

"ΛΟΓΟΠΛΟΗΓΗΣΗ"

Δεκέμβριος 1996

Τεύχος 1

Επιστημονικός Υπεύθυνος:

Καθηγητής Γιώργος Καραγιάννης

Υπεύθυνη Έκδοσης:

Δρ. Ιωάννα Μαλαγαρδή

Συνεργάτες:

Γιώργος Μικρός

Αναστάσιος Πατρικάκος

Δέσποινα Σκούταρη

Γραφίστας:

Άρτεμις Γλάρου

Διεύθυνση:

Ινστιτούτο Επεξεργασίας του Λόγου

Επιδαύρου & Αρτέμιδος

151 25 Παράδεισος Αμαρουσίου

τηλ.: 6800959

fax: 6854270

e-mail: ioanna@ilsp.gr

http:// www.ilsp.gr

Την ευθύνη των κειμένων έχουν οι συγγραφείς.

Η χρηματοδότηση της έκδοσης αυτής έγινε από τα προγράμματα της ΓΓΕΤ "Γραφεία Διαμεσολάβησης" και "Ανθρώπινα Δίκτυα".

Η "Λογοπλοήγηση" διανέμεται δωρεάν.

Πίνακας Περιεχομένων

I. Εισαγωγικό Σημείωμα	σελ. 2
II. Κείμενα Μελών του "Ανθρώπινου Δικτύου Γλωσσικής Τεχνολογίας"	σελ. 2
1. Ινστιτούτο Επεξεργασίας του Λόγου	σελ. 2
2. Ιόνιο Πανεπιστήμιο	σελ. 3
3. Εργαστήριο Τεχνολογίας Γνώσεων και Λογισμικού Ινστιτούτο Πληροφορικής και Τηλεπικοινωνιών ΕΚΕΦΕ "ΔΗΜΟΚΡΙΤΟΣ"	σελ. 4
4. Εθνικό Μετσόβιο Πολυτεχνείο	σελ. 5
5. Δημοκρίτειο Πανεπιστήμιο Θράκης	σελ. 5
6. Πανεπιστήμιο Πατρών	σελ. 6
7. Υπουργείο Εσωτερικών Δημόσιας Διοίκησης και Αποκέντρωσης Γενική Γραμματεία Δημόσιας Διοίκησης Υπηρεσία Ανάπτυξης Πληροφορικής	σελ. 7
8. INTRASOFT A.E.	σελ. 7
9. KNOWLEDGE A.E.	σελ. 8
III. Ειδήσεις για τη Γλωσσική Τεχνολογία	σελ. 8

I. Εισαγωγικό Σημείωμα

Με την έκδοση του πρώτου Ενημερωτικού Δελτίου με τον τίτλο "Λογοπλοήγηση" στο πλαίσιο του έργου "Ανθρώπινο Δίκτυο Γλωσσικής Τεχνολογίας" εγκαινιάζεται μία νέα δραστηριότητα η οποία αποσκοπεί στην προώθηση της συνεργασίας μεταξύ των ομάδων και των εργαστηρίων που έχουν ενδιαφέρον για τη Γλωσσική Τεχνολογία. Παράλληλα επιδιώκεται η πληρέστερη ενημέρωση των χρηματοδοτικών αρχών για την πρόοδο των εργασιών των μελών του δικτύου σε εθνικά και ευρωπαϊκά προγράμματα. Τέλος η "Λογοπλοήγηση" θα προσπαθήσει να ευαισθητοποιήσει τη Βιομηχανία σε σχέση με την εκμετάλλευση της υπάρχουσας τεχνογνωσίας, έτσι ώστε να αναπτυχθούν με ταχύ ρυθμό χρήσιμες εφαρμογές και ταυτόχρονα να εισαχθεί η Ελληνική γλώσσα σε πολυγλωσσικά προϊόντα.

Στο τεύχος αυτό παρουσιάζονται οι δραστηριότητες εργαστηρίων των συμμετεχόντων φορέων στο δίκτυο. Σε τεύχη που πρόκειται να εκδοθούν μελλοντικά θα υπάρχει ποικιλία θεματικών ενοτήτων όπως:

- Παρουσίαση Συστημάτων / Προϊόντων
- Παρουσίαση Καινοτομιών από Ερευνητές
- Παρουσίαση Ομιλιών των Ημερίδων και Σεμιναρίων στο πλαίσιο των εκδηλώσεων του Δικτύου.
- Γλωσσάριο Όρων Γλωσσικής Τεχνολογίας και Πληροφορικής
- Ειδήσεις και Πληροφόρηση σε θέματα Γλωσσικής Τεχνολογίας (διαγωνισμοί/ήμερομηνίες, συνέδρια, σεμινάρια, συναντήσεις εργασίας κ.λπ.)

Τέλος πληροφορούμε τους ενδιαφερομένους ότι η "Λογοπλοήγηση" θα είναι διαθέσιμη και μέσω Internet από τη σελίδα του ΙΕΛ: <http://www.ilsp.gr>

II. Κείμενα Μελών του "Ανθρώπινου Δικτύου Γλωσσικής Τεχνολογίας"

1. Ινστιτούτο Επεξεργασίας του Λόγου

Η Γλωσσική Τεχνολογία στο ΙΕΛ

Καθηγητής Γεώργιος Καραγιάννης
Ινστιτούτο Επεξεργασίας του Λόγου

Τρέχουσες Δραστηριότητες σε θέματα Γλωσσικής Τεχνολογίας



Το ΙΕΛ αναπτύσσει επιστημονική και τεχνολογική έρευνα στους τομείς της Υπολογιστικής Γλωσσολογίας, της Ανάλυσης, Σύνθεσης και Αναγνώρισης Φωνής για την παραγωγή εργαστηριακών προτύπων και προϊόντων τεχνολογίας του λόγου. Παράλληλα, το ΙΕΛ δραστηριοποιείται στην ανάπτυξη και δημιουργία προγραμμάτων εκπαιδευτικού λογισμικού που καλύπτουν ένα ευρύ φάσμα προϊόντων, όπως οπτικούς δίσκους για τη διδασκαλία της Νέας Ελληνικής ως μητρικής ή ξένης γλώσσας ("λογομάθεια", "φιλογλωσσία") σε συνδυασμό με θέματα τέχνης και πολιτισμού. Επίσης το Ινστιτούτο δραστηριοποιείται στην κατασκευή ηλεκτρονικών λεξικών για κάθε χρήση ακόμη και για εκπαι-

δευτικούς σκοπούς, όπως είναι η "λεξιπαιδεία".

Οι ερευνητές του Ινστιτούτου εργάζονται αναπτύσσοντας μεθοδολογίες και εργαλεία επεξεργασίας της γλώσσας και της φωνής. Επίσης συγκεντρώνεται ένα πλήθος γλωσσικών πόρων από κείμενα, λεξικά και γραμματικές. Άλλες σημαντικές δραστηριότητες του Ινστιτούτου στο χώρο της Γλωσσικής Τεχνολογίας είναι οι εξής:

- Η κατασκευή μεταφραστικών μνημών τόσο για το λειτουργικό σύστημα UNIX όσο και κάτω από τα Windows.
- Η κατασκευή ενός ολοκληρωμένου περιβάλλοντος με αντιπροσωπευτικά κείμενα διαφόρων γνωστικών περιοχών (υπογλωσσών) και τύπων κειμένων συνοδευόμενων από σχετικά υπολογιστικά εργαλεία επεξεργασίας των κειμένων. Τα κείμενα

αυτά θα διατίθενται σε όλους τους ενδιαφερόμενους ερευνητές στην Ευρώπη, μαζί με το πρωτότυπο περιβάλλον χειρισμού που σχεδιάζεται.

- Η κατασκευή ενός νέου προϊόντος που δίνει τη δυνατότητα αναζήτησης σε βάσεις δεδομένων πληροφοριών με χρήση αποκλειστικά φωνής (αναγνώριση φωνής). Το προϊόν αυτό θα δώσει τη δυνατότητα επίσκεψης και άντλησης πληροφοριών από βάσεις δεδομένων εντός και εκτός εργασίμου ωραρίου.

Μελλοντικές Δραστηριότητες σε θέματα Γλωσσικής Τεχνολογίας

Το ΙΕΛ στοχεύει στην περαιτέρω ανάπτυξη εργαλείων για τη γλωσσική εκπαίδευση, ιδιαίτερα των παιδιών, μέσω Η/Υ, στη δημιουργία ηλεκτρονικών μονόγλωσσων και δίγλωσσων λεξικών υπό μορφή οπτικών δίσκων (CD-ROM), στη δημιουργία προτύπων αναγνώρισης, σύνθεσης και συμπίεσης φωνής, στην κατασκευή συστημάτων διόρθωσης λαθών κειμένων κ.τ.λ. Την Άνοιξη του 1997 θα είναι διαθέσιμα τα εξής προϊόντα του ΙΕΛ:

Λογομάθεια - III: Η νέα έκδοση της "λογομάθειας" με όλη τη διδακτέα ύλη βελτιωμένη στο επίπεδο της εργονομίας με πλήθος νέων ασκήσεων και βραβείων.

Οργανολόγιο: Το νέο CD του ΙΕΛ που δημιουργήθηκε σε συνεργασία με το Μουσείο Λαϊκών Οργάνων για την παρουσίαση των παραδοσιακών λαϊκών οργάνων με χρήση πολυμέσων.

Λεξιπαιδεία: Το πολύγλωσσο λεξικό πολυμέσων του ΙΕΛ με πλήθος παιχνιδιών και με τη δυνατότητα προφοράς κάθε λέξης σε έξη συνολικά ευρωπαϊκές γλώσσες.

Επίσης, το Ινστιτούτο αναλαμβάνει τεχνολογικές αξιολογήσεις ερευνητικών εργασιών στις συγκεκριμένες περιοχές, την αξιοποίηση των αποτελεσμάτων τους, καθώς και την επιμόρφωση σε θέματα Υπολογιστικής Γλωσσολογίας και Επεξεργασίας Φωνής με τη μορφή συνεχούς εκπαίδευσης.

Ενδιαφέροντα σε θέματα Γλωσσικής Τεχνολογίας

Τα ενδιαφέροντα του ΙΕΛ επικεντρώνονται στις εξής περιοχές:

- Τεχνολογία φωνής (Σύνθεση φωνής, Αναγνώριση φωνής, Αποθορυβοποίηση).

- Τεχνολογίες για εφαρμογές σε συστήματα γραφείου (μεταφραστικές μνήμες, διόρθωση λαθών, ανάκτηση πληροφορίας).
- Ηλεκτρονική Λεξικογραφία και συλλογές κειμένων και όρων.
- Εκπαιδευτική Τεχνολογία με έμφαση στη διδασκαλία της Ελληνικής ως μητρικής, ως δεύτερης μητρικής και ως ξένης γλώσσας καθώς και στην ανάδειξη της πολιτιστικής ταυτότητας στον ελληνικό χώρο με τη χρήση της τεχνολογίας των πολυμέσων.
- Θέματα Μεταφραστικής Τεχνολογίας με έμφαση σε συγκεκριμένες υπογλώσσες και ελεγχόμενες γλώσσες. Το ΙΕΛ αναζητεί έναν βιομηχανικό εταίρο για την παραγωγή ενός βιομηχανικού προτύπου Μηχανικής Μετάφρασης με βάση τα εργαστηριακά πρότυπα που έχει αναπτύξει.

2. Ιόνιο Πανεπιστήμιο

Μετάφραση και Γλωσσική Τεχνολογία Μια συνεργασία στο πλαίσιο του προγράμματος Διάλογος

Επικ. Καθηγήτρια Μαρία Τσούτσουρα
Τμήμα Ξένων Γλωσσών, Μετάφρασης και Διερμηνείας

Σε μια προσπάθεια συγχρονισμού με τη νεότερη γλωσσική τεχνολογία και συμβολής στη δημιουργία εργαλείων για τον μεταφραστή, τετραμελής ερευνητική ομάδα του Τμήματος Ξένων Γλωσσών, Μετάφρασης και Διερμηνείας του Ιονίου Πανεπιστημίου συνεργάστηκε για το πρόγραμμα Διάλογος- ΕΠΕΤ II με το Ινστιτούτο Επεξεργασίας του Λόγου.

Η δραστηριότητα αναπτύχθηκε σε δύο κυρίως τομείς:

- Διερεύνηση και περιγραφή της διάδοσης των ηλεκτρονικών μέσων, ώστε να προσδιοριστούν συγκεκριμένα εργαλεία για τη δραστηριότητα του μεταφραστή, αλλά και του χρήζοντα μεταφραστικών υπηρεσιών στην Ελλάδα.
- Έρευνα γύρω από τα διαλογικά ή αλληλεπιδραστικά (interactive) μοντέλα αυτόματης μετάφρασης.

Τα μέχρι στιγμής συμπεράσματα έδειξαν πως, ενώ μεν η ηλεκτρονική επεξεργασία του κειμένου έχει επιβληθεί γενικά στους συγγραφείς και στους μεταφραστές ειδικότερα, οι χρήστες

αγνοούν το εύρος των πιθανών εφαρμογών, την ποικιλία και την αποτελεσματικότητα την οποία μπορεί πραγματικά να εξασφαλίσει η γλωσσική τεχνολογία, τόσο για την ποιότητα όσο και για την ταχύτητα παραγωγής των κειμένων.

Η ελλιπής κατοχύρωση του επαγγέλματος του μεταφραστή οδηγεί σε ανασφάλεια και σε σύνδρομα συσπείρωσης, που εμποδίζουν τη συνεργασία με τους ερευνητικούς φορείς. Είναι χαρακτηριστικό πως στη χώρα μας δεν υπάρχουν μεγάλες μεταφραστικές εταιρείες, ενώ οι ελεύθεροι επαγγελματίες (free-lance) μεταφραστές σπάνια ασκούν το επάγγελμα με ορισμένη ειδικότητα ή με πλήρη και μοναδική απασχόληση.

Τα προβλήματα που αναφύονται πολλαπλασιάζονται από το γεγονός ότι η ελληνική, παρότι ευρωπαϊκή, δεν είναι μια γλώσσα ευρείας διάδοσης και η επικοινωνία της με άλλες σχετικά σπάνιες γλώσσες, με τις οποίες όμως καλείται όλο και περισσότερο να συναλλαγεί, τουλάχιστον στο πλαίσιο των μεσογειακών και βαλκανικών σχέσεων, δεν είναι ικανοποιητικά εξασφαλισμένη.

Για τις μελλοντικές δραστηριότητες προβλέπουμε πως απαιτούνται:

- Προγραμματισμός για καλλιέργεια σε ευρεία κλίμακα της γλωσσομάθειας των νέων δυνάμει μεταφραστών, τόσο σε ποιοτικό βάθος όσο και σε ποικιλία γλωσσών, και μάλιστα με γνώμονα τις δυνατότητες που παρέχει η σύγχρονη τεχνολογία.
- Επαρκής εκπαίδευση των μεταφραστών, ώστε να αντιλαμβάνονται την μεταφραστική εργασία ως μεθοδικά κλιμακούμενη δραστηριότητα που αναπτύσσεται σε διακεκριμένες φάσεις, εξειδικεύσεις και ποιότητες, με συγκεκριμένους στόχους και προσδιορισμένη αποτελεσματικότητα. Είναι σημαντικό να καταστεί συνείδηση στους ενδιαφερομένους ότι τα εργαλεία του μεταφραστή δεν μπορούν παρά να αναφέρονται σε κάποιες από τις φάσεις αυτές και όχι σε ένα συγκεχυμένο και ασαφές σύνολο δράσεων.
- Συστηματική ενημέρωση των μεταφραστών για τις δυνατότητες που παρέχουν τα ηλεκτρονικά μέσα: ελεγκτές, μεταφραστικές μνήμες, γλωσσάρια και λεξικά, αρχεία, ευρετήρια, βάσεις δεδομένων κ.ά.
- Διαμόρφωση στελεχών ικανών να αντιμετωπίσουν

τις απαιτήσεις της γλωσσικής τεχνολογίας. Εξειδικευμένοι γλωσσολόγοι, φιλόλογοι και προγραμματιστές οφείλουν να συνεργάζονται μεταξύ τους αποδοτικά για την έρευνα και παραγωγή των ηλεκτρονικών εργαλείων, ώστε να ικανοποιηθούν οι ανάγκες της σύγχρονης επικοινωνίας, ακόμα και αν αυτό προϋποθέτει, όπως σε όλες τις σύγχρονες χώρες, αναδιάρθρωση των γλωσσικών επιστημών.

3. Εργαστήριο Τεχνολογίας Γνώσεων και Λογισμικού Ινστιτούτο Πληροφορικής και Τηλεπικοινωνιών ΕΚΕΦΕ "ΔΗΜΟΚΡΙΤΟΣ"

Δρ. Κωνσταντίνος Σπυρόπουλος
Διευθυντής Έρευνας

Τρέχουσες Δραστηριότητες σε θέματα Γλωσσικής Τεχνολογίας

Συμμετοχή στο ερευνητικό πρόγραμμα ECRAN (LE-2110) που έχει σαν στόχο την ανάπτυξη μιας νέας τεχνολογίας για την αυτοματοποιημένη εξαγωγή πληροφορίας από κείμενα καθώς επίσης και την προώθηση αυτής της τεχνολογίας στην αγορά για την εμπορική αξιοποίησή της. Βρίσκεται επίσης σε εξέλιξη μία διδακτορική εργασία, με θέμα την οργάνωση λεξικών και γραμματικών της Ελληνικής γλώσσας με στόχο την αποτελεσματική αξιοποίησή τους σε εφαρμογές εξαγωγής πληροφορίας από κείμενα.

Δημιουργία και αξιοποίηση βάσεων ορολογικής γνώσης και αυτόματη παραγωγή πολύγλωσσων μηνυμάτων σε προϊόντα λογισμικού. Η δραστηριότητα αυτή αποτελεί συνέχεια του ερευνητικού προγράμματος GLOSSASOFT (LRE-61003) και ενός διδακτορικού με θέμα "Βάσεις Ορολογικής Γνώσης και Αξιοποίησή τους σε Πολύγλωσσες Εφαρμογές Λογισμικού".

Μελλοντικές Δραστηριότητες σε θέματα Γλωσσικής Τεχνολογίας

Με βάση την υπάρχουσα εμπειρία, το Εργαστήριο προσβλέπει στην υλοποίηση ερευνητικών έργων και συνεργασιών στα παρακάτω θέματα Γλωσσικής Τεχνολογίας:

- αυτοματοποιημένη δημιουργία μεγάλων βάσεων ορολογικής γνώσης από κείμενα,
- δημιουργία Ελληνικής γλώσσας ελεγχόμενης δομής (controlled language) και ανάπτυξη εργαλείων για την αποτελεσματική αξιοποίησή της,
- ανάπτυξη μεθόδων παραγωγής πολύγλωσσων διαπροσωπικών (multilingual interfaces) για διεθνοποιημένα προϊόντα λογισμικού,
- δημιουργία λεξικών για συγκεκριμένες θεματικές περιοχές (θεματικά λεξικά) από γενικά λεξικά,
- ανάπτυξη μεθόδων εξαγωγής πληροφοριών από κείμενα της Ελληνικής σε συγκεκριμένες θεματικές περιοχές με χρήση θεματικών λεξικών.

Ενδιαφέροντα σε θέματα Γλωσσικής Τεχνολογίας

- Αξιοποίηση υπαρχόντων και ανάπτυξη νέων εργαλείων γλωσσικής τεχνολογίας για την τοπική προσαρμογή προϊόντων λογισμικού (π.χ. εργαλεία διαχείρισης ορολογίας, μετάφρασης κειμένων, παραγωγής ελεγχόμενης γλώσσας, λεξικών, κ.λπ.).
- Αξιοποίηση υπαρχόντων και ανάπτυξη νέων μεθόδων και τεχνικών αναπαράστασης γνώσης και παραγωγής φυσικής γλώσσας για την παραγωγή πολύγλωσσων μηνυμάτων σε εφαρμογές λογισμικού.
- Ανάπτυξη εργαλείων που αξιοποιούν μεθόδους και τεχνικές τεχνητής νοημοσύνης και γλωσσικής τεχνολογίας για την αυτοματοποιημένη εξαγωγή πληροφορίας από κείμενα της Ελληνικής γλώσσας.

4. Εθνικό Μετσόβιο Πολυτεχνείο

Υπεύθυνος: Καθηγητής Γεώργιος Παπακωνσταντίνου
 Εργαστήριο Υπολογιστικών Συστημάτων
 Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Η ομάδα Γλωσσικής Τεχνολογίας του Εργαστηρίου Ψηφιακών Συστημάτων και Υπολογιστών επικεντρώνει την έρευνά της σε ερευνητικά και αναπτυξιακά θέματα συντακτικής ανάλυσης με έμφαση (i)στη συντακτική ανάλυση φυσικής γλώσσας με κατηγορικές γραμματικές (attribute grammars) και άλλες προσεγγίσεις (π.χ. γραμματικές Eurotra και λογικό προγραμματισμό), (ii)στην εφαρμογή τεχνικών παραλληλοποίησης στη διαδικασία της συντακτικής ανάλυσης, καθώς και (iii)στην ανάπτυξη τεχνικών

συντακτικής ανάλυσης που υποβοηθούνται από τη σημασιολογία, για την εφαρμογή της εξαγωγής πληροφορίας (information extraction). Επίσης, γίνεται έρευνα στο πεδίο της μορφολογικής ανάλυσης με εφαρμογή σε εκπαιδευτικά περιβάλλοντα.

Στις τρέχουσες δραστηριότητες της ομάδας συγκαταλέγεται η συμμετοχή στο έργο "ΔΙΑΛΟΓΟΣ", σκοπός του οποίου είναι η μελέτη της αυτόματης παραλληλοποίησης συντακτικών αναλυτών για το συμβολισμό Eurotra. Επίσης, αναπτύσσονται εργαλεία παράλληλης συντακτικής ανάλυσης κατηγορικών γραμματικών και παράλληλου λογικού προγραμματισμού.

Η εξαγωγή πληροφορίας και η εισαγωγή της σημασιολογίας στη συντακτική ανάλυση, αποτελούν νέα ερευνητικά πεδία για την ομάδα. Στο άμεσο μέλλον σχεδιάζεται η διεύρυνση των δραστηριοτήτων της ομάδας, με στόχο την κατασκευή εκπαιδευτικού λογισμικού για το συντακτικό της νεοελληνικής γλώσσας, το οποίο θα έχει τη δυνατότητα να διδάξει συντακτικό μέσα από περιβάλλον πολυμέσων, εκτελώντας συντακτική ανάλυση (μη προκαθορισμένων) προτάσεων που θα δίνονται από μαθητές.

5. Δημοκρίτειο Πανεπιστήμιο Θράκης

Δραστηριότητες στον Τομέα της Γλωσσικής Τεχνολογίας του Εργαστηρίου Ανάλυσης Ηλεκτρικών Κυκλωμάτων του Δ.Π. Θράκης και της Μονάδος Υπηρεσιών Ψηφιακής Επεξεργασίας Εικόνων και Πολυμέσων

Αν. Καθηγητής Χριστόδουλος Χαμζάς
 Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Το Εργαστήριο Ανάλυσης Ηλεκτρικών Κυκλωμάτων (ΑΗΚ) και η Εργαστηριακή Μονάδα Παροχής Υπηρεσιών Ψηφιακής Επεξεργασίας Εικόνων και Πολυμέσων έχει αναπτύξει τα τελευταία χρόνια σημαντική δραστηριότητα σε θέματα γλωσσικής τεχνολογίας με έμφαση σε θέματα οπτικής αναγνώρισης χαρακτήρων (ΟΑΧ) και συμπίεσης δεδομένων.

Ο κύριος σκοπός των τεχνικών της ΟΑΧ είναι η μετατροπή εικόνων κειμένου σε αναγνωρίσιμους ASCII χαρακτήρες. Ως γνωστό ένα σύστημα οπτικής αναγνώρισης χαρακτήρων αποτελείται από διάφορα στάδια και τεχνικές και συγκεκριμένα τεχνικές προεπεξεργασίας εγγράφων, εξαγωγής χαρακτηριστικών και τεχνικές τελικής ταξινόμησης των χαρακτήρων.

Μερικά από τα θέματα με τα οποία ασχολήθηκε το Εργαστήριο αναφέρονται:

- Στη μετατροπή μιας εικόνας εγγράφου με 256 επίπεδα του γκρι σε δυαδική. Η τεχνική που αναπτύχθηκε βασίζεται στην εύρεση της βέλτιστης τιμής κατωφλιού του ιστογράμματος της εικόνας του εγγράφου. Για το σκοπό αυτό χρησιμοποιήθηκε η τμηματική προσέγγιση του ιστογράμματος με ρητές συναρτήσεις σε συνδυασμό με μια τεχνική εύρεσης ελαχίστου μονοδιάστατων συναρτήσεων.
- Στην εύρεση της στροφής των εγγράφων ώστε αυτά να μπορούν αυτόματα να επανέρχονται στη σωστή μηδενική τους κλίση. Η τεχνική που αναπτύχθηκε εδώ βασίζεται στη συσχέτιση των pixels πάνω σε συγκεκριμένες κάθετες ευθείες του εγγράφου.
- Την τμηματοποίηση των εγγράφων και το διαχωρισμό των χαρακτήρων. Εδώ αναπτύχθηκε μια τεχνική βασισμένη σε εξομάλυνση Run-Length. Για το τελικό στάδιο της εξαγωγής των χαρακτήρων αναπτύχθηκε μια γρήγορη τεχνική που βασίζεται στη διάτρηση των χαρακτήρων.
- Στην ανάπτυξη ενός συστήματος οπτικής αναγνώρισης χαρακτήρων με έμφαση στην αναγνώριση ελληνικών γραμμάτων βασισμένου σε δυαδικά χαρακτηριστικά, δυαδικό ταξινομητή (δένδρο) και μιας νέας τεχνικής σύγκρισης των χαρακτήρων στους τελικούς κόμβους του δένδρου ταξινόμησης.
- Στην ανάπτυξη ενός συστήματος οπτικής αναγνώρισης χαρακτήρων βασισμένου στη χρησιμοποίηση σειριακού ταξινομητή. Τα σύνολα χαρακτηριστικών, που χρησιμοποιήθηκαν αναφέρονται στις χαρακτηριστικές κλίσεις των χαρακτήρων, τις ροπές Zernike και τις Gaussian μάσκες.
- Στην ανάπτυξη μιας νέας τεχνικής γρήγορης εύρεσης του μετασχηματισμού Hough μιας δυαδικής εικόνας. Η τεχνική αυτή έχει άμεσες εφαρμογές στην εύρεση της στροφής και της τμηματοποίησης εγγράφων.

Στον χώρο της συμπίεσης:

- Έχει συμβάλει σημαντικά στη δημιουργία του νέου διεθνούς προτύπου JBIG για αποθήκευση

δυναμικών εικόνων.

- Έχει εισαγάγει έναν προσαρμοζόμενο εκτιμητή πιθανοτήτων ο οποίος ήδη χρησιμοποιείται στα πρότυπα JBIG και JPEG.

Στα μελλοντικά ενδιαφέροντα του Εργαστηρίου Ανάλυσης Ηλεκτρικών Κυκλωμάτων περιλαμβάνονται θέματα όπως:

- Η χρησιμοποίηση των Neuro-Fuzzy τεχνικών στην ανάλυση και οπτική αναγνώριση χαρακτήρων.
- Η ανάπτυξη τεχνικών οπτικής αναγνώρισης χαρακτήρων με έμφαση στην ελληνική γλώσσα. Εδώ προβλέπεται η χρησιμοποίηση ελληνικού λεξικού και γραμματικής ως post-processing τεχνικές.
- Η χρήση του εκτιμητή πιθανοτήτων για συμπίεση ηχητικών σημάτων.

Ενδιαφέροντα Μελών Εργαστηρίου

Τα κύρια ενδιαφέροντα των μελών του Εργαστηρίου αναφέρονται στην περιοχή των πολυμέσων, της ανάλυσης εικόνων και της συμπίεσης δεδομένων.

6. Πανεπιστήμιο Πατρών

Καθηγητής Γεώργιος Κοκκινάκης

Εργαστήριο Ενσύρματης Τηλεπικοινωνίας

Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών

Τρέχουσες δραστηριότητες σε θέματα γλωσσικής τεχνολογίας

Συμμετοχή στα ερευνητικά προγράμματα:

EU, ACCeSS: LE1- 1802- Automated Call Center through Speech Understanding System. (1996-1998)

Coordinator: Daimler Benz

EU, SpeechDat: LE2-4001- Speech Databases for Creation of Voice Driven Teleservices. (1996- 1998)

Cordinator: Siemens A.G.

EU, Socrates: Lingua ActionD TM-LD-1995-I-GR-30 (1/1)-An Electronic Dictionary for the Pronunciation and Use of Modern Greek Language for Foreigners. (1995-1997)

Coordinator: University of Patras

EU, EDUCATION, TRAINING & YOUTH: 96-06-MDD- 0137-00-Electronic Dictionary for the Pronunciation and Usage of the Grecanic Dialect of South Italy. (1996- 1998)

Coordinator: University of Patras

ΓΓΕΤ ΕΠΕΤ II No. 715 ΔΙΑΛΟΓΟΣ: Βελτίωση της Επικοινωνίας Ανθρώπου - Μηχανής με χρήση Γλωσσικής Τεχνολογίας. (1994- 1997)

Coordinator: Ινστιτούτο Επεξεργασίας του Λόγου (ΙΕΛ)

Εκπόνηση 14 διδακτορικών διατριβών στην περιοχή τεχνολογίας ομιλίας και φυσικής γλώσσας. Εκπόνηση 25 περίπου διπλωματικών εργασιών κατ' έτος.

Μελλοντικές δραστηριότητες

Συνέχιση της υφισταμένης δραστηριότητας στις περιοχές: Σύνθεση ομιλίας, αναγνώριση ομιλίας, αναγνώριση ομιλητή, κωδικοποίηση ομιλίας, αντιμετώπιση θορύβου, ηλεκτρονική λεξικογραφία.

Ενδιαφέροντα εργαστηρίου

Τα ενδιαφέροντα του Εργαστηρίου συμπίπτουν με τις αναφερθείσες περιοχές δραστηριότητας.

7. Υπουργείο Εσωτερικών Δημόσιας Διοίκησης και Αποκέντρωσης Γενική Γραμματεία Δημόσιας Διοίκησης

Υπηρεσία Ανάπτυξης Πληροφορικής

Γιώργος Τσακογιάννης
Προϊστάμενος της Υπηρεσίας Ανάπτυξης Πληροφορικής

Η Υπηρεσία Ανάπτυξης Πληροφορικής (ΥΑΠ) έχει ως σκοπό της την εφαρμογή της κυβερνητικής πολιτικής για την ανάπτυξη της πληροφορικής στον δημόσιο τομέα και τον συντονισμό των σχετικών δραστηριοτήτων. Οι επιτελικές και συντονιστικές αρμοδιότητες της υπηρεσίας περιλαμβάνουν, μεταξύ άλλων, τον μεσοπρόθεσμο σχεδιασμό για την περαιτέρω ανάπτυξη της πληροφορικής στον δημόσιο τομέα, την παρακολούθηση της εξέλιξης των έργων πληροφορικής, την καθιέρωση προτύπων και μεθοδολογιών και την εκπροσώπηση της ελληνικής Δημόσιας Διοίκησης σε επιτροπές φορέων στην Ελλάδα και στο εξωτερικό.

Με τη συμμετοχή της στο ανθρώπινο δίκτυο για τη διάδοση των γλωσσικών τεχνολογιών στη χώρα μας, η ΥΑΠ αναμένει την ενημέρωση των στελεχών της σε θέματα αυτόματης επεξεργασίας φυσικής γλώσσας,

ηλεκτρονικών λεξικών, εκπαιδευτικής τεχνολογίας και αυτόματης μετάφρασης γλωσσών.

Ιδιαίτερη σημασία αποδίδεται από την υπηρεσία σε τρεις τομείς:

- αυτόματη μετάφραση κειμένων στα ελληνικά
- ανάκτηση πληροφορίας από κείμενα, και
- δημιουργία εργαλείων για την ελεγχόμενη γλώσσα της Δημόσιας Διοίκησης.

Ήδη βρισκόμαστε σε συνεννοήσεις με το Ινστιτούτο Επεξεργασίας του Λόγου (ΙΕΛ) και το Εθνικό Κέντρο Δημόσιας Διοίκησης (ΕΚΔΔ) σε μια προσπάθεια καθορισμού ενός πλαισίου συνεργασίας το οποίο θα μας οδηγήσει στη δημιουργία και καθιέρωση εργαλείων γλωσσικής τεχνολογίας στην ελληνική Δημόσια Διοίκηση.

8. INTRASOFT A.E.

Τομείς Δραστηριοτήτων σε θέματα Γλωσσικής Τεχνολογίας

Δρ. Νίκος Τζωρτζόπουλος
Διευθυντής Νομικής Πληροφορικής ΝΟΜΟΣ

Οι κύριοι τομείς δραστηριοτήτων της INTRASOFT, οι οποίες αφορούν τη Γλωσσική τεχνολογία είναι οι εξής:

1. RENOS (LRE62-074)

Reduction of NOise and Silence in FTRS for legal texts

Ερευνητικό έργο, χρηματοδοτούμενο από την Ευρωπαϊκή Ένωση, με θέμα τη μείωση του "θορύβου" και της "σιωπής" σε Full Text Retrieval Systems πάνω σε νομικά κείμενα. Κατά τη διάρκεια του έργου αναπτύχθηκε μεθοδολογία η οποία αφορούσε: α) Στατιστική και νομική επεξεργασία νομικών κειμένων, β) Ενοποίηση νομικών όρων (simple - compound) σε ομάδες συνωνύμων (concept groups), γ) Εφαρμογή γραμματικής στην αναγνώριση των όρων, δ) retrieval μηχανισμό στηριγμένο στα concept groups.

2. HAGIS (LE-1470 TELEMATICS)

HAzardous Goods Information Service

Πιλοτικό έργο χρηματοδοτούμενο από την Ευρωπαϊκή Ένωση με στόχο τη δημιουργία διεθνούς βάσης δεδομένων με πληροφορίες σχετικές με το διεθνές, εθνικό και κοινοτικό δίκαιο της μεταφοράς επικίνδυνων αγαθών. Στο πλαίσιο του HAGIS, η μεθοδολογία και τα ερ-

γαλεία που αναπτύχθηκαν στο RENOS θα χρησιμοποιηθούν και θα βελτιωθούν.

Μετά το πέρας της πιλοτικής φάσης, το HAGIS προορίζεται για εμπορική εκμετάλλευση μέσω της επέκτασής του σε όλους τους τομείς του δικαίου των μεταφορών και σε όλες τις Ευρωπαϊκές χώρες με την έννοια της multinational και multilingual database.

3. ΔΙΑΛΟΓΟΣ (ΕΠΕΤ-II)

Η συμμετοχή της INTRASOFT στο έργο αφορά την προσφορά σώματος νομικών κειμένων (corpus) για γλωσσολογική επεξεργασία, και την ανάπτυξη εργαλείων για αποδοτική ιεραρχική κατηγοριοποίηση κειμένων.

4. ΝΟΜΟΣ

Βάση δεδομένων νομικών πληροφοριών (Ελληνική και Κοινοτική Νομοθεσία και Νομολογία- Μέγεθος: 5GB) , στους μηχανισμούς της οποίας αναπτύσσονται και ενσωματώνονται σταδιακά μεθοδολογίες και εργαλεία για τη βελτίωση της διαδικασίας ανάκτησης πληροφοριών από τους χρήστες. Οι μελλοντικές δραστηριότητες της INTRASOFT σχετικά με τη γλωσσική τεχνολογία, αφορούν την προώθηση και ενίσχυση της ενσωμάτωσης των μεθοδολογιών και εργαλείων που έχουν αναπτυχθεί, στο retrieval περιβάλλον της ΝΟΜΟΣ, με στόχο την μείωση του "θορύβου" και της "σιωπής" στην αναζήτηση πληροφοριών και επομένως στη δημιουργία φιλικού και αποδοτικού περιβάλλοντος για τους χιλιάδες χρήστες της ΝΟΜΟΣ.

9. KNOWLEDGE A.E.

Δραστηριότητες της Εταιρείας σε θέματα Γλωσσικής Τεχνολογίας

Δρ. Αναστάσιος Τσοπάνογλου
Τομέας Επικοινωνίας Ανθρώπου- Μηχανής

Η KNOWLEDGE A.E είναι η πρώτη ιδιωτική εταιρεία στην Ελλάδα που έχει επενδύσει σε θέματα Γλωσσικής Τεχνολογίας. Η μεγάλη εμπειρία που έχει αποκτήσει, κυρίως από την συμμετοχή της σε μεγάλα κοινωνικά και εθνικά Ερευνητικά/Αναπτυξιακά προγράμματα, την κατατάσσουν ανάμεσα στις γνωστότερες ευρωπαϊκές εταιρείες του χώρου.

Η ανάπτυξη συστημάτων Επικοινωνίας Ανθρώπου-Μηχανής, κυρίως μέσω κέντρων κλήσης (Call Centers), αποτελεί μία από τις κυριότερες δραστηριότητες της εταιρείας. Τα συστήματα Επικοινωνίας Ανθρώπου-Μη-

χανής, χρησιμοποιώντας τεχνικές αυτόματης αναγνώρισης και κατανόησης ομιλίας, επιτυγχάνουν την ανάκληση και διαχείριση πληροφοριών μέσω του τηλεφωνικού δικτύου χωρίς την ανάγκη παρέμβασης χειριστή.

Επιλεγμένα ερευνητικά προγράμματα Γλωσσικής Τεχνολογίας

ACCeSS (LE-1 1802). Σκοπός του προγράμματος είναι η δημιουργία αυτόματου κέντρου εξυπηρέτησης πελατών και απευθύνεται κυρίως σε ασφαλιστικές εταιρείες. Το σύστημα κατανοεί τις προφορικές εντολές του χρήστη, που μεταδίδονται μέσω τηλεφωνικών γραμμών, χρησιμοποιώντας σύγχρονες μεθόδους αναγνώρισης και κατανόησης φωνής και αποκρίνεται στις εντολές αυτές με συνθετική ομιλία.

VASME (TRANS.PL-00010). Στόχος του προγράμματος είναι η ανάπτυξη ενός πληροφοριακού συστήματος λήψης δεδομένων και απευθύνεται σε πλοία που πλησιάζουν σε λιμάνια ή σταθμούς πλοήγησης. Η προσπέλαση του συστήματος και η μετάδοση των πληροφοριών γίνεται με χρήση ομιλίας μέσω VHF καναλιών.

ΔΙΑΛΟΓΟΣ (ΕΠΕΤ-II). Δύο είναι οι στόχοι του προγράμματος: ο πρώτος, είναι η δημιουργία αυτόματου συστήματος διεκπεραίωσης τραπεζικών εργασιών μέσω τηλεφωνικού δικτύου με χρήση τεχνικών αναγνώρισης και κατανόησης ομιλίας και ο δεύτερος, είναι η ανάπτυξη αυτομάτου συστήματος δημιουργίας εμπορικών επιστολών.

SPEECHDAT II (LE-2 4001). Σκοπό έχει την δημιουργία βάσης δεδομένων ομιλίας από μεγάλο αριθμό ομιλητών (5000 ομιλητές) για 17 ομιλούμενες ευρωπαϊκές γλώσσες.

III.Ειδήσεις για τη Γλωσσική Τεχνολογία

Ημερομηνίες για τους Διαγωνισμούς Γλωσσικής Τεχνολογίας του Δεκεμβρίου 1996

Σε ότι αφορά τους Διαγωνισμούς αναφέρουμε τα εξής:

- Η τέταρτη πρόσκληση υποβολής προτάσεων της δράσης Language Engineering του προγράμματος TELEMATICS APPLICATIONS PROGRAMME πρόκειται να ανακοινωθεί στις 17 Δεκεμβρίου 1996 με πιθανή καταληκτική ημερομηνία στις 17 Μαρτίου 1997.
- Η καταληκτική ημερομηνία της έβδομης πρόσκλησης υποβολής προτάσεων του προγράμματος INFORMATION TECHNOLOGIES (ESPRIT 4) είναι η 16η Δεκεμβρίου 1996.