



ΠΛΗΡΩΜΕΝΟ  
ΤΕΛΟΣ  
Τεχ. Ισορροπία  
ΚΕΜΠ ΑΘ  
Αριθμός Άδειας  
3861

ΕΝΤΥΠΟ ΚΛΕΙΣΤΟ ΑΡ. ΑΔΕΙΑΣ ..... 1618 ..... ΚΣΗΠΙΑ

Κωδικός: 7007



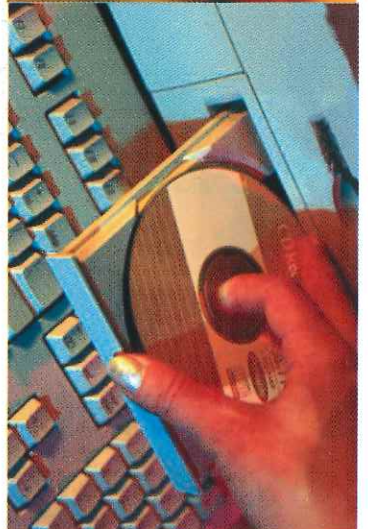
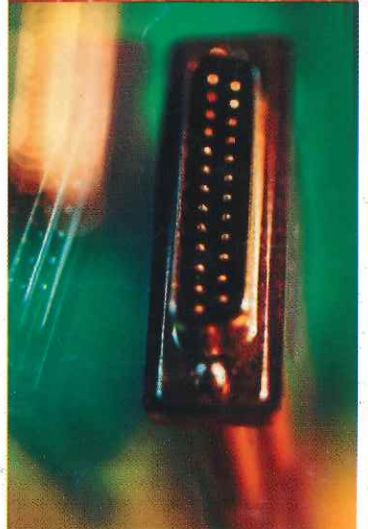
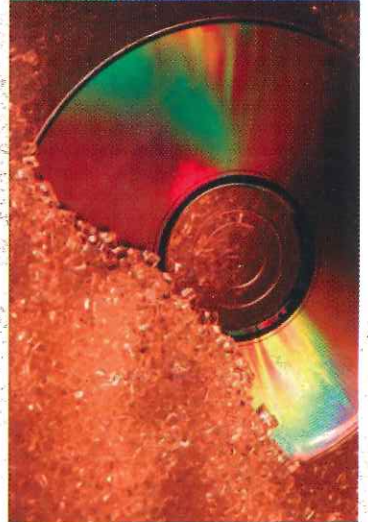
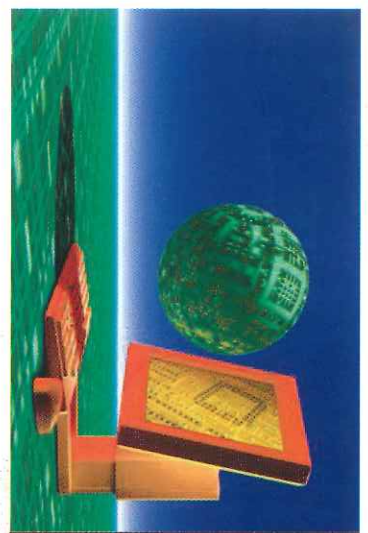
**N LINE**

ΤΕΥΧΟΣ Νο 6

Το επίσημο περιοδικό του  
Πανελληνίου Συλλόγου Διπλωματούχων Μηχανικών Η/Υ & Πληροφορικής

[www.online-mag.gr](http://www.online-mag.gr)

[www.computer-engineers.gr](http://www.computer-engineers.gr)



Δρ. Αριστείδης Βαγγελάτος\*,  
Μηχανικός Ηλεκτρονικών Υπολογιστών και Πληροφορικής,  
Ε.Α. Ινστιτούτο Τεχνολογίας Υπολογιστών  
Επαχάλακου 13, 11851 Θησείο, Αθήνα  
vagelat@cti.gr

Δρ. Γιώργος Ορφανός\*,  
Μηχανικός Ηλεκτρονικών Υπολογιστών και  
Πληροφορικής,  
Neurosoft A.E. Κωφίδου 24, 14231, Ν. Ιωνία, Αθήνα  
orphan@neurosoft.gr

\* Και οι δύο συγγραφείς έχουν  
εκπονήσει τη διδακτορική τους  
διατριβή σε θέματα Γλωσσικής  
Τεχνολογίας για την Ελληνική  
Γλώσσα, στο Πανεπιστήμιο Πατρών.

## Η Γλωσσική Τεχνολογία (ΓΤ) στην υπηρεσία του Παγκόσμιου Ιστού

Ο παγκόσμιος ιστός δεν είναι πλέον διαχειρίσιμος από τον απλό άνθρωπο που αναζητά μια συγκεκριμένη πληροφορία.

Οι τρέχουσες στατιστικές αναφέρουν ότι περίπου 935 εκατομμύρια άνθρωποι παγκοσμίως έχουν πρόσβαση στον παγκόσμιο ιστό (το αντίστοιχο νούμερο στη χώρα μας ανάγεται σε περίπου 2,7 εκατομμύρια συνδεδεμένους χρήστες). Ο πληθυσμός αυτός έχει πρόσβαση σε περισσότερες από 3 δισεκατομμύρια σελίδες κειμένου. Όπως γίνεται εύκολα κατανοητό, κανείς δεν μπορεί να εγγυηθεί πλέον για την αξιοπιστία του συνόλου της πληροφορίας αυτής, αλλιώς ακόμα περισσότερο, είναι προφανές ότι η πληροφορία αυτή δεν είναι διαχειρίσιμη από τον απλό χρήστη στα πλαίσια των καθημερινών του αναζητήσεων.

Η πλέον επιτυχημένη παρομοίωση της κατάστασης που επικρατεί σήμερα στο διαδίκτυο είναι η ακόλουθη εικόνα: ας φανταστούμε ένα διψασμένο και καταπονημένο άνθρωπο (που έχει περπατήσει πάνω από 10 ώρες στην έρημο κάτω από τον καυτό ήλιο) να βρίσκει ξαφνικά μπροστά του έναν πυροσβεστικό κρουνό και να προσπαθεί να πει νερό από αυτόν. Πράγματι, ο όγκος της πληροφορίας («νερό») που υπάρχει στο διαδίκτυο είναι τόσο τραγικά τεράστιος, που πρέπει κανείς («διψασμένος») να σπαταλήσει ώρες μπροστά σε έναν υπολογιστή, εάν θέλει να βρει τελικά κάτι αξιόλογο (χωρίς το αποτέλεσμα να είναι πάντα εγγυημένης ποιότητας ούτε και ακρίβειας).

### Οι σημερινές μηχανές αναζήτησης είναι αρκούντως χαζές!

Δε μπορούμε παρά να δεχτούμε ότι τα περασμένα 10 χρόνια, η εκθετική αύξηση (σε ετήσια βάση) της πληροφορίας που δημοσιεύτηκε στο διαδίκτυο, οδήγησε σε μια νέα πραγματικότητα: αντί για την επίσκεψη στη βιβλιοθήκη, την παραγγελία κάποιου βιβλίου ή την αναμονή για το αντίτυπο ενός συγγράμματος που ευγενικά προσφέρθηκε να μας αποστείλει ο συγγραφέας, σήμερα είμαστε σε θέση να βρούμε το σύνολο σχεδόν της προς αναζήτηση πληροφορίας απλά με μερικά κλικ μπροστά στον προσωπικό μας υπολογιστή.

Είναι έτσι απλά όμως τα πράγματα; «5.480.000 περίπου ιστοσελίδες βρέθηκαν για human language technology & WWW» μας απάντησε γνωστή μηχανή αναζήτησης σε σχετικό ερώτημα που θέσαμε! Υπέροχα! Και τώρα τι; Εάν το

ενδιαφέρον μας περιορίζεται στο πόσες «σχετικές» σελίδες υπάρχουν, τότε θα πρέπει αν είμαστε ευχαριστημένοι! Εάν όμως θέλουμε πραγματικά κάτι να αθιεύσουμε μέσα σε αυτές, τότε τα πράγματα δεν είναι πολύ απλά! (Η αξιοπιστία της πληροφορίας που βρέθηκε είναι ένα θέμα που χρήζει ιδιαίτερης συζήτησης από μόνη της...). Εμείς όμως απλά θέλαμε να βρούμε τις «έγκυρες» σελίδες που αναφέρονταν στην χρήση Γλωσσικής Τεχνολογίας (ΓΤ) στον παγκόσμιο ιστό. Και μόνο για να κάνουμε μια γρήγορη ανάγνωση στις 20 με 30 πρώτες σελίδες θα χρειαστεί το υπόλοιπο της ημέρας... Ας το αφήσουμε γι' αργότερα!

### Ο παγκόσμιος ιστός θα πρέπει να μάθει να συνεννοείται!

Έστω ότι ήρθε η ώρα να προγραμματίσουμε τις Πασχαλινές διακοπές μας. Άνοιξη γαρ και επιλέγουμε να πάμε ένα ταξίδι σε κάποια Ευρωπαϊκή πρωτεύουσα. Το αποφασίσαμε μεν σχεδόν μόλις τέλειωσαν οι Χριστουγεννιάτικες αργίες, έχει φτάσει όμως παραμονή της Μεγάλης Εβδομάδας και (ως γνήσιοι απόγονοι των γονέων μας) δεν έχουμε ακόμα κάνει κράτηση για την πτήση. Και τότε κατανοούμε ότι εκείνες οι φθηνές πτήσεις που μας έλεγε ο πράκτορας τα Χριστούγεννα έχουν εξαντληθεί και το κόστος έχει ουσιαστικά πενταπλασιαστεί. Η μόνη λύση είναι κάποιες φθηνές εταιρείες οι οποίες προσφέρουν τη δυνατότητα κρατήσεων μέσω του web και μόνο.

Έτσι ξεκινάμε ένα πρωί να αναζητούμε μια μια αυτές τις εταιρείες και να ψάχνουμε για την πτήση που μας ενδιαφέρει. Εάν είμαστε τυχεροί, μετά από δυο-τρεις ώρες προσπάθειας (αναζήτηση διαφορετικών συνδυασμών πτήσεων και ημερομηνιών), βρίσκουμε μια συμφέρουσα περίπτωση, η οποία όμως δεν φτάνει στο κυρίως αεροδρόμιο της πόλης που θέλουμε αλλιώς σε ένα κοντινό. Άντε λίγη προσπάθεια ακόμα για να βρούμε δρομολόγια τρένων, ρεωφορείων ή αν υπάρχει δυνατότητα ενοικίασης αυτοκινήτου ώστε να φτάσουμε στον τελικό προορισμό... Και ο χρόνος περνά.

Εμείς αυτό που θέλαμε ήταν να βρούμε «αεροπορική πτήση για 2 άτομα με επιστροφή, από Αθήνα προς Δουβλίνο, για τις ημερομηνίες αυτές (μια μέρα πριν ή μετά δεν μας πειράζει) που να φεύγει και να φτάνει στον προορισμό κάποιες λογικές ώρες!» Υπάρχει δυνατότητα να απαντήσει κανείς σε αυτήν την ερώτηση;

Για να γίνει δυνατό αυτό, το internet θα πρέπει να μάθει να

Δρ. Αριστείδης Βαγγελάτος\*,  
Μηχανικός Ηλεκτρονικών Υπολογιστών και Πληροφορικής,  
Ε.Α. Ινστιτούτο Τεχνολογίας Υπολογιστών  
Επτακάλκου 13, 11851 Θησείο, Αθήνα  
vagelat@cti.gr

Δρ. Γιώργος Ορφανός\*,  
Μηχανικός Ηλεκτρονικών Υπολογιστών και  
Πληροφορικής,  
Neurosoft A.E. Κωφίδου 24, 14231, Ν. Ιωνία, Αθήνα  
orphan@neurosoft.gr

\* Και οι δύο συγγραφείς έχουν  
εκπνήσει τη διδακτορική τους  
διατριβή σε θέματα Γλωσσικής  
Τεχνολογίας για την Ελληνική  
Γλώσσα, στο Πανεπιστήμιο Πατρών.

## Η Γλωσσική Τεχνολογία (ΓΤ) στην υπηρεσία του Παγκόσμιου Ιστού

Ο παγκόσμιος ιστός δεν είναι πλέον διαχειρίσιμος από τον απλό άνθρωπο που αναζητά μια συγκεκριμένη πληροφορία.

Οι τρέχουσες στατιστικές αναφέρουν ότι περίπου 935 εκατομμύρια άνθρωποι παγκοσμίως έχουν πρόσβαση στον παγκόσμιο ιστό (το αντίστοιχο νούμερο στη χώρα μας ανάγεται σε περίπου 2,7 εκατομμύρια συνδεδεμένους χρήστες). Ο πληθυσμός αυτός έχει πρόσβαση σε περισσότερες από 3 δισεκατομμύρια σελίδες κειμένου. Όπως γίνεται εύκολα κατανοητό, κανείς δεν μπορεί να εγγυηθεί πλέον για την αξιοπιστία του συνόλου της πληροφορίας αυτής, αλλά ακόμα περισσότερο, είναι προφανές ότι η πληροφορία αυτή δεν είναι διαχειρίσιμη από τον απλό χρήστη στα πλαίσια των καθημερινών του αναζητήσεων.

Η πλέον επιτυχημένη παρομοίωση της κατάστασης που επικρατεί σήμερα στο διαδίκτυο είναι η ακόλουθη εικόνα: ας φανταστούμε ένα διψασμένο και καταπονημένο άνθρωπο (που έχει περπατήσει πάνω από 10 ώρες στην έρημο κάτω από τον καυτό ήλιο) να βρίσκει ξαφνικά μπροστά του έναν πυροσβεστικό κρουνό και να προσπαθεί να πιει νερό από αυτόν. Πράγματι, ο όγκος της πληροφορίας («νερό») που υπάρχει στο διαδίκτυο είναι τόσο τραγικά τεράστιος, που πρέπει κανείς («διψασμένος») να σαπαλήσει ώρες μπροστά σε έναν υπολογιστή, εάν θέλει να βρει τελικά κάτι αξιόλογο (χωρίς το αποτέλεσμα να είναι πάντα εγγυημένης ποιότητας ούτε και ακρίβειας).

### Οι σημερινές μηχανές αναζήτησης είναι αρκούντως χαζές!

Δε μπορούμε παρά να δεχτούμε ότι τα περασμένα 10 χρόνια, η εκθετική αύξηση (σε ετήσια βάση) της πληροφορίας που δημοσιεύτηκε στο διαδίκτυο, οδήγησε σε μια νέα πραγματικότητα: αντί για την επίσκεψη στη βιβλιοθήκη, την παραγγελία κάποιου βιβλίου ή την αναμονή για το αντίτυπο ενός συγγράμματος που ευγενικά προσφέρθηκε να μας αποστείλει ο συγγραφέας, σήμερα είμαστε σε θέση να βρούμε το σύνολο σχεδόν της προς αναζήτηση πληροφορίας απλά με μερικά κλικ μπροστά στον προσωπικό μας υπολογιστή.

Είναι έτσι απλά όμως τα πράγματα; «5.480.000 περίπου ιστοσελίδες βρέθηκαν για human language technology & WWW» μας απάντησε γνωστή μηχανή αναζήτησης σε σχετικό ερώτημα που θέσαμε! Υπέροχα! Και τώρα τι; Εάν το

ενδιαφέρον μας περιορίζεται στο πόσες «σχετικές» σελίδες υπάρχουν, τότε θα πρέπει αν είμαστε ευχαριστημένοι! Εάν όμως θέλουμε πραγματικά κάτι να διευδύσουμε μέσα σε αυτές, τότε τα πράγματα δεν είναι πολύ απλά! (Η αξιοπιστία της πληροφορίας που βρέθηκε είναι ένα θέμα που χρήζει ιδιαίτερης συζήτησης από μόνη της...). Εμείς όμως απλά θέλαμε να βρούμε τις «έγκυρες» σελίδες που αναφέρονταν στην χρήση Γλωσσικής Τεχνολογίας (ΓΤ) στον παγκόσμιο ιστό. Και μόνο για να κάνουμε μια γρήγορη ανάγνωση στις 20 με 30 πρώτες σελίδες θα χρειαστεί το υπόλοιπο της ημέρας... Ας το αφήσουμε γι' αργότερα!

### Ο παγκόσμιος ιστός θα πρέπει να μάθει να συνεννοείται!

Έστω ότι ήρθε η ώρα να προγραμματίσουμε τις Πασχαλινές διακοπές μας. Άνοιξη γαρ και επιλέγουμε να πάμε ένα ταξίδι σε κάποια Ευρωπαϊκή πρωτεύουσα. Το αποφασίσαμε μεν σχεδόν μόλις τέλειωσαν οι Χριστουγεννιάτικες αργίες, έχει φτάσει όμως παραμονή της Μεγάλης Εβδομάδας και (ως γνήσιοι απόγονοι των γονέων μας) δεν έχουμε ακόμα κάνει κράτηση για την πτήση. Και τότε κατανοούμε ότι εκείνες οι φθηνές πτήσεις που μας έλεγε ο πράκτορας τα Χριστούγεννα έχουν εξαντληθεί και το κόστος έχει ουσιαστικά πενταπλασιαστεί. Η μόνη λύση είναι κάποιες φθηνές εταιρείες οι οποίες προσφέρουν τη δυνατότητα κρατήσεων μέσω του web και μόνο.

Έτσι ξεκινάμε ένα πρωί να αναζητούμε μια μια αυτές τις εταιρείες και να ψάχνουμε για την πτήση που μας ενδιαφέρει. Εάν είμαστε τυχεροί, μετά από δυο-τρεις ώρες προσπάθειας (αναζήτηση διαφορετικών συνδυασμών πτήσεων και ημερομηνιών), βρίσκουμε μια συμφέρουσα περίπτωση, η οποία όμως δεν φτάνει στο κυρίως αεροδρόμιο της πόλης που θέλουμε αλλά σε ένα κοντινό. Άντε λίγη προσπάθεια ακόμα για να βρούμε δρομολόγια τρένων, λεωφορείων ή αν υπάρχει δυνατότητα ενοικίασης αυτοκινήτου ώστε να φτάσουμε στον τελικό προορισμό... Και ο χρόνος περνά.

Εμείς αυτό που θέλαμε ήταν να βρούμε «αεροπορική πτήση για 2 άτομα με επιστροφή, από Αθήνα προς Δουβλίνο, για τις ημερομηνίες αυτές (μία μέρα πριν ή μετά δεν μας πειράζει) που να φεύγει και να φτάνει στον προορισμό κάποιες λογικές ώρες!»! Υπάρχει δυνατότητα να απαντήσει κανείς σε αυτήν την ερώτηση;

Για να γίνει δυνατό αυτό, το internet θα πρέπει να μάθει να

μιλά και να συνεννοείται! Θα πρέπει να μπορεί να απαντήσει άμεσα και με αξιοπιστία στην ερώτηση «που μπορώ να βρω ένα κατάστημα που να πουλά ποντίκια για τον υπολογιστή μου στην περιοχή X» χωρίς να μας παραπέμπει σε καταστήματα που πουλούν ζώα, ή σε μαγαζιά που είναι σε άλλη χώρα!

Η Γλωσσική Τεχνολογία είναι μεταξύ των τεχνολογιών που υπόσχονται να βοηθήσουν προς την κατεύθυνση αυτή. Σε πολλά ερευνητικά κέντρα στον κόσμο γίνεται αυτή τη στιγμή έρευνα και κατασκευάζονται πρωτότυπα που θα οδηγήσουν σε μια νέα πραγματικότητα στα επόμενα χρόνια.

### Γλωσσική Τεχνολογία

Τι είναι όμως η Γλωσσική Τεχνολογία με πιο απλά λόγια; Όπως διαπιστώνουμε, αποτελείται από δύο όρους: τη γλώσσα και την τεχνολογία.

Η γλώσσα (φυσική) αποτελεί για τον άνθρωπο το κύριο μέσο ανταλλαγής και καταγραφής πληροφοριών, καθώς και μέσο έκφρασης συναισθημάτων. Από την πολυπλοκότητα και το εύρος των σκέψεων, των ιδεών και των συναισθημάτων που μπορούν να εκφραστούν μέσω της γλώσσας, συμπεραίνουμε ότι πρόκειται για ένα μέσο σύνθετο και πολύπλοκο. Σύνηθες χαρακτηριστικό της γλωσσικής επικοινωνίας μεταξύ των ανθρώπων είναι η παρερμηνεία και η παρεξήγηση, γιατί η ερμηνεία των λεγόμενων προϋποθέτει γνώση τόσο για το ίδιο το γλωσσικό σύστημα όσο και για τον κόσμο και τις έννοιες που εκφράζονται μέσω της γλώσσας.

Η τεχνολογία στη συγκεκριμένη περίπτωση είναι η τεχνολογία που παράγεται από την Πληροφορική. Η Πληροφορική μελετά την αναπαράσταση, επεξεργασία και ανταλλαγή πληροφορίας σε φυσικά και τεχνητά συστήματα. Έχει υπολογιστικό, γνωστικό και κοινωνικό προσανατολισμό. Κεντρική έννοια είναι η μετάφραση της

πληροφορίας \_ είτε υπολογιστικά είτε επικοινωνιακά, είτε από φυσικούς οργανισμούς είτε από τεχνουργήματα. Στα φυσικά και τεχνητά συστήματα, η πληροφορία οργανώνεται σε διάφορα επίπεδα, π.χ. στα βιολογικά μόρια και στα ηλεκτρονικά κυκλώματα, στα νευρικά συστήματα και στους ηλεκτρονικούς υπολογιστές, στις ανθρώπινες κοινωνίες και στα καταμετρημένα υπολογιστικά συστήματα.

Η Πληροφορική έρχεται να συμπληρώσει και να προαγάγει όλες τις υπόλοιπες επιστήμες (που παραδοσιακά μελετούν τη φύση, τον άνθρωπο και τα κατασκευάσματα του ανθρώπου)\_ κοινός παρονομαστής όλων είναι η πληροφορία. Η φυσική γλώσσα, το βασικό συστατικό των ανθρώπινων πληροφοριακών συστημάτων, δεν μπορεί παρά να αποτελεί αντικείμενο μελέτης της Πληροφορικής, και πιο συγκεκριμένα της Υπολογιστικής Γλωσσολογίας, ενός κλάδου που συνδυάζει την Πληροφορική με τη Γλωσσολογία.

Τα παραγόμενα της Υπολογιστικής Γλωσσολογίας σε σχέση με εργαλεία, υποδομές, κωδικοποιήσεις, εφαρμογές και γενικότερα ό,τι έχει σχέση με εφαρμοσμένες πρακτικές, έχεις στις μέρες μας πάρει το όνομα «Γλωσσική Τεχνολογία».

### Εφαρμογές της Γλωσσικής Τεχνολογίας

Τα συστήματα που σήμερα ενσωματώνουν Γλωσσική Τεχνολογία ανήκουν στα παρακάτω πεδία εφαρμογών: Αυτόματη μετάφραση (machine translation). Το πρόβλημα της επικοινωνίας ανθρώπων με διαφορετικές μητρικές γλώσσες είναι πολύ παλαιότερο από το πρόβλημα της επικοινωνίας ανθρώπου-υπολογιστή. Τα σημερινά επιτυχημένα συστήματα αυτόματης μετάφρασης εφαρμόζονται πάνω σε κείμενα συγκεκριμένου θεματικού περιεχομένου (π.χ. τεχνικά κείμενα) και παράγουν πρόχειρες μεταφράσεις, οι οποίες κατά κανόνα στη συνέχεια ελέγχονται από ανθρώπους-μεταφραστές.



Ανάκτηση πληροφορίας (information retrieval). Εφόσον μεγάλο κομμάτι της πληροφορίας που διαχειρίζονται οι υπολογιστές έχει τη μορφή ηλεκτρονικών κειμένων, η ανάκτηση (από μια συλλογή) των κειμένων που είναι σχετικά με μια ερώτηση είναι πολύ σημαντική εφαρμογή. Οι ερωτήσεις κατά κανόνα αποτελούνται από λέξεις, φράσεις ή και προτάσεις που ο χρήστης περιμένει να βρει στα κείμενα που αναζητεί. Παρότι τα παραδοσιακά συστήματα βασίζουν την ανάκτηση πληροφορίας σε τεχνικές ταυτοποίησης αλφαριθμητικών (string matching) σε συνδυασμό με στατιστικές μεθόδους, τα τελευταία χρόνια υπάρχει μια ισχυρή τάση υπέρ της ενσωμάτωσης γλωσσικής επεξεργασίας στα συστήματα αυτά, με στόχο η ανάκτηση της πληροφορίας να βασιστεί όχι στις λέξεις αλλά στις έννοιες που εκφράζονται μέσα από τις λέξεις.

Συμπύκνωση/εξαγωγή πληροφορίας (summarization, information extraction). Τα κείμενα σε φυσική γλώσσα είναι χρονοβόρες πηγές πληροφορίας. Τις περισσότερες φορές ο αναγνώστης πρέπει να διαβάσει ένα μεγάλο κομμάτι τους, ή και ολόκληρα, προκειμένου να αντλήσει τις πληροφορίες που θέλει ή για να διαπιστώσει ότι έχασε το χρόνο του. Η αύξηση του πλήθους και του όγκου των ηλεκτρονικών κειμένων έχει ενεργοποιήσει την έρευνα πάνω στην αυτόματη εξαγωγή περιλήψεων\_ τα πρώτα αξιόλογα συστήματα εξαγωγής περιλήψεων έχουν ήδη ενσωματωθεί σε κειμενογράφους. Πέρα όμως από τις περιλήψεις, είναι επιπλέον επιθυμητό από τα κείμενα να εξάγονται μόνο κομμάτια που αναφέρονται πάνω σε ένα θέμα και να παρουσιάζονται στο χρήστη με δομημένη μορφή, π.χ. υπό τη μορφή πίνακα.

Υποστήριξη της συγγραφής και διόρθωσης κειμένων (proofing tools). Η πιο διαδεδομένη χρήση του υπολογιστή, ιδιαίτερα του προσωπικού υπολογιστή, είναι για τη συγγραφή κειμένων. Προς αρωγή του χρήστη στη συγγραφή κειμένων έχει αναπτυχθεί πληθώρα ηλεκτρονικών λεξικών και γλωσσικών εργαλείων, τα οποία είτε είναι αυτόνομες εφαρμογές είτε είναι ενσωματωμένα σε κειμενογράφους: ερμηνευτικά, μεταφραστικά και θησαυρικά λεξικά, συλλαβιστές, ορθογραφικοί, γραμματικοί και υφολογικοί διορθωτές. Προς το παρόν, εύρωστη συμπεριφορά έχουν μόνο τα γλωσσικά εργαλεία που εξετάζουν το κείμενο σε επίπεδο λέξης, δηλ. οι συλλαβιστές και οι ορθογραφικοί διορθωτές. Παράλληλα γίνεται έρευνα και βρισκόμαστε κοντά στην ανάπτυξη εύρωστων γραμματικών διορθωτών.

Φυσικές διεπαφές ανθρώπου-υπολογιστή (man-machine interfaces). Ο καταλληλότερος τρόπος επικοινωνίας του ανθρώπου με τον υπολογιστή είναι μέσω της φυσικής γλώσσας, ειδικά για ανθρώπους που δεν έχουν εμπειρία από υπολογιστές. Η αλληλεπίδραση αυτή προς το παρόν καθίσταται εφικτή μόνο όταν βασίζεται σε συγκεκριμένο σενάριο, το οποίο προδιαγράφει τη δομή των προτάσεων και το λεξιλόγιο που μπορεί σε κάθε βήμα να χρησιμοποιήσει ο χρήστης, δηλ. όταν ο διάλογος είναι καθοδηγούμενος από

τον υπολογιστή και έχει περιορισμένο θεματικό πεδίο.

#### Η Γλωσσική Τεχνολογία αρωγός σε ένα πιο διαχειρίσιμο WWW

Εφόσον το μεγαλύτερο κομμάτι της ανθρώπινης γνώσης είναι καταγεγραμμένο μέσω της φυσικής γλώσσας, η πρόσβαση σε αυτή τη γνώση θα ήταν εξαιρετικά εύκολη, ιδιαίτερα στα πλαίσια του Διαδικτύου, αν η φυσική γλώσσα ήταν κατανοητή από τον υπολογιστή. Τότε, όλοι μας θα είχαμε λύσει πολλά από τα σημερινά προβλήματα που αναφέρθηκαν και παραπάνω. Αρκεί να φανταστούμε τον υπολογιστή:

- να απαντά ερωτήσεις στο τηλέφωνο
- να επεξεργάζεται το κείμενο μιας ιστοσελίδας και να στέλνει μια περίληψή της σε πιθανούς ενδιαφερόμενους
- να μεταφράζει μια καθημερινή εφημερίδα από τα Αγγλικά στα Ελληνικά
- να εξάγει γνώση από περιοδικά/βιβλία/ιστοσελίδες και να δημιουργεί ένα έμπειρο σύστημα που να καθοδηγεί τη λήψη αποφάσεων.

Η ΓΤ έχει προσφέρει εφαρμογές που ικανοποιούν ως ένα βαθμό τις παραπάνω επιθυμίες. Σίγουρα αυτές οι εφαρμογές κάνουν πιο διαχειρίσιμο τον παγκόσμιο ιστό και έχουν αρχίσει να εισάγονται δειλά σε υπηρεσίες που μέσω αυτού προσφέρονται. Απέχουμε όμως πολύ από το να μπορεί ο υπολογιστής να καταλαβαίνει αυτό που καταλαβαίνει ο άνθρωπος ακούγοντας ή διαβάζοντας ένα κείμενο και επομένως να μπορεί να προσφέρει στον άνθρωπο με ακρίβεια τις πληροφορίες που ζητάει.

#### Η ελληνική περίπτωση

Μιλώντας για φυσική γλώσσα, μοιραία κάποια στιγμή έρχεται η ερώτηση «μα για ποια ακριβώς γλώσσα μιλάμε τελικά;». Η απάντηση στο ερώτημα αυτό (Ελληνική Γλώσσα) αποτελεί και την ειδοποιό διαφορά όσον αφορά την έρευνα και ανάπτυξη σε θέματα ΓΤ: πολλά πράγματα έχουν ήδη υλοποιηθεί για την Αγγλική (μήπως τα περισσότερα δεν γίνονται πρώτα σε αυτή τη γλώσσα;) και αντίστοιχα για άλλες ευρείας χρήσης γλώσσες. Τι γίνεται όμως με τα Ελληνικά; Η πραγματικότητα είναι ότι και σε αυτόν τον τομέα, ακολουθούμε από κάποια απόσταση τις εξελίξεις της πρώτης γραμμής (αιχμή της ερευνητικής δραστηριότητας). Και αυτό είναι εν μέρει δικαιολογημένο! Αυτό όμως που θα πρέπει να γίνει απολύτως κατανοητό είναι ότι σε αυτή την περίπτωση, κανείς μα κανείς άλλος δεν πρόκειται να μας βοηθήσει: οι ίδιοι εμείς θα πρέπει να υλοποιήσουμε τις υποδομές και τα απαραίτητα εργαλεία που χρειάζονται. Και είναι ιδιαίτερα αισιόδοξο το γεγονός ότι υπάρχουν σήμερα στην Ελλάδα 3-4 πανεπιστημιακά/ερευνητικά κέντρα όπως και 2-3 εταιρείες που δραστηριοποιούνται με αξιώσεις στα ζητήματα της Γλωσσικής Τεχνολογίας για την ελληνική γλώσσα.

Ανάκτηση πληροφορίας (information retrieval). Εφόσον μεγάλο κομμάτι της πληροφορίας που διαχειρίζονται οι υπολογιστές έχει τη μορφή ηλεκτρονικών κειμένων, η ανάκτηση (από μια συλλογή) των κειμένων που είναι σχετικά με μια ερώτηση είναι πολύ σημαντική εφαρμογή. Οι ερωτήσεις κατά κανόνα αποτελούνται από λέξεις, φράσεις ή και προτάσεις που ο χρήστης περιμένει να βρει στα κείμενα που αναζητεί. Παρότι τα παραδοσιακά συστήματα βασίζονται στην ανάκτηση πληροφορίας σε τεχνικές ταυτοποίησης αλφαριθμητικών (string matching) σε συνδυασμό με στατιστικές μεθόδους, τα τελευταία χρόνια υπάρχει μια ισχυρή τάση υπέρ της ενσωμάτωσης γλωσσικής επεξεργασίας στα συστήματα αυτά, με στόχο η ανάκτηση της πληροφορίας να βασιστεί όχι στις λέξεις αλλά στις έννοιες που εκφράζονται μέσα από τις λέξεις.

Συμπύκνωση/εξαγωγή πληροφορίας (summarization, information extraction). Τα κείμενα σε φυσική γλώσσα είναι χρονοβόρες πηγές πληροφορίας. Τις περισσότερες φορές ο αναγνώστης πρέπει να διαβάσει ένα μεγάλο κομμάτι τους, ή και ολόκληρα, προκειμένου να αντλήσει τις πληροφορίες που θέλει ή για να διαπιστώσει ότι έχασε το χρόνο του. Η αύξηση του πλήθους και του όγκου των ηλεκτρονικών κειμένων έχει ενεργοποιήσει την έρευνα πάνω στην αυτόματη εξαγωγή περιλήψεων\_ τα πρώτα αξιόλογα συστήματα εξαγωγής περιλήψεων έχουν ήδη ενσωματωθεί σε κειμενογράφους. Πέρα όμως από τις περιλήψεις, είναι επιπλέον επιθυμητό από τα κείμενα να εξαγονται μόνο κομμάτια που αναφέρονται πάνω σε ένα θέμα και να παρουσιάζονται στο χρήστη με δομημένη μορφή, π.χ. υπό τη μορφή πίνακα.

Υποστήριξη της συγγραφής και διόρθωσης κειμένων (proofing tools). Η πιο διαδεδομένη χρήση του υπολογιστή, ιδιαίτερα του προσωπικού υπολογιστή, είναι για τη συγγραφή κειμένων. Προς αρωγή του χρήστη στη συγγραφή κειμένων έχει αναπτυχθεί πληθώρα ηλεκτρονικών λεξικών και γλωσσικών εργαλείων, τα οποία είτε είναι αυτόνομες εφαρμογές είτε είναι ενσωματωμένα σε κειμενογράφους: ερμηνευτικά, μεταφραστικά και θησαυρικά λεξικά, συλλαβιστές, ορθογραφικοί, γραμματικοί και υφολογικοί διορθωτές. Προς το παρόν, εύρωστη συμπεριφορά έχουν μόνο τα γλωσσικά εργαλεία που εξετάζουν το κείμενο σε επίπεδο λέξης, δηλ. οι συλλαβιστές και οι ορθογραφικοί διορθωτές. Παράλληλα γίνεται έρευνα και βρισκόμαστε κοντά στην ανάπτυξη εύρωστων γραμματικών διορθωτών.

Φυσικές διεπαφές ανθρώπου-υπολογιστή (man-machine interfaces). Ο καταλληλότερος τρόπος επικοινωνίας του ανθρώπου με τον υπολογιστή είναι μέσω της φυσικής γλώσσας, ειδικά για ανθρώπους που δεν έχουν εμπειρία από υπολογιστές. Η αλληλεπίδραση αυτή προς το παρόν καθίσταται εφικτή μόνο όταν βασίζεται σε συγκεκριμένο σενάριο, το οποίο προδιαγράφει τη δομή των προτάσεων και το λεξιλόγιο που μπορεί σε κάθε βήμα να χρησιμοποιήσει ο χρήστης, δηλ. όταν ο διάλογος είναι καθοδηγούμενος από

τον υπολογιστή και έχει περιορισμένο θεματικό πεδίο.

**Η Γλωσσική Τεχνολογία αρωγός σε ένα πιο διαχειρίσιμο WWW**

Εφόσον το μεγαλύτερο κομμάτι της ανθρώπινης γνώσης είναι καταγραμμένο μέσω της φυσικής γλώσσας, η πρόσβαση σε αυτή τη γνώση θα ήταν εξαιρετικά εύκολη, ιδιαίτερα στα πλαίσια του Διαδικτύου, αν η φυσική γλώσσα ήταν κατανοητή από τον υπολογιστή. Τότε, όλοι μας θα είχαμε λύσει πολλά από τα σημερινά προβλήματα που αναφέρθηκαν και παραπάνω. Αρκεί να φανταστούμε τον υπολογιστή:

- να απαντά ερωτήσεις στο τηλέφωνο
- να επεξεργάζεται το κείμενο μιας ιστοσελίδας και να στέλνει μια περίληψή της σε πιθανούς ενδιαφερόμενους
- να μεταφράζει μια καθημερινή εφημερίδα από τα Αγγλικά στα Ελληνικά
- να εξαγει γνώση από περιοδικά/βιβλία/ιστοσελίδες και να δημιουργεί ένα έμπειρο σύστημα που να καθοδηγεί τη λήψη αποφάσεων.

Η ΓΤ έχει προσφέρει εφαρμογές που ικανοποιούν ως ένα βαθμό τις παραπάνω επιθυμίες. Σίγουρα αυτές οι εφαρμογές κάνουν πιο διαχειρίσιμο τον παγκόσμιο ιστό και έχουν αρχίσει να εισάγονται δειλά σε υπηρεσίες που μέσω αυτού προσφέρονται. Από εμάς όμως πολύ από το να μπορεί ο υπολογιστής να καταλαβαίνει αυτό που καταλαβαίνει ο άνθρωπος ακούγοντας ή διαβάζοντας ένα κείμενο και επομένως να μπορεί να προσφέρει στον άνθρωπο με ακρίβεια τις πληροφορίες που ζητάει.

**Η ελληνική περίπτωση**

Μιλώντας για φυσική γλώσσα, μοιραία κάποια στιγμή έρχεται η ερώτηση «για ποια ακριβώς γλώσσα μιλάμε τελικά;». Η απάντηση στο ερώτημα αυτό (Ελληνική Γλώσσα) αποτελεί και την ειδοποιό διαφορά όσον αφορά την έρευνα και ανάπτυξη σε θέματα ΓΤ: πολλά πράγματα έχουν ήδη υλοποιηθεί για την Αγγλική (μήπως τα περισσότερα δεν γίνονται πρώτα σε αυτή τη γλώσσα;) και αντίστοιχα για άλλες ευρείας χρήσης γλώσσες. Τι γίνεται όμως με τα Ελληνικά; Η πραγματικότητα είναι ότι και σε αυτόν τον τομέα, ακολουθούμε από κάποια απόσταση τις εξελίξεις της πρώτης γραμμής (αιχμή της ερευνητικής δραστηριότητας). Και αυτό είναι εν μέρει δικαιολογημένο! Αυτό όμως που θα πρέπει να γίνει απολύτως κατανοητό είναι ότι σε αυτή την περίπτωση, κανείς μα κανείς άλλος δεν πρόκειται να μας βοηθήσει: οι ίδιοι εμείς θα πρέπει να υλοποιήσουμε τις υποδομές και τα απαραίτητα εργαλεία που χρειάζονται. Και είναι ιδιαίτερα αισιόδοξο το γεγονός ότι υπάρχουν σήμερα στην Ελλάδα 3-4 πανεπιστημιακά/ερευνητικά κέντρα όπως και 2-3 εταιρείες που δραστηριοποιούνται με αξιώσεις στα ζητήματα της Γλωσσικής Τεχνολογίας για την ελληνική γλώσσα.

# συναλλαγή

πρόγραμμα παρακολούθησης οικονομικών συναλλαγών

**Κρατείστε τις συναλλαγές σας, μόνοι σας!**

- ένα «φορητό» πρόγραμμα χρήσιμο όσο η ατζέντα σας  
 - λειτουργεί με απλές προτάσεις  
 - το πρόγραμμα και τα δεδομένα βρίσκονται πάντα μαζί σας  
 σ' ένα memory stick 64MB

**Χαρακτηριστικά:**

- κωδικός πρόσβασης • ευέλικτη περιγραφή των συναλλαγών σύμφωνα με τις ανάγκες σας • παρακολούθηση επιταγών - γραμματίων
- πρόχειρες αποδείξεις • καρτέλες ονομάτων (δ/νοσεις, τηλέφωνα κλπ.) • ομαδοποίηση συναλλαγών με τον καθορισμό ιδιοτήτων • χρωματιστική παρουσίαση αποτελεσμάτων • γραφήματα • εξαγωγή αποτελεσμάτων σε RTF αρχεία κειμένου

"ΚΑΤΑΣΚΕΥΑΣΤΙΚΗ ΕΤΑΙΡΕΙΑ"		Ημερ/νία: 19/11/2004	No 15
Λ. ΜΕΣΟΓΕΙΩΝ 429 ΑΓΙΑ ΠΑΡΑΣΚΕΥΗ ΤΗΛ. 210-6019150 - FAX 210-6018014		Απόδειξη Πληρωμής	
		1.200,00 €	
Ο/Η παρακάτω υπογεγραμμένος/η Γεώργιος Γεωργίου			
επάγγελμα		διεύθυνση	
Έλαβε το ποσό των χιλιά διακόσια € με την επιταγή υπ' αριθμόν 123-4567890			
Απολογία:	Εργασίες:	Πληρωσε Ο/Η	Έλαβε Ο/Η
Εργο: Οικοδομή Αγ. Παρασκευής	Εργασία: Πλακάκια		
Παρατηρήσεις: προκαταβολή			

**ειδική προσφορά για τους αναγνώστες του Δελτίου ΣΠΜΕ 99 ευρώ**

- στην τιμή δεν συμπεριλαμβάνεται ο ΦΠΑ 18%, και τυχόν μεταφορικά έξοδα.
- το πρόγραμμα λειτουργεί σε οποιονδήποτε υπολογιστή διαθέτει USB θύρα και Windows 98 και άνω