοματοποίησης όσον αφορά την περιγραφή ενός βίντεο. Τα εργαλεία αυτά εξασφαλίζουν μια μαζική, γρήγορη και συνεπή ανάλυση πληροφοριακού υλικού δίνοντας τη δυνατότητα της αυτόματης εξαγωγής των ιδιαίτερων χαρακτηριστικών. Τα χαρακτηριστικά αυτά μπορούν στη συνέχεια να χρησιμοποιηθούν στην ταξινόμηση και τον ευρετηριασμό των πληροφοριών και του υλικού που τις εμπεριέχει. Με τον τρόπο αυτό οι διάφορες πληροφορίες γίνονται εύκολα προσβάσιμες από τον εκάστοτε ενδιαφερόμενο, δίχως αυτές να χάνονται στον τεράστιο όγκο υλικού, ελαχιστοποιώντας ταυτόχρονα τον χρόνο αναζήτησής τους.

1.1 Οργάνωση μεταπτυχιακής διατριβής

Η παρούσα διατριβή στο Κεφάλαιο 1 ξεκινάει με την εισαγωγή στο πρόβλημα της συσσώρευσης πολυμεσικού υλικού και της αναγκαιότητας της ύπαρξης τεχνικών περιγραφής τους. Γίνεται αναφορά στις τεχνικές περιγραφής για το σύνολο των πληροφοριών που εμπεριέχει ένα βίντεο και τέλος παρουσιάζεται η εφαρμογή που υλοποιήθηκε και τα ιδιαίτερά της χαρακτηριστικά.

Στο κεφάλαιο 2 γίνεται μια επισκόπηση της βιβλιογραφίας σχετικής με την ανάλυση βίντεο και αντίστοιχων εφαρμογών. Οι εφαρμογές αυτές κάνουν μια παρόμοια με την εν λόγω εργασία προσέγγιση στο θέμα της ανάλυσης και πολυεπίπεδης επεξεργασίας αρχείων βίντεο.

Το πρώτο επίπεδο της ανάλυσης του βίντεο αφορά στην ανάλυση του κειμένου όπως προκύπτει από την επεξεργασία του ακουστικού υλικού και τον σχηματισμό προτάσεων. Ο τρόπος, η δομή και τα χαρακτηριστικά της ανάλυσης του σταδίου αυτού περιγράφονται στο Κεφάλαιο 3.

Για την εν λόγω εφαρμογή, το δεύτερο επίπεδο ανάλυσης αφορά στην περαιτέρω επεξεργασία των αποτελεσμάτων του πρώτου σταδίου, οι διεργασίες αυτές περιγράφονται στο Κεφάλαιο 4. Ειδικότερα, επιστρατεύεται ο Stanford POS-Tagger και γίνεται εφικτός η αναγνώριση τι μέρος του λόγου είναι οι λέξεις κάθε πρότασης.

Τα πρώτα αποτελέσματα της εκτίμησης για τη θέση αλλαγής ενός γεγονότος ή θέματος γίνονται μετά την τρίτου επιπέδου ανάλυση και σύγκριση της συνάφειας των προτάσεων, με την βοήθεια του WordNet. Στο κεφάλαιο 5 περιγράφονται όλα τα χαρακτηριστικά της διαδικασίας αυτής, των τεχνικών και αλγορίθμων που υλοποιούνται όπως και των αποτελεσμάτων που απορρέουν.

Μια δεύτερη εκτίμηση για την ενδεχόμενη αλλαγή γεγονότος ή θεματολογίας γίνεται στο τέταρτο στάδιο ανάλυσης το οποίο αφορά στην ανάλυση των πλαισίων βίντεο. Ανάλυση η οποία περιγράφεται στο Κεφάλαιο 6 και περιλαμβάνει τους ιδιαίτερους χειρισμούς της ροής των πλαισίων βίντεο αλλά και τεχνικές ανάλυσης εικόνας με την χρήση των λειτουργιών της πλατφόρμας JMF.

Το Κεφάλαιο 7 αφιερώνεται στην διεπαφή χρήστη, η οποία υλοποιείται για την εν λόγω εφαρμογή. Γίνεται ιδιαίτερη μνεία στις λογής λειτουργίες που προσφέρει και δίνεται πλήρης περιγραφή της εσωτερικής δομής και της αρθρωτής αρχιτεκτονικής σχεδιασμού της.

Το όλο εγχείρημα της πολυεπίπεδης ανάλυσης στην παρούσα εργασία περιλαμβάνει πλήθος εργαλείων λογισμικού και τεχνικών. Όλα μαζί σχεδιάστηκαν συντονίστηκαν και υλοποιήθηκαν με χρήση των κανόνων της γλώσσας προγραμματισμού Java. Το σύνολο του απαιτούμενου λογισμικού και ο τρόπος ενορχήστρωσής τους περιγράφονται στο Κεφάλαιο 8.

Τα συμπεράσματα της όλης διαδικασίας έρευνας, μελέτης, σχεδιασμού και υλοποίησης περιλαμβάνονται στο Κεφάλαιο 9, ενώ στη συνέχεια ακολουθεί παράτημα με λεπτομέρειες αναφορικά με την εσωτερική δομή του συστήματος αλλά και ειδικών αλγορίθμων που αναπτύχθηκαν.

1.2 Χρησιμότητα Περιγραφής βίντεο

Η μέγιστη αξιοποίηση πολυμεσικού υλικού καθιστά απαραίτητη την περιγραφή του περιεχομένου του. Συγκεκριμένα όσον αφορά σε ακολουθίες βίντεο, οι πληροφορίες που μπορεί να εμπεριέχονται ενδέχεται να είναι πολλές και ποικίλες και ο ευρετηριασμός τους καθίσταται μέγιστης σημασίας για την αποδοτική εύρεση και ανάκτησή τους.

Για παράδειγμα, ένα βίντεο ειδήσεων χρονικής διάρκειας 30 λεπτών εμπεριέχει πληροφορίες για πλήθος θεμάτων επικαιρότητας. Ένα πολύ σημαντικό στοιχείο εξαρχής είναι ο τόπος και ο χρόνος παραγωγής του εκάστοτε βίντεο. Όταν αναζητούνται πληροφορίες για ένα γεγονός σε συγκεκριμένο χρόνο και τόπο, τότε τα παραπάνω στοιχεία ίσως να είναι αρκετά δίχως την ανάγκη περαιτέρω ανάλυσης του βίντεο. Η πλειοψηφία όμως των αναζητήσεων αφορούν σε πληθώρα χαρακτηριστικών τα οποία εμπεριέχονται στο βίντεο, όπως ονόματα, ημερομηνίες κ.α. Για το λόγο αυτό μια περιγραφή του τι λέγεται και σε τι αφορά κάθε βίντεο κρίνεται απαραίτητη. Επίσης στο βίντεο ειδήσεων του παραδείγματος τα θέματα που θίγονται είναι κατά κανόνα πολλά όπως και οι κατηγορίες πληροφοριών, πολιτικά, αθλητικά, καιρός κτλ. Μια συνεπής περιγραφή θα πρέπει να εμπεριέχει όλα τα παραπάνω, την κατηγορία δηλαδή του θέματος και το τι αυτό αφορά. Μεγάλης χρησιμότητας όσον αφορά στην περιγραφή των πληροφοριών ενός

βίντεο κρίνονται και οι χρονικές επισημάνσεις. Σε ένα βίντεο μεγάλης χρονικής διάρκειας ενδέχεται να δίνεται στην περιγραφή ότι μια πληροφορία υπάρχει, χωρίς όμως χρονικές επισημάνσεις, ο χρήστης θα πρέπει να την αναζητήσει σε όλη τη διάρκεια του βίντεο.

Οι μη αυτοματοποιημένες τεχνικές περιγραφής αφορούν στην περιγραφή που ο ίδιος ο παραγωγός προσδίδει στο εκάστοτε βίντεο. Τέτοιες τεχνικές επικεντρώνονται κυρίως στην επίδοση λέξεων «κλειδιών» ή άλλων χαρακτηριστικών του βίντεο δίχως χρονικές επισημάνσεις. Για να είναι μια τέτοια περιγραφή επαρκής απαιτείται μεγάλη προσπάθεια, συχνά απαγορευτική για μεγάλης διάρκειας βίντεο, ενώ αναγκαστικά περιορίζεται στα υποκειμενικά χαρακτηριστικά που ο καθένας αναγνωρίζει και καταγράφει.

Για τους παραπάνω λόγους έχουν επινοηθεί τεχνικές και εργαλεία για την αυτόματη επεξεργασία και εξαγωγή πληροφοριών και χρονικών επισημάνσεων. Με τον τρόπο αυτό διευκολύνεται η ανάλυση του πληροφοριακού υλικού και γίνεται εφικτός ο ευρετηριασμός και η ανάκτηση πληροφοριών.

1.3 Τεχνικές περιγραφής βίντεο

Στην προηγούμενη ενότητα περιγράφηκε η χρησιμότητα των διάφορων τεχνικών ανάλυσης πολυμεσικού υλικού και το ποια είναι η ανάγκη στην αυτόματη εξαγωγή ιδιαίτερων χαρακτηριστικών που καλούνται να καλύψουν. Επικεντρώνοντας στις τεχνικές που αφορούν την ανάλυση ακολουθιών βίντεο και δεδομένου ότι ένα βίντεο εμπεριέχει ταυτόχρονα ήχο και εικόνα, υπάρχουν αντίστοιχες τεχνικές για την πολυεπίπεδη ανάλυσή του.

1.3.1 Τεχνικές ανάλυσης ήχου

Οι τεχνικές αυτές εμπίπτουν με τεχνικές ανάλυσης και αναγνώρισης ομιλίας και στόχο έχουν την καταγραφή των λέξεων που ακούγονται στο βίντεο σε κείμενο (λεκτική περιγραφή-transcript). Τέτοιες τεχνικές αφορούν στο πρώτο επίπεδο ανάλυσης του ακουστικού υλικού και το έργο που καλούνται να φέρουν εις πέρας είναι από τα δυσκολότερα δεδομένης της εξάρτησης της ποιότητας του προς επεξεργασία υλικού από διάφορους παράγοντες. Τέτοιοι παράγοντες αφορούν στα διαφορετικά τεχνικά μέσα καταγραφής αλλά κυρίως στην

διαφορετικότητα των φυσικών γλωσσών[1,2] και της άμεσης εξάρτησης με τον εκάστοτε εκφωνητή.

Στην εν λόγω εργασία, το υπό εξέταση dataset περιλαμβάνει τόσο τα βίντεο προς ανάλυση όσο και τις αντίστοιχες λεκτικές τους περιγραφές. Έχοντας έτοιμη την λεκτική περιγραφή το σύστημα προχωρά απευθείας στην ανάλυση του επόμενου επιπέδου που αφορά στην επεξεργασία της λεκτικής περιγραφής. Παρόλα αυτά η αρχιτεκτονική του συστήματος δίνει την δυνατότητα της μελλοντικής ενσωμάτωσης εργαλείου αναγνώρισης ήχου (Audio Recognizer) για την παραγωγή από το ίδιο το εργαλείο της λεκτικής περιγραφής.

1.3.2 Τεχνικές ανάλυσης λεκτικής περιγραφής

Το επόμενο στάδιο αφορά στην ανάλυση της λεκτικής περιγραφής (transcript). Η λεκτική περιγραφή εκτός από τις λέξεις που ακούγονται στο βίντεο είναι δυνατό να εμπεριέχει και πολύ σημαντικές πληροφορίες για το πότε ειπώθηκε η κάθε λέξη και το πότε υπάρχουν παύσεις στην εκφώνηση. Η χρησιμότητα των πληροφοριών αυτών έγκειται στην δυνατότητα εξαγωγής χρονικών επισημάνσεων αλλά και σε μια αρχική ομαδοποίηση των λέξεων. Για την συγκεκριμένη εφαρμογή έχει αναπτυχθεί ειδικά σχεδιασμένος αναλυτής (parser) ο οποίος σαρώνει την εκάστοτε λεκτική περιγραφή και εξάγει τις παραπάνω πληροφορίες αποθηκεύοντάς τις σε ειδικές δομές του συστήματος.

Ο τρόπος χειρισμού των παραπάνω πληροφοριών για την εξαγωγή των επιθυμητών συμπερασμάτων ποικίλει από εφαρμογή σε εφαρμογή. Για την συγκεκριμένη εφαρμογή χρησιμοποιούνται 2 εργαλεία. Το πρώτο αφορά στην αναγνώριση τι μέρος του λόγου είναι καθεμίας από τις λέξεις που ακούγονται στο βίντεο, τέτοια εργαλεία ονομάζονται Part-Of-Speech Taggers. Το δεύτερο αφορά στην εύρεση της συνάφειας μεταξύ των προτάσεων του βίντεο. Ουσιαστικά χρησιμοποιείται μια βάση δεδομένων λέξεων και συσχετίσεων μεταξύ τους η οποία είναι ικανή να δώσει μια μετρική συνάφειας, ανάλογα με την εκάστοτε χρησιμοποιούμενη τεχνική σύγκρισης.

1.3.3 Τεχνικές ανάλυσης εικόνας

Ένα από τα σημαντικότερα στάδια ανάλυσης ενός βίντεο είναι η ανάλυση των πλαισίων βίντεο από τα οποία συνίσταται (video frames)[3]. Καθένα από τα πλαίσια αυτά στην ουσία είναι μια εικόνα από την οποία μπορούν να εξαχθούν λογής χαρακτηριστικά. Στο παραπάνω παράδειγμα ενός βίντεο ειδήσεων μια εικόνα ενδέχεται να εμπεριέχει τον τίτλο του θέματος με τις λέξεις κλειδιά, αλλά και άλλα στοιχεία όπως η θέση και το πλήθος των εκφωνητών μπορεί να αποτελούν χρήσιμες πληροφορίες.

Εκτός από την αυτόνομη εξέταση ενός πλαισίου βίντεο, μεγάλης σπουδαιότητας πληροφορίες εξάγονται από την ανάλυση της ροής των πλαισίων σε ένα βίντεο. Τέτοιες πληροφορίες αφορούν στον εντοπισμό κίνησης, αλλαγής σκηνής, επεισοδίων κ.α. Για την εν λόγω εφαρμογή, Στα πλαίσια της εργασίας έχει αναπτυχθεί και ενσωματωθεί ειδικός μηχανισμός ανάλυσης των πλαισίων βίντεο. Ο μηχανισμός αυτό υλοποιεί ειδικό αλγόριθμο, ο οποίος εξετάζει την ροή των πλαισίων βίντεο και εξάγει πληροφορίες για το πότε έχουμε αλλαγή στο σκηνικό σύμφωνα με ένα προκαθορισμένο όριο αλλά και «ευαισθησία».

1.4 Η αναπτυχθείσα εφαρμογή

Σκοπός της εργασίας αυτής είναι η ανάπτυξη μιας εφαρμογής η οποία θα αποτελέσει μια επεκτάσιμη πλατφόρμα για την επεξεργασία πολυμεσικού υλικού, κάνοντας χρήση διάφορων τεχνικών. Η αναγκαιότητα της επεκτασιμότητας μια τέτοιας εφαρμογής έγκειται στην ολοένα βελτιστοποιούμενη αποδοτικότητα των υπαρχόντων τεχνικών και εργαλείων για την επιμέρους ανάλυση των στοιχείων ενός βίντεο, αλλά και στην δημιουργία καινούριων.

Στην τρέχουσα έκδοση της εφαρμογής, ενορχηστρώνεται ο σημασιολογικός εντοπισμός γεγονότων ενός βίντεο, με την ανάπτυξη λειτουργιών ανάλυσης τόσο των πλαισίων βίντεο όσο και της αντίστοιχης λεκτικής περιγραφής. Γίνεται δηλαδή μια εκτίμηση 2 σταδίων για το ποια είναι τα σημεία του βίντεο όπου έχουμε αλλαγή γεγονότος ή θέματος γενικότερα.

Για την επίτευξη του παραπάνω στόχου, η εφαρμογή υλοποιεί μια πολυεπίπεδη ανάλυση ενσωματώνοντας και χρησιμοποιώντας συνδυαστικά ευρέως διαδεδομένα εργαλεία ανάλυσης πολυμεσικού

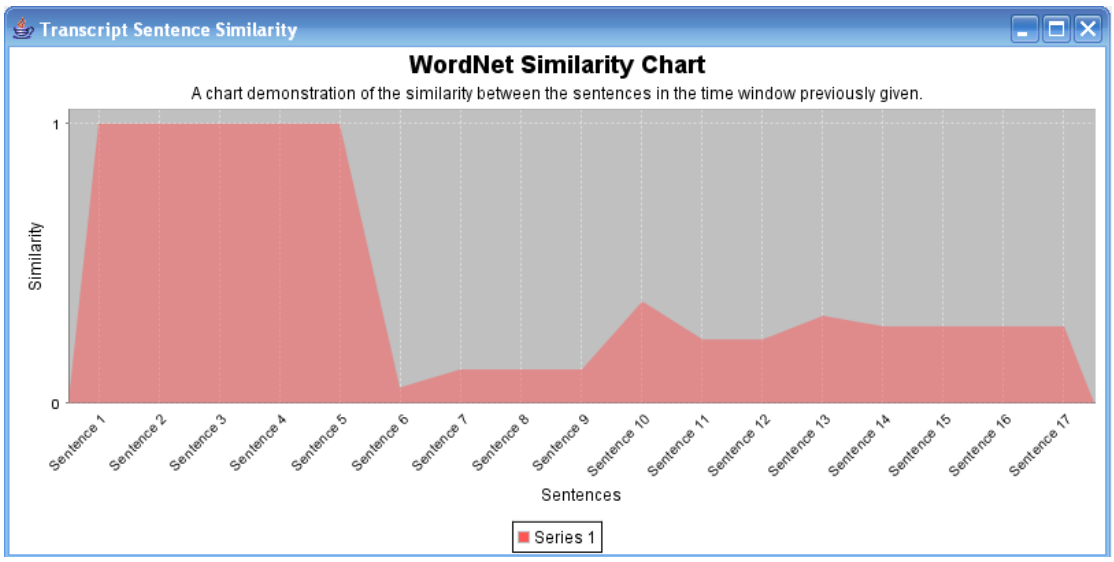
υλικού (Stanford POS-Tagger, WordNet, JMF). Ταυτόχρονα αναπτύχθηκαν τεχνικές για την βελτιστοποίηση των επιδόσεων του συστήματος τόσο όσον αφορά την ταχύτητα απόκρισης αλλά και την ακρίβεια των αποτελεσμάτων.

Για την συγκεκριμένη εφαρμογή η συλλογή των βίντεο που χρησιμοποιείται ως dataset (TRECVID'03) περιέχει βίντεο με ειδήσεις από το CNN και BBC με τις αντίστοιχες λεκτικές τους περιγραφές (Transcripts). Για τη λειτουργία του εργαλείου και τον εντοπισμού θεμάτων σε βίντεο, και την πρώτου επιπέδου ανάλυση απαραίτητη είναι η λεκτική περιγραφή του βίντεο (transcript), το αποτέλεσμα δηλαδή της επεξεργασίας του βίντεο μέσω ενός Audio Recognizer.

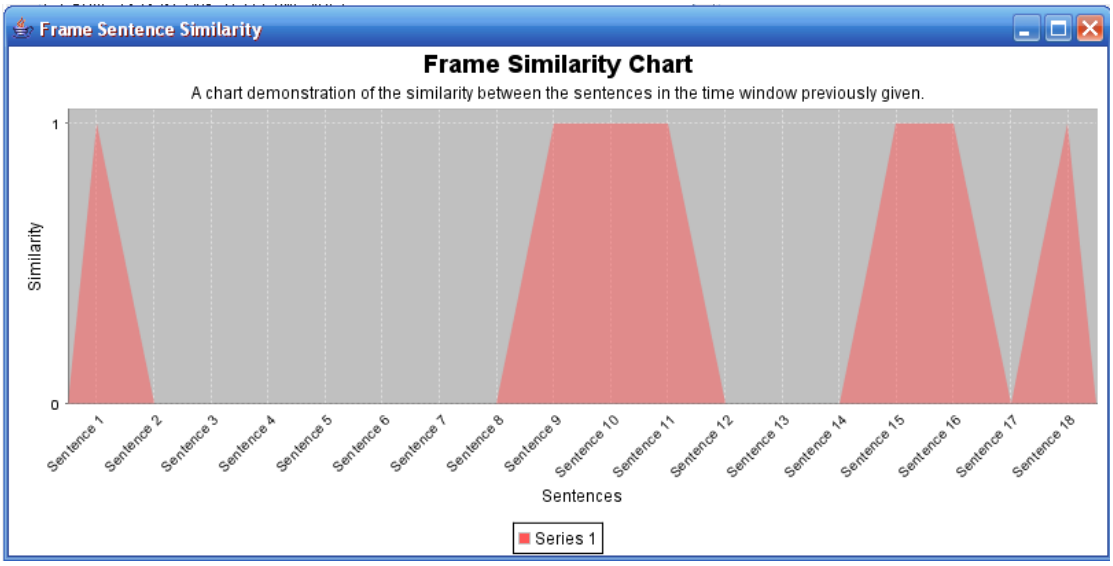
Στο πρώτο επίπεδο ανάλυσης ειδικά σχεδιασμένος αναλυτής (parser) σαρώνει την λεκτική περιγραφή και εξάγει πληροφορίες που αφορούν στο τι ειπώθηκε και πότε αλλά και πότε έχουμε παύσεις και ποιας διάρκειας. Οι παύσεις αυτές δίνουν ένα πρώτο στοιχείο της ομαδοποίησης των λέξεων σε προτάσεις. Επίσης χαρακτηριστικά μεγάλες παύσεις στην εκφώνηση μπορεί να υποδηλώνουν αλλαγή θέματος. Όλες οι παραπάνω πληροφορίες αποθηκεύονται σε ειδικές δομές του συστήματος, ώστε να είναι προσβάσιμες στα επόμενα επίπεδα ανάλυσης.

Στο δεύτερο επίπεδο ανάλυσης γίνεται χρήση του Stanford POS-Tagger. Με τη βοήθεια του παραπάνω εργαλείου οι προτάσεις όπως έχουν σχηματιστεί από το προηγούμενο επίπεδο ανάλυσης επεξεργάζονται για την εύρεση των μερών του λόγου των λέξεων που τις αποτελούν. Με τον τρόπο αυτό, το σύστημα είναι σε θέση να απομονώσει από τις λέξεις αυτές τα ουσιαστικά. Τα ουσιαστικά εμπεριέχουν όλη την ουσία, το νόημα δηλαδή των προτάσεων και η ανάλυση του επόμενου επιπέδου θα επικεντρωθεί σε αυτά.

Με διαθέσιμα τα ουσιαστικά και ομαδοποιημένα ανά πρόταση, στο τρίτο επίπεδο ανάλυσης η εφαρμογή ενσωματώνει το WordNet. Το εργαλείο αυτό είναι σε θέση να δώσει μια μετρική συνάφειας μεταξύ λέξεων και ειδικότερα ουσιαστικών. Με την χρήση του παραπάνω εργαλείου μέσω συγκεκριμένης τεχνικής καταλήγουμε στο να έχουμε μια εκτίμηση της συνάφειας των προτάσεων της λεκτικής περιγραφής ανά 2. Η εκτίμηση αυτή, ως αποτέλεσμα της ανάλυσης των τριών πρώτων επίπεδων, αποτελεί την πρώτου σταδίου εκτίμηση του εντοπισμού γεγονότων και αποτυπώνεται από την εφαρμογή τόσο σε ειδικό πλαίσιο κειμένου, όσο και σε ειδικό διάγραμμα.

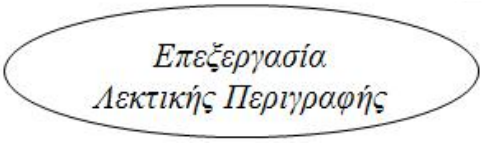


Το τέταρτο και τελευταίο επίπεδο ανάλυσης αφορά στην ανάλυση του βίντεο καθεαυτού και συγκεκριμένα επιλεγμένων πλαισίων βίντεο. Για την επίτευξη της διαχείρισης της ροής των πλαισίων βίντεο αλλά και την επιμέρους ανάλυση καθενός από αυτά, το σύστημα ενσωματώνει την πλατφόρμα Java Media Framework (JMF). Κάνοντας χρήση των δυνατοτήτων της παραπάνω πλατφόρμας στην εν λόγω εργασία αναπτύχθηκαν τεχνικές επιλογής ορισμένων από τα πλαίσια βίντεο, των «κρίσιμων» όσον αφορά την αλλαγή γεγονότος ή θεματολογίας. Συγκεκριμένα, απομονώνονται τα πλαίσια τα οποία βρίσκονται μεταξύ των προτάσεων, κατά τις παύσεις δηλαδή εκφώνησης, μιας και μια πιθανή αλλαγή δεν μπορεί να λαμβάνει χώρα κατά τη διάρκεια της εκφώνησης μιας πρότασης. Στη συνέχεια τα «κρίσιμα» αυτά πλαίσια εξετάζονται ανά δύο για τον εντοπισμό της αλλαγής του σκηνικού, σύμφωνα με ένα ορισμένο όριο και με προσαρμοζόμενη ευαισθησία. Μεταξύ των προτάσεων των οποίων παρατηρείται αλλαγή του σκηνικού πέραν το ορίου θεωρούμε ότι έχουμε μια πιθανή αλλαγή γεγονότος ή θέματος. Τα αποτελέσματα της παραπάνω ανάλυσης οδηγούν στην δεύτερη σταδίου εκτίμηση του εντοπισμού γεγονότων η οποία επίσης αποτυπώνεται σε ειδικό διάγραμμα.

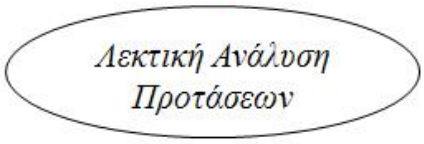


Ένα από τα χαρακτηριστικά της εφαρμογής είναι η δυνατότητα ειδικών ρυθμίσεων, οι οποίες προσφέρονται στον χρήστη ακόμη και κατά την διάρκεια εκτέλεσης (tunability). Τέτοιες ρυθμίσεις αφορούν σε καθοριστικές λειτουργίες του συστήματος όπως το όριο αλλαγής προτάσεων, το όριο και η ευαισθησία της σύγκρισης πλαισίων βίντεο, αλλά και του συντονισμού χρονικά του βίντεο με την λεκτική του περιγραφή. Στην εν λόγω υλοποίηση η εφαρμογή κάνει χρήση συγκεκριμένων εργαλείων με ειδικά ανεπτυγμένες τεχνικές και στρατηγικές. Παρόλα αυτά, δεδομένου ότι η γλώσσα ανάπτυξης του εργαλείου τόσο σε επίπεδο λειτουργικότητας όσο και σε επίπεδο διεπαφής χρήστη είναι η Java, έχει υιοθετηθεί μια modular αρχιτεκτονική η οποία επιτρέπει ρυθμίσεις, βελτιστοποιήσεις, αλλαγές ακόμη και επεκτάσεις σε καθένα από τα επίπεδα ανάλυσης όπως περιγράφηκαν παραπάνω.

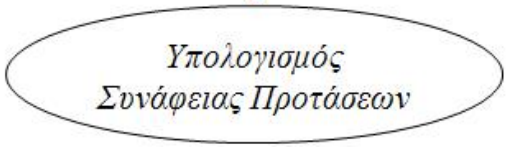
Τα επίπεδα του υλοποίησης συστήματος είναι πέντε, τα τέσσερα πρώτα συμπίπτουν με τα 4 επίπεδα ανάλυσης, όπως ήδη περιγράφηκαν. Το πέμπτο επίπεδο αφορά στο γραφικό περιβάλλον διαχείρισης που προσφέρεται στον χρήστη και δίνει την δυνατότητα ενορχήστρωσης της όλης διαδικασίας ανάλυσης. Τα επίπεδα και έχουν ως εξής:



1. Σάρωση λεκτικής περιγραφής και σχηματισμός προτάσεων



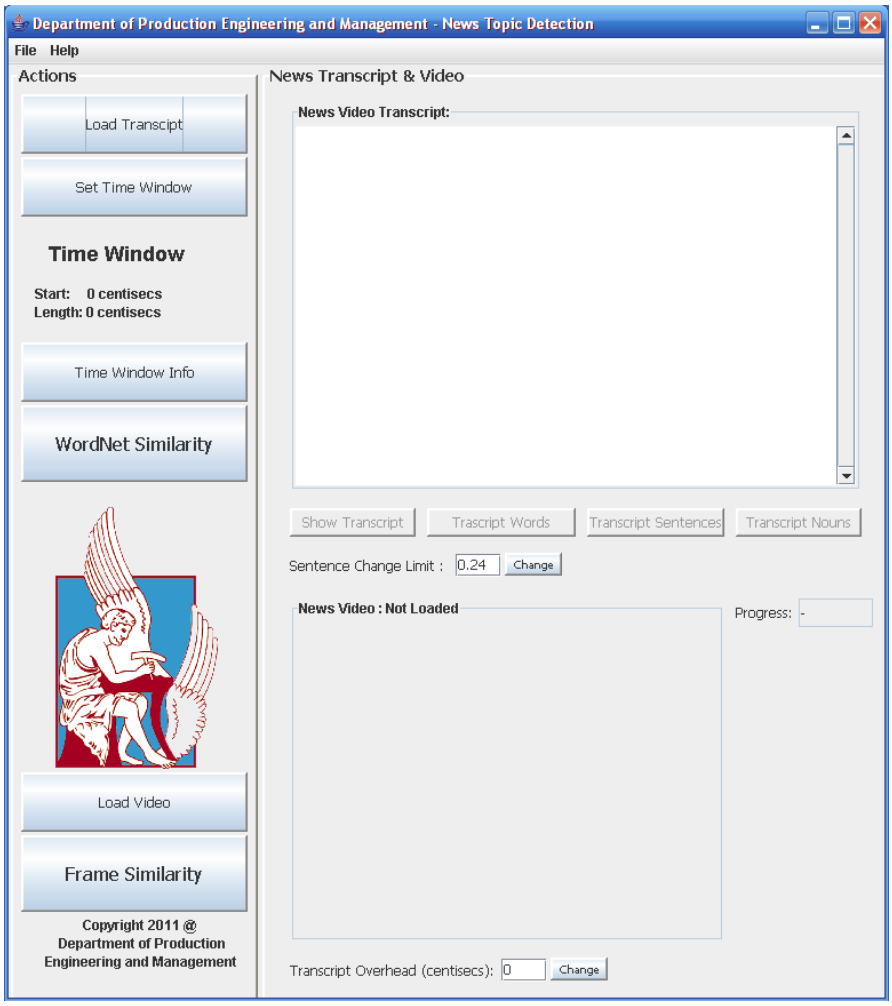
2. Εντοπισμός μερών του λόγου λέξεων (POS-Tagger)



3. Υπολογισμός Συνάφειας προτάσεων ανά 2 (WordNet)

4. Υπολογισμός ομοιότητας Πλαισίων Βίντεο (Java Media Framework)

5. Γραφικό Περιβάλλον Διεπαφής Χρήστη



Κεφάλαιο 2

Επισκόπηση βιβλιογραφίας

Η ολοένα αυξανόμενη χρήση και ζήτηση οπτικοακουστικού υλικού σε συνδυασμό με την ανάπτυξη των τεχνολογιών αναπαραγωγής και αποθήκευσής του, έχει ως συνέπεια την ανάγκη ανάπτυξης ιδιαίτερων τεχνολογιών αναφορικά με την αναπαράσταση, μοντελοποίηση, ευρετηριασμό και αναπαράσταση πολυμεσικού υλικού[4]. Συγκεκριμένα, είναι επιτακτική η ανάγκη ύπαρξης αποδοτικών τεχνικών για τον ευρετηριασμό, ανάκτηση και συμπίεση πληροφοριών όσον αφορά τα ιδιαίτερα χαρακτηριστικά του εκάστοτε υλικού, όπως και αξιόπιστους αλγόριθμους αναζήτησης, οι οποίοι επιτρέπουν την πρόσβαση σε βάσεις δεδομένων πολυμεσικού υλικού.

Ο ευρετηριασμός και ανάκτηση πολυμεσικού υλικού βάσει περιεχομένου περιλαμβάνει τέσσερις βασικές διεργασίες [5,6,7], την ανάλυση περιεχομένου (video content analysis), την ανάλυση της δομής (video structure parsing), την σύνοψη (summarization) και τον ευρετηριασμό (indexing).

2.1 Video content analysis

Το βασικό πρόβλημα στην ανάλυση του περιεχομένου σε ένα βίντεο είναι η δυσκολία αντιστοίχισης οπτικών χαρακτηριστικών τα οποία μπορούν να εξαχθούν, όπως χρώμα, σχήμα, μέγεθος, κίνηση κτλ, με συγκεκριμένες έννοιες όπως εξωτερικός ή εσωτερικός χώρος, άνθρωπος, σκηνή χορού κτλ. Αν και το οπτικό περιεχόμενο είναι η βασική πηγή πληροφοριών κατά την ανάλυση ενός βίντεο, χρήσιμες πληροφορίες μπορούν να εξαχθούν και από άλλες πηγές, όπως κείμενο που εμφανίζεται στο video (πχ τίτλοι ειδήσεων) και ο ήχος και ομιλίες που το συνοδεύουν. Ο συνδυασμός της ανάλυσης των παραπάνω κρίνεται ιδιαίτερα αποδοτικός σε κάθε είδους εφαρμογές, οι οποίες μάλιστα επεκτείνονται σε όλους τους τομείς της καθημερινότητας όπως εκπαίδευση, τουρισμό, επιστήμες, περιβάλλον κ.α.. Παράδειγμα τέτοιων εφαρμογών είναι το ‘Informedia’ [8] ή το ‘AT&T_s Pictorial Transcripts’ [9]

2.1.1 Η εφαρμογή 'Informedia'

Στόχος της εφαρμογής είναι να επιτύχει την κατανόηση οπτικοακουστικού υλικού και σε συνδυασμό με λειτουργίες αναζήτησης, ανάκτησης, οπτικοποίησης και σύνοψης του περιεχομένου, να δημιουργήσει κατάλληλες συλλογές ευρετηρίων[10,11].

Οι τεχνολογίες που αναπτύχθηκαν κατά την κατασκευή της εφαρμογής συνδυάζουν την κατανόηση ομιλίας, εικόνας και φυσικών γλωσσών για το αυτοματοποιημένο κατακερματισμό και ευρετηριασμό σε βίντεο με στόχο την υψηλού επιπέδου αναζήτηση και ανάκτηση εικόνας. Βασικό χαρακτηριστικό της εφαρμογής αποτελεί η χρήση αναλυτή ομιλίας, ο οποίος ανεξάρτητα με τον εκφωνητή, καταφέρνει μεταφράζοντας το ηχητικό μέρος ενός βίντεο να κατασκευάσει ένα σύστημα εξαγωγής πληροφοριών κειμένου. Επίσης χαρακτηριστικό της εφαρμογής αυτής αποτελεί η επιτάχυνση της παρουσίασης του εκάστοτε βίντεο με χρήση ειδικής τεχνικής η οποία κάνει εφικτή την προεπισκόπηση των κύριων σημείων σε διάρκεια της τάξης του 5 με 20% της αρχικής.

2.1.2 Η εφαρμογή 'AT&T_s Pictorial Transcripts'

Η συγκεκριμένη εφαρμογή αφορά στην αυτοματοποιημένη αρχειοθέτηση και την επιλεκτική ανάκτηση πληροφοριών κειμένου, εικόνας και ήχου που εμπεριέχονται σε αρχεία βίντεο. Η επεξεργασία του βίντεο αφορά στην αναπαράσταση των οπτικών χαρακτηριστικών με χρήση ενός υποσυνόλου των πλαισίων βίντεο. Η γλωσσική επεξεργασία αφορά στην ανάλυση του περιλαμβανόμενου κειμένου, δημιουργεί πίνακα περιεχομένων και συσχετίσεων με άλλο πολυμεσικό υλικό. Πληροφορίες ήχου και βίντεο συμπιέζονται και ευρετηριάζονται βάσει της χρονικής τους συσχέτισης με τα επιλεχθέντα πλαίσια βίντεο και κείμενο. Τα χαρακτηριστικά που απορρέουν χρησιμοποιούνται για την αυτόματη παραγωγή υπερκείμενης αναπαράστασης των περιεχομένων του προγράμματος. Έτσι παρέχεται μια πλήρης αναπαράσταση των πληροφοριών που εμπεριέχονται στο εκάστοτε βίντεο.

Η εφαρμογή επίσης χρησιμεύει και για τον ευρετηριασμό μέσω κειμένου και εικόνας του πλήρους βίντεο. Το σύστημα, πλήρως αυτοματοποιημένα, δημιουργεί μια αναπαράσταση των προγραμμάτων

της τηλεόρασης σε HyperText Markup Language (HTML), κάνοντας τα διαθέσιμα μέσω διαδικτύου δευτερόλεπτα μετά την εκπομπή τους.

2.2 Video structure parsing

Σημαντικό βήμα κατά την ανάλυση της δομής του βίντεο αφορά στον διαχωρισμό του βίντεο σε μεμονωμένες σκηνές (scene). Μια σκηνή αποτελείται από μια αλληλουχία λήψεων βίντεο (shot) τα οποία ομαδοποιούνται λόγω της θέσης ή του θεματικού τους περιεχομένου. Ο εντοπισμός των παραπάνω σκηνών βίντεο είναι διαδικασία αντίστοιχη με τον εντοπισμό παραγράφων κατά την ανάλυση κειμένου, με τη διαφορά ότι απαιτεί υψηλότερου επιπέδου ανάλυση περιεχομένου. Αντίστοιχα με τις λέξεις ή προτάσεις ενός κειμένου, οι λήψεις είναι καλή επιλογή βασικής μονάδας αναφορικά με τον ευρετηριασμό ενός βίντεο καθώς μπορούν να αποτελέσουν βάση για την κατασκευή ενός πίνακα περιεχομένων του εκάστοτε βίντεο. Οι αλγόριθμοι εντοπισμού των ορίων των λήψεων που βασίζονται στις οπτικές πληροφορίες που περιέχουν τα πλαίσια βίντεο (video frames) ομαδοποιούν πλαίσια με παρόμοια οπτικά χαρακτηριστικά [5]. Η ομαδοποίηση των λήψεων σε μέρη σύμφωνα με το εννοιολογικό τους περιεχόμενο σε γεγονότα ή θέματα δεν είναι ωστόσο συχνά εφικτή χωρίς να ληφθούν υπόψη πληροφορίες και από άλλα χαρακτηριστικά του βίντεο πέραν τον οπτικών. Η πολυεπίπεδη ανάλυση με αλγόριθμους που περιλαμβάνουν την ανάλυση όχι μόνο των πλαισίων βίντεο αλλά και κείμενο, ήχο και ομιλία, έχει αποδειχθεί αποδοτική στην εκπλήρωση του παραπάνω στόχου [7].

2.3 Video summarization

Η διαδικασία της σύνοψης (summarization) ενός βίντεο αφορά στην σύντομη παρουσίαση οπτικών πληροφοριών αναφορικά με την δομή του βίντεο[12]. Αυτή η αφαιρετική διαδικασία είναι παρόμοια με την εξαγωγή λέξεων κλειδιών για την δημιουργία περίληψης κατά την ανάλυση κειμένου. Συγκεκριμένα, γίνεται εξαγωγή ενός υποσυνόλου από τα δεδομένα του βίντεο όπως πλαίσια βίντεο «κλειδιά», ή εισαγωγικά πλαίσια σε λήψεις, σκηνές κτλ. Ο τρόπος της επιλογής των πληροφοριών αυτών είναι πολύ σημαντικός δεδομένου του τεράστιου όγκου πληροφοριών που εμπεριέχονται ακόμη και σε ολιγόλεπτα βίντεο. Το αποτέλεσμα της διαδικασίας αυτής αποτελεί την βάση όχι μόνο της αναπαράστασης βίντεο βάσει περιεχομένου, αλλά και της

βάσει περιεχομένου αναζήτησης. Συνδυάζοντας τις πληροφορίες για την δομή που εξάγονται κατά την ανάλυση του βίντεο και τα πλαίσια «κλειδιά» όπως επιλέγονται από την παραπάνω διαδικασία, γίνεται δυνατή η κατασκευή ενός οπτικού πίνακα περιεχομένων για ένα βίντεο.

2.4 Video indexing

Τα χαρακτηριστικά δομής και περιεχομένου που βρέθηκαν κατά την ανάλυση του περιεχομένου, την ανάλυση όλων των στοιχείων του βίντεο και της αφαιρετικής διαδικασίας της σύνοψης αλλά και τα χαρακτηριστικά που προσδίδονται «χειρωνακτικά» από το χρήστη ονομάζονται μεταδεδομένα. Βάσει των χαρακτηριστικών αυτών, είναι εφικτή η κατασκευή ευρετηρίου και ενός πίνακα περιεχομένων μέσω, για παράδειγμα, μιας διαδικασίας ομαδοποίησης των πλαισίων βίντεο σε διαφορετικές οπτικές κατηγορίες ή δομές ευρετηριασμού. Όπως και σε αντίστοιχα συστήματα ευρετηρίων, είναι απαραίτητη η υιοθέτηση συγκεκριμένων σχημάτων και εργαλείων για την κατάλληλη χρήση των ευρετηρίων και των μεταδεδομένων του περιεχομένου. Τέτοια εργαλεία δίνουν την δυνατότητα δημιουργίας επερωτήσεων, της αναζήτησης και την περιήγηση σε βάσεις δεδομένων βίντεο. Σε ερευνητικό επίπεδο έχει αναπτυχθεί πλήθος σχημάτων και εργαλείων για τον ευρετηριασμό και την δημιουργία επερωτήσεων. Παρόλα αυτά, αξιόπιστα και αποδοτικά εργαλεία, πειραματικά δοκιμασμένα σε μεγάλα σύνολα δεδομένων δεν έχουν ακόμη καθιερωθεί.

Κεφάλαιο 3

Επεξεργασία Λεκτικής Περιγραφής (Transcript)

Η πρώτη διεργασία που καλείται το σύστημα να επιτελέσει αφορά στην επεξεργασία και ανάλυση της λεκτικής περιγραφής του βίντεο, στο εξής transcript. Το transcript στην ουσία είναι το παράγωγο ενός audio ή speech recognizer στον οποίο δίνεται ως είσοδος το προς επεξεργασία βίντεο. Περιλαμβάνει πληροφορίες για τις λέξεις που ακούγονται στο βίντεο: πότε ακούγονται και με ποια διάρκεια όπως και πότε έχουμε παύσεις κατά την εκφώνηση. Κατά την ανάπτυξη του συστήματος επιλέχθηκε ένα συγκεκριμένο dataset (TRECVID'03) το οποίο περιλαμβάνει τα προς ανάλυση βίντεο και τα αντίστοιχα transcripts. Οι πληροφορίες που περιέχει το transcript θα αξιοποιηθούν για να γίνει μια εκτίμηση της ομαδοποίησης των λέξεων σε προτάσεις όπως και για τον συγχρονισμό της ανάλυσης του transcript με την ανάλυση των πλαισίων του βίντεο (video frames).

3.1 Δομή Λεκτικής Περιγραφής

Ένα ενδεικτικό μέρος ενός transcript έχει ως εξής:

```
<DOCSET type=ASRTEXT fileid=19980501_1830_1900_ABC_WNT
collect_date=19980501_1830 collect_src=ABC src_lang=ENGLISH
content_lang=NATIVE proc_remarks="Byblos English ASR">
<X Bsec=0.00 Dur=33.20 Conf=NA>
<W recid=1 Bsec=33.20 Dur=0.39 Clust=NA Conf=NA> HELPFUL
<W recid=2 Bsec=33.61 Dur=0.11 Clust=NA Conf=NA> AND
<W recid=3 Bsec=33.72 Dur=0.10 Clust=NA Conf=NA> IT
<W recid=4 Bsec=33.82 Dur=0.61 Clust=NA Conf=NA> STARS
<W recid=5 Bsec=34.43 Dur=0.25 Clust=NA Conf=NA> IN
<W recid=6 Bsec=34.72 Dur=0.09 Clust=NA Conf=NA> THE
<W recid=7 Bsec=34.81 Dur=0.26 Clust=NA Conf=NA> WHITE
<W recid=8 Bsec=35.07 Dur=0.35 Clust=NA Conf=NA> HOUSE
<W recid=9 Bsec=35.42 Dur=0.09 Clust=NA Conf=NA> AND
<W recid=10 Bsec=35.51 Dur=0.35 Clust=NA Conf=NA> VERY
<W recid=11 Bsec=35.89 Dur=0.36 Clust=NA Conf=NA> POINTED
<W recid=12 Bsec=36.25 Dur=0.59 Clust=NA Conf=NA> MESSAGE
<X Bsec=36.84 Dur=0.32 Conf=NA>
<W recid=13 Bsec=37.16 Dur=0.34 Clust=NA Conf=NA> ABOUT
<W recid=14 Bsec=37.50 Dur=0.28 Clust=NA Conf=NA> HIS
<W recid=15 Bsec=37.78 Dur=0.84 Clust=NA Conf=NA>INVESTIGATION
<W recid=16 Bsec=38.65 Dur=0.60 Clust=NA Conf=NA> AND
<W recid=17 Bsec=39.25 Dur=0.04 Clust=NA Conf=NA> I
<W recid=18 Bsec=39.31 Dur=0.51 Clust=NA Conf=NA> ASSUME
<W recid=19 Bsec=39.82 Dur=0.13 Clust=NA Conf=NA> WHEN
```

```
<W recid=20 Bsec=39.95 Dur=0.37 Clust=NA Conf=NA> YOU
<W recid=21 Bsec=40.35 Dur=0.18 Clust=NA Conf=NA> GO
<X Bsec=40.53 Dur=1.87 Conf=NA>
<W recid=22 Bsec=42.40 Dur=0.61 Clust=NA Conf=NA> YEAH
<X Bsec=43.01 Dur=0.29 Conf=NA>
…
<W recid=4324 Bsec=1828.27 Dur=0.50 Clust=NA Conf=NA> JACKPOT
<W recid=4325 Bsec=1828.79 Dur=0.19 Clust=NA Conf=NA> WILL
<W recid=4326 Bsec=1828.98 Dur=0.19 Clust=NA Conf=NA> BE
<W recid=4327 Bsec=1829.17 Dur=0.39 Clust=NA Conf=NA> HEAVILY
</DOCSET>
```

Η πρώτη γραμμή του transcript ξεκινά με το '`<DOCSET`' και περιέχει πληροφορίες που αφορούν σε ιδιαίτερα χαρακτηριστικά όπως ο κωδικός που προσδίδεται στο transcript (file_id), η ημερομηνία και ώρα του βίντεο της συλλογής (collect_date), το όνομα της συλλογής (collect_src), η γλώσσα περιγραφής (src_lang) και ο τύπος της (content_lang) .

Η επόμενη γραμμή ξεκινά το χαρακτηριστικό '`<X`' το οποίο υποδηλώνει παύση εκφώνησης. Η γραμμή αυτή περιέχει την πολύ σημαντική πληροφορία του πόση ώρα διαρκεί η εισαγωγή του βίντεο δίχως να έχει εκφωνηθεί κάποια λέξη. Το να προηγείται της εκφώνησης του εκάστοτε εκφωνητή μια εισαγωγή με τίτλους και μουσική, αποτελεί μια πολύ συνήθη πρακτική στα βίντεο. Στο συγκεκριμένο transcript παρατηρούμε ότι η εισαγωγή ξεκινά τη χρονική στιγμή 0 ('`Bsec=0.00`') και έχει διάρκεια 33,20 δευτερόλεπτα ('`Dur=33.20`'). Η σπουδαιότητα της πληροφορίας αυτής έγκειται στο συγχρονισμό της ανάλυσης του transcript με την ανάλυση των πλαισίων του βίντεο (video frames) όπως θα περιγραφεί παρακάτω.

Οι επόμενες γραμμές του transcript αφορούν στο κυρίως μέρος του, όπου περιγράφονται τα λεχθέντα στο βίντεο. Συγκεκριμένα, αν η εκάστοτε γραμμή ξεκινάει με '`<W`' τότε αφορά στην περιγραφή μιας λέξης. Η περιγραφή αυτή περιέχει έναν κωδικό για την λέξη ('`recid`'), την χρονική στιγμή που ξεκίνησε να λέγεται ('`Bsec`') και το πόσο διήρκησε ('`Dur`'). Στην συνέχεια, στο τέλος της γραμμής, ακολουθεί η λέξη αυτή καθαυτή. Στην περίπτωση που η γραμμή ξεκινάμε '`<X`' τότε έχουμε παύση εκφώνησης και περιγράφεται η χρονική στιγμή που ξεκίνησε η παύση και πόσο διήρκησε, όπως περιγράφηκε παραπάνω για την παύση κατά την εισαγωγή του βίντεο.

Οι λέξεις του βίντεο όπως αναπαριστώνται στο transcript παρατηρείται ότι δεν έχουν ακριβώς την συνοχή και συνάφεια όπως ακούγονται στο βίντεο. Αυτό οφείλεται στα λάθη και τις ανακρίβειες κατά την ανάλυση

του βίντεο από τον Ηχητικό Αναλυτή (Audio ή Speech Recognizer) ο οποίος είναι υπεύθυνος για την εξαγωγή του transcript. Τέτοια λάθη είναι ως ένα βαθμό αναπόφευκτα δεδομένης της πολυπλοκότητας των φυσικών γλωσσών και του ότι υπάρχει άμεση εξάρτηση με το πόσο καθαρά εκφράζεται ο εκάστοτε εκφωνητής, την μουσική υπόκρουση, την ποιότητα του ήχου κτλ.

Στο σημείο αυτό είναι σημαντικό να τονισθεί η σπουδαιότητα της ακρίβειας των δεδομένων του transcript. Ενδεχόμενα λάθη στο επίπεδο αυτό αναπόφευκτα οδηγούν σε λάθη στα επόμενα επίπεδα ανάλυσης και επεξεργασίας αλλοιώνοντας την ακρίβεια των υπολογισμών και των συμπερασμάτων τους. Η αναγκαιότητα της modular αρχιτεκτονικής του συστήματος η οποία τηρείται, έχει αρχίσει ήδη να διαφαίνεται μιας και μια μελλοντική πιο ποιοτική εκδοχή της λεκτικής περιγραφής μπορεί να υποστηριχθεί με μηδενικές ή ελάχιστες αλλαγές στο σύστημα όπως θα περιγραφεί στην επόμενη υποενότητα.

3.2 Ανάλυση Λεκτικής Περιγραφής

Για την ανάλυση του transcript έπρεπε να ληφθεί υπόψη ότι δεν υπάρχει κάποια τυποποιημένη μορφή (format) με τον κάθε Audio Recognizer να προσδίδει τα δικά του ιδιαίτερα χαρακτηριστικά στο αρχείο που εξάγει. Για το λόγο αυτό, με γνώμονα την modular αρχιτεκτονική του συστήματος, αναπτύχθηκε η ειδική συνάρτηση ‘***parseTranscript()***’.

Η συνάρτηση αυτή, σύμφωνα με τις ιδιαιτερότητες του transcript καταφέρνει και εξάγει πληροφορίες που αφορούν όχι μόνο στις λέξεις που ειπώθηκαν αλλά επίσης σε ποιες χρονικές στιγμές και με ποια διάρκεια. Επιπρόσθετα, εκτός του τι ειπώθηκε και πότε εξάγονται πληροφορίες για τις παύσεις στην εκφώνηση, τόσο στην αρχή αλλά και κατά την διάρκεια του βίντεο. Όλες οι παραπάνω πληροφορίες αποθηκεύονται σε κατάλληλες δομές του συστήματος ώστε να είναι προσβάσιμες ανά πάσα στιγμή για την ανάλυση των επόμενων επιπέδων.

Για την επίτευξη των παραπάνω, η συνάρτηση ‘*parseTranscript()*’ κάνει χρήση των ιδιαίτερων χαρακτηριστικών την δομής του transcript. Συγκεκριμένα, όπως αναλύθηκε στην ενότητα 3.1, η δομή του transcript ακολουθεί ένα συγκεκριμένο πρότυπο (pattern). Στην πρώτη γραμμή έχουμε διάφορες πληροφορίες για την συγκεκριμένη λεκτική ανάλυση.

Στις επόμενες αν η γραμμή ξεκινάει με '<W' έχουμε την περιγραφή μιας λέξης με την λέξη αυτή καθαυτή στο τέλος της γραμμής, ενώ αν ξεκινάει με '<X' τότε έχουμε την περιγραφή μιας παύσης. Τέλος, οι γραμμές στο transcript ακολουθούν χρονική αλληλουχία και με τον τρόπο τέτοιο ώστε να εξετάζεται το βίντεο καθ' όλη τη διάρκεια του δίχως κενά, ανά πάσα στιγμή είτε εκφωνείται μια λέξη είτε πρόκειται για παύση.

Για παράδειγμα, απομονώνοντας μερικές γραμμές από το παραπάνω απόσπασμα λεκτικής περιγραφής έχουμε:

```
<W recid=11 Bsec=35.89 Dur=0.36 Clust=NA Conf=NA> POINTED
<W recid=12 Bsec=36.25 Dur=0.59 Clust=NA Conf=NA> MESSAGE
<X Bsec=36.84 Dur=0.32 Conf=NA>
<W recid=13 Bsec=37.16 Dur=0.34 Clust=NA Conf=NA> ABOUT
```

Οι πληροφορίες που μπορούμε να αποσπάσουμε είναι:

- η ενδέκατη λέξη που ακούγεται στο βίντεο είναι η 'POINTED' (recid=11)
- η χρονική στιγμή που ξεκίνησε να λέγεται η λέξη 'POINTED' είναι η 35,89 δευτερόλεπτα(Bsec=35.89)
- η χρονική διάρκεια που ακούγονταν η λέξη 'POINTED' είναι 0,36 δευτερόλεπτα (Dur=0.36)
- αμέσως μετά, δίχως παύση ειπώθηκε η λέξη 'MESSAGE' και έχουμε τις αντίστοιχες πληροφορίες και για τη λέξη αυτή
- αμέσως μετά τη λέξη 'MESSAGE', τη χρονική στιγμή 36,84 δευτερόλεπτα (Bsec=36.84) ξεκίνησε μια παύση στην εκφώνηση η οποία διήρκησε 0,32 δευτ. (Dur=0.32)
- αμέσως μετά την παύση ειπώθηκε η λέξη 'ABOUT'
- κτλ

Επίσης παρατηρούμε την χρονική αλληλουχία των γραμμών δίχως κενά καθ' όλη τη διάρκεια του βίντεο όπως περιγράφηκε παραπάνω αφού έχουμε Bsec=35.89 + Dur=0.36 = Bsec=36.25 και Bsec=36.25 + Dur=0.59 = Bsec=36.84 κτλ.

3.3 Εντοπισμός Προτάσεων

Η ανάλυση του transcript, όπως περιγράφηκε παραπάνω, είναι το πρώτο βήμα του εγχειρήματος Εντοπισμού Θεμάτων και ίσως το πιο καθοριστικό. Οι πληροφορίες που εξάγονται θα χρησιμοποιηθούν σε καθένα από τα επόμενα επίπεδα του συστήματος και καθορίζουν και

την ακρίβεια των αποτελεσμάτων του συστήματος. Συγκεκριμένα από τις παύσεις που περιγράφονται γίνεται μια πρώτη εκτίμηση για τον διαχωρισμό των λεγόμενων σε προτάσεις. Για να γίνει αλλαγή από μια πρόταση σε μια άλλη ο εκάστοτε εκφωνητής κάνει μια παύση. Για το χρησιμοποιηθέν dataset έχει υπολογιστεί ένας μέσος όρος των 0,24 secs, τιμή η οποία χρησιμοποιείται αρχικά από το σύστημα για την εναλλαγή αυτή, ενώ ταυτόχρονα δίνεται στο χρήστη η δυνατότητα να επέμβει αλλάζοντάς τον.

Η μεταβλητή του συστήματος, η οποία αφορά στο παραπάνω όριο αλλαγής πρότασης, ονομάζεται '***Sentence_change_limit***'. Το σύστημα χρησιμοποιεί ως default την τιμή 0,24 secs ενώ το γραφικό περιβάλλον δίνει την δυνατότητα αλλαγής κατά τη διάρκεια εκτέλεσης.

Η τεχνική του να χρησιμοποιούνται οι παύσεις για τον καθορισμό των προτάσεων κρίνεται αρκετά αποδοτική δεδομένου ότι δεν είναι δυνατό να υφίσταται αλλαγή πρότασης σε φυσική γλώσσα, δίχως κάποιας χρονικής διάρκειας παύση. Το κλειδί στον επιτυχή διαχωρισμό είναι η εύρεση του κατάλληλου ορίου, ένα πολύ μικρό όριο θα δίνει περισσότερες προτάσεις από όσες είναι στην πραγματικότητα, ενώ ένα μεγάλο όριο θα κάνει συγχώνευση προτάσεων. Ο καθορισμός ενός τέτοιου ορίου δεν θα είναι υπόθεση απλώς μαθηματικών υπολογισμών και μέσων τιμών δεδομένου ότι ο κάθε εκφωνητής έχει τον δικό του τρόπο ομιλίας και έκφρασης. Για το λόγο αυτό η δυνατότητα αλλαγής τιμής κατά τη διάρκεια εκτέλεσης, για τη συγκεκριμένη μεταβλητή κρίθηκε απαραίτητη.

3.4 Συγχρονισμός Λεκτικής Περιγραφής με βίντεο

Σε προηγούμενη ενότητα αναφέρθηκε ότι κατά την εκκίνηση ενός βίντεο και πριν την εκφώνηση λέξεων, συχνά ακούγεται μουσική με την ταυτόχρονη εμφάνιση διάφορων λογότυπων. Ο Audio Recognizer το διάστημα αυτό το εκφράζει στο transcript ως παύση του εκφωνητή. Το διάστημα αυτό σε αρκετά βίντεο παρατηρείται να ορίζεται ως μεγαλύτερο στο transcript από ότι στην πραγματικότητα, με αποτέλεσμα η λεκτική περιγραφή να είναι πιο «μπροστά» σε σχέση με το βίντεο.

Το παραπάνω φαινόμενο μη συγχρονισμού στην περίπτωση που δεν αντιμετωπιστεί θα δημιουργήσει πρόβλημα σε επόμενο επίπεδο όπου θα πρέπει να συγκριθούν τα αποτελέσματα της ανάλυσης του transcript

με την ανάλυση του βίντεο. Πρόκειται για διαφορά της τάξης των μερικών εκατοστών του δευτερολέπτου, παρόλα αυτά κρίνεται καθοριστική για την ακρίβεια των αποτελεσμάτων σε επόμενα βήματα ανάλυσης από το σύστημα.

Η μεταβλητή του συστήματος η οποία είναι υπεύθυνη για τον συγχρονισμό transcript και βίντεο ονομάζεται '***Overhead***'. Η προκαθορισμένη τιμή που χρησιμοποιείται είναι η '0' εκατοστά του δευτερολέπτου (centisecs) και το γραφικό περιβάλλον δίνει την δυνατότητα αλλαγής της κατά τη διάρκεια εκτέλεσης. Για παράδειγμα για τιμή της μεταβλητής '*Overhead*' στο 16 εννοείται ότι το transcript είναι «μπροστά» σε σχέση με το βίντεο κατά 16 εκατοστά του δευτερολέπτου. Στην περίπτωση που το transcript είναι «πίσω» η μεταβλητή '*Overhead*' θα πρέπει να πάρει αρνητική τιμή.

Η επιλογή του αριθμού 16 για το παραπάνω παράδειγμα δεν είναι τυχαία. Στα βίντεο του dataset παρατηρείται το transcript να είναι μπροστά κατά 16-20 εκατοστά του δευτερολέπτου. Η μεταβλητή '*Overhead*' καλείται να γεφυρώσει τη διαφορά αυτή κατά την ανάλυση των πλαισίων του βίντεο (video frames). Συγκεκριμένα, σύμφωνα με τις πληροφορίες που εξάγουμε για τις προτάσεις από το transcript, σε επόμενο επίπεδο το σύστημα θα κληθεί να εξετάσει ομοιότητες σε πλαίσια βίντεο σε χρονικά διαστήματα που καθορίζονται από το transcript. Μετακυλώντας τα χρονικά διαστήματα αυτά προς τα αριστερά, σύμφωνα πάντα με την τιμή της μεταβλητής '*Overhead*', καταφέρνουμε το επιθυμητό αποτέλεσμα συγχρονισμού. Περισσότερες λεπτομέρειες δίνονται παρακάτω στην ενότητα περιγραφής της ανάλυσης των πλαισίων βίντεο.

Για παράδειγμα στο απόσπασμα της λεκτικής περιγραφής που παρατίθεται παραπάνω παρατηρούμε ότι η λέξη '`POINTED`' κατά το transcript ξεκινάει την χρονική στιγμή 35,89 δευτερόλεπτα, ενώ στο αντίστοιχο βίντεο (19980501_ABC.mpg) η ίδια λέξη ακούγεται την χρονική στιγμή 35,73 δευτερόλεπτα. Έχουμε μια διαφορά της τάξης των 16 εκατοστών του δευτερολέπτου και για το συγχρονισμό κατά τη διάρκεια της ανάλυσης θα πρέπει να έχουμε '*Overhead=16*'.

Κεφάλαιο 4

Λεκτική Ανάλυση Προτάσεων (POS-Tagging)

Από την παραπάνω πρώτου επιπέδου ανάλυση της λεκτικής περιγραφής (transcript) έχουμε καταλήξει σε μια εκτίμηση για το ποια είναι η ομαδοποίηση των λέξεων σε προτάσεις. Το σύστημα σε δεύτερο επίπεδο χρησιμοποιεί τις προτάσεις αυτές για μια περαιτέρω ανάλυση, η οποία θα οδηγήσει στην αναγνώριση τι μέρος του λόγου είναι οι λέξεις κάθε πρότασης. Για να γίνει αυτό εφικτό, το σύστημα κάνει χρήση ενός ειδικού λογισμικού το οποίο αναγνωρίζει και σηματοδοτεί τι μέρος του λόγου είναι η κάθε λέξη που δέχεται ως είσοδο (POS-tagger). Τα αποτελέσματα της ανάλυσης του επιπέδου αυτού θα χρησιμοποιηθούν για την ανάλυση του τρίτου και τελευταίου επιπέδου, όπως περιγράφεται σε επόμενη ενότητα, με την ακρίβεια τους να παίζει καθοριστικό ρόλο στην ορθότητα των τελικών αποτελεσμάτων.

4.1 Stanford POS-Tagger

Το σύστημα στο δεύτερο επίπεδο ανάλυσης κάνει χρήση ενός POS-tagger [13,14,15] για την αναγνώριση τι μέρος του λόγου είναι κάθε λέξη του transcript.

Ένας Part-Of-Speech Tagger (POS Tagger) είναι ειδικό λογισμικό το οποίο διαβάζει κείμενο σε μια συγκεκριμένη γλώσσα και καθορίζει για κάθε λέξη του κειμένου τι μέρος του λόγου είναι, όπως ουσιαστικό, ρήμα, επίρρημα κτλ Το χρησιμοποιούμενο από σύστημα λογισμικό αποτελεί μια υλοποίηση σε Java του Stanford POS Tagger, στο εξής tagger. Ο συγκεκριμένος tagger διατελεί σε 3 διαφορετικές λειτουργίες, σε λειτουργία αναγνώρισης των μερών του λόγου (tagging), σε λειτουργία εκπαίδευσης (training) και σε λειτουργία δοκιμής των επιδόσεών του. Ενώ η κύρια λειτουργία του tagger είναι η πρώτη, παρόλα αυτά οι υπόλοιπες 2 είναι πολύ σημαντικές όσον αφορά τις επιδόσεις του tagger.

Κατά τη λειτουργία εκπαίδευσης ο χρήστης μπορεί να εισάγει στον tagger δεδομένα εκπαίδευσής του και να πάρει ως έξοδο ένα εκπαιδευμένο μοντέλο (trained model). Σύμφωνα με το μοντέλο αυτό μπορεί να καθοδηγήσει στο εξής τον tagger ώστε να λειτουργεί και να καθορίζει τις εκάστοτε λέξεις με τον τρόπο που θέλουμε εμείς. Ένα

τέτοιο μοντέλο είναι πάντα απαραίτητο ώστε αυτός να γνωρίζει τους κανόνες με τους οποίους αναγνωρίζει τις εκάστοτε λέξεις. Το σύστημα για της ανάγκες της εν λόγω εφαρμογής χρησιμοποιεί ένα από τα προκαθορισμένα για την Αγγλική γλώσσα μοντέλα που είναι διαθέσιμα μαζί με το λογισμικό του tagger.

Κατά τη λειτουργία δοκιμής επιδόσεων ο tagger δίνει τη δυνατότητα στο χρήστη να αξιολογήσει τις επιδόσεις του με το εκάστοτε μοντέλο το οποίο έχουμε ορίσει να ακολουθεί. Για τα προκαθορισμένα μοντέλα Αγγλικής γλώσσας έχουμε:

1. Μοντέλο : bidirectional-distsim-wsj-0-18.tagger Ακρίβεια 97.32%
2. Μοντέλο : left3words-wsj-0-18.tagger Ακρίβεια 96.92%

Το μοντέλο που χρησιμοποιεί το σύστημα είναι το 'bidirectional-distsim-wsj-0-18.tagger' λόγω της μεγαλύτερης ακρίβειας αν και το μοντέλο 'left3words-wsj-0-18.tagger' επιτυγχάνει καλύτερες ταχύτητες.

4.2 Βέλτιστη Λειτουργία POS - Tagger

Δοκιμάζοντας ενδελεχώς τις επιδόσεις και την ακρίβεια των αποτελεσμάτων του παραπάνω tagger, παρατηρείται ότι το βέλτιστο αποτέλεσμα αφορά στην περίπτωση που δοθεί στον tagger κείμενο με σωστή σύνταξη[16]. Οι επιδόσεις πέφτουν δραματικά στην περίπτωση που δοθούν άναρχες λέξεις, ενώ μια ομαδοποίηση σε προτάσεις δίνει σαφώς καλύτερα αποτελέσματα. Αυτό συμβαίνει επειδή ο Stanford POS-Tagger υλοποιείται βασιζόμενος στην γραμματική της εκάστοτε γλώσσας, ώστε να αναγνωρίζει τα κύρια σημεία του κειμένου προς ανάλυση. Στις περιπτώσεις που η σύνταξη δεν είναι καλή ο tagger αδυνατεί να διαχωρίσει κυρίως τα ρήματα και έχει την τάση να τα ορίζει ως ουσιαστικά.

Πειραματικά παρατηρούμε ότι

Είσοδος	***Έξοδος***	***Επίδοση***
John is working	John/NNP is/VBZ working/VBG	Καλή
working	working/VBG	Καλή

Playing	playing/NN	Κακή
She ran to the station quickly	She/PRP ran/VBD to/TO the/DT station/NN quickly/RB	Καλή
talk and jump	talk/NN and/CC jump/NN	Κακή (λανθασμένη σύνταξη)

4.3 Καθορισμός μερών του λόγου (POS-tagging)

Για την καλύτερη διαχείριση του tagger με γνώμονα την modular αρχιτεκτονική του συστήματος και την δυνατότητα μελλοντικής βελτιστοποίησης ή επέκτασης, έχει υλοποιηθεί αποκλειστικά μια συνάρτηση όσον αφορά τον καθορισμό των μερών του λόγου των λέξεων του transcript, η '***tagSentences()***'.

Η συνάρτηση παίρνει ως όρισμα τις προτάσεις προς επεξεργασία και επιστρέφει τις προτάσεις αυτές με ενσωματωμένα διακριτικά [17] για κάθε λέξη ανάλογα με το τι μέρος του λόγου είναι η καθεμία.

Για παράδειγμα η πρόταση:

This sentence demonstrates the POS_Tagger capabilities

θα επιστραφεί ως:

This/DT sentence/NN demonstrates/VBZ the/DT POS_Tagger/NNP capabilities/NNS

Όπου
DT : singular determiner/quantifier (this, that)
NN : singular or mass noun
VBZ : verb, 3rd. singular present
NNS : plural noun
NNP : Proper noun, singular

Για την καλύτερη λειτουργία του tagger η συνάρτηση tagSentences() στην τρέχουσα έκδοση καλεί τον tagger ανά πρόταση. Η συγκεκριμένη τεχνική κατά τη διάρκεια των δοκιμών, όπως περιγράφεται στην προηγούμενη ενότητα, δίνει τα καλύτερα αποτελέσματα. Στο σημείο αυτό φαίνεται η σπουδαιότητα του ορθού διαχωρισμού των προτάσεων

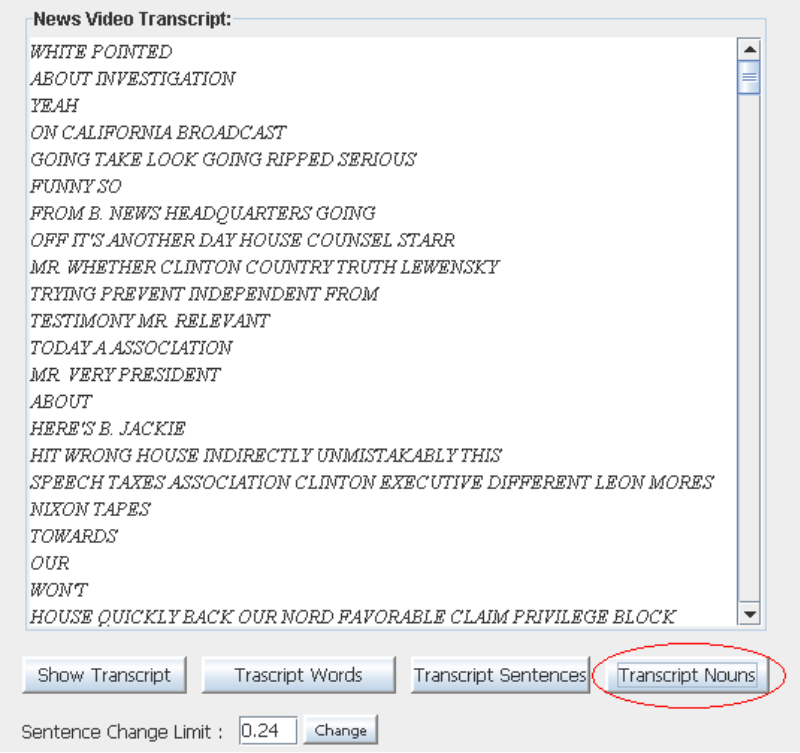
στο πρώτο στάδιο, η αρτιότητα του οποίου εξασφαλίζει την βέλτιστη αποτελεσματικότητα του tagger. Η αποτελεσματικότητα του με τη σειρά της θα κρίνει την απόδοση των λειτουργιών του επόμενου σταδίου.

4.4 Απόδοση ανάλυσης μερών του λόγου

Παραπάνω εξετάστηκε ο τρόπος της βέλτιστης χρησιμοποίησης του tagger μέσω σωστής σύνταξης του προς ανάλυση κειμένου. Ταυτόχρονα παρατέθηκαν οι προσπάθειες προς την κατεύθυνση αυτή μέσω του όσο το δυνατό ορθολογικού διαχωρισμού σε προτάσεις, στην πρώτου σταδίου ανάλυση της λεκτικής περιγραφής του βίντεο.

Για την εν λόγω εφαρμογή, το κείμενο προς λεκτική ανάλυση προκύπτει από την επεξεργασία του βίντεο από έναν Audio Recognizer. Η ποιότητα του κειμένου αυτού εξαρτάται από την ποιότητα του βίντεο αλλά και των δυνατοτήτων ανάλυσης του Audio Recognizer.

Εξετάζοντας τα αποτελέσματα της αναγνώρισης των μερών του λόγου στην εν λόγω εφαρμογή, παρατηρούνται αρκετά λάθη τα οποία αφορούν κυρίως στην αναγνώριση ρημάτων ως ουσιαστικών. Το ίδιο φαινόμενο παρατηρήθηκε στις δοκιμές όπως παρουσιάζονται στην ενότητα 4.2, και καταλογίζεται όχι στην ανικανότητα του tagger, αλλά στην πολύ κακή σύνταξη του κειμένου που έχουμε στη διάθεσή μας.



Κεφάλαιο 5

Συνάφεια προτάσεων

Στα 2 προηγούμενα κεφάλαια αναλύθηκε ο τρόπος που έγινε ο διαχωρισμός των προτάσεων από τις λέξεις του transcript και η αναγνώριση τι μέρος του λόγου είναι καθεμία από αυτές. Στο κεφάλαιο αυτό περιγράφεται η χρήση των παραπάνω για την ανάλυση της λεκτικής περιγραφής σε τρίτο και τελευταίο επίπεδο. Η ανάλυση αυτή αφορά στην εξαγωγή των ουσιαστικών κάθε πρότασης και στη συνέχεια την σύγκριση αυτών. Η σύγκριση αυτή πραγματοποιείται με τη βοήθεια ειδικού λογισμικού (WordNet) το οποίο με κατάλληλη χρήση μπορεί να δώσει μια μετρική για το πόσο το περιεχόμενο μιας πρότασης σε σύγκριση με μια άλλη είναι συναφές. Η παραπάνω τεχνική στηρίζεται στην παραδοχή ότι τα κύρια σημεία του «νοήματος» μια πρότασης αντιπροσωπεύονται από τα ουσιαστικά της και άρα από αυτά απορρέει το μέγεθος της συνάφειας. Στόχος των παραπάνω είναι η δημιουργία ενός διαγράμματος συνάφειας μεταξύ των προτάσεων έτσι ώστε στα σημεία έντονων μεταβολών να έχουμε μια εκτίμηση αλλαγής θέματος.

5.1 Εξαγωγή Ουσιαστικών κάθε πρότασης

Το σύστημα στην παρούσα κατάσταση έχει στη διάθεσή του όλες τις πληροφορίες του transcript «αποκωδικοποιημένες» μιας και έχει γίνει ο διαχωρισμός των προτάσεων και η αναγνώριση τι μέρος του λόγου είναι οι λέξεις κάθε πρότασης από τον POS-tagger. Για την εξαγωγή των ουσιαστικών υπεύθυνη είναι η συνάρτηση του συστήματος ‘getWindowNouns()’ η οποία εντοπίζει τις λέξεις των προτάσεων που τις έχουν δοθεί ως όρισμα και έχουν επισημανθεί ως ουσιαστικά από τον POS-tagger και τις επιστρέφει σε κατάλληλη δομή ώστε να είναι εύκολα προσβάσιμες.

Στο σημείο αυτό το σύστημα «εμπιστεύεται» την κρίση του POS-tagger και δεν ελέγχει για το αν έχει γίνει εντοπισμός σημασιολογίας που προκύπτει από συνδυασμό λέξεων. Για παράδειγμα σε προτάσεις με τις λέξεις «white house» και «Clinton» θα είναι λανθασμένος ο αποκλεισμός του «white» ως επίθετο. Αυτό συμβαίνει γιατί η σύγκριση που θα γίνει σε επόμενο στάδιο θα δώσει υποδεέστερα αποτελέσματα αν συγκριθούν «house» και «Clinton» και όχι «white house» το οποίο έχει σαφώς μεγαλύτερη συσχέτιση.

Για την αποφυγή λαθών όπως τα παραπάνω, σε επόμενη έκδοση της συνάρτησης ‘getWindowNouns()’ υπάρχει η δυνατότητα ενσωμάτωσης κατάλληλου λεξικού (thesaurus) το οποίο αναγνωρίζει τη σημασιολογία συνδυασμού λέξεων. Ένα τέτοιο εργαλείο μάλιστα θα μπορούσε να προσδίδει επιπρόσθετες πληροφορίες στις υπό εξέταση λέξεις κατηγοριοποιώντας για παράδειγμα σε Πολιτικού περιεχομένου, Αθλητικού, Πολιτιστικού κτλ δίνοντας ένα ακόμη στοιχείο στο διαχωρισμό των θεμάτων.

5.2 Σύγκριση προτάσεων

Για την εξαγωγή συμπερασμάτων συνάφειας ή μη μεταξύ των προτάσεων του transcript θα πρέπει να γίνει η «νοηματική» σύγκριση μεταξύ των προτάσεων. Η σύγκριση αυτή μπορεί να γίνει εφικτή χρησιμοποιώντας τα ουσιαστικά της κάθε πρότασης, δεδομένου ότι τα ουσιαστικά είναι αυτά που κρύβουν το νόημα ή αλλιώς την «ουσία» μιας πρότασης. Το εργαλείο το οποίο θα μας δώσει μια μετρική για την συνάφεια 2 ουσιαστικών είναι το WordNet, για τον ακριβή τρόπο λειτουργίας του εργαλείου αυτού γίνεται αναφορά σε επόμενη ενότητα (5.3). Στην ενότητα αυτή θα αναλυθεί ο τρόπος που χρησιμοποιείται η συνάφεια ανά 2 ουσιαστικά ώστε να καταλήξουμε στον υπολογισμό της συνάφειας 2 προτάσεων.

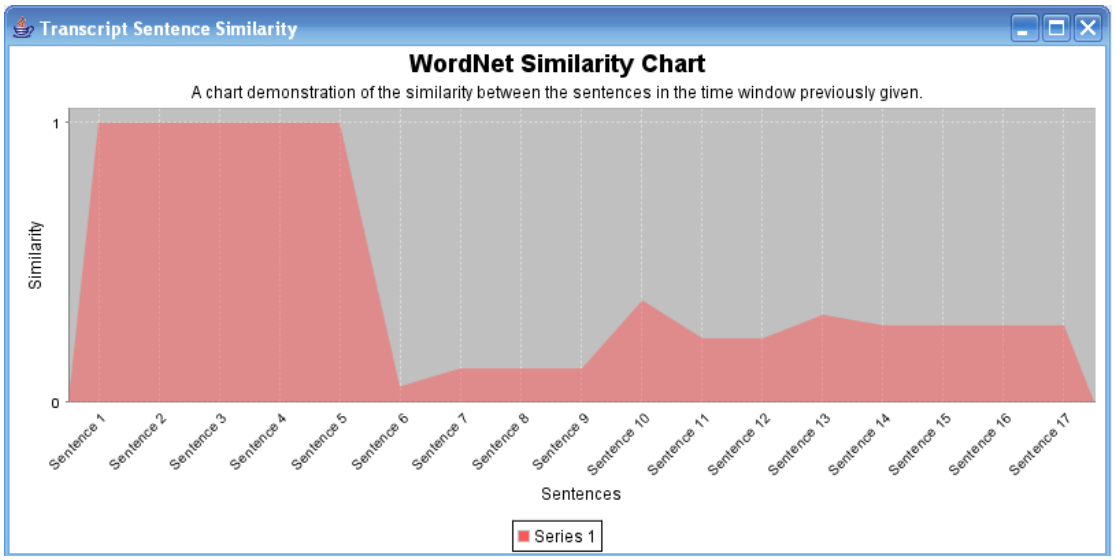
Στην βιβλιογραφία περιγράφονται προσπάθειες που αφορούν στην επεξεργασία κειμένου με χρήση του WordNet [18], Επεξεργαστές Φυσικών Γλωσσών(NLP)[19] και άλλων τεχνικών με σκοπό την εξαγωγή πληροφοριών για τον εντοπισμό των θεμάτων (topic detection), της δημιουργίας περίληψης (summarization)[20] κ.α.

Για την εργασία αυτή το σύστημα στην τρέχουσα έκδοση χρησιμοποιεί αποκλειστικά το WordNet. Συγκεκριμένα, μέσω της συνάρτησης ‘getNounSimilarity()’ και μιας απλής τεχνικής εξάγεται μια μετρική που αφορά στο πόσο συναφείς είναι οι προτάσεις του transcript ανά 2.

Για τον υπολογισμό της παραπάνω μετρικής για κάποιο αριθμό προτάσεων η συνάρτηση ‘getNounSimilarity()’ αρχικά δέχεται ως είσοδο τα ουσιαστικά δομημένα ανά πρόταση. Οι προτάσεις συγκρίνονται ανά 2, η 1η με τη 2η, η 2η με την 3η κτλ ενώ για τον υπολογισμό της συνάφειας υπολογίζεται και αθροίζεται η συνάφεια καθενός ουσιαστικού της μιας πρότασης με κάθε ουσιαστικό της δεύτερης, ενώ

το αποτέλεσμα διαιρείται με τον αριθμό των αθροισμάτων (κανονικοποίηση). Με την ολοκλήρωση της διαδικασίας η συνάρτηση επιστρέφει σε κατάλληλη δομή μια τιμή για κάθε ζεύγος προτάσεων.

Στο σημείο αυτό, μετά από τις διεργασίες όπως περιγράφηκαν στα παραπάνω και στο τρέχον κεφάλαιο, οι τιμές που προκύπτουν από τη σύγκριση των προτάσεων αναπαριστώνται από το σύστημα διαγραμματικά δίνοντας στο χρήστη μια πρώτη εκτίμηση για την νοηματική ομαδοποίηση των προτάσεων.



Στο παραπάνω διάγραμμα στον άξονα Υ έχουμε αποτύπωση της συνάφειας όπως προκύπτει από την σύγκριση των εκάστοτε 2 προτάσεων, όπου το εύρος το τιμών κυμαίνεται μεταξύ 0 και 1. Στον άξονα Χ αποτυπώνεται ο αύξων αριθμός της 2ης πρότασης του εξετασθέντους ζεύγους, εντός του διαστήματος υπό εξέταση. Για το συγκεκριμένο διάγραμμα η τιμή συνάφειας των προτάσεων 4 και 5 είναι 1, για τις 5 και 6 είναι 0,056, για τις 9 και 10 είναι 0,364 κτλ.

5.3 WordNet

Μια από τις βασικότερες λειτουργίες του συστήματος στην τρέχουσα κατάσταση είναι η ομαδοποίηση συναφών προτάσεων, η οποία κυρίως επιτυγχάνεται με την χρήση του WordNet.

5.3.1 Τι είναι το WordNet

Το WordNet είναι μια υπερμεγέθης βάση δεδομένων Αγγλικών λεκτικών[21]. Ουσιαστικά, ρήματα, επίθετα και επιρρήματα ομαδοποιούνται σε ομάδες με νοηματική συνάφεια, με την καθεμία από αυτές να εκφράζει μια διαφορετική έννοια. Οι ομάδες αυτές διασυνδέονται τόσο νοηματικά όσο και λεξικογραφικά. Με τον τρόπο αυτό προκύπτει ένα δίκτυο συναφών λέξεων και εννοιών. Το WordNet διατίθεται ως ελεύθερο λογισμικό και η δομή του είναι κατάλληλη για τον υπολογισμό μέσω γλωσσολογικής ανάλυσης αλλά και για την επεξεργασία φυσικών γλωσσών.

Το WordNet επιφανειακά μοιάζει με ένα λεξικό συνωνύμων, το οποίο ομαδοποιεί λέξεις βάσει του νοήματός τους. Παρόλα αυτά υπάρχουν βασικές διαφοροποιήσεις σε σχέση με ένα απλό λεξικό. Αρχικά, το WordNet διασυνδέει όχι απλά λέξεις λεξικογραφικά, αλλά συγκεκριμένες έννοιες λέξεων. Αυτό έχει ως αποτέλεσμα λέξεις οι οποίες στο δίκτυο του WordNet βρίσκονται κοντά, να έχουν καθορισμένη, ξεκάθαρη σημασιολογία. Επίσης το WordNet σηματοδοτεί τις σημασιολογικές σχέσεις μεταξύ λέξεων, ενώ σε ένα λεξικό συνωνύμων η ομαδοποίηση των λέξεων δεν ακολουθεί κανένα κανόνα πέραν της νοηματικής συνάφειας.

(Πηγή: http://wordnet.princeton.edu/wordnet/)

5.3.2 Η Δομή του WordNet

Η κύρια σχέση μεταξύ των λέξεων του WordNet είναι τα συνώνυμα, όπως για παράδειγμα μεταξύ των λέξεων «shut» και «close» ή «car» και «automobile». Τα συνώνυμα είναι λέξεις οι οποίες δηλώνουν την ίδια έννοια και μπορούν να εναλλάσσονται σε ποικίλα νοηματικά πλαίσια και συγκεντρώνονται σε μη διατεταγμένες ομάδες. Καθεμιά από τις 117.000 ομάδες του WordNet ενώνεται με άλλες ομάδες μέσω μικρού αριθμού «ιδεολογικών σχέσεων». Επιπρόσθετα, κάθε ομάδα περιλαμβάνει ένα σύντομο ορισμό, ενώ στις περισσότερες περιπτώσεις, μια ή περισσότερες προτάσεις αναπαριστούν την χρησιμότητα των μελών της ομάδας. Λέξεις οι οποίες έχουν περισσότερα από ένα διαφορετικά νοήματα αντιστοιχούν σε περισσότερες από μια ομάδες. Με τον τρόπο αυτό κάθε λέξη με συγκεκριμένο νόημα αντιστοιχεί αποκλειστικά σε συγκεκριμένη ομάδα.

5.3.3 Υπολογισμός Συνάφειας μέσω WordNet

Η συνάφεια μέσω του WordNet υπολογίζεται με βάση τη συνάφεια και την σχετικότητα όπως προκύπτει από τις δομές και το περιεχόμενο του WordNet. Οι διάφορες μετρικές συνάφειας χρησιμοποιούν πληροφορίες που βρίσκονται σε μια δενδροειδή (is-A) ιεραρχία εννοιών (ή ομάδων εννοιών), και υπολογίζουν το κατά πόσο μια έννοια είναι ίδια (ή παρόμοια) με μια άλλη. Για παράδειγμα, μια τέτοια μετρική θα μπορούσε να δείξει ότι ένα 'αυτοκίνητο' μοιάζει περισσότερο με μια 'βάρκα' από ότι με ένα 'δένδρο', δεδομένου ότι το 'αυτοκίνητο ' και η 'βάρκα' έχουν τον κοινό «πρόγονο» 'όχημα' στην ιεραρχία του WordNet.

Το WordNet κρίνεται χρήσιμο συγκεκριμένα για τον υπολογισμό συνάφειας λόγω της δενδροειδούς ιεραρχίας (is-A relatedness) ουσιαστικών και ρημάτων που εμπεριέχει. Στην έκδοση 2.0, υπάρχουν εννέα διαφορετικές ιεραρχίες ουσιαστικών οι οποίες περιλαμβάνουν 80.000 έννοιες, και 554 ιεραρχίες ρημάτων οι οποίες συνίσταται από 13.500 έννοιες. Οι δενδροειδής αυτές ιεραρχίες αφορούν αποκλειστικά έννοιες συγκεκριμένου κάθε φορά μέρους του λόγου (POS) και οι μετρικές που μπορούν να εξαχθούν περιορίζονται σε μεταξύ τους συγκρίσεις, δηλαδή ουσιαστικά με ουσιαστικά (π.χ. 'cat' και 'dog') και ρήματα με ρήματα (π.χ. 'run' και 'walk').

Η βάση δεδομένων του WordNet ενώ περιλαμβάνει επίθετα και επιρρήματα, δεν εμπεριέχει μια δενδροειδή ιεραρχία οργάνωσής τους, οπότε η εξαγωγή μετρικών συνάφειας δεν μπορεί να εφαρμοστεί. Παρόλα αυτά, οι έννοιες μπορεί να σχετίζονται μεταξύ τους με πολλούς τρόπους πέραν του να είναι παρόμοιες μεταξύ τους. Για παράδειγμα, μια 'ρόδα' είναι κομμάτι του 'αυτοκινήτου', η 'νύχτα' είναι το αντίθετο του 'μέρα', το 'χιόνι' φτιάχνεται από 'νερό', ένα 'μαχαίρι' χρησιμοποιείται για να κοπεί το 'ψωμί' κτλ. Με τον ίδιο τρόπο και το WordNet πέρα από τις δενδροειδείς (is-A) ιεραρχίες, περιλαμβάνει και σχέσεις 'είναι μέρος του'(has-part), 'φτιάχνεται από'(is-made-of), και 'είναι γνώρισμα του'(is-attribute-of). Επιπρόσθετα, κάθε έννοια περιλαμβάνει μια μικρή περιγραφή και ενδεχομένως ένα παράδειγμα χρήσης. Όλες οι παραπάνω πληροφορίες μπορούν να χρησιμοποιηθούν για την εξαγωγή πληροφοριών συνάφειας. Με τον τρόπο αυτό μπορούν να υπάρξουν πιο ευέλικτες μετρικές επιτρέποντας τον υπολογισμό

συνάφειας λέξεων που είναι διαφορετικό μέρος του λόγου (για παράδειγμα το ρήμα ‘φόνος’ με το ουσιαστικό ‘όπλο’).

5.3.4 Απόδοση Υπολογισμών μέσω WordNet

Για το τεράστιο πλήθος λέξεων, εννοιών και ιεραρχιών συσχέτισης που φιλοξενεί η βάση δεδομένων του WordNet έγινε μνεία σε προηγούμενη ενότητα. Αυτός είναι και ο λόγος που η χρήση της πλατφόρμας του WordNet έχει μεγάλη απήχηση σε αντίστοιχες εφαρμογές μέτρησης σημασιολογικής συνάφειας[22,23,24].

Παρόλα αυτά, δεδομένου ότι το προς επεξεργασία κείμενο είναι αποτέλεσμα της ανάλυσης από ένα Audio Recognizer, δημιουργούνται προβλήματα λόγω της παραποιημένης σύνταξης αλλά και της συχνά μη ορθής απόδοσης των εκφωνηθέντων. Το ίδιο πρόβλημα επίσης μειώνει την απόδοση του POS-tagging όπως περιγράφηκε στην ενότητα 4.4, ο οποίος συνακόλουθα σηματοδοτεί λανθασμένα πλήθος λέξεων ακόμη και ρημάτων, ως ουσιαστικά.

Τα παραπάνω προβλήματα στην ποιότητα του κειμένου προς ανάλυση δημιουργούν επιπρόσθετα προβλήματα κατά την διάρκεια του υπολογισμού συνάφειας προτάσεων. Η συνάφεια αυτή ανάγεται στην σύγκριση συνάφειας μεταξύ ουσιαστικών, τα οποία όμως για την εν λόγω εφαρμογή στην πραγματικότητα συχνά δεν είναι καν ουσιαστικά. Κατά την σύγκριση τέτοιων λέξεων το WordNet εύλογα επιστρέφει ότι αδυνατεί να τις αναγνωρίζει. Στην περίπτωση αυτή το σύστημα αδυνατεί με τη σειρά του να προχωρήσει στην εξαγωγή κάποιας μετρικής, οπότε το εκάστοτε προβληματικό ζεύγος ουσιαστικών αποκλείεται από τον συνυπολογισμό.

Κεφάλαιο 6

Ανάλυση Πλαισίων Βίντεο

Οι λειτουργίες που επιτελούσε το σύστημα, όπως περιγράφονταν μέχρι το σημείο αυτό, αφορούσαν στην επεξεργασία και ανάλυση της λεκτικής περιγραφής του transcript. Η πολυεπίπεδη αυτή ανάλυση αφορούσε στην εύρεση και την ομαδοποίηση των λέξεων σε προτάσεις,

την αναγνώριση των μερών του λόγου για τις προτάσεις αυτές, την απομόνωση των ουσιαστικών, και τέλος την εύρεση της συνάφειας των προτάσεων μέσω της σύγκρισης των ουσιαστικών. Τα αριθμητικά δεδομένα από την παραπάνω διαδικασία αναπαρίστανται σε ειδικό διάγραμμα συνάφειας προτάσεων, όπου μπορεί κανείς να διακρίνει μια ομαδοποίηση των προτάσεων και μια εκτίμηση για το πού αλλάζει η θεματολογία.

Ολοκληρώνοντας την παραπάνω διαδικασία το σύστημα ξεκινά ένα δεύτερο κύκλο εργασιών για να καταλήξει σε μια δεύτερη εκτίμηση του πότε έχουμε αλλαγή θέματος κατά την εκφώνηση. Οι εργασίες αυτές αφορούν στην επεξεργασία του βίντεο και συγκεκριμένα της ακολουθίας πλαισίων βίντεο, στο εξής video frames, με στόχο να εντοπισθούν αλλαγές στο σκηνικό κατά την εκφώνηση, οι οποίες θα δώσουν άλλο ένα στοιχείο αλλαγής θεματολογίας. Με το πέρας των εργασιών αυτών του δεύτερου κύκλου, τα δεδομένα και πάλι αναπαριστώνται σε ένα ειδικό διάγραμμα, αυτή τη φορά συνάφειας των video frames από όπου μπορεί κανείς να διακρίνει μια δεύτερη εκτίμηση για το πού έχουμε αλλαγή θέματος.

Η υλοποίηση της ανάλυσης του συγκεκριμένου επίπεδου γίνεται εφικτή με την ενσωμάτωση της πλατφόρμας διαχείρισης πολυμεσικού υλικού Java Media Framework (JMF)[25]. Πρόκειται για μια Java βιβλιοθήκη, η οποία επιτρέπει την διαχείριση πολυμεσικών αρχείων σε Java εφαρμογές. Οι κλάσεις που αφορούν στην συγκεκριμένη υλοποίηση κάνουν χρήση των λειτουργιών της πλατφόρμας αυτής και μέσω ειδικά ανεπτυγμένων για την εργασία αυτή αλγορίθμων, επιτυγχάνεται η διαχείριση της ροής των video frames αλλά και η ανάλυση αυτών, όταν απαιτείται.

6.1 Κλάση 'ReadFromVideo'

Σε προηγούμενες ενότητες έχει γίνει ήδη μνεία για τα modular χαρακτηριστικά κατά την ανάπτυξη των επιμέρους υποσυστημάτων της εφαρμογής. Η τεχνική αυτή εφαρμόζεται και στην επεξεργασία των video frames και μάλιστα με χρήση της γλώσσας προγραμματισμού Java, ένα χαρακτηριστικό που προσδίδει στο σύστημα προστιθέμενη αξία όσον αφορά στη διαλειτουργικότητα και επεκτασιμότητα. Η κύρια κλάση που αναλαμβάνει την διαχείριση των video frames είναι η ReadFromVideo. Οι εργασίες που επιτελεί αφορούν στη διαχείριση της ροής των video frames, την επιλογή ορισμένων από αυτών σύμφωνα με

συγκεκριμένα κριτήρια και επεξεργασία αυτών και τέλος στην επικοινωνία με το γραφικό περιβάλλον της διεπαφής χρήστη.

6.1.1 Διαχείριση Ροής Πλαισίων Βίντεο

Η κλάση ReadFromVideo, όταν καλείται από το σύστημα, ζητά ως όρισμα την θέση (path) του βίντεο. Στη συνέχεια ακολουθεί η επεξεργασία των video frames ένα προς ένα. Στα πρώιμα στάδια ανάπτυξης του συστήματος δεν εφαρμοζόταν επιλογή μεταξύ των frames για το ποια εμπεριέχουν χρήσιμη πληροφορία και ποια όχι, με αποτέλεσμα να μην εξαιρείται κανένα από την διαδικασία ανάλυσης. Η ανάλυση αυτή αφορούσε στον εντοπισμό διαφορών μεταξύ των frames ανά 2 σε βαθμό τέτοιο που θα μπορούσε να θεωρηθεί αλλαγή θεματολογίας. Λεπτομέρειες για την ανάλυση αυτή θα δοθούν σε επόμενη ενότητα.

Ο αριθμός των video frames σε ένα βίντεο ποικίλλει ανάλογα με την ποιότητα της εικόνας, παρόλα αυτά ισχύει ένα ελάχιστο της τάξης των 10 video frames ανά δευτερόλεπτο βίντεο έτσι ώστε το ανθρώπινο μάτι να μην εντοπίζει ασυνέχεια στην ακολουθία των video frames. Στο χρησιμοποιούμενο dataset ο αριθμός των video frames κυμαίνεται μεταξύ των 50 και 55 χιλιάδων. Η ανάλυση ενός τόσο μεγάλου αριθμού video frames επιβεβαιώθηκε ως απαγορευτική κατά τις πρώιμες δοκιμές του συστήματος, όπου έστω και η ελάχιστη επεξεργασία ανά frame οδηγούσε σε αρκετά λεπτά χρόνου εκτέλεσης από το σύστημα.

Με αφορμή τα παραπάνω, στα επόμενα στάδια ανάπτυξης υιοθετήθηκε ο διαχωρισμός των video frames σε κρίσιμα και μη. Τα μη κρίσιμα frames προσπερνιούνται από το σύστημα δίχως περαιτέρω επεξεργασία ενώ για τα κρίσιμα frames και μόνο επιτελούνται οι διεργασίες ανάλυσης. Η τεχνική αυτή οδήγησε σε δραματική μείωση του χρόνου εκτέλεσης του συστήματος για την ανάλυση του βίντεο δίχως καμία απώλεια στην ακρίβεια των αποτελεσμάτων, αυξάνοντας κατακόρυφα την ευχρηστία και αποδοτικότητα της εφαρμογής

6.1.2 Κρίσιμα Πλαίσια Βίντεο

Ο απώτερος στόχος της εφαρμογής είναι ο εντοπισμός των σημείων, όπου αλλάζει η θεματολογία. Σύμφωνα με αυτό, και δεδομένου ότι από τις αναλύσεις των προηγούμενων επιπέδων έχουμε μια καλή εκτίμηση

των προτάσεων μπορούμε να περιορίσουμε την αναζήτηση της αλλαγής θεματολογίας μεταξύ των προτάσεων αυτών.

Σύμφωνα με την παραπάνω παραδοχή, το εύρος των video frames που καλούμαστε να εξετάσουμε περιορίζονται αποκλειστικά στα διαστήματα μεταξύ των προτάσεων. Με τον τρόπο αυτό μπορούμε να παραλείψουμε τα frames του βίντεο κατά τη διάρκεια εκφώνησης μιας ομάδας λέξεων (πιθανότατα μιας πρότασης), και να επικεντρώσουμε την ανάλυση στα **κρίσιμα πλαίσια βίντεο** (ΠΑΡΑΤΗΜΑ Β) που είναι αυτά που βρίσκονται στα διαστήματα μεταξύ των παύσεων του εκφωνητή.

Το κάθε video frame εμπεριέχει διάφορες πληροφορίες από τις οποίες ιδιαίτερη χρησιμότητα έχουν αυτές που αφορούν στην χρονική στιγμή εμφάνισής του στο βίντεο. Οι πληροφορίες αυτές σε συνδυασμό με τις πληροφορίες που έχουμε συλλέξει στα προηγούμενα στάδια ανάλυσης του transcript αναφορικά με την χρονική διάρκεια, την αρχή και το τέλος χρονικά της κάθε πρότασης κάνουν εφικτό τον παραπάνω διαχωρισμό.

Σύμφωνα με τα παραπάνω, παρατίθεται ο αλγόριθμος, ο οποίος υλοποιεί τον διαχωρισμό των video frames σε κρίσιμα και μη, με τις επεξηγήσεις που αφορούν στο περιεχόμενο των χρησιμοποιούμενων μεταβλητών και συναρτήσεων να ακολουθούν.

```
if( frame_time < (((Float)superFrame.T_Det.Sentence_end.get(curr_sentence_id)).floatValue()
                - trans_video_overhead) && curr_sentence_id < to_sentence_id)
    return;
else if( frame_time < (((Float)superFrame.T_Det.Sentence_start.get(curr_sentence_id+1)).floatValue()
                    -trans_video_overhead) && curr_sentence_id < to_sentence_id)
    useFrameData(inBuffer);
else {
    if(curr_sentence_id < to_sentence_id){
        superFrame.progressField.setText(String.format("%.2f",(float)curr_sentence_id
                                        /(float)num_of_sentences*100)+"%");
        curr_sentence_id++;
    }
    else {
        superFrame.progressField.setText("Completed");
        System.out.println("End of selected frames");
        tidyClose();
        thisFrame.dispose();
        return;
    }
    return;
}
```

frame_time : Η χρονική στιγμή που εμφανίζεται στο βίντεο το τρέχον πλαίσιο βίντεο

Η ροή του αλγορίθμου έχει ως εξής:
Για όσο είμαστε εντός του εύρους των υπό εξέταση προτάσεων
Για κάθε πλαίσιο κειμένου

-Αν η χρονική στιγμή κατά την οποία εμφανίζεται στο βίντεο είναι μικρότερη (νωρίτερα) σε σχέση με το τέλος της υπό εξέταση πρότασης τότε αγνόησε το συγκεκριμένο πλαίσιο βίντεο
-Αλλιώς αν δεν ισχύει η παραπάνω συνθήκη και ταυτόχρονα εμφανίζεται πριν την αρχή της επόμενης από την υπό εξέταση πρόταση, τότε προχώρα στην περαιτέρω ανάλυση του πλαισίου (κλήση συνάρτησης (useFrameData())
-Σε κάθε άλλη περίπτωση
Αν είμαστε ακόμη εντός του εύρους των υπό εξέταση προτάσεων, προχωράμε στο επόμενο ζεύγος προτάσεων, αλλιώς τερματίζεται η ανάλυση

Στις δύο συνθήκες των 'if' παρατηρείται η διαμεσολάβηση της μεταβλητής 'trans_video_overhead'. Η μεταβλητή αυτή ουσιαστικά φέρει την τιμή της μεταβλητής '*Overhead*' και επιτελεί την λειτουργία του συγχρονισμού του transcript με το βίντεο όπως περιγράφεται στην ενότητα 3.4.

Η συνάρτηση η οποία αναλαμβάνει την ανάλυση των εκάστοτε video frames εντός κρίσιμης περιοχής, αναφέρθηκε παραπάνω ότι είναι η '***useFrameData()***'. Η συνάρτηση αυτή δέχεται ως όρισμα το τρέχον κάθε φορά video frame και φροντίζει για την δημιουργία μιας εικόνας αντιγράφου του. Στη συνέχεια, φροντίζει την σύγκρισή ανά 2 τέτοιων frames μέσω των εικόνων τους. Για να το επιτύχει αυτό, η ίδια συνάρτηση δημιουργεί στιγμιότυπο της κλάσης '*ImageCompare*' (ενότητα 6.2) δίνοντας ως όρισμα τις υπό σύγκριση εικόνες. Τέλος συλλέγει τα αποτελέσματα της σύγκρισής τους και ανάλογα διακόπτει ή αφήνει να συνεχιστεί η ανάλυση. (πλήρης υλοποίηση παρατίθεται στο ΠΑΡΑΡΤΗΜΑ)

6.1.3 Επικοινωνία με γραφικό περιβάλλον

Η όλη διαδικασία διαχείρισης των πλαισίων βίντεο δεδομένου του μεγάλου πλήθους τους αλλά και της πολυπλοκότητας επιλογής και ανάλυσής τους κρίνεται ιδιαίτερα χρονοβόρα. Όπως έχει ήδη επισημανθεί, ένα από τα κριτήρια κατά το σχεδιασμό της εφαρμογής

είναι η ταχύτητα απόκρισης του συστήματος, παρόλα αυτά διεργασίες όπως το POS-tagging των λέξεων, μέτρηση της συνάφειας των προτάσεων και η σύγκριση των πλαισίων απαιτούν αρκετό χρόνο εκτέλεσης.

Ιδιαίτερα όσον αφορά την επεξεργασία των πλαισίων βίντεο η κλάση '*ReadFromVideo*' έχει σχεδιαστεί έτσι ώστε ο χρήστης να ενημερώνεται από το γραφικό περιβάλλον για την πορεία των διεργασιών.

Για την παραπάνω ενημέρωση του χρήστη, το γραφικό περιβάλλον χρησιμοποιεί 2 μεταβλητές, την μεταβλητή '*videoPanel*' και την μεταβλητή '*progressField*'. Οι μεταβλητές αυτές είναι διαθέσιμες προς ανανέωση κατά τη διάρκεια των διεργασιών της '*ReadFromVideo*'. Η τελευταία με την σειρά της φροντίζει στην πρώτη μεταβλητή να καταγράφει ένα αντίγραφο του εκάστοτε υπό ανάλυση πλαισίου βίντεο, το οποίο και εμφανίζεται άμεσα στον χρήστη. Στη δεύτερη μεταβλητή αποθηκεύει έναν αριθμό που αντιπροσωπεύει τον ποσοστιαίο βαθμό ολοκλήρωσης των συνολικών διεργασιών που επιτελεί.

Ο υπολογισμός του ποσοστιαίου βαθμού ολοκλήρωσης γίνεται εφικτός λόγω τις μοντελοποίησης του προβλήματος των «κρίσιμων πλαισίων βίντεο» όπως περιγράφηκε στην ενότητα 6.1.2 με βάση την ομαδοποίηση σε προτάσεις. Κατά την ανάλυση ολόκληρου ή μέρους του βίντεο από την ανάλυση των πρώτων επιπέδων γνωρίζουμε ήδη τον ακριβή αριθμό των προτάσεων που αντιστοιχούν. Επιπρόσθετα η ανάλυση γίνεται ανά πρόταση οπότε καταφέρνουμε μια εκτίμηση του αριθμού των αναλυθέντων προτάσεων προς το σύνολό τους ως εξής:

```
superFrame.progressField.setText(String.format("%.2f",(float)curr_sentence_id
                                                /(float)num_of_sentences*100)+"%");
```

Όπου 'curr_sentence_id' ο αριθμός των ήδη αναλυθέντων προτάσεων και 'num_of_sentences' το σύνολο των προτάσεων για την συγκεκριμένη ανάλυση.

6.2 Κλάση ‘ImageCompare’

Σε προηγούμενες ενότητες του τρέχοντος κεφαλαίου αναφέρθηκε ο τρόπος διαχείρισης της ροής των πλαισίων βίντεο, ο διαχωρισμός τους σε κρίσιμα και μη και ο τρόπος σύγκρισης των κρίσιμων με σκοπό την εύρεση αυτών που έχουν σημαντική διαφορά μεταξύ τους. Η κλάση η οποία υλοποιεί την σύγκριση αυτή είναι η ‘***ImageCompare***’.

Η σύγκριση 2 πλαισίων βίντεο στην ουσία ανάγεται στην σύγκριση 2 εικόνων οι οποίες είναι αντίγραφα των υπό εξέταση πλαισίων. Με στόχο την καλύτερη σχέση επίδοσης του συστήματος και αποτελέσματος σύγκρισης, η σχεδίαση της κλάσης είναι τέτοια που επιτρέπει την ρύθμιση της ακρίβειας των υπολογισμών μέσω ειδικών παραμέτρων. Επίσης το μαθηματικό μοντέλο το οποίο χρησιμοποιείται για τον υπολογισμό και την επιστροφή κάποιας μετρικής που αφορά στην εν λόγω σύγκριση, υλοποιείται αποκλειστικά από μια συνάρτηση με γνώμονα την εύκολη αλλαγή, αναβάθμιση ή επέκταση σε αυτό.

6.2.1 Παράμετροι Σύγκρισης πλαισίων βίντεο

Οι παράμετροι που αφορούν στις ιδιαιτερότητες της σύγκρισης των εκάστοτε 2 video frames είναι τέσσερις (4). Οι δύο αφορούν στον αριθμό των πλαισίων στα οποία θα «κοπεί» καθεμιά από τις υπό

εξέταση εικόνες και άλλες δύο αφορούν στην «ευαισθησία» κατά τον εντοπισμό διαφορών.

Οι δύο πρώτες παράμετροι ονομάζονται στο σύστημα '*compareX*' και '*compareY*' και ορίζουν σε πόσα μέρη κατά τον άξονα Χ και Υ αντίστοιχα έχουμε επιμερισμό των αρχικών εικόνων. Συγκεκριμένα οι τιμές που χρησιμοποιούνται από το σύστημα είναι οι 8 και 6 αντίστοιχα οπότε έχουμε τον εξής διαμερισμό:

1	2	3	4	5	6	7	8
2							
3							
4							
5							
6							

Κάθε μια από τις 48 (=8x6) υποπεριοχές (blocks) που νοητά δημιουργούνται θα αποτελέσει μονάδα για την σύγκριση που θα γίνει σύμφωνα με το εκάστοτε μαθηματικό μοντέλο (ενότητα 6.2.2). Στο σημείο αυτό θα πρέπει να τονίσουμε ότι αναφορικά με την σχέση απόδοσης του συστήματος και της ακρίβειας των αποτελεσμάτων που αναφέρθηκε παραπάνω, ισχύει ότι όσο περισσότερες είναι οι υποπεριοχές στις οποίες χωρίζεται η κάθε εικόνα, τόσο μεγαλύτερη ακρίβεια αποτελεσμάτων επιτυγχάνουμε, αλλά και τόσο επιβαρύνουμε το σύστημα με πλήθος υπολογισμών.

Οι άλλες δύο παράμετροι είναι οι '*factorA*' και '*factorD*' οι οποίες αρχικοποιούνται από το σύστημα με τις τιμές 10 και 10 αντίστοιχα. Η επεξήγηση της χρησιμότητας των δύο αυτών παραμέτρων θα γίνει σε επόμενη ενότητα ταυτόχρονα με την ανάλυση του μαθηματικού μοντέλου σύμφωνα με το οποίο πραγματοποιείται η σύγκριση των εικόνων.

6.2.2 Μαθηματικό Μοντέλο Σύγκρισης

Για τον υπολογισμό μιας μετρικής αναφορικά με την σύγκριση 2 εικόνων υπεύθυνη είναι η συνάρτηση '*compare()*' της κλάσης '*ImageCompare*'. Η συνάρτηση αυτή υλοποιεί το μαθηματικό μοντέλο της σύγκρισης, το οποίο για τις ανάγκες της εν λόγω εφαρμογής έχει ως εξής:

```
for (int y = 0; y < comparey; y++) {
        if (debugMode > 0) System.out.print("|");
        for (int x = 0; x < comparex; x++) {
                int b1 = getAverageBrightness(img1.getSubimage(x*blocksx,
                                y*blocksy, blocksx - 1, blocksy - 1));
                int b2 = getAverageBrightness(img2.getSubimage(x*blocksx,
                                y*blocksy, blocksx - 1, blocksy - 1));
                int diff = Math.abs(b1 - b2);
                if (diff > factorA)
                        this.match = false;

                if (debugMode == 1)
                    System.out.print((diff > factorA ? "X" : " "));
                if (debugMode == 2)
                    System.out.print(diff + (x < comparex - 1 ? "," : ""));
                sum_diff += diff;
        }
        if (debugMode > 0) System.out.println("|");
}
System.out.println("Kanonikopoihmeno Metro diaforas tvn frames :"
                                        +sum_diff/(comparex*comparey));
```

Επεξηγώντας την παραπάνω υλοποίηση έχουμε:

Για καθεμία από τις υποπεριοχές των εικόνων (blocks)

-Υπολόγισε τη μέση φωτεινότητα του συγκεκριμένου block και για τις 2 εικόνες

-Υπολόγισε την απόλυτη διαφορά της φωτεινότητας του συγκεκριμένου block για τις 2 εικόνες

-Αν η διαφορά είναι μεγαλύτερη από την τιμή της μεταβλητής factorA, τότε σηματοδότησε ότι έχουμε διαφορετικές εικόνες

-Συνάθροισε την τιμή της διαφοράς

Στο τέλος εκτύπωσε το παραπάνω άθροισμα δια του αριθμού των blocks που χωρίζεται η κάθε εικόνα

Η χρησιμότητα της παραμέτρου *'factorA,'* όπως φαίνεται στην παραπάνω υλοποίηση, αφορά στο να ορίζουμε το όριο σύμφωνα με το οποίο το μοντέλο θα κρίνει ότι ένα block της μιας εικόνας θα διαφέρει σε φωτεινότητα από το αντίστοιχο της άλλης.

Η μεταβλητή 'sum_diff' συναθροίζει όλες τις διαφορές που προκύπτουν συγκρίνοντας την φωτεινότητα των εικόνων ανά block. Το άθροισμα αυτό αν το διαιρέσουμε με τον αριθμό των blocks καταλήγουμε σε μια κανονικοποιημένη μετρική της διαφοράς φωτεινότητας μεταξύ των υπό εξέταση εικόνων.

Στον παραπάνω κώδικα παρατηρούμε την χρήση άλλη μιας μεταβλητής, της ‘debugMode’. Η μεταβλητή αυτή είναι υπεύθυνη για την καταγραφή πληροφοριών ελέγχου στην έξοδο του συστήματος. Οι πληροφορίες αυτές είναι πολύ χρήσιμες για τις δοκιμές κατά την ανάπτυξη του συστήματος αλλά και κατά τη διάρκεια κανονικής εκτέλεσης για την επιβεβαίωση της ορθής λειτουργίας.

Οι πληροφορίες ελέγχου για το συγκεκριμένο κομμάτι κώδικα αφορούν στην αναπαράσταση της διαφοράς των τιμών της φωτεινότητας μεταξύ των blocks των εκάστοτε 2 video frames, στην κανονικοποιημένη μετρική της διαφοράς και άλλων πληροφοριών για τα frames ως εξής:

```
  Frame ID #: 2800
  Time: 94.31secs
Comparison between frame: 2799 and 2800
|0,1,0,0,0,1,1,0|
|0,0,0,0,0,0,0,0|
|0,0,0,0,0,1,0,0|
|0,0,0,0,0,0,0,0|
|0,1,1,1,0,0,0,0|
|1,0,0,0,1,0,1,0|
Kanonikopoihmeno Metro diaforas tvn frames :0.20833333
Match: true

  Frame ID #: 2872
  Time: 96.71secs
Comparison between frame: 2871 and 2872
|31,33,29,22,11,11,36,46|
|45,61,47,25,24,40,45,44|
|63,63,50,51,30,5,45,46|
|57,68,55,60,41,5,50,55|
|62,57,60,64,27,17,52,40|
|58,57,63,53,28,13,34,44|
Kanonikopoihmeno Metro diaforas tvn frames :42.145832
Match: false
```

Στην τρέχουσα έκδοση της εφαρμογής έχει επιλεχθεί η απλοποίηση των υπολογισμών με γνώμονα τη μέγιστη ταχύτητα απόκρισης του συστήματος και για την σύγκριση των video frames λαμβάνουμε υπόψη αποκλειστικά την τιμή της μεταβλητής ‘match’. Η μεταβλητή αυτή όπως φαίνεται στην παραπάνω υλοποίηση έχει την τιμή ‘*true*’ στην περίπτωση που η σύγκριση μεταξύ των blocks δεν ξεπερνά για κανένα από αυτά το δοθέν όριο (factorA). Σε κάθε άλλη περίπτωση παίρνει την τιμή ‘*false*’.

Η συνάρτηση ‘getAverageBrightness()’ είναι υπεύθυνη για τον υπολογισμό της φωτεινότητας. Παίρνει ως όρισμα μια εικόνα, ενώ για τις ανάγκες της σύγκρισης, όπως υλοποιείται παραπάνω, καλείται με όρισμα την εκάστοτε υπό εξέταση υποπεριοχή. Η υλοποίησή είναι σχετικά απλή, γίνεται εντός της κλάσης ‘*ImageCompare*’ και παρατίθεται.

```
protected int getAverageBrightness(BufferedImage img) {
    Raster r = img.getData();
    int total = 0;
    for (int y = 0; y < r.getHeight(); y++) {
        for (int x = 0; x < r.getWidth(); x++) {
            total += r.getSample(r.getMinX() + x, r.getMinY() + y, 0);
        }
    }
    return (int)(total / ((r.getWidth()/factorD)*(r.getHeight()/factorD)));
}
```

Επεξηγώντας την παραπάνω υλοποίηση έχουμε:

-Από την εικόνα που δίνεται ως όρισμα φτιάξε ένα πίνακα με τα εικονοστοιχεία της (Raster).

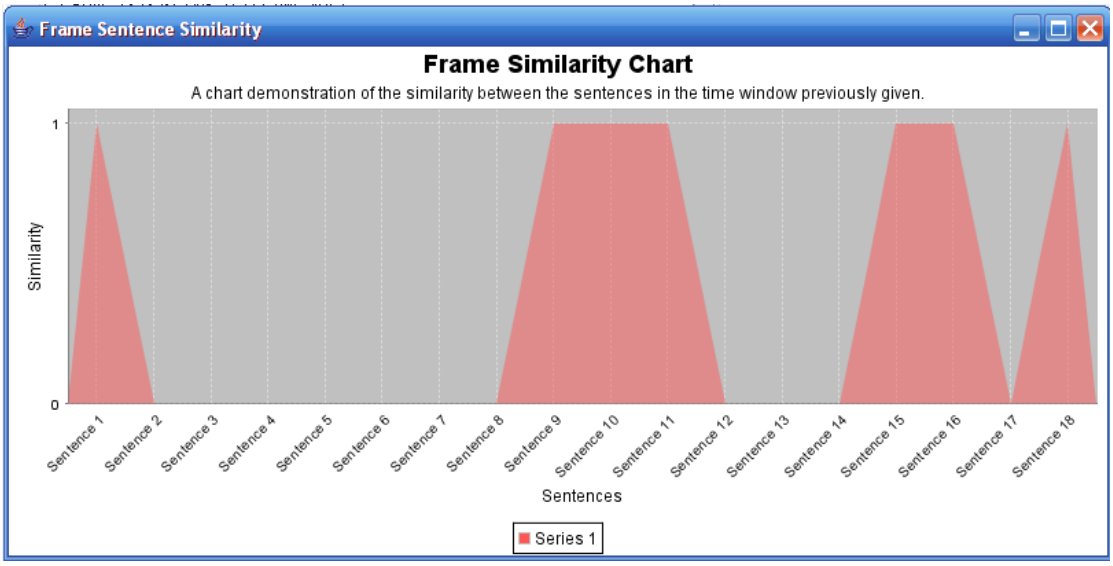
-Άθροισε όλες τις τιμές των εικονοστοιχείων όπως περιγράφονται στον παραπάνω πίνακα.

-Επέστρεψε το παραπάνω άθροισμα κανονικοποιημένο ως προς μέγεθος του παραπάνω πίνακα και την παράμετρο factorD

Από την επεξήγηση του κώδικα της συνάρτησης ‘getAverageBrightness()’ βλέπουμε την χρησιμότητα της τελευταίας από τις παραμέτρους, όπως περιγράφηκαν στην ενότητα 6.2.1, την ‘*factorD*’, η οποία αφορά στην ρύθμιση της φωτεινότητας.

6.3 Διαγραμματική απεικόνιση αποτελεσμάτων

Με το πέρας των διεργασιών ανάλυσης των πλαισίων βίντεο καταφέρνουμε να έχουμε μια εικόνα για το αν έχουμε αλλαγή στο σκηνικό στα διαστήματα των παύσεων εκφώνησης. Τα αποτελέσματα αυτά αναπαριστώνται σε ειδικό διάγραμμα, αντίστοιχο με το διάγραμμα που προέκυψε από την ανάλυση του transcript. Στο διάγραμμα αυτό στα σημεία που δεν έχει εντοπιστεί αλλαγή έχουμε τιμή 0 (μηδέν), αλλιώς έχει εντοπιστεί σε κάποιο block αλλαγή φωτεινότητας πέραν του ορίου που έχει δοθεί και δίνεται η τιμή 1.



Στο παραπάνω διάγραμμα στον άξονα Υ έχουμε αποτύπωση της συνάφειας όπως προκύπτει από την σύγκριση των πλαισίων βίντεο μεταξύ των παύσεων των εκάστοτε 2 προτάσεων, οι δυνατές τιμές είναι το 0 και 1. Στον άξονα Χ αποτυπώνεται ο αύξων αριθμός της 2ης πρότασης του εξετασθέντος ζεύγους, εντός του διαστήματος υπό εξέταση. Για το συγκεκριμένο διάγραμμα αποτυπώνεται η αλλαγή φωτεινότητας πέραν του ορίου μεταξύ των προτάσεων 0 και 1, 8 και 9, 9 και 10, 10 και 11, 14 και 15, 15 και 16, 17 και 18.

Κεφάλαιο 7

Γραφικό περιβάλλον Διεπαφής Χρήστη

Η λειτουργικότητα της εφαρμογής, όπως έχει ήδη αναφερθεί, έχει σχεδιαστεί, μοντελοποιηθεί και υλοποιηθεί πλήρως στα πλαίσια της γλώσσας προγραμματισμού Java. Το ίδιο ισχύει και για την ανάπτυξη του γραφικού περιβάλλοντος χρήστη, δίνοντας στην εφαρμογή τα ιδιαίτερα χαρακτηριστικά φορητότητας, επεκτασιμότητας και ευχρηστίας που συνοδεύονται από την παραπάνω επιλογή.

Το ειδικά σχεδιασμένο γραφικό περιβάλλον από την πλευρά του χρήστη δίνει την δυνατότητα της χρήσης καθεμιάς από τις λειτουργίες του συστήματος ταυτόχρονα με την εξέταση των αποτελεσμάτων των επιλογών του. Από την πλευρά του προγραμματιστή η όλη λειτουργικότητα του συστήματος έχει συσσωρευτεί στην κλήση

συγκεκριμένων συναρτήσεων. Η υλοποίηση του γραφικού περιβάλλοντος έχει σαφέστατη δομή, οργανωμένη με γνώμονα την εναρμόνιση με αυτήν της υλοποίησης των λειτουργιών της εφαρμογής.

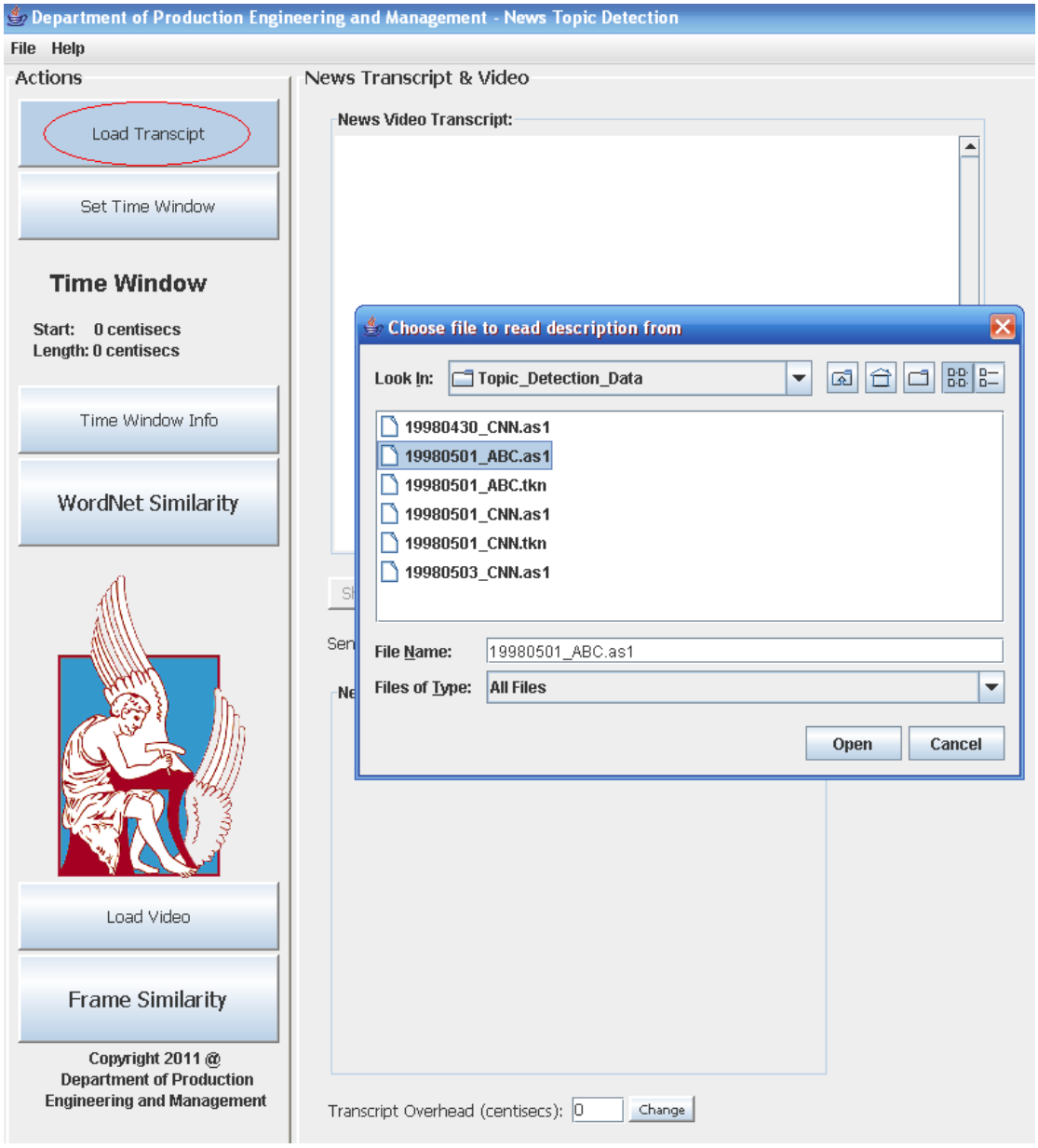
7.1 Χρήση Γραφικού Περιβάλλοντος Διεπαφής

Καθένα από τα επίπεδα ανάλυσης της λεκτικής περιγραφής του βίντεο και του βίντεο καθεαυτού αντιπροσωπεύεται επακριβώς από αντίστοιχες επιλογές στο γραφικό περιβάλλον. Οι επιλογές αυτές συνοπτικά έχουν ως εξής:

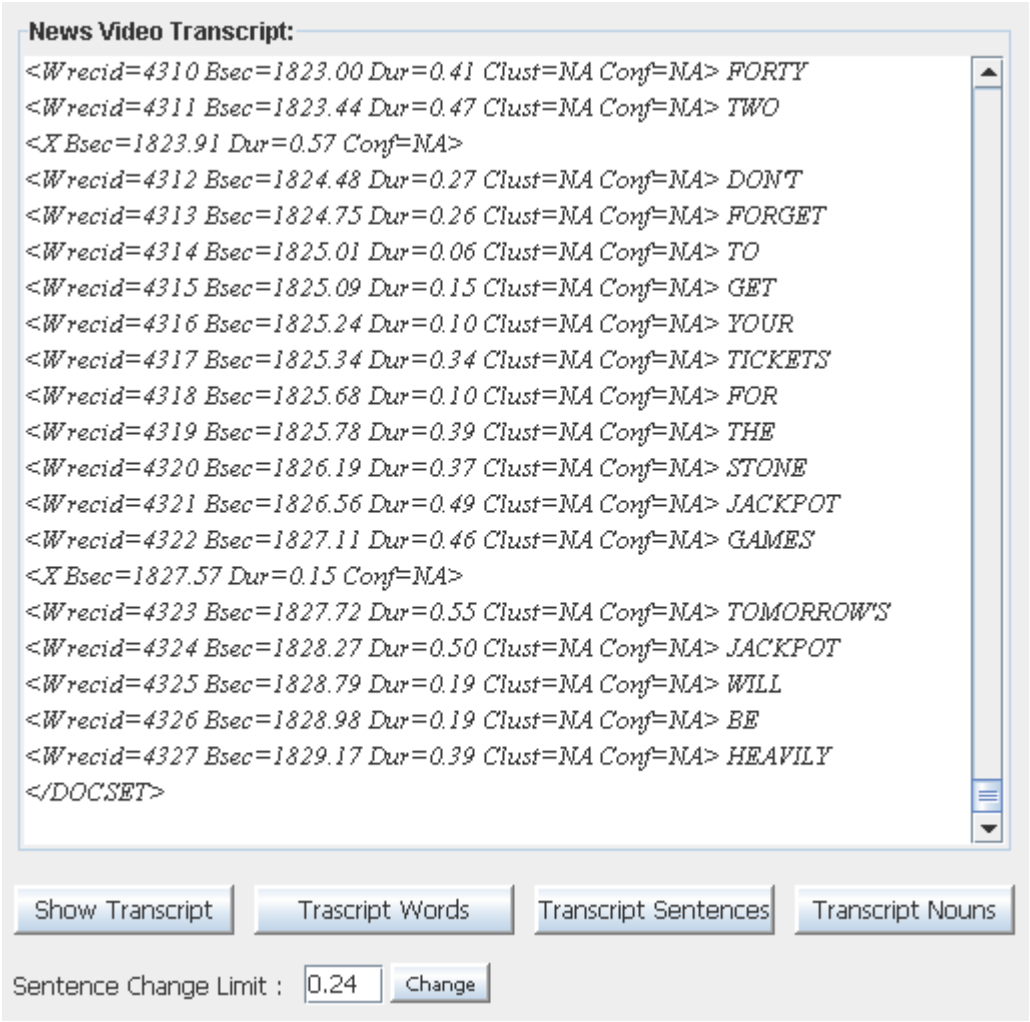
- Εξαγωγή των λέξεων, προτάσεων και ουσιαστικών από την ανάλυση της λεκτικής περιγραφής
- Υπολογισμός Συνάφειας μεταξύ των προτάσεων της λεκτικής περιγραφής
- Εύρεση των σημείων όπου αλλάζει το σκηνικό κατά τη διάρκεια των παύσεων εκφώνησης

7.1.1 Επιλογές Ανάλυσης λεκτικής περιγραφής

Κατά την εκκίνηση της εφαρμογής ο χρήστης αρχικά καλείται να επιλέξει ποια είναι η λεκτική περιγραφή του βίντεο προς ανάλυση. Με το πάτημα του αντίστοιχου κουμπιού (‘Load Transcript’) στο γραφικό περιβάλλον εμφανίζεται ένα ειδικό παράθυρο αναζήτησης στο σκληρό δίσκο ή σε άλλες μανάδες αποθήκευσης του μηχανήματος όπου «τρέχει» η εφαρμογή.



Στη συνέχεια η εφαρμογή εμφανίζει την περιγραφή αυτή σε ειδικό πλαίσιο και ενεργοποιούνται οι επιλογές για την ανάλυση της λεκτικής περιγραφής.



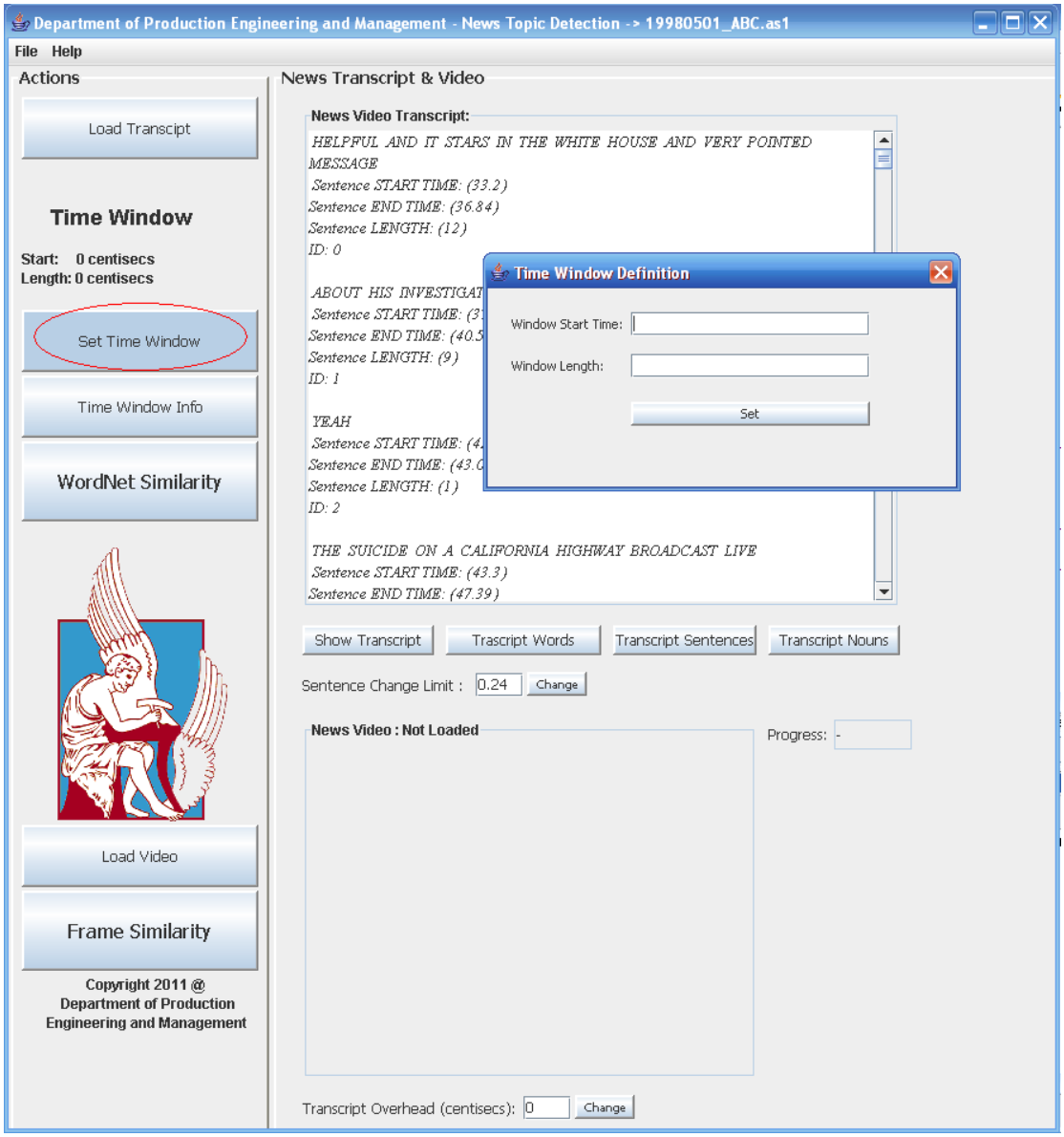
Οι επιλογές αυτές αφορούν στην εξαγωγή των λέξεων της λεκτικής περιγραφής (κουμπί 'Transcript Words'), στην ομαδοποίηση των λέξεων σε προτάσεις (κουμπί 'Transcript Sentences') και στην εξαγωγή των ουσιαστικών ανά πρόταση (κουμπί 'Transcript Nouns'). Με το πάτημα καθεμίας από τις παραπάνω επιλογές το σύστημα εμφανίζει τα αντίστοιχα αποτελέσματα στο παραπάνω πλαίσιο.

Το 'Sentence Change Limit' αφορά στο κάτω όριο της παύσης στην εκφώνηση σύμφωνα με το οποίο γίνεται η ομαδοποίηση των λέξεων, κατά το πάτημα του κουμπιού 'Transcript Sentences'. Το όριο αυτό δίνεται η δυνατότητα στο χρήστη να το αλλάξει κατά τη διάρκεια εκτέλεσης της εφαρμογής πατώντας το αντίστοιχο κουμπί ('Change'). Η λειτουργία του παραπάνω ορίου κατά τον εντοπισμό των προτάσεων έχει αναλυθεί εκτενώς στην ενότητα 3.3.

Κουμπί	Λειτουργία	Περιγραφή
Load transcript	Φόρτωση λεκτικής περιγραφής	-Άνοιγμα παράθυρου αναζήτησης αρχείου στο C:\. -Εμφάνιση του transcript στο ειδικό πλαίσιο
Change (Sentence Change Limit)	Καθορισμός τιμής ορίου αλλαγής προτάσεων	Άνοιγμα ειδικού παράθυρου για τον ορισμό νέας τιμής σε δευτερόλεπτα
Show transcript	Εμφάνιση λεκτικής περιγραφής	Εμφάνιση του transcript στο ειδικό πλαίσιο
Transcript Words	Εμφάνιση Λέξεων λεκτικής περιγραφής	-Επεξεργασία transcript για εξαγωγή λέξεων -Εμφάνιση λέξεων στο ειδικό πλαίσιο
Load Sentences	Εμφάνιση προτάσεων λεκτικής Περιγραφής	-Ομαδοποίηση λέξεων σε προτάσεις σύμφωνα με το όριο αλλαγής προτάσεων. -Εμφάνιση των προτάσεων και πληροφοριών χρονικής διάρκειας στο ειδικό πλαίσιο
Load Nouns	Εμφάνιση ουσιαστικών λεκτικής Περιγραφής	-Tagging των προτάσεων με χρήση του Stanford POS-Tagger -Επιλογή ουσιαστικών από τα αποτελέσματα του tagging και εμφάνιση στο ειδικό πλαίσιο

7.1.2 Επιλογές Υπολογισμού Συνάφειας Προτάσεων

Το επόμενο βήμα μετά την επιλογή και ανάλυση της λεκτικής ανάλυσης του βίντεο, αφορά στον υπολογισμό της συνάφειας των προτάσεων. Η λειτουργία αυτή, λόγω της πολυπλοκότητας αλλά και της συμμετοχής διαφόρων εργαλείων στους υπολογισμούς (ενότητα 3.1) δεσμεύει αρκετό χρόνο εκτέλεσης. Για το λόγο αυτό, ο χρήστης καλείται να ορίσει το χρονικό διάστημα εντός της διάρκειας του βίντεο για το οποίο επιθυμεί να εκτελέσει την παραπάνω λειτουργία, δίχως να δεσμεύεται και για την επιλογή ολόκληρης τη διάρκειας του βίντεο.



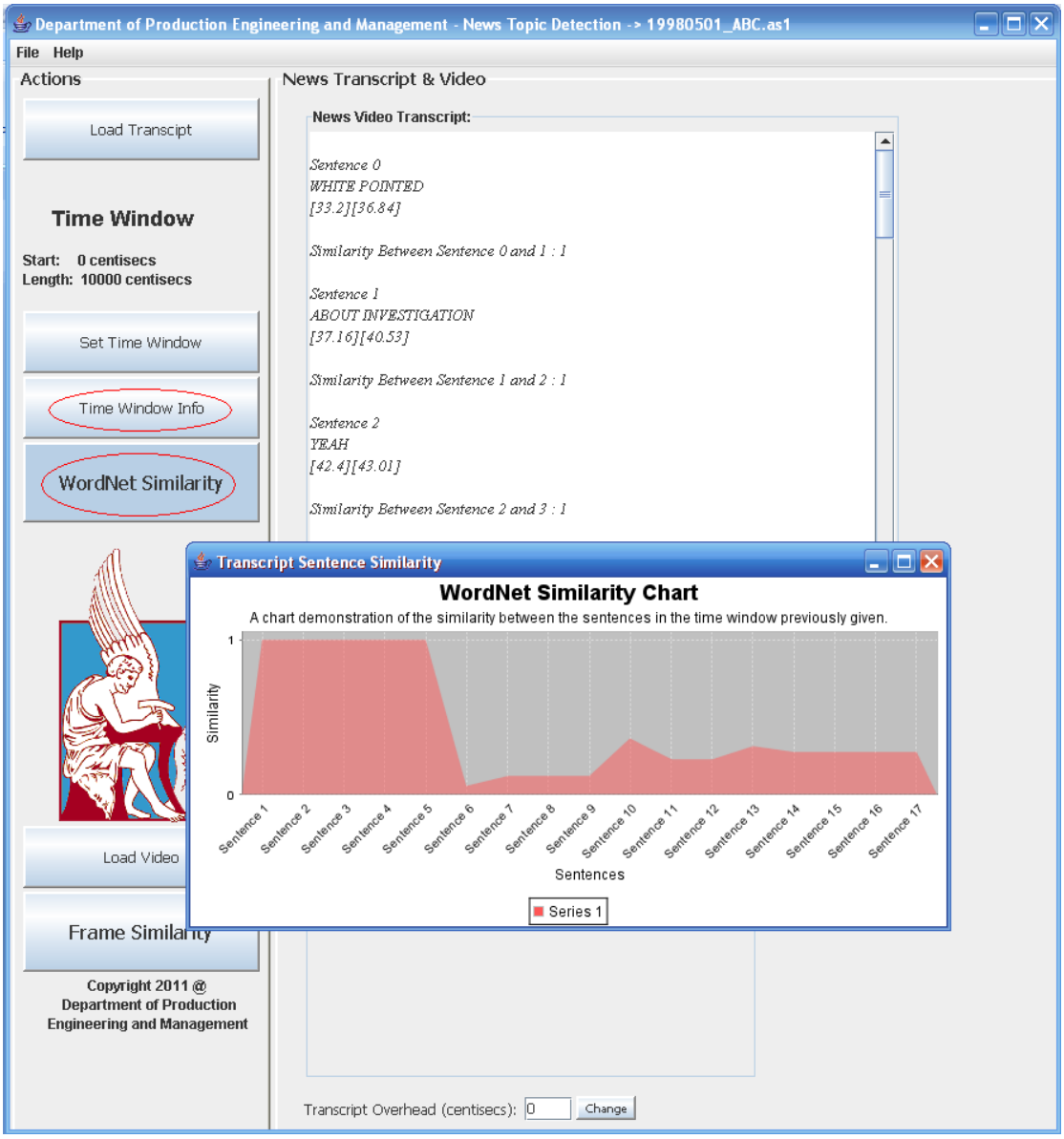
Για την επιλογή του επιθυμητού διαστήματος το σύστημα έχει σχεδιαστεί ώστε ο χρήστης ουσιαστικά να ορίζει ένα χρονικό «παράθυρο» (Time Window). Συγκεκριμένα, μέσω κατάλληλου γραφικού εργαλείου ζητούνται 2 τιμές οι οποίες αφορούν στην χρονική στιγμή εκκίνησης του χρονικού «παράθυρου» ('Window Start Time') και την χρονική διάρκειά του ('Window Length'). Οι τιμές αυτές αντιπροσωπεύουν εκατοστά του δευτερολέπτου δεδομένου ότι η ανάλυση του βίντεο και της λεκτικής του περιγραφής γίνεται με λεπτομέρεια αυτής της τάξης μεγέθους.

Για παράδειγμα ορίζοντας 'Window Start Time=200' και 'Window Length=10000' το σύστημα καλείται να περιορίσει την ανάλυση μεταξύ του 2^{ου} και 102^{ου} δευτερολέπτου του βίντεο.

Η παραπάνω τεχνική εξυπηρετεί στο να μπορεί ο χρήστης να κάνει στοχευμένη ανάλυση συνάφειας, με λεπτομέρεια της τάξης του εκατοστού του δευτερολέπτου, δίχως να χάνεται χρόνος για την ανάλυση του συνόλου των προτάσεων σε κάθε δοκιμή.

Μετά την επιλογή της λεκτικής περιγραφής και του ορισμού του επιθυμητού 'Time Window', το σύστημα μπορεί να προχωρήσει στον υπολογισμό της συνάφειας των προτάσεων που περιλαμβάνονται εντός των χρονικών ορίων. Ο χρήστης, στο σημείο αυτό, έχει να επιλέξει μεταξύ 2 διαφορετικών τρόπων αναπαράστασης των αποτελεσμάτων. Με την παρουσίασή τους σε μορφή κειμένου στο ειδικό πλαίσιο όπως έγινε και σε προηγούμενες λειτουργίες (κουμπί 'Time Window Info'), είτε με την διαγραμματική τους αναπαράσταση (κουμπί 'WordNet Similarity').

Καθεμιά από τις παραπάνω αναπαραστάσεις έχει τα δικά της πλεονεκτήματα. Η αναπαράσταση σε κείμενο βοηθά στην εξέταση των αποτελεσμάτων πρόταση ανά πρόταση, αφού περιλαμβάνει ακριβείς πληροφορίες για το περιεχόμενο και την χρονική υπόσταση καθεμιάς από τις προτάσεις αλλά και για το αποτέλεσμα των συγκρίσεων ανά 2. Η αναπαράσταση σε διάγραμμα από την άλλη δίνει μια συνολική επισκόπηση των αποτελεσμάτων των συγκρίσεων βοηθώντας στον οπτικό εντοπισμό των σημείων πιθανούς αλλαγής θέματος.

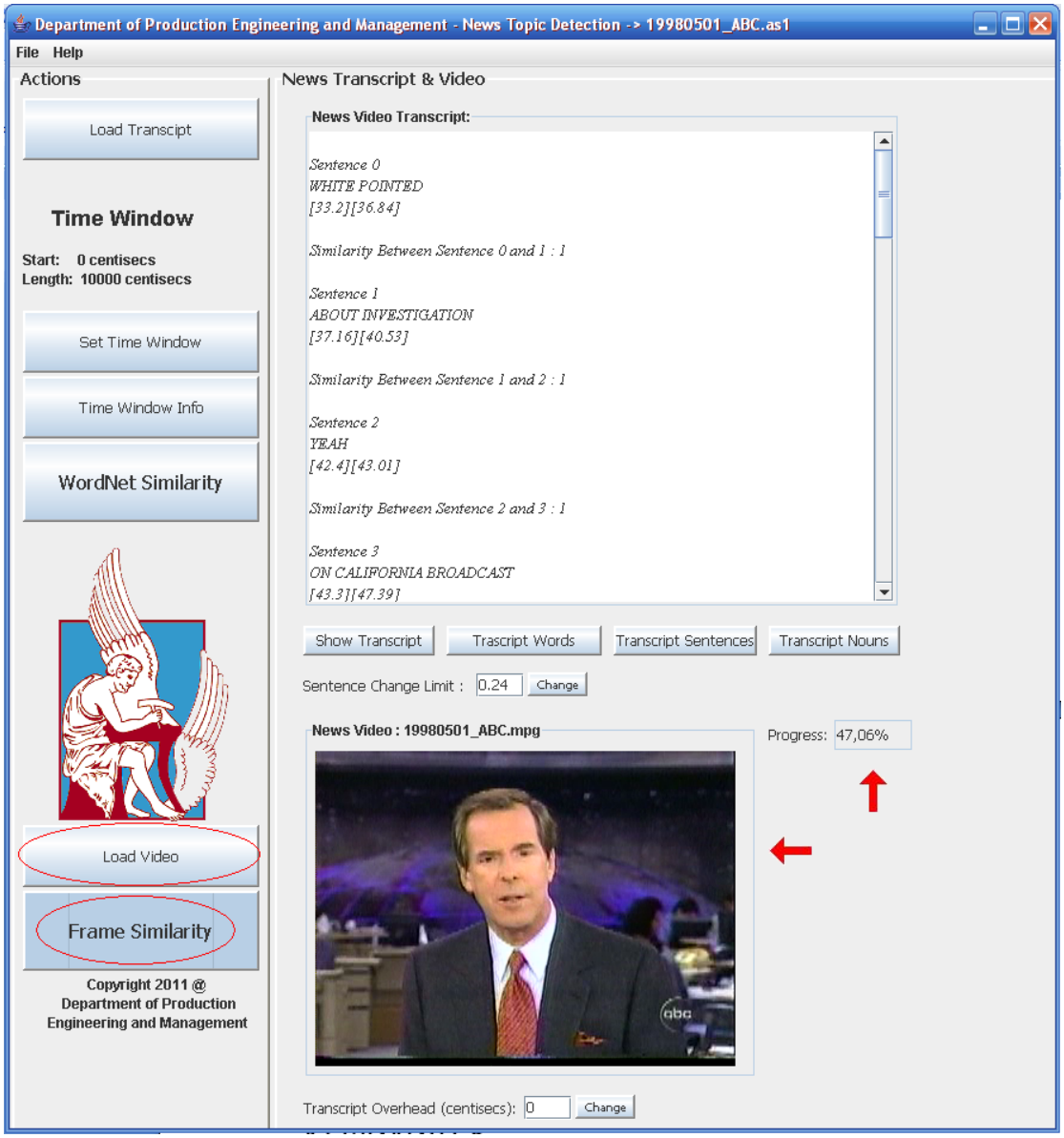


Κουμπί	Λειτουργία	Περιγραφή
Set Time Window	Ορισμός Χρονικού Παραθύρου	-Άνοιγμα ειδικού παράθυρου διαλόγου για τον ορισμό σε εκατοστά του δευτερολέπτου του χρονικού «παράθυρου» ανάλυσης των προτάσεων.
Time Window Info	Εμφάνιση σε κείμενο αποτελεσμάτων συνάφειας	Υπολογισμός της συνάφειας των προτάσεων ανά 2 με χρήση του WordNet και εμφάνιση λεπτομερών αποτελεσμάτων στο ειδικό πλαίσιο.
WordNet Similarity	Εμφάνιση σε διάγραμμα αποτελεσμάτων συνάφειας	Υπολογισμός της συνάφειας των προτάσεων ανά 2 με χρήση του WordNet και εμφάνιση διαγράμματος αποτελεσμάτων

7.1.3 Επιλογές Ανάλυσης πλαισίων βίντεο

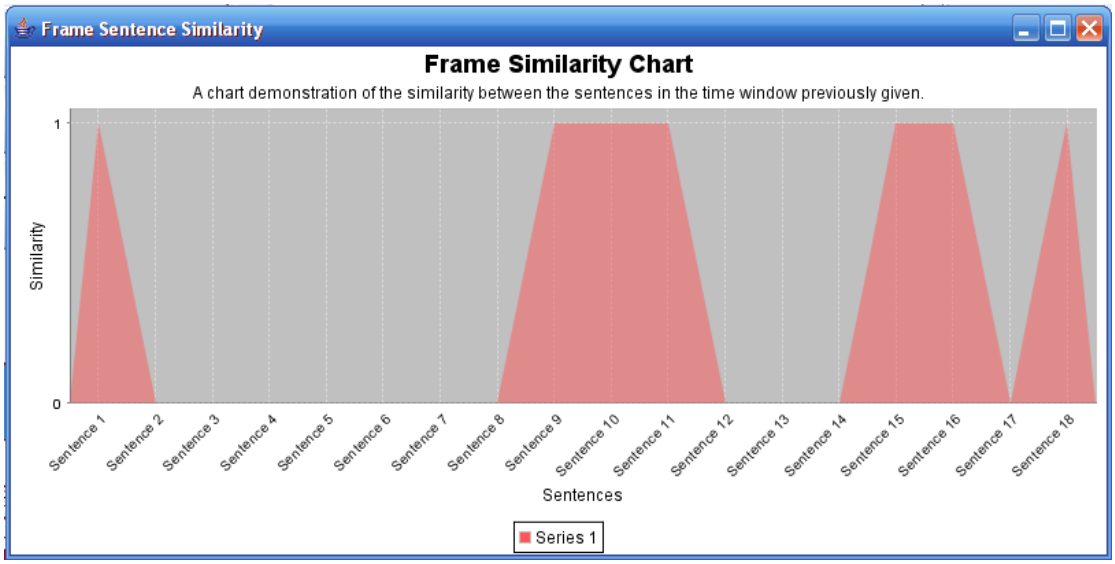
Για την ανάλυση των πλαισίων βίντεο (video frames) βασική προϋπόθεση είναι η εκτέλεση των λειτουργιών ανάλυσης της λεκτικής περιγραφής (ενότητα 7.1.1) και του ορισμού του 'Time Window'. Αυτό οφείλεται στο ότι η ανάλυση των video frames χρησιμοποιεί τις πληροφορίες για την χρονική διάρκεια κάθε πρότασης και ομοίως με τη συνάφεια προτάσεων, η ανάλυση περιορίζεται στα χρονικά πλαίσια που ορίζει ο χρήστης.

Στη συνέχεια ο χρήστης καλείται να επιλέξει το «φόρτωμα» του υπό εξέταση βίντεο. Με το πάτημα του κουμπιού 'Load Video' εμφανίζεται ειδικό παράθυρο αναζήτησης του εκάστοτε βίντεο. Μετά την επιτυχή «φόρτωση» η διαδικασία ανάλυσης μπορεί να ξεκινήσει άμεσα με το πάτημα του κουμπιού 'Frame Similarity'. Η ανάλυση αυτή αφορά στην σύγκριση των 'video frames`, τα οποία βρίσκονται εντός συγκεκριμένων χρονικών διαστημάτων. Περισσότερα για την ανάλυση των video frames στην ενότητα 6.1.2.



Κατά τη διάρκεια της ανάλυσης των video frames από το σύστημα, το γραφικό περιβάλλον φροντίζει για την ενημέρωση του χρήστη για την εξέλιξη της διαδικασίας. Συγκεκριμένα, σε ειδικό πλαίσιο παρουσιάζονται τα εκάστοτε υπό εξέταση video frames με την ταυτόχρονη εμφάνιση μιας ένδειξης για το ποσοστό της ολοκληρωμένης διαδικασίας (‘Progress’).

Με το πέρας της ανάλυσης η ένδειξη ‘Progress’ παίρνει την τιμή ‘Completed’ και εμφανίζεται διάγραμμα με τα αποτελέσματα. Το διάγραμμα αυτό σε συνδυασμό με το διάγραμμα της συνάφειας των προτάσεων από παραπάνω δίνει στον χρήστη άμεσα μια διπλή οπτική για το πού μπορεί να λαμβάνει στο βίντεο χώρα μια πιθανή αλλαγή θέματος.



Πληροφορίες για το πώς αναλύεται το παραπάνω διάγραμμα στην ενότητα 6.2.2.

Κουμπί	Λειτουργία	Περιγραφή
Load Video	Φόρτωση βίντεο	Άνοιγμα παράθυρου αναζήτησης αρχείου στο 'Documents and Settings'.
Change (Transcript Overhead)	Καθορισμός τιμής διαστήματος συγχρονισμού	Άνοιγμα ειδικού παράθυρου για τον ορισμό νέας τιμής σε εκατοστά του δευτερόλεπτου
Frame Similarity	Εμφάνιση αποτελεσμάτων σύγκρισης πλαισίων βίντεο	-Υπολογισμός της συνάφειας των «κρίσιμων» πλαισίων ανά 2 -εμφάνιση πληροφοριών προόδου διαδικασίας και των υπό εξέταση πλαισίων -εμφάνιση διαγράμματος αποτελεσμάτων

7.2 Εσωτερική Δομή Γραφικού Περιβάλλοντος Διεπαφής

Για καθεμία από τις επιλογές που προσφέρει το γραφικό περιβάλλον στο χρήστη, το σύστημα εσωτερικά καλεί τις συναρτήσεις που υλοποιούν την αντίστοιχη λειτουργικότητα. Ο σχεδιασμός γραφικού περιβάλλοντος και εσωτερικών λειτουργιών είναι τέτοιος που επιτρέπει την εύκολη μεταξύ τους αντιστοίχιση, προσθέτοντας ακόμη ένα στοιχείο στα modular χαρακτηριστικά του συστήματος.

Οι λειτουργίες του γραφικού περιβάλλοντος, όπως ήδη αναφέρθηκε, αφορούν στην διαχείριση της επεξεργασίας της λεκτικής περιγραφής, της συνάφειας μεταξύ προτάσεων και της επεξεργασίας των πλαισίων βίντεο. Για την εσωτερική υλοποίηση των παραπάνω συνδυάζονται, η κλάση '*Frame1*', κλάση η οποία είναι συνολικά υπεύθυνη για την δημιουργία και εμφάνιση του γραφικού περιβάλλοντος, με την κλάση '*TopicDetection*', που υλοποιεί τις συναρτήσεις επεξεργασίας της λεκτικής περιγραφής και τις δομές αποθήκευσης των πληροφοριών που προκύπτουν. Επίσης η κλάση '*Frame1*' χρησιμοποιεί την κλάση '*ReadFromVideo*' για την διαχείριση της ροής των πλαισίων βίντεο. Η τελευταία σε συνεργασία με την κλάση '*ImageCompare*' υλοποιεί την σύγκριση των πλαισίων βίντεο. Τέλος τα διαγράμματα αναπαράστασης των τελικών αποτελεσμάτων υλοποιούνται από την κλάση '*Chart*'.

Παρακάτω παρατίθενται λεπτομέρειες για την εσωτερική δομή των λειτουργιών του γραφικού περιβάλλοντος.

7.2.1 Επεξεργασία λεκτικής περιγραφής

- Φόρτωση λεκτικής περιγραφής ('Load transcript')

  -Άνοιγμα παράθυρου αναζήτησης αρχείου στο C:\.

```
T_Det.transcript_str = readTranscriptFromFile();
```

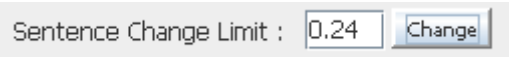
  T_Det : Στιγμιότυπο της κλάσης TopicDetection
  transcript_str : μεταβλητή τύπου String όπου φυλάσσεται το transcript
  readTranscriptFromFile() : συνάρτηση που ανοίγει παράθυρο αναζήτησης του transcript

  -Εμφάνιση του transcript στο ειδικό πλαίσιο

```
disptext.setText(T_Det.transcript_str);
```

  disptext : Ειδικό τύπου πλαίσιο όπου εμφανίζεται το transcript και είναι τύπου JTextArea

- Καθορισμός τιμής ορίου αλλαγής προτάσεων ('Change')



  Με το πάτημα του κουμπιού η μεταβλητή 'Sentence_change_limit' παίρνει την τιμή του παραθύρου δίπλα (σε δευτερόλεπτα) ως εξής:

```
T_Det.Sentence_change_limit = (new Float(changeSentenceLimit.getText()).floatValue());
```

- Εμφάνιση λεκτικής περιγραφής (‘Show transcript’)

Εμφάνιση του transcript στο ειδικό πλαίσιο, ομοίως με παραπάνω

```
disptext.setText(T_Det.transcript_str);
```

- Εμφάνιση Λέξεων λεκτικής περιγραφής (‘Transcript Words’)

-Επεξεργασία transcript για εξαγωγή λέξεων

```
T_Det.transcript_words_str = T_Det.extractWords(T_Det.transcript_str);
```

extractWords() : Η συνάρτηση της κλάσης TopicDetection, παίρνει ως όρισμα το transcript ως έχει και επιστρέφει μόνο τις λέξεις που περιέχει.

-Εμφάνιση λέξεων στο ειδικό πλαίσιο

```
disptext.setText(T_Det.transcript_words_str);
```

- Εμφάνιση προτάσεων λεκτικής Περιγραφής (‘Load Sentences’)

-Ομαδοποίηση λέξεων σε προτάσεις σύμφωνα με το όριο αλλαγής προτάσεων.

```
T_Det.parseTranscript(T_Det.transcript_str);
T_Det.transcript_sentences_str = T_Det.printSentences();
```

parseTranscript() : Η συνάρτηση που με χρήση του ‘Sentence_Change_Limit’ ομαδοποιεί τις λέξεις του transcript σε προτάσεις και τις αποθηκεύει μαζί με τις πληροφορίες για την χρονική στιγμή αρχής και τέλους τους, σε ειδικές δομές τις κλάσης TopicDetection.

printSentences() : Ειδική συνάρτηση που επιστρέφει σε String τις προτάσεις του transcript σύμφωνα με τις πληροφορίες που εκμαίευσε η συνάρτηση ‘parseTrancript()’.

-Εμφάνιση των προτάσεων και πληροφοριών χρονικής διάρκειας στο ειδικό πλαίσιο

```
disptext.setText(T_Det.transcript_sentences_str);
```

- Εμφάνιση ουσιαστικών λεκτικής Περιγραφής (‘Load Nouns’)

-Tagging των προτάσεων με χρήση του Stanford POS-Tagger

```
sentences_taggedwords = T_Det.tagSentences(0, T_Det.Sentences.size() - 1);
```

tagSentences() : Η συνάρτηση που κάνει χρήση της δομής τύπου Vector T_Det.Sentences με τις προτάσεις του transcript και του POS-tagger για να επισημάνει τι μέρος του λόγου είναι καθεμία από τις λέξεις.

-Επιλογή ουσιαστικών από τα αποτελέσματα του tagging

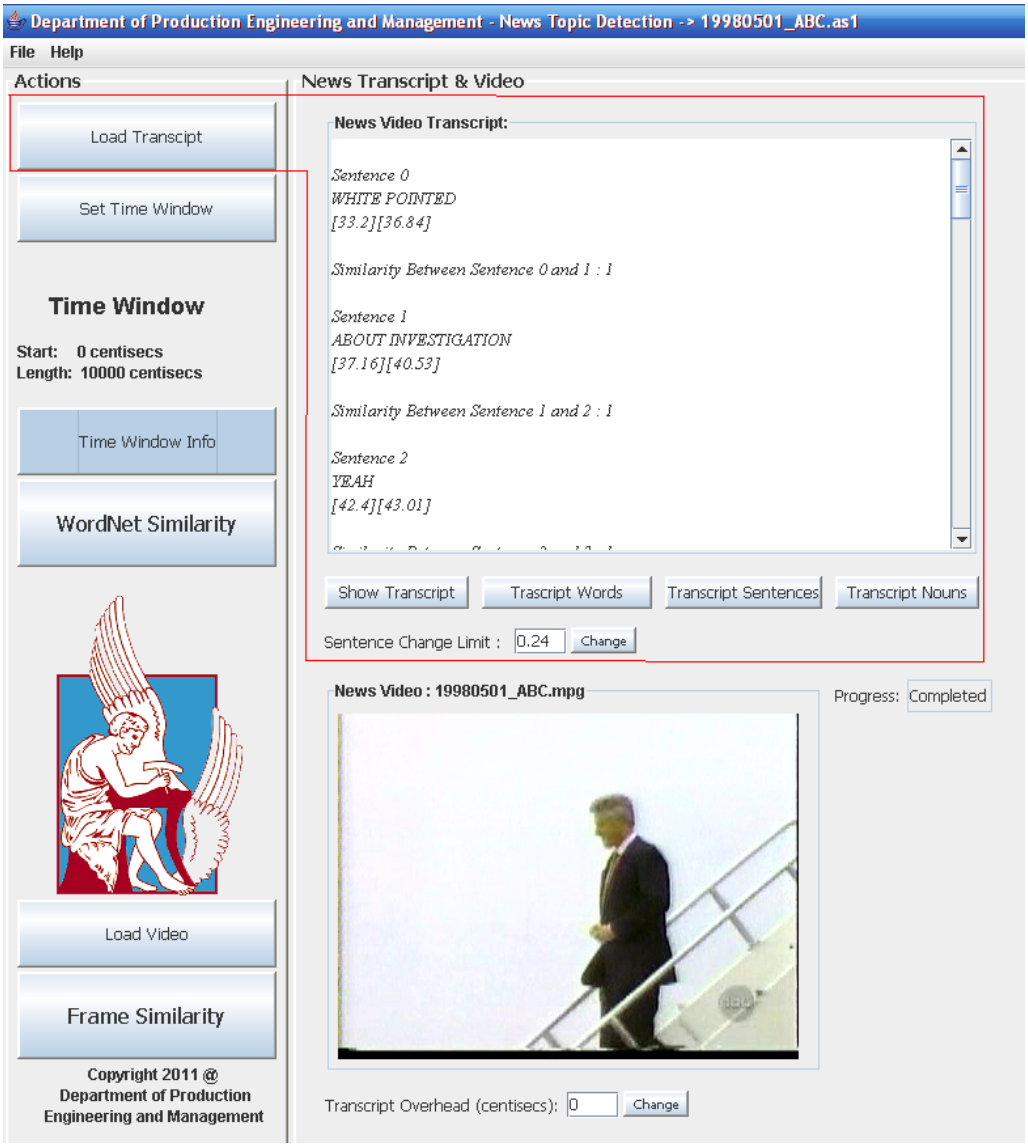
```
sentences_nouns = T_Det.getWindowNouns(sentences_taggedwords);
```

getWindowNouns() : Συνάρτηση που κάνει χρήση της δομής 'Vector sentences_taggedWords' από την προηγούμενη διαδικασία για να επιλέξει και να επιστρέψει αποκλειστικά τα ουσιαστικά ανά πρόταση.

-Εμφάνιση ουσιαστικών του transcript στο ειδικό πλαίσιο

```
disptext.setText(trans_nouns);
```

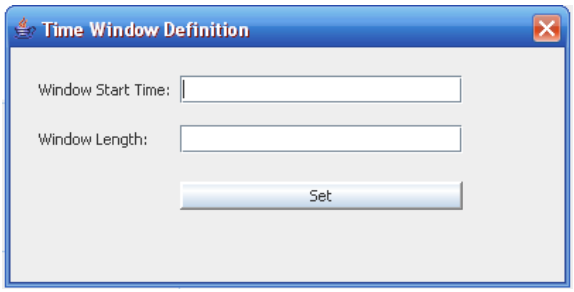
trans_nouns : String με τα ουσιαστικά του 'Vector sentences_nouns' από την προηγούμενη διαδικασία



7.2.2 Συνάφεια μεταξύ προτάσεων

- Ορισμός Χρονικού Παραθύρου (‘Set Time Window’)

  Άνοιγμα ειδικού παράθυρου διαλόγου για τον ορισμό σε εκατοστά του δευτερολέπτου του χρονικού «παράθυρου» ανάλυσης των προτάσεων.




```
LoginDialog ld=new LoginDialog(this,"Time Window Definition",true);

T_Det.StartTime = (new Integer(loginInfo[0])).intValue();
T_Det.TimeWindow = (new Integer(loginInfo[1])).intValue();
```

Δημιουργείται στιγμιότυπο της ειδικού σκοπού κλάσης ‘LoginDialog’ η οποία εμφανίσει το παραπάνω παράθυρο και είναι υπεύθυνη για την καταχώρηση τιμών στις μεταβλητές ‘T_Det.StartTime’ και ‘T_Det.TimeWindow’ για την αρχή και την διάρκεια του επιθυμητού χρονικού «παράθυρου».

- Εμφάνιση σε κείμενο αποτελεσμάτων συνάφειας (‘Time Window Info’)

  Υπολογισμός της συνάφειας των προτάσεων ανά 2 με χρήση του WordNet και εμφάνιση λεπτομερών αποτελεσμάτων στο ειδικό πλαίσιο.

```
disptext.setText(T_Det.getTimeWindowInfo());
```

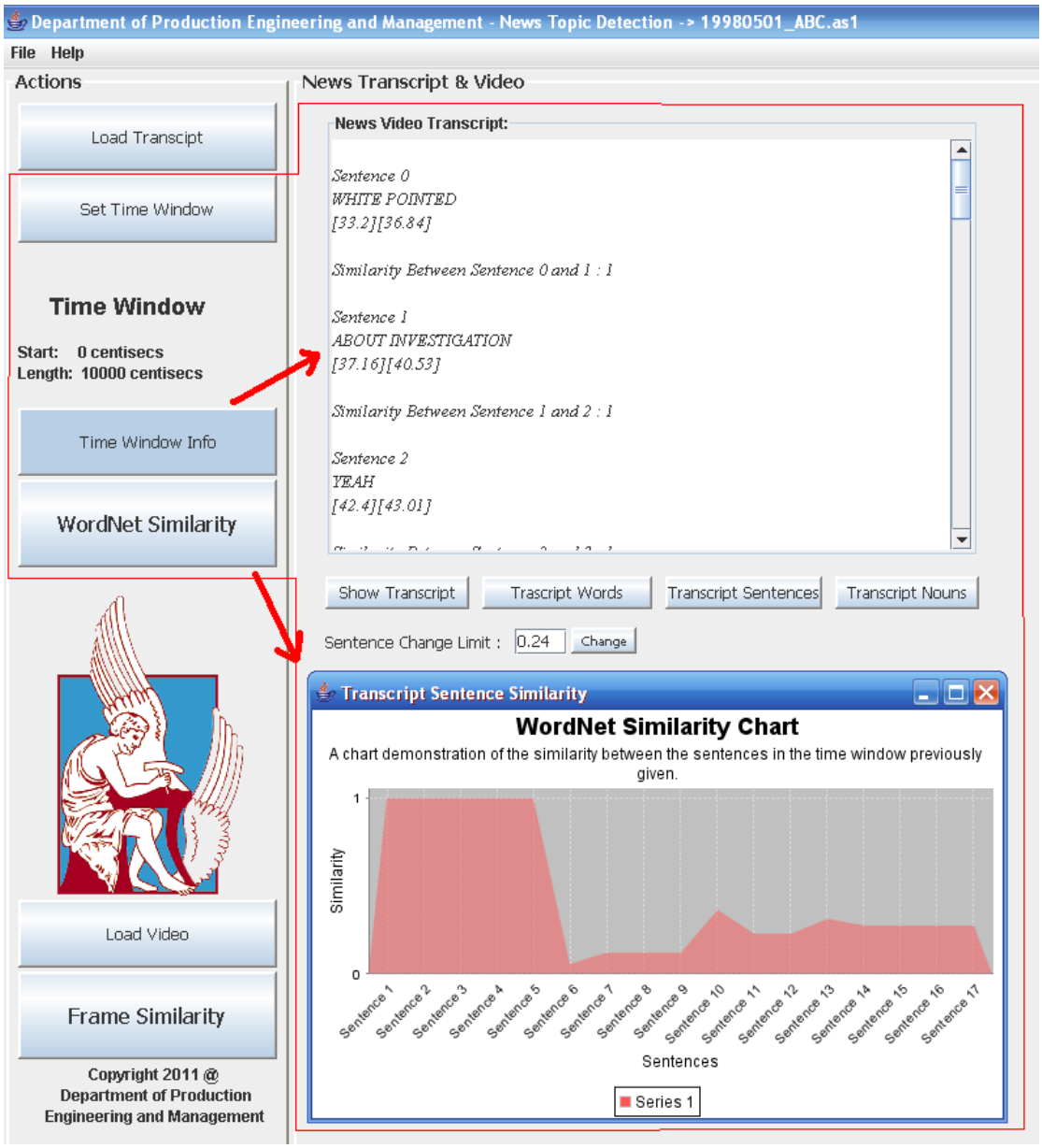
getTimeWindowInfo () : Συνάρτηση που κάνει χρήση των συναρτήσεων ‘tagSentences()’ και ‘getWindowNouns()’ για την εξαγωγή των ουσιαστικών ανά προτάσεις του χρονικού «παράθυρου» όπως ορίστηκε παραπάνω. Στη συνέχεια με την συνάρτηση ‘getNounSimilarity()’ κάνει χρήση του WordNet για τον υπολογισμό της συνάφειας των παραπάνω προτάσεων ανά 2 και επιστρέφει String με αποτελέσματα.

- Εμφάνιση σε διάγραμμα αποτελεσμάτων συνάφειας (‘WordNet Similarity’)

  Υπολογισμός της συνάφειας των προτάσεων ανά 2 με χρήση του WordNet και εμφάνιση διαγράμματος αποτελεσμάτων

```
demo = new Chart("Transcript Sentence Similarity",
                 num, "WordNet Similarity Chart");
```

Υπολογισμός της συνάφειας των προτάσεων του χρονικού «παράθυρου» όπως παραπάνω και αποθήκευση των αποτελεσμάτων στην δομή ‘double num[][]’. Έπειτα δημιουργία στιγμιότυπου της κλάσης Chart με όρισμα την ‘num’ για την εμφάνιση ειδικού διαγράμματος με τα αποτελέσματα.



7.2.3 Επεξεργασία πλαισίων βίντεο

- Φόρτωση λεκτικής περιγραφής (‘Load transcript’)

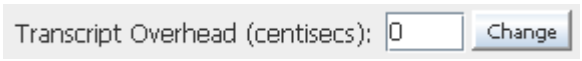
Άνοιγμα παράθυρου αναζήτησης αρχείου στο ‘Documents and Settings’.

```
videoPath = selectVideoPath();
```

Videopath : Μεταβλητή τύπου String όπου αποθηκεύεται το ‘path’ στο σκληρό δίσκο για το βίντεο

selectVideoPath() : Συνάρτηση που ανοίγει παράθυρο αναζήτησης του βίντεο

- Καθορισμός τιμής διαστήματος συγχρονισμού (‘Change’)



Με το πάτημα του κουμπιού η μεταβλητή ‘Overhead’ παίρνει την τιμή (σε εκατοστά του δευτερολέπτου) του παράθυρου δίπλα ως εξής:

```
T_Det.Overhead = (new Float(changeOverhead.getText()).intValue());
```

- Εμφάνιση αποτελεσμάτων σύγκρισης πλαισίων βίντεο (‘Frame Similarity’)

Υπολογισμός της συνάφειας των «κρίσιμων» πλαισίων βίντεο ανά 2 και εμφάνιση διαγράμματος αποτελεσμάτων

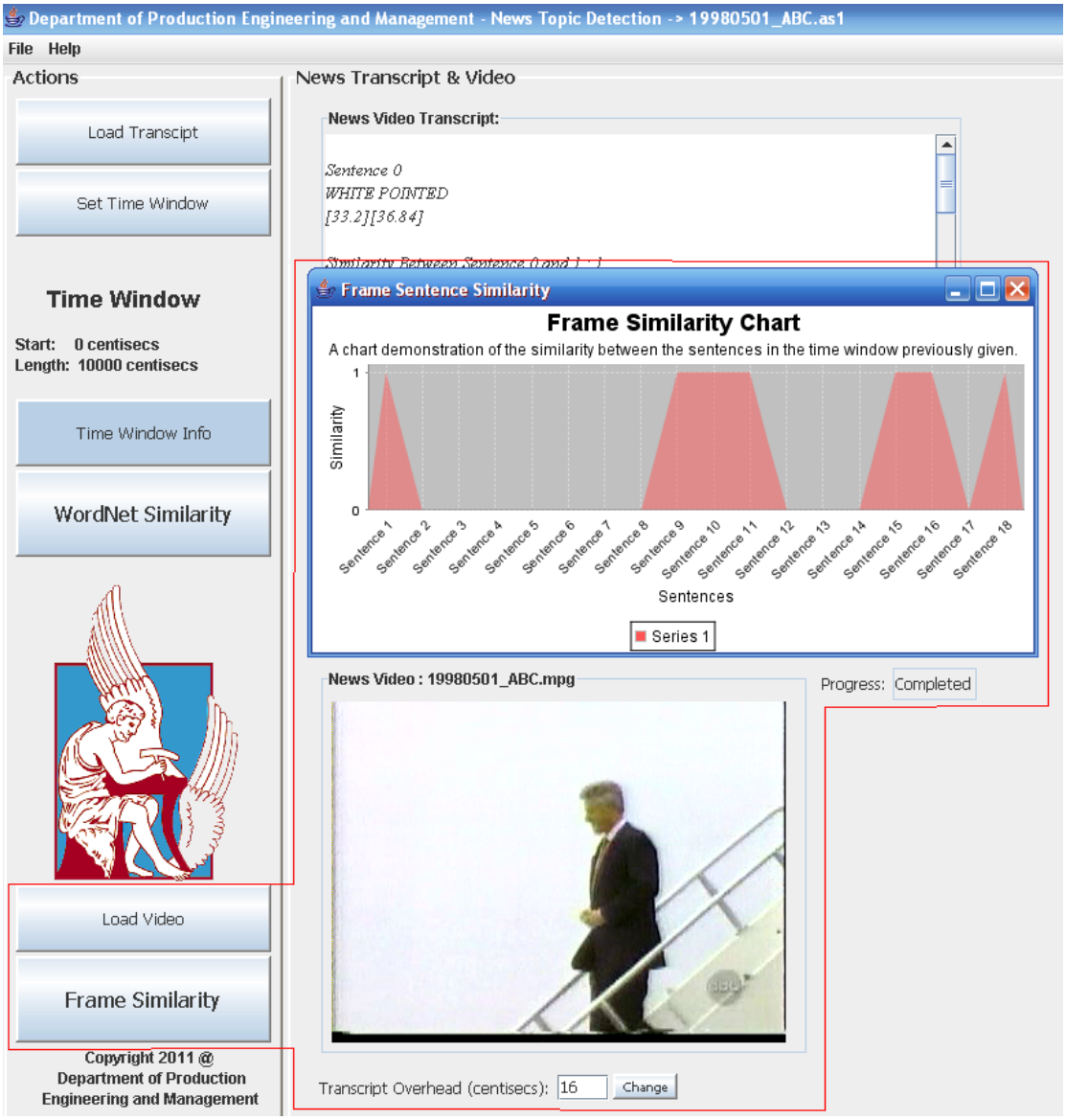
```
video_processing = new ReadFromVideo("file:" + videoPath,this);
videoPanel.add(video_processing.currPanel);
```

Video_processing : Στιγμιότυπο της κλάσης ReadFromVideo στην οποία ανατίθεται η διαχείριση της ροής των πλαισίων βίντεο.

videoPanel : ειδικό πλαίσιο εμφάνισης του εκάστοτε προς επεξεργασία πλαίσιο βίντεο (‘currPanel’)

```
demo = new Chart("Frame Sentence Similarity",
                 Video_Frame_Similarity,"Frame Similarity Chart");
```

Η ReadFromVideo εσωτερικά, με την βοήθεια της κλάσης ImageCompare υπολογίζει τις ομοιότητες στα «κρίσιμα» πλαίσια βίντεο στο υπό επεξεργασία χρονικό διάστημα του βίντεο. Τα αποτελέσματα τα αποθηκεύει στην ειδική δομή ‘Video_Frame_Similarity’. Τέλος δημιουργεί στιγμιότυπο της Chart, η οποία δημιουργεί διάγραμμα για την αναπαράσταση των παραπάνω αποτελεσμάτων.



Κεφάλαιο 8

Απαιτούμενο Λογισμικό

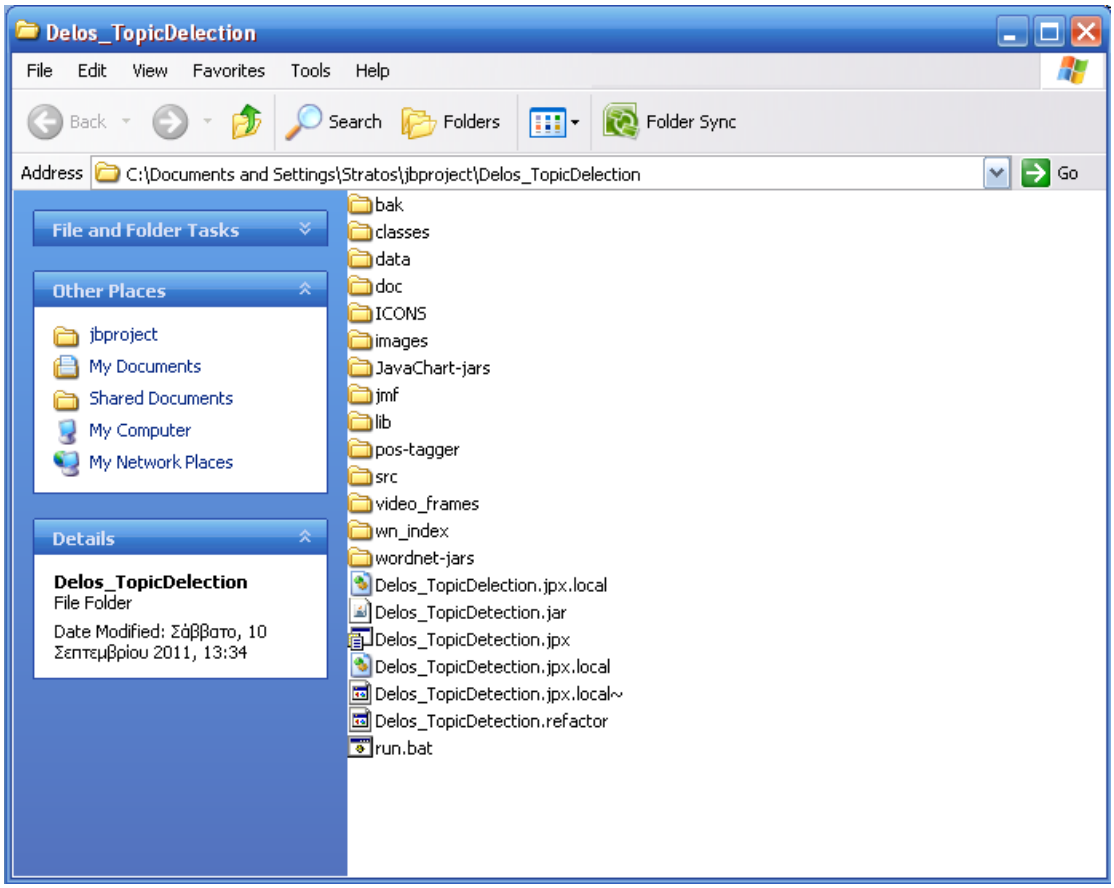
Ο σχεδιασμός της συγκεκριμένης εφαρμογής έχει γίνει με χρήση των κανόνων και των ιδιαίτερων χαρακτηριστικών της γλώσσας προγραμματισμού Java. Για την υλοποίηση έχει χρησιμοποιηθεί το εργαλείο ανάπτυξης λογισμικού JBuilder στην έκδοση 2006. Για εργαλεία Stanford POS-Tagger και WordNet που ενσωματώθηκαν επιλέχθηκαν οι υλοποιήσεις τους σε Java. Τέλος, επιλέχθηκε η πλατφόρμα Java Media Framework (JMF) για τις λειτουργίες ανάλυσης και διαχείρισης της ροής πλαισίων βίντεο.

Η χρήση της πλατφόρμας ανάπτυξης JBuilder προσδίδει στο σύστημα όλα τα πλεονεκτήματα μιας άρτιας εσωτερικής οργάνωσης στο αναπτυχθέν λογισμικό, της εύκολης ενσωμάτωσης νέων εργαλείων και της δυνατότητας χειρισμού ειδικών ρυθμίσεων. Χαρακτηριστική είναι η ευκολία στην ανάπτυξη και στη δοκιμαστική λειτουργία του υπό ανάπτυξη λογισμικού, αλλά και η απεξάρτηση της τελικής εφαρμογής από την πλατφόρμα ανάπτυξης.

8.1 Εσωτερική οργάνωση κώδικα

8.1.1 Εσωτερική δομή αρχείων και φακέλων

Ειδικός φάκελος του συστήματος εμπεριέχει όλα τα απαιτούμενα αρχεία και τις βιβλιοθήκες για την λειτουργία της εφαρμογής (ΠΑΡΑΡΤΗΜΑ Α).



- Φάκελος ‘classes’
  Περιέχει τα «εκτελέσιμα» αρχεία της εφαρμογής. Τα αρχεία αυτά είναι αποτέλεσμα της μεταγλώττισης (compilation) του κώδικα ανάπτυξης.

- Φάκελος ‘data’
  Περιέχει αρχεία για την δοκιμαστική λειτουργία του συστήματος

- Φάκελος ‘ICONS’
  Περιέχει εικόνες που χρησιμοποιεί το GUI

- Φάκελος ‘images’ και ‘video_frames’
  Φάκελοι προσωρινής αποθήκευσης των αποτελεσμάτων της ανάλυσης των πλαισίων βίντεο.

- Φάκελος ‘JavaChart-jars’
  Περιέχει τις βιβλιοθήκες για την διαγραμματική απεικόνιση των αποτελεσμάτων ανάλυσης.

- Φάκελος ‘jmf’
  Περιέχει τις βιβλιοθήκες της πλατφόρμα Java Media Framework που ενσωματώθηκε.

- Φάκελος ‘lib’
  Περιέχει βοηθητικές βιβλιοθήκες κυρίως για την ανάπτυξη του GUI.

- Φάκελος ‘pos-tagger’
  Περιέχει τις βιβλιοθήκες για το εργαλείο Stanford POS-Tagger.

- Φάκελος ‘src’
  Περιέχει τον κώδικα ανάπτυξης της εφαρμογής.

- Φάκελος ‘wn_index’
  Περιέχει βοηθητικά αρχεία για την λειτουργία του WordNet.

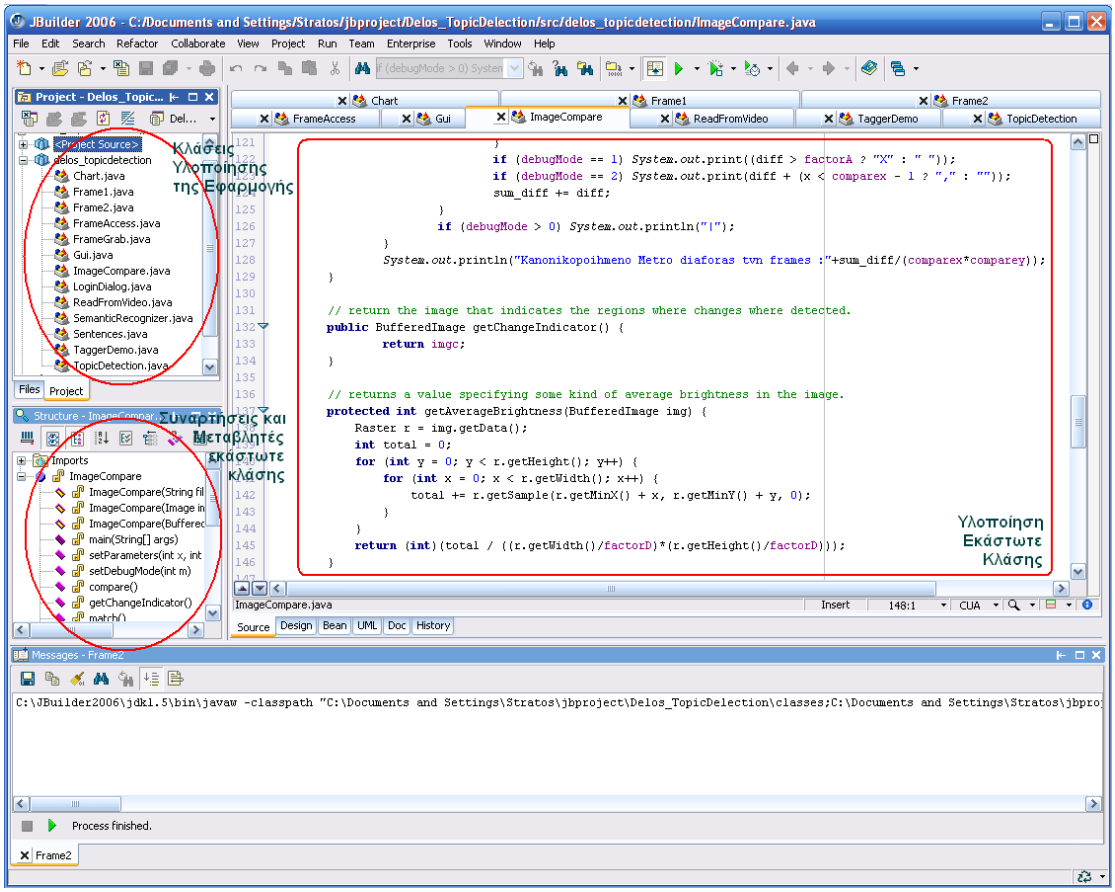
- Φάκελος ‘wordnet-jars’
  Περιέχει τις βιβλιοθήκες για το εργαλείο WordNet.

- Αρχείο ‘Delos_TopicDetection.jpx’
  Το αρχείο διαχείρισης της εργασίας από τον JBuilder

- Αρχείο ‘Delos_TopicDetection.jar’
  Αρχείο αυτόνομης εκτέλεσης της εφαρμογής.

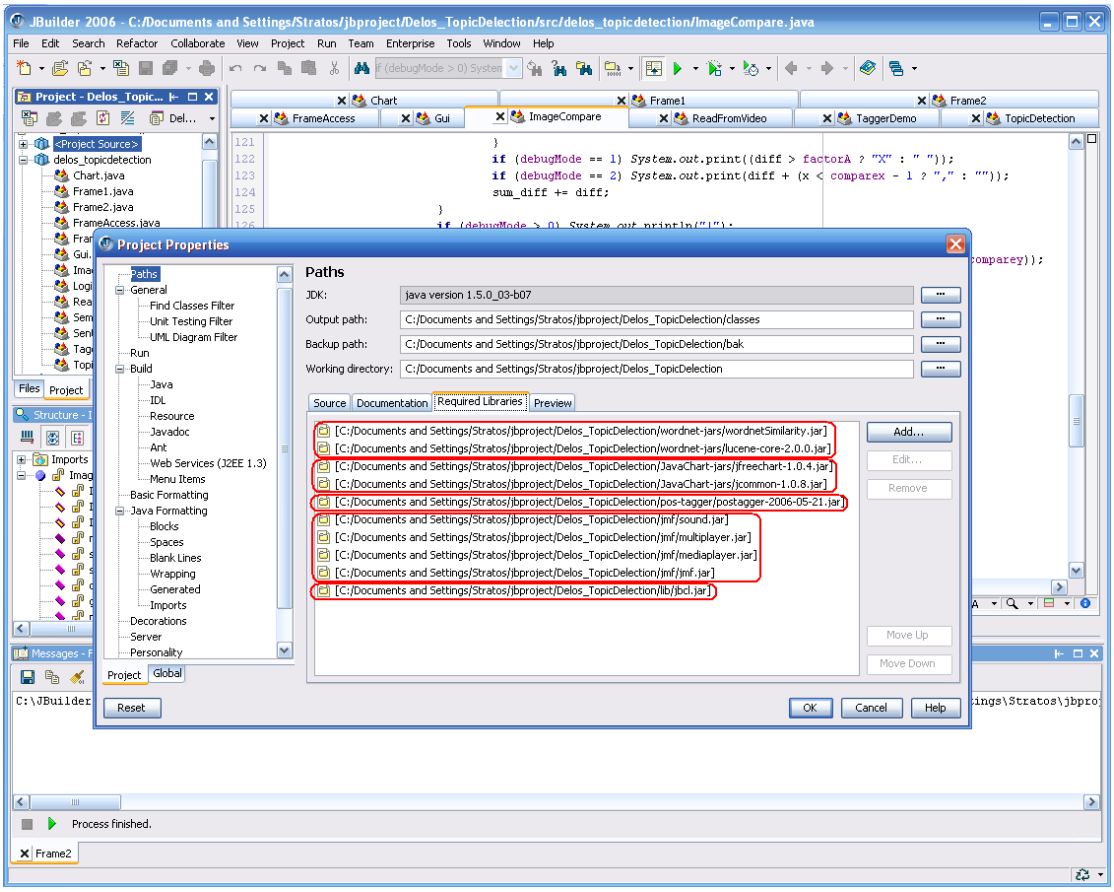
8.1.2 Δομή περιβάλλοντος ανάπτυξης

Η συγκεκριμένη πλατφόρμα διαθέτει εύχρηστο περιβάλλον διαχείρισης των κλάσεων (αρχείων) υλοποίησης της εφαρμογής. Ενδεικτικά παρατίθεται ένα στιγμιότυπο του γραφικού αυτού περιβάλλοντος.

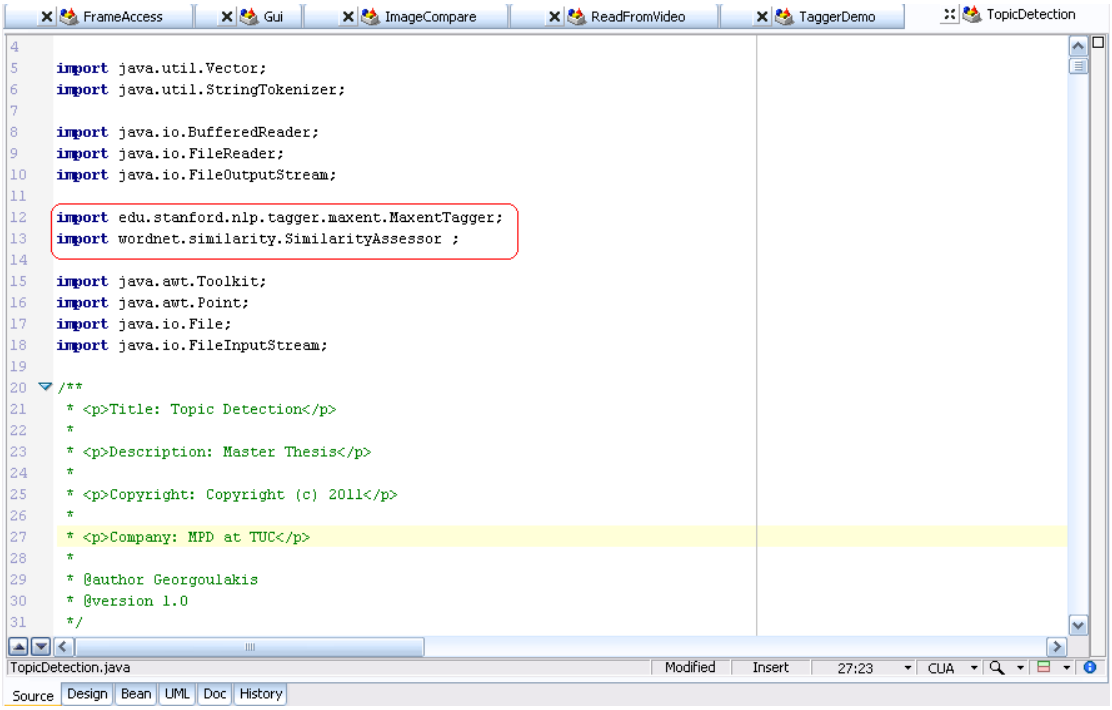


8.2 Ενσωμάτωση εργαλείων

Η ενσωμάτωση εργαλείων κατά τη διάρκεια ανάπτυξης στη γλώσσα προγραμματισμού Java αφορά στην εισαγωγή και χρήση κατάλληλων βιβλιοθηκών που υλοποιούν τα εκάστοτε εργαλεία. Στην εν λόγω εργασία έχουν ενσωματωθεί βιβλιοθήκες για τα Stanford POS-Tagger, WordNet, και JMF. Επίσης γίνεται χρήση βοηθητικών βιβλιοθηκών για την διαγραμματική απεικόνιση των αποτελεσμάτων και της καλύτερης διαχείρισης του γραφικού περιβάλλοντος διεπαφής χρήστη. Σε προηγούμενη ενότητα είδαμε πού βρίσκονται εσωτερικά στο σύστημα καθεμιά από αυτές ενώ η ενσωμάτωσή τους γίνεται ως εξής:

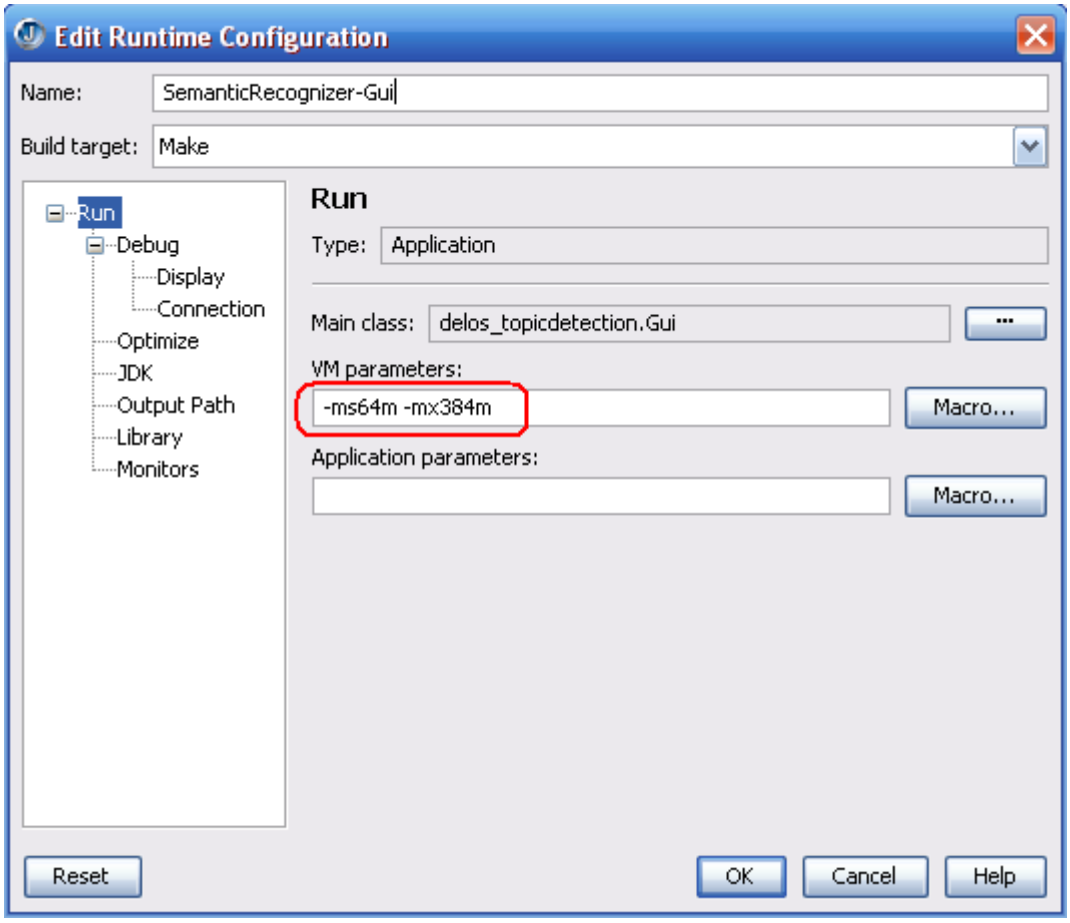


Μετά την ενσωμάτωση των κατάλληλων βιβλιοθηκών οι διάφορες λειτουργίες του εκάστοτε εργαλείου γίνονται διαθέσιμες κατά την διάρκεια ανάπτυξης της εφαρμογής μέσω κατάλληλης κλήσης αυτών από της κλάσεις (αρχεία) υλοποίησης. Για παράδειγμα για μια από τις βασικές κλάσεις υλοποίησης της εφαρμογής την ‘TopicDetection’ έχουμε:



8.3 Χειρισμός ειδικών ρυθμίσεων

Συχνά για την λειτουργία ορισμένων εργαλείων απαιτούνται ειδικές ρυθμίσεις. Συγκεκριμένα, για την αποδοτική λειτουργία του POS-Tagger απαιτείται η δέσμευση περισσότερης μνήμης κατά τη διάρκεια εκτέλεσης. Στην εν λόγω εφαρμογή είναι απαραίτητη η ρύθμιση '-ms64m -mx384m' η οποία δεσμεύει αρχικά 64MB μνήμης και μπορεί να φτάσει μέχρι και τα 384MB κατά τη διάρκεια εκτέλεσης.



Κεφάλαιο 9

Συμπεράσματα

9.1 Αξιολόγηση Συστήματος

Τα εργαλεία που χρησιμοποιούνται για τους υπολογισμούς που διενεργούνται από το σύστημα είναι δοκιμασμένα και με ευρεία απήχηση, με την αποδοτικότητά τους να εξαρτάται από την ποιότητα του υλικού που καλούνται να επεξεργαστούν. Η τεχνική που χρησιμοποιείται για την εύρεση της αλλαγής θεματολογίας κατά την ανάλυση των τριών πρώτων σταδίων είναι ενθαρρυντική με την εφαρμογή να προσφέρει πληθώρα ρυθμίσεων και την δυνατότητα άμεσης δοκιμής των αποτελεσμάτων που αποφέρει καθεμιά από αυτές.

Η πλατφόρμα ανάπτυξης JMF δίνει την δυνατότητα πλήρους διαχείρισης πολυμεσικού υλικού. Η τρέχουσα έκδοση κάνει χρήση

απλής τεχνικής όσον αφορά την ανάλυση των πλαισίων βίντεο σε σχέση με τις τεχνικές που περιγράφονται στην βιβλιογραφία δίνοντας την βάση για μελλοντικές επεκτάσεις.

Η πρωτοτυπία της εφαρμογής έγκειται στην επεκτασιμότητα της εφαρμογής σε κάθε επίπεδο ανάπτυξης. Το γραφικό περιβάλλον κρίνεται εύχρηστο, πλήρες και ευέλικτο στη χρήση. Επίσης ευέλικτο είναι το σύστημα καθεαυτό, με τη modular αρχιτεκτονική στη σχεδίαση κάθε επιπέδου και τη χρήση της γλώσσας προγραμματισμού Java να κάνει τη διαφορά.

9.2 Ρυθμιστικοί Παράγοντες

Η πρόβλεψη της ύπαρξης παραγόντων παρέμβασης από τον χρήστη δίνουν την δυνατότητα εύκολου πειραματισμού και εξαγωγής συμπερασμάτων ακόμη και κατά τη διάρκεια εκτέλεσης. Τέτοιοι παράγοντες αφορούν:

- Το όριο διάρκειας παύσης για σχηματισμό προτάσεων
- Τη χρονική διάρκεια για τον συγχρονισμό βίντεο και λεκτικής περιγραφής
- Την ευαισθησία κατά την σύγκριση πλαισίων βίντεο
- Τα κριτήρια σύγκρισης των πλαισίων βίντεο
- Τον αλγόριθμο διαχείρισης της ροής των πλαισίων βίντεο
- Τον αλγόριθμο για την εξαγωγή μετρικής συνάφειας των προτάσεων

9.3 Μελλοντικές Επεκτάσεις

Μια από τις βασικές επεκτάσεις του συστήματος αφορά στην ενσωμάτωση ενός εργαλείου αναγνώρισης ομιλίας (Speech Recognition) ή κάτι αντίστοιχο για την αυτόματη παραγωγή ποιοτικότερης λεκτικής περιγραφής. Μια καλύτερη λεκτική περιγραφή θα δώσει καλύτερα αποτελέσματα από το πρώτο κιόλας επίπεδο ανάλυσης με ακόμη πιο θεαματικά αποτελέσματα στα επόμενα, όπως περιγράφεται στις αντίστοιχες ενότητες (Κεφάλαιο 4 και 5).

Επίσης η επέκταση του αλγόριθμου εξαγωγής της μετρικής αναφορικά με την συνάφεια των προτάσεων θα έδινε καλύτερα αποτελέσματα

στην πρώτου σταδίου εκτίμηση για την θέση αλλαγής γεγονότος ή θέματος.

Τέλος, ιδιαίτερα ενδιαφέρουσες επεκτάσεις αφορούν στην εξαγωγή λέξεων «κλειδιών» δεδομένου ότι ήδη έχει υλοποιηθεί ο μηχανισμός εξαγωγής και ομαδοποίησης των ουσιαστικών. Επίσης ενδιαφέρουσα είναι και η κατηγοριοποίηση των εντοπισθέντων γεγονότων ή θεμάτων με την βοήθεια των παραπάνω λέξεων «κλειδιών» και ενός συστήματος αντιστοίχισης σε κατηγορίες.

Αναφορές

[1] Karat, Clare-Marie; Vergo, John; Nahamoo, David (2007). "Conversational Interface Technologies". In Sears, Andrew; Jacko, Julie A.. The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications (Human Factors and Ergonomics). Lawrence Erlbaum Associates Inc. ISBN 978-0805858709.

[2] Managing editors: Giovanni Battista Varile, Antonio Zampolli. (1997). Cole, Annie; Mariani, Joseph; Uszkoreit, Hans et al.. eds. Survey of the state of the art in human language technology. Cambridge Studies In Natural Language Processing. XII–XIII. Cambridge University Press. ISBN 0-521-59277-1.

[3] Real time video scene detection and classi®cation John M. Gauch, Susan Gauch, Sylvain Bouix, Xiaolan Zhu Electrical Engineering and Computer Science, The University of Kansas, Lawrence, KS, USA*

[4] Elsevier (2003) 'Video retrieval and summarization' (Internet). Available from: www.elsevier.com/locate/cviu. Accessed 7.12.11

[5] R. Bole, B. Yeo, M. Yeung, Video query: research directions, IBM Journal of Research and Development 42 (2) (1998) 233–252.

[6] N. Dimitrova, H.-J. Zhang, B. Shahraray, I. Sezan, T. Huang, A. Zakhor, Applications of videocontent analysis and retrieval, IEEE Multimedia 9 (3) (2002) 42–55.

[7] M. Lew, N. Sebe, P. Gardner, Video indexing and understanding, in: M. Lew (Ed.), Principles of Visual Information Retrieval, Springer, Berlin, 2001, pp. 163–196.

[8] A. Hauptmann, T.D. Ng, R. Baron, W. Lin, Chen M., M. Derthick, M. Christel, R. Jin, R. Yan, Video classification and retrieval with the Informedia digital video library system, Text Retrieval Conference (TREC02), 2002.

[9] B. Shahraray, Multimedia information retrieval using pictorial transcripts, in: B. Furth (Ed.), Handbook of Multimedia Computing, CRC Press, Boca Raton, FL, 1999, pp. 345–359.

[10] A. Hauptmann, R.V. Baron, M.-Y. Chen, M. Christel, P. Duygulu, C. Huang, R. Jin, W.-H. Lin, T. Ng, N. Moraveji, N. Papernick, C.G.M. Snoek, G. Tzanetakis, J. Yang, R. Yan, and H.D. Wactlar (2004) 'Informedia at TRECVID 2003: Analyzing and Searching Broadcast News Video'

[11] Informedia Site : http://www.informedia.cs.cmu.edu/ Accessed: 8.12.11

[12] Yong Rui, Ziyou Xiong, Regunathan Radhakrishnan, Ajay Divakaran, Thomas S. Huang(2004) 'A Unified Framework for Video Summarization, Browsing & Retrieval'

[13] Stanford Log-linear Part-Of-Speech Tagger Site : http://nlp.stanford.edu/software/tagger.shtml Accessed: 8.12.11

[14] Kristina Toutanova and Christopher D. Manning. 2000. Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger. In Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-2000), pp. 63-70.

[15] Kristina Toutanova, Dan Klein, Christopher Manning, and Yoram Singer. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In Proceedings of HLT-NAACL 2003, pp. 252-259.

[16] Its Shakthydoss's Technical Blog / POS-Tagger : https://shakthydoss.wordpress.com/tag/the-stanford-pos-tagger/ Accessed: 8.12.11

[17 Brown_Corpus at Wikipedia: http://en.wikipedia.org/wiki/Brown_Corpus#Part-of-speech_tags_used Accessed: 8.12.11

[18] Carlos N. Silla Jr.1_, Celso A. A. Kaestner1 , Alex A. Freitas (2003) 'A non-linear topic detection method for text summarization using Wordnet'

[19] Hsin-Hsi Chen, Lun-Wei Ku (2002)'An NLP & IR approach to topic detection'

[20] Kezban Demirtas, Nihan Kesim Cicekli, Ilyas Cicekli, (2010)'Automatic categorization and summarization of documentaries'

[21] Ted Pedersen, Siddharth Patwardhan, Jason Michelizzi (2004) 'WordNet::Similarity: measuring the relatedness of concepts'

[22] Michael Pucher (2004) 'Performance Evaluation of WordNet-based Semantic Relatedness Measures for Word Prediction in Conversational Speech'

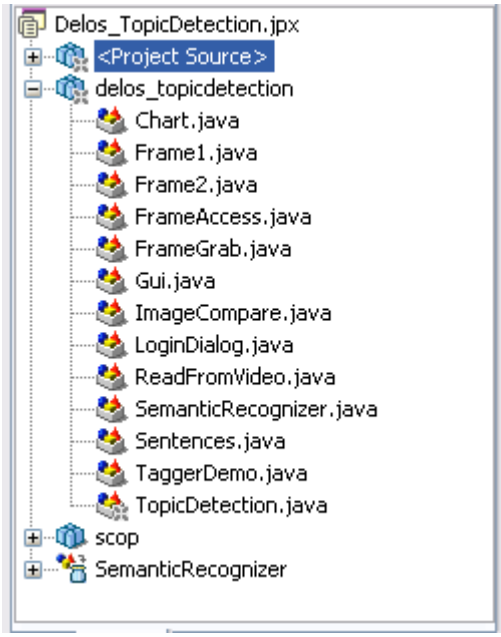
[23] Euripides G.M. Petrakis, Giannis Varelas, Angelos Hliaoutakis, Paraskevi Raftopoulou (2006) 'Design and Evaluation of Semantic Similarity Measures for Concepts Stemming from the Same or Different Ontologies'

[24] Evaluation of Wordnet-based Similarity / Site : http://dissertations.ub.rug.nl/FILES/faculties/arts/2010/t.van.de.cruys/05c5.pdf Accessed: 9.12.11

[25] Tutorial : Getting started with the JavaTM Media Framework Site: http://www.ee.iitm.ac.in/~tgvenky/JMFBook/Tutorial.pdf Accessed: 3.2.11

ΠΑΡΑΡΤΗΜΑ Α

Εσωτερική Δομή Κώδικα Υλοποίησης της Εφαρμογής



Οι βασικές κλάσεις υλοποίησης της εφαρμογής είναι:

- Κλάση ‘Frame2’
  Βοηθητική κλάση για την κλήση του GUI μέσω της συνάρτησης ‘Frame1’

- Κλάση ‘Frame1’
  Υλοποίηση του GUI

- Κλάση ‘LoginDialog’
  Βοηθητική κλάση για την υλοποίηση παράθυρου διαλόγου στο GUI

- Κλάση ‘TopicDetection’
  Κλάση υλοποίησης των λειτουργιών ανάλυσης των τριών πρώτων επιπέδων και αποθήκευσης των εξαγόμενων πληροφοριών σε κατάλληλες δομές

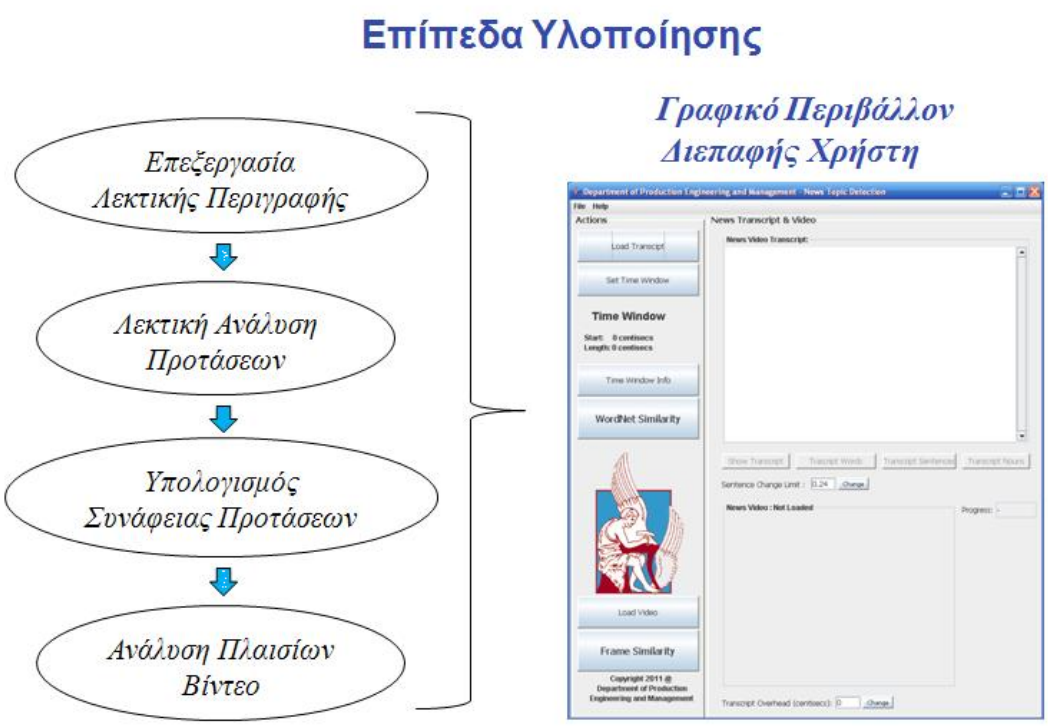
- Κλάση ‘Sentences’

Βοηθητική κλάση της ‘TopicDetection’ για την ομαδοποίηση των λέξεων σε προτάσεις

- Κλάση ‘ReadFromVideo’
  Κλάση υλοποίησης των λειτουργιών του τέταρτου επιπέδου ανάλυσης. Διαχείριση της ροής των πλαισίων βίντεο και της σύγκρισης ορισμένων από αυτά.

- Κλάση ‘ImageCompare’
  Υλοποίηση αλγορίθμου σύγκρισης. Βοηθητική της κλάσης ‘ReadFromVideo’.

Οι υπόλοιπες κλάσεις αναπτύχθηκαν για τον πειραματισμό της χρήσης διάφορων λειτουργιών και δεν χρησιμεύουν στην τελική έκδοση της εφαρμογής.



ΠΑΡΑΡΤΗΜΑ Β

'Critical' Video Frames Analysis Function

```
private void useFrameData(Buffer inBuffer) {
    try {
        printDataInfo(inBuffer);
        if (inBuffer.getData() != null) {
            if(showParsedFrames) {
                if (outvid == null)
                    outvid = new int[imgWidth * imgHeight];

                outdataBuffer(outvid, (byte[]) inBuffer.getData());
                setImage(outvid);  // Edw pairnei timh to outputImage

                if (frame_Prev==null && frame_Curr==null && outputImage!=null)
                    frame_Prev = outputImage;
                else if (frame_Curr==null && outputImage!=null){
                    frame_Curr = outputImage;
                    long cur_frame_id = inBuffer.getSequenceNumber();
                    System.out.println("Comparison between frame: " + (cur_frame_id-1) + " and "+cur_frame_id);
                    ImageCompare image_comp = new ImageCompare(frame_Prev,frame_Curr);
                    System.out.println("Match: " + image_comp.match());
                    if(!image_comp.match()){
                        Video_Frame_Similarity[0][curr_sentence_id-from_sentence_id] = 0;
                        superFrame.progressField.setText(String.format("%.2f",(float)curr sentence id/(
                                                                    float)num_of_sentences*100)+"%");
                        curr_sentence_id++;
                    }
                    frame_Prev = outputImage;
                    frame_Curr=null;
                }
            }
            if (saveFrame_to_JPG){
                if (sunjava)
                    saveJpeg(outputImage, "image_" + countFr + ".jpg");
                else{
                    if (e == null)
                    {
                        initJpeg( (RGBFormat) inBuffer.getFormat());
                    }
                    byte[] b = fetchJpeg(inBuffer);
                    String filename = "image_" + countFr + ".jpg";
                    makeFile(filename, b);
                }
            }
        }
    } catch (Exception e) { System.out.println(e);}
}
```