

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΚΟΙΝΩΝΙΑ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΠΡΑΞΗ

«ΕΠΕΞΕΡΓΑΣΙΑ ΕΙΚΟΝΩΝ, ΗΧΟΥ ΚΑΙ ΓΛΩΣΣΑΣ»

στο πλαίσιο του ΜΕΤΡΟΥ 3.3

«Έρευνα και Τεχνολογική Ανάπτυξη στην Κοινωνία της Πληροφορίας»

ΕΡΓΟ - 9: ΙΑΤΡΟΛΕΞΗ

ΠΑΡΑΔΟΤΕΟ

Π21: Ημερίδα Διάχυσης Αποτελεσμάτων

Ημερομηνία:	20.12.2007
Έκδοση:	Final
Τύπος:	Εμπιστευτικό

ΠΕΡΙΕΧΟΜΕΝΑ

1	ΕΙΣΑΓΩΓΗ	3
2	ΤΟ ΠΡΟΓΡΑΜΜΑ ΤΗΣ ΗΜΕΡΙΔΑΣ	4
3	ΑΝΑΛΥΤΙΚΗ ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΘΕΜΑΤΩΝ	5
3.1	ΕΙΣΑΓΩΓΗ	5
3.2	ΕΝΑΡΚΤΗΡΙΑ ΣΥΝΕΔΡΙΑ	5
3.3	ΠΡΟΣΚΕΚΛΗΜΕΝΟΙ ΟΜΙΛΗΤΕΣ	6
3.4	ΤΟ ΕΡΓΟ ΙΑΤΡΟΛΕΞΗ	9

1 ΕΙΣΑΓΩΓΗ

Η παρούσα αναφορά αντιστοιχεί στο παραδοτέο Π21: «**Ημερίδα διάχυσης αποτελεσμάτων**» του έργου ΙΑΤΡΟΛΕΞΗ.

Η Ημερίδα που διοργανώθηκε είχε τον τίτλο «Γλωσσική τεχνολογία στο Βιοϊατρικό χώρο» και πραγματοποιήθηκε τη Δευτέρα 10 Δεκεμβρίου 2007 στο ξενοδοχείο Saint George Lecabettus Hotel (Κλεομένους 2, Κολωνάκι). Βασικός στόχος της Ημερίδας ήταν να παρουσιαστούν στο κοινό τα παραγόμενα αποτελέσματα (γλωσσικοί πόροι και υπολογιστικά εργαλεία) που υλοποιήθηκαν στα πλαίσια του ερευνητικού έργου ΙΑΤΡΟΛΕΞΗ.

Στην ανακοίνωση της διοργάνωσης της Ημερίδας η ανταπόκριση ήταν άμεση. Τα άτομα που δήλωσαν πρόθεση συμμετοχής στην Ημερίδα ήταν πάνω από 25, ενώ την ημέρα της διοργάνωσης οι συμμετοχές ξεπέρασαν τις 70.

Αξίζει να σημειωθεί ότι η δημόσια παρουσίαση των αποτελεσμάτων του ΙΑΤΡΟΛΕΞΗ έφερε την ομάδα του έργου σε άμεση επαφή με ειδικούς από διάφορους ερευνητικούς χώρους, όπως μηχανικούς Η/Υ, γλωσσολόγους, λεξικογράφους, ορολόγους, μεταφραστές, ιατρούς, εκπροσώπους φαρμακευτικών εταιρειών κτλ. αλλά και απλούς χρήστες που ενδιαφέρονται για θέματα που άπτονται της εξαγωγής βιοϊατρικής πληροφορίας.

Τέλος, η συμμετοχή προσκεκλημένων ομιλητών προερχόμενων από άλλα ερευνητικά και πανεπιστημιακά ιδρύματα που έχουν ασχοληθεί με ανάλογα ερευνητικά προγράμματα προσέδωσε ιδιαίτερο κύρος στη διοργάνωση, ενώ παράλληλα έδωσε την ευκαιρία για μια πολυδιάστατη προσέγγιση του θέματος, για επικοινωνιακή συζήτηση και ανταλλαγή απόψεων πάνω σε θέματα κοινού ενδιαφέροντος.

2 ΤΟ ΠΡΟΓΡΑΜΜΑ ΤΗΣ ΗΜΕΡΙΔΑΣ

Η έναρξη της Ημερίδας έγινε την προγραμματισμένη ώρα (9.30π.μ.) με πολύ μικρή καθυστέρηση. Το πρόγραμμα, όπως είχε αναρτηθεί εκ των προτέρων στον ιστοχώρο του έργου (www.iatrolexi.gr), ήταν το εξής:

ΠΡΟΓΡΑΜΜΑ

09.30 – 10.00 Προσέλευση - Εγγραφή συνέδρων, καφές
10.00 – 10.10 Χαιρετισμοί (ΓΓΕΤ, Neurosoft, ΕΑΙΤΥ)
10.10 – 10.40 «*Το έργο ΙΑΤΡΟΛΕΞΗ*», Δρ. Αρ. Βαγγελάτος, Ερευνητικός Υπεύθυνος έργου, ΕΑΙΤΥ

Προσκεκλημένοι Ομιλητές

10.40 – 11.10 «*Σώματα Κειμένων και Εξόρυξη Πολυλεκτικών Ιατρικών Όρων: η C-value μέθοδος*», Κ. Φραντζή, Επ. Καθηγήτρια, Παν/μιο Αιγαίου
11.10 – 11.40 «*Πολυμεσικότητα και πολυγλωσσία : Οι σύγχρονες περιπέτειες της τεχνολογίας περιεχομένου*», Στέλιος Πιπερίδης, Ερευνητής, ΙΕΛ - ΕΚ "Αθηνά"
11.40 – 12.10 «*Χαρακτηρισμός ιστοχώρων με περιεχόμενο σχετικό με θέματα υγείας: τρέχουσες προσεγγίσεις και το έργο MedIEQ*», Δρ. Βαγγέλης Καρκαλέτσης, Δ/ντής Έρευνας, Ινστιτούτο Πληροφορικής & Τηλεπικοινωνιών, ΕΚΕΦΕ "Δημόκριτος"

12.10 – 12.40 Διάλειμμα για Καφέ

Βασικοί άξονες έργου

12.40 – 13.10 «*Οντολογία ΙΑΤΡΟΛΕΞΗ: Σχεδιασμός και Ανάπτυξη*» Έλενα Μάντζαρη, Γλωσσολόγος
13.10 – 13.40 «*Όροι της βιοϊατρικής: συλλογή και επεξεργασία του γλωσσικού υλικού*», Δρ. Μ. Πανταζάρα, Γλωσσολόγος
13.40 – 14.10 «*Ιατρική ορολογία και κυβερνοχώρος*», Β. Παπαπαναγιώτου, Ιατρός - Καρδιολόγος

14.10 – 15.00 Διάλειμμα για ελαφρύ γεύμα

Τεχνική Περιγραφή έργου

15.00 – 15.45 Τεχνική Παρουσίαση του έργου, Δρ. Γ. Ορφανός, Μηχανικός Η/Υ και Πληροφορικής
15.45 – 16.30 Συζήτηση

3 ΑΝΑΛΥΤΙΚΗ ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΘΕΜΑΤΩΝ

3.1 ΕΙΣΑΓΩΓΗ

Σύμφωνα με το καθορισμένο πρόγραμμα, η Ημερίδα διαρθρώθηκε σε τρία βασικά μέρη:

- α) εναρκτήρια συνεδρία: καλωσόρισμα, χαιρετισμοί και ομιλία,
- β) εισηγήσεις των προσκεκλημένων ομιλητών,
- γ) το έργο ΙΑΤΡΟΛΕΞΗ: βασικοί άξονες και τεχνική παρουσίαση.

Στη συνέχεια θα παρουσιαστούν πιο αναλυτικά τα θέματα που συζητήθηκαν στην Ημερίδα.

3.2 ΕΝΑΡΚΤΗΡΙΑ ΣΥΝΕΔΡΙΑ

Ο ερευνητικός υπεύθυνος του έργου και ερευνητής στο EAITY κ. Άρης Βαγγελάτος καλωσόρισε τους προσκεκλημένους ομιλητές και τους συμμετέχοντες στην Ημερίδα του ΙΑΤΡΟΛΕΞΗ.

Ακολούθησαν οι επίσημοι χαιρετισμοί εκ μέρους των φορέων υλοποίησης του έργου.

Εκ μέρους του EAITY, χαιρετισμό απεύθυνε ο Διευθυντής του Τομέα Ηλεκτρονικής Διακυβέρνησης κ. Χρήστος Μανωλόπουλος, ενώ την εταιρία NEUROSOFT εκπροσώπησε ο κ. Χρήστος Τσαλίδης, μέλος του Δ.Σ. της εταιρίας.

Ακολούθησε η γενική παρουσίαση του έργου ΙΑΤΡΟΛΕΞΗ από τον ερευνητικό υπεύθυνο του έργου κ. Α. Βαγγελάτο. Μετά από μία εισαγωγή στον κόσμο του WWW και του Web 2.0 όπου αφενός η ταχύτητα των εξελίξεων είναι ιλιγγιώδης και αφετέρου η αναζήτηση βιοϊατρικής πληροφορίας γίνεται ολοένα και πιο σύνθετη, ευρεία και αναγκαία, ο κ. Βαγγελάτος τόνισε την ανάγκη για ανάπτυξη υποδομής στο βιοϊατρικό τομέα με στόχο την εξαγωγή πληροφορίας. Επισημάνθηκαν οι βασικοί στόχοι του έργου και έγινε μια σύντομη επισκόπηση των τελικών παραδοτέων.

3.3 ΠΡΟΣΚΕΚΛΗΜΕΝΟΙ ΟΜΙΛΗΤΕΣ

Την πρόσκληση εκ μέρους του έργου ΙΑΤΡΟΛΕΞΗ για συμμετοχή στην Ημερίδα αποδέχτηκαν τρεις εκλεκτοί συνάδελφοι ερευνητές που προέρχονται από γνωστά ερευνητικά και ανώτατα εκπαιδευτικά ιδρύματα της χώρας. Πιο συγκεκριμένα, με την παρουσία τους μας τίμησαν οι:

- κ. Κατερίνα Φραντζή, Επίκουρη Καθηγήτρια στο Πανεπιστήμιο Αιγαίου
- κ. Στέλιος Πιπερίδης, Ερευνητής στο Ινστιτούτο Επεξεργασίας του Λόγου (ΙΕΛ) – Ερευνητικό Κέντρο «Αθηνά»
- κ. Βαγγέλης Καρκαλέτσης, Διευθυντής Έρευνας στο Εθνικό Κέντρο Έρευνας Φυσικών Επιστημών «ΔΗΜΟΚΡΙΤΟΣ».

1) ΟΜΙΛΙΑ κ. ΦΡΑΝΤΖΗ: «Σώματα Κειμένων και Εξόρυξη Πολυλεκτικών Ιατρικών Όρων: η C-value μέθοδος»

Η επεξεργασία πληροφορίας, ιδιαίτερα δε στους τομείς της βιολογίας και της ιατρικής, δημιουργεί αυξανόμενη ανάγκη για ορολογικούς πόρους. Τα τρία βήματα που είναι απαραίτητα για την απόκτηση ορολογίας (term identification) είναι: α) η αναγνώριση των όρων (term recognition), β) η κατηγοριοποίηση των όρων (term categorization) και γ) το ταίριασμα των όρων σε οντολογίες και βάσεις δεδομένων (term mapping). Η αυτόματη αναγνώριση όρων (automatic term recognition – ATR) θεωρείται το πρώτο βήμα για την απόκτηση ορολογίας, αφού εντοπίζει τους όρους και τις ποικιλίες τους μέσα στα κείμενα και διαχωρίζει τους όρους από τους μη όρους. Η αυτόματη αναγνώριση όρων όμως έχει να αντιμετωπίσει σύνθετα προβλήματα όπως οι πολυλεκτικοί όροι (multi-word terms), οι φωλιασμένοι ή ένθετοι όροι (nested terms), η δημιουργία ακρώνυμων, η μορφολογική ποικιλία κτλ.

Η C/NC-value είναι μία υβριδική μέθοδος που συνδυάζει στατιστικές μετρήσεις με γλωσσικές πληροφορίες και μπορεί να εφαρμοστεί σε κείμενα οποιουδήποτε γνωστικού τομέα για αναγνώριση πολυλεκτικών και φωλιασμένων/ένθετων όρων. Αφού πρώτα γίνει μία μορφοσυντακτική προεπεξεργασία, ακολουθεί η εφαρμογή γλωσσικών φίλτρων με στόχο τον εντοπισμό κυρίως ονοματικών φράσεων. Η μέθοδος χρησιμοποιεί τις στατιστικές

μετρήσεις των τιμών C και NC με στόχο το συνδυασμό τόσο εγγενών γλωσσικών πληροφοριών για τον όρο, όπως τα μορφοσυντακτικά πρότυπα σχηματισμού του, όσο και εξωγενών πληροφοριών, δηλαδή αυτών που προέρχονται από το γλωσσικό του περιβάλλον. Πιο συγκεκριμένα, μετράει τη συνολική συχνότητα της ακολουθίας λέξεων στο σώμα κειμένων, τη συχνότητα της ακολουθίας λέξεων ως μέρος μεγαλύτερων υποψήφιων όρων, τον αριθμό αυτών των μεγαλύτερων όρων και το μήκος του string (σε αριθμό λέξεων). Τα δεδομένα εξόδου του συστήματος έχουν τη μορφή καταλόγου προτεινόμενων όρων ταξινομημένων με βάση τη C/NC-value που έχουν λάβει.

2) ΟΜΙΛΙΑ κ. ΠΙΠΕΡΙΔΗ: «Πολυμεσικότητα και πολυγλωσσία: Οι σύγχρονες περιπέτειες της τεχνολογίας περιεχομένου»

Δύο είναι οι βασικοί στόχοι της τεχνολογίας ανάλυσης πολυμεσικού περιεχομένου:

α) η ανάπτυξη συστημάτων επεξεργασίας περιεχομένου με σκοπό να βοηθήσουμε τους πολίτες να αντιμετωπίσουν την έκρηξη του ψηφιακού περιεχομένου που διατίθεται από διαφορετικές πλατφόρμες (radio, TV, World Wide Web), διαφορετικά μέσα (φωνή (speech), κείμενο (text), εικόνα (image), video) και διαφορετικές γλώσσες

και β) η ανάπτυξη τεχνολογίας για τη σημασιολογική ευρετηρίαση/δεικτοδότηση, κατηγοριοποίηση, περίληψη και διασύνδεση πολυμεσικού, πολυγλωσσικού περιεχομένου.

Ποικίλα είναι τα προβλήματα που συνδέονται με τη διαμεσικότητα (cross-mediality) όπως:

- Συγκέντρωση και συγχρονισμός (time-alignment) μονομεσικών μεταδεδομένων
- Συγκέντρωση, συγκριτική αξιολόγηση και εκτίμηση εμπιστοσύνης/βεβαιότητας της ορθότητας/ακρίβειας των μονομεσικών δεδομένων
- Συγκέντρωση, σύγκριση και εφαρμογή μηχανισμών συμπερασμού (inference mechanisms)
- Χρήση πολυμεσικών οντολογιών
- Διαμεσικότητα εντός του ίδιου πολυμεσικού «εγγράφου», αλλά και μεταξύ διαφορετικών πολυμεσικών «εγγράφων».

Τέλος, επισημάνθηκαν ορισμένοι κίνδυνοι που εγκυμονεί η διαμεσικότητα και προτάθηκαν κάποιες μέθοδοι για τη διαχείρισή τους.

3) ΟΜΙΛΙΑ κ. ΚΑΡΚΑΛΕΤΣΗ: «Χαρακτηρισμός ιστοχώρων με περιεχόμενο σχετικό με θέματα υγείας: τρέχουσες προσεγγίσεις και το έργο MedIEQ»

Το έργο MedIEQ (Quality Labelling of Medical Web Content using Multilingual Information Extraction, www.medieq.org) αποσκοπεί στην αυτοματοποίηση της διαδικασίας ποιοτικού χαρακτηρισμού ιστοχώρων με περιεχόμενο σχετικό με ιατρικά θέματα και θέματα υγείας, αξιοποιώντας και αναπτύσσοντας προηγμένες τεχνολογίες πολυγλωσσικής αναζήτησης και εξαγωγής πληροφορίας. Πιο συγκεκριμένα, το έργο απαιτεί εξειδικευμένες τεχνολογίες για: α) την απόδοση χαρακτηρισμών ποιότητας που να είναι αυτόματα επεξεργάσιμοι, β) τη συντήρηση των χαρακτηρισμών ποιότητας, γ) τον εντοπισμό ιστοχώρων χωρίς χαρακτηρισμό ποιότητας και δ) τη βελτιστοποίηση της προσβασιμότητας στους χαρακτηρισμούς ποιότητας από τους χρήστες. Επιμέρους στόχοι του έργου είναι:

- να αναπτύξει ένα σχήμα για τον ποιοτικό χαρακτηρισμό ιστοχώρων σχετικών με θέματα υγείας και να παράσχει τα εργαλεία που να υποστηρίζουν τη δημιουργία, τη συντήρηση και την προσβασιμότητα στα δεδομένα που έχουν χαρακτηριστεί σύμφωνα με το σχήμα
- να εξειδικεύσει μία μεθοδολογία για την ανάλυση περιεχομένου των ιατρικών ιστοχώρων σύμφωνα με το σχήμα MedIEQ
- να ενσωματώσει αυτές τις τεχνολογίες σε ένα πρότυπο σύστημα ποιοτικού χαρακτηρισμού
- να επιδείξει το παραγόμενο πρότυπο σε επτά διαφορετικές γλώσσες (ισπανικά, καταλανικά, γερμανικά, αγγλικά, ελληνικά, τσεχικά και φιλανδικά), καθώς και δύο εφαρμογές ποιοτικού χαρακτηρισμού (πιστοποίηση τρίτου, ταξινόμηση).

3.4 ΤΟ ΕΡΓΟ ΙΑΤΡΟΛΕΞΗ

Η παρουσίαση του ΙΑΤΡΟΛΕΞΗ από τα μέλη της ομάδας του έργου πραγματοποιήθηκε σε δύο μέρη (θεωρητικό και τεχνικό).

Στο πρώτο μέρος αναπτύχθηκαν οι βασικοί θεωρητικοί και μεθοδολογικοί άξονες του έργου, και πιο συγκεκριμένα: ο σχεδιασμός και η ανάπτυξη της οντολογίας (κ. Έλενα Μάντζαρη), θέματα που σχετίζονται με τη συλλογή και την επεξεργασία της βιοϊατρικής ορολογίας από τη σκοπιά των γλωσσολόγων (κ. Μαβίνα Πανταζάρα) και τέλος η αναζήτηση της ιατρικής πληροφορίας στον κυβερνοχώρο από τη σκοπιά των ιατρών (κ. Βασίλης Παπαπαναγιώτου).

Στο δεύτερο μέρος έγινε η τεχνική παρουσίαση του έργου που περιελάμβανε και την επίδειξη των εργαλείων και των πόρων που υλοποιήθηκαν στα πλαίσια του έργου (κ. Γιώργος Ορφανός).

1) ΟΜΙΛΙΑ κ. ΜΑΝΤΖΑΡΗ «Οντολογία του ΙΑΤΡΟΛΕΞΗ: Σχεδιασμός και ανάπτυξη»

Πρόκειται για συνοπτική παρουσίαση των όσων αναπτύχθηκαν εκτενώς στο παραδοτέο Π2: «Μοντέλο Αναπαράστασης Οντολογίας». Μετά από μία σύντομη ανασκόπηση του όρου «οντολογία» και των βασικών σχετικών εννοιών, παρουσιάστηκε η μεθοδολογία ανάπτυξης της οντολογίας του ΙΑΤΡΟΛΕΞΗ, καθώς και οι βασικές αρχές σχεδιασμού της. Τεκμηριώθηκαν τα κριτήρια επιλογής του συγκεκριμένου μοντέλου, και επιπλέον κατονομάστηκαν και επεξηγήθηκαν οι οντότητες του μοντέλου (έννοιες, σημασιολογικός τύπος, σημασιολογικές σχέσεις και σχεδιότυπο). Ως γλώσσα αναπαράστασης της οντολογίας επιλέχθηκε η OWL-Lite, ενώ ως εργαλείο διαχείρισης της οντολογίας υιοθετήθηκε το Protégé.

2) ΟΜΙΛΙΑ κ. ΠΑΝΤΑΖΑΡΑ «Όροι της βιοϊατρικής: συλλογή και επεξεργασία του γλωσσικού υλικού»

Πρόκειται για συνοπτική παρουσίαση των όσων αναπτύχθηκαν εκτενώς στα παραδοτέα Π5: «Μονολεκτικοί όροι βιοϊατρικής», Π6: «Κανόνες αναγνώρισης πολυλεκτικών όρων βιοϊατρικής» και Π7: «Πολυλεκτικοί όροι βιοϊατρικής». Πιο συγκεκριμένα, έγινε αναφορά τόσο στη συλλογή των όρων (πηγές και μέθοδοι) όσο και στη γλωσσική επεξεργασία των

συλλεχθέντων μονολεκτικών και πολυλεκτικών όρων (κανονικοποίηση και κωδικοποίηση). Τέλος, δόθηκε μεγαλύτερη έμφαση στο κομμάτι που αφορούσε την επεξεργασία των πολυλεκτικών όρων της βιοϊατρικής, η οποία περιελάμβανε μορφοσυντακτική, ορολογική, λεξιλογική και σημασιολογική ανάλυση, με στόχο να αναπτυχθεί το σύστημα αναγνώρισης των όρων μέσα τα κείμενα.

3) ΟΜΙΛΙΑ κ. ΠΑΠΑΠΑΝΑΓΙΩΤΟΥ «Ιατρική ορολογία και κυβερνοχώρος»

Ποιοι αναζητούν ιατρική πληροφορία στο διαδίκτυο και γιατί; Χρήστες που τους απασχολούν ποικίλα ιατρικά θέματα κυρίως για τον εαυτό τους ή συγγενικά τους πρόσωπα, καθώς και επαγγελματίες υγείας, φοιτητές κτλ. Τόσο οι καταναλωτές-χρήστες όσο και οι παραγωγοί αυτής της πληροφορίας έχουν ανάγκη αφενός από καλύτερα μέσα για την αναζήτηση της πληροφορίας, αφετέρου από ποιότητα και αξιοπιστία των δεδομένων. Ο ιατρός αποκτά πλέον νέο ρόλο: θα πρέπει να είναι σε θέση να αντιμετωπίσει έναν καλά ενημερωμένο ή και παραπληροφορημένο ασθενή, να προτείνει αξιόπιστους ιστοχώρους, αλλά και να χρησιμοποιεί ο ίδιος το διαδίκτυο για τη δική του εξέλιξη και εκπαίδευση.

Στα πλαίσια του έργου ΙΑΤΡΟΛΕΞΗ, οι ιατροί ανέλαβαν να επιμεληθούν τους όρους που συλλέχθηκαν από τις πηγές και να ελέγξουν την απόδοση δύσκολων περιπτώσεων στην ελληνική γλώσσα (ιοί, βακτήρια, μικροοργανισμοί, χημικές ουσίες και φάρμακα, όροι της σύγχρονης βιοϊατρικής τεχνολογίας, λατινικές ονομασίες, εκκλαϊκευμένες ονομασίες κτλ.). Συζητήθηκαν ορισμένες προβληματικές κατηγορίες όρων και παρουσιάστηκαν οι συγκεκριμένες επιλογές που υιοθετήθηκαν για τις ανάγκες του έργου.

4) ΟΜΙΛΙΑ κ. ΟΡΦΑΝΟΥ «Τεχνική παρουσίαση του έργου ΙΑΤΡΟΛΕΞΗ»

Στόχος ήταν να δοθεί μία συνολική εικόνα σχετικά με τους στόχους, τη μεθοδολογία και τα εργαλεία που αναπτύχθηκαν στα πλαίσια του έργου. Ως τεχνολογικοί στόχοι του έργου είχαν οριστεί οι εξής: α) η βελτίωση του τρόπου αναζήτησης/ανάκλησης βιοϊατρικής πληροφορίας, β) η δημιουργία της βασικής υποδομής από πλευράς γλωσσικής τεχνολογίας για περαιτέρω εφαρμογές (π.χ. κατηγοριοποίηση κειμένων, εξόρυξη πληροφορίας) και γ) η ευρωστία και πληρότητα του συστήματος. Η προσέγγιση που υιοθετήθηκε ήταν κατά βάση «γλωσσολογική» (και όχι «στατιστική»). Η μεθοδολογία περιελάμβανε τα εξής βήματα: α) εμπλουτισμός ή/και ανάπτυξη γλωσσικών πόρων (Ορολογικό Λεξικό, Συλλογή κειμένων, Οντολογία), β) προσαρμογή ή/και ανάπτυξη εργαλείων υποδομής (Μηχανισμός Κλίσης

Πολυλεκτικών Όρων, Μορφοσυντακτικός Αναλυτής, Σημασιολογικός Σχολιαστής, Συλλογέας Συνάψεων), γ) ανάπτυξη εργαλείων για τους χρήστες (Ορθογράφος, Περιηγητής Οντολογίας, Μηχανή Αναζήτησης).

Τέλος, πραγματοποιήθηκε επίδειξη τόσο των εργαλείων υποδομής όσο και των εργαλείων που απευθύνονται στους χρήστες και πιο συγκεκριμένα παρουσιάστηκαν τα εξής:

- Μετατροπή κειμένων σε επεξεργάσιμη μορφή
- Συλλογή άγνωστων λέξεων
- Κλίση όρων
- Διαχείριση της οντολογίας
- Μορφοσυντακτικός και σημασιολογικός σχολιασμός κειμένων
- Δεικτοδότηση κειμένων
- Ορθογράφος
- Περιηγητής Οντολογίας
- Μηχανή Αναζήτησης