

Ανάπτυξη Οντολογίας Βιοϊατρικών Όρων

**Α. Βαγγελάτος², Γ. Ορφανός², Χ. Τσαλίδης²,
Χ. Καλαμαρά³**

www.iatrolexi.cti.gr

¹ Ερευνητικό Ακαδημαϊκό Ινστιτούτο Τεχνολογίας Υπολογιστών (ΕΑΙΤΥ)

² Neurosoft A.E.

³ Ευρωκλινική Αθηνών



Έργο: ΙΑΤΡΟΛΕΞΗ

Ανάπτυξη παντού. Ανάπτυξη για όλους.

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



ΕΑΙΤΥ



Ερευνητικό Ακαδημαϊκό Ινστιτούτο Τεχνολογίας Υπολογιστών

ΝΠΙΔ μη κερδ. χαρακτήρα, εποπτευόμενο από το ΥπΕΠΘ, <http://www.cti.gr>

Έδρα: Πάτρα (Πανεπιστήμιο Πατρών),

Παράρτημα: Αθήνα

Προσωπικό: 200 εργαζόμενοι

Neurosoft A.E.

Ελληνική Εταιρεία Παραγωγής Λογισμικού

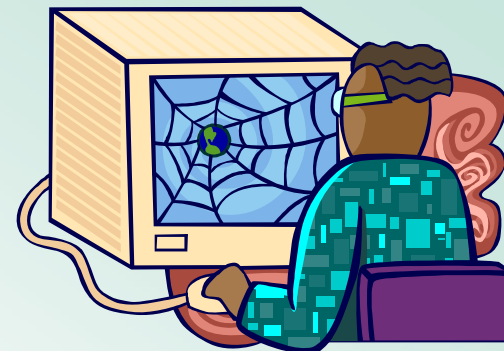
www.neurosoft.gr

Τμήμα με εμπειρία στην Επεξεργασία Φυσικής Γλώσσας

(Παραγωγή ορθογράφου, θησαυρού, κλπ)

www.neurolingo.gr

Εισαγωγή (I)



Το **μέγεθος** της ιατρικής - βιοϊατρικής πληροφορίας που παράγεται καθημερινά στο σύγχρονο κόσμο είναι τεράστιο και ίσως δύσκολο ακόμα και να εκφραστεί με τρόπο απλό.

Οι πηγές είναι πολλές (φορείς υγείας, εκπαίδευση, ερευνητικοί φορείς, θεσμικοί φορείς, κλπ).

Τα μέσα διακίνησης, άλλαξαν με την πάροδο του χρόνου και η ψηφιακή μορφή τείνει να επικρατήσει ολοσχερώς.

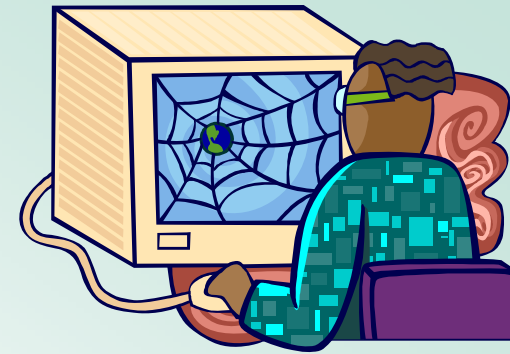
Ο κυβερνοχώρος (WWW) και σε αυτή την περίπτωση είναι το μέσο που υπερिशύει με βάση τα μεγέθη.

Το αναπόφευκτο ερώτημα είναι αν και με ποιο τρόπο είναι δυνατό να αξιοποιηθεί πλέον αυτή η υπερπληθώρα πληροφορίας και να μετατραπεί σε γνώση.

Εισαγωγή (II)



Η βιοϊατρική πληροφορία που υπάρχει σήμερα διαθέσιμη σε ψηφιακή μορφή, έχει ήδη χαρακτηριστεί **«ζούγκλα πληροφορίας»** (information jungle - G. Eysenbach), μορφής με βασικό πλέον (αρνητικό) χαρακτηριστικό την **ιδιαίτερη δυσκολία αξιοποίησής της**. Ο στόχος της επιστημονικής – ερευνητικής προσπάθειας είναι η μετάβαση από την κατάσταση αυτή σε μια νέα όπου η πληροφορία θα είναι δομημένη με τρόπο ώστε να μπορεί να χαρακτηριστεί ως συλλογή – **δεξαμενή γνώσης (knowledge repository)** και η οποία θα επιτρέπει τη χρήση προηγμένων τεχνολογιών διαχείρισης γνώσης (knowledge management).



Εισαγωγή (III)

Σήμερα, ο απλός χρήστης (είτε ερευνητής, είτε ασθενής, είτε άλλης κατηγορίας) έχει ιδιαίτερα απλοϊκά εργαλεία ως υποβοήθηση στην προσπάθειά του να πλοηγηθεί ή να αναζητήσει ιατρική πληροφορία.

Απλές μηχανές αναζήτησης γενικού σκοπού με μόνη δυνατότητα κάποιες λογικές εκφράσεις.

Ειδικά δε για την Ελληνική γλώσσα με τις ιδιαιτερότητές της (πλούσια μορφολογία, μακρόχρονη εξέλιξη, κλπ) τα προβλήματα είναι ακόμα πιο μεγάλα.



Εισαγωγή (IV)



Στις μέρες μας, η αναζήτηση ιατρικής πληροφορίας στον κυβερνοχώρο είναι ένας από τους συνηθέστερους λόγους πλοήγησης: μελέτες δείχνουν ότι 50 – 75% των χρηστών του World Wide Web το έχουν χρησιμοποιήσει για την αναζήτηση ιατρικών πληροφοριών με συχνότητα πάνω από 5 φορές το μήνα κατά μέσο όρο.

Με βάση στοιχεία του 2005, το NHS Direct (Ιατρική πύλη του συστήματος Υγείας της Μεγάλης Βρετανίας - www.nhsdirect.nhs.uk) δέχεται πάνω από ένα εκατομμύριο επισκέψεις κάθε μήνα, με αύξηση περίπου 58% σε σχέση με την προηγούμενη χρονιά!

Έργο ΙΑΤΡΟΛΕΞΗ

Αντικείμενο του έργου ΙΑΤΡΟΛΕΞΗ είναι η δημιουργία της απαραίτητης γλωσσικής υποδομής για την Ελληνική γλώσσα, που θα επιτρέψει σε πρώτη φάση την καλύτερη διαχείριση και επεξεργασία της βιοϊατρικής πληροφορίας που υπάρχει σε ψηφιακή μορφή.

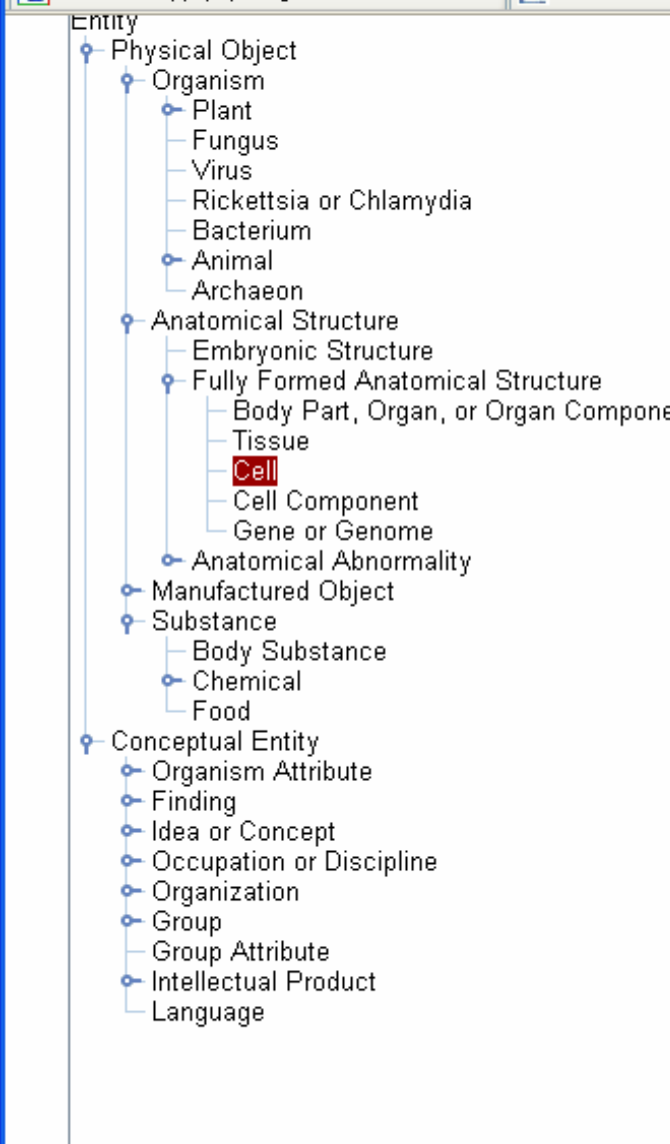
Αυτό θα καταστεί δυνατό με τη δημιουργία και διάθεση στους χρήστες εξελιγμένων εργαλείων για την αναζήτηση, τη συσχέτιση και το χαρακτηρισμό των βιοϊατρικών κειμένων.

Οντολογία Βιοϊατρικών Όρων

Πιο συγκεκριμένα, το έργο αποσκοπεί στη δημιουργία περιβάλλοντος ανάπτυξης, πιστοποίησης και αξιοποίησης **οντολογίας βιοϊατρικών όρων** με ιδιαίτερο χαρακτηριστικό τη δυνατότητα διαχείρισης πολυλεκτικών όρων.

Η Οντολογία αποτελεί τον πλέον ενδεδειγμένο και αξιοποιήσιμο μηχανισμό αποτύπωσης γνώσης μίας θεματικής περιοχής. Καθορίζει τον τρόπο αναπαράστασης με τον οποίο οι έννοιες τα αντικείμενα και οι μεταξύ τους σχέσεις αποτυπώνονται.

Οι κυριότεροι λόγοι για την κατασκευή μιας οντολογίας είναι 1) Αναπαράσταση της γνώσης με ένα μοντέλο εξίσου καταληπτό σε ανθρώπους και μηχανές. 2) Καταγραφή των παραδοχών. 3) Επαναχρησιμοποίηση της αποτυπωμένης γνώσης. 4) Ανάλυση της γνώσης. 5) Προηγμένη ευρετηρίαση, ανάκτηση και παρουσίαση πληροφορίας, δυνατότητα εξόρυξης δεδομένων.


 Concepts
 Relations
 Raw records
Semantic Type: Cell**TUI:** T025**Definition:** The fundamental structural and functional unit of living organisms.

Παραγόμενα του έργου

- Μεθοδολογία Ανάπτυξης Οντολογιών
- Υπολογιστικοί πόροι και εργαλεία
- Σώμα (corpus) βιοϊατρικών κειμένων
- Εφαρμογές Ιστού





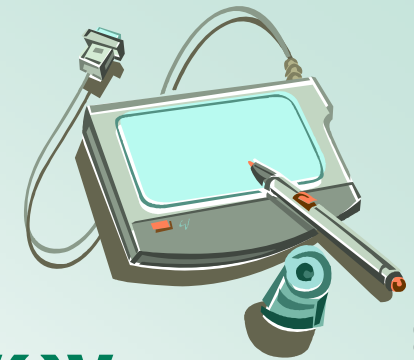
Μεθοδολογία Ανάπτυξης Οντολογιών

- καθορισμός της αρχικής ταξινόμιας πάνω στην οποία θα βασιστεί η οντολογία
- συλλογή κειμένων της θεματικής περιοχής-στόχου
- ανίχνευση του ειδικού λεξιλογίου που χρησιμοποιούν τα κείμενα
- καθορισμός μορφοσυντακτικών κανόνων περιγραφής των όρων
- εξαγωγή υποψήφιων όρων
- εμπλουτισμό της οντολογίας με επιλεγμένους όρους και σχέσεις



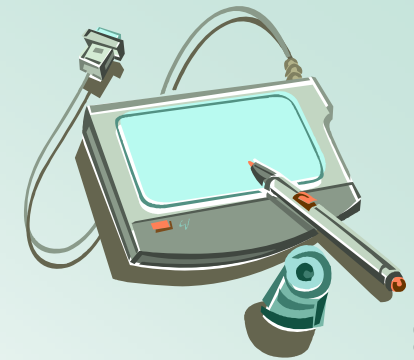
Υπολογιστικοί πόροι και εργαλεία

- Μορφολογικό Λεξικό εμπλουτισμένο με βιοϊατρικούς όρους.
- Μορφοσυντακτικός Σχολιαστής (Morphosyntactic Tagger)
- Αναγνωριστής Ονοματικών Φράσεων (Noun-Phrase Chunker), ο οποίος θα ανιχνεύει υποψήφιους πολυλεκτικούς όρους σε κείμενα με χρήση μορφοσυντακτικών κανόνων.
- Περιβάλλον Ανάπτυξης Οντολογιών βασισμένο στο state-of-the-art σύστημα Protégé (<http://protege.stanford.edu>)
- Σημασιολογικός Σχολιαστής (Semantic Tagger), ο οποίος θα επισυνάπτει πληροφορία από την οντολογία σε όρους που αναγνωρίζονται σε κείμενα.
- Μηχανισμός δεικτοδότησης βιοϊατρικών κειμένων.



Σώμα Βιοϊατρικών κειμένων

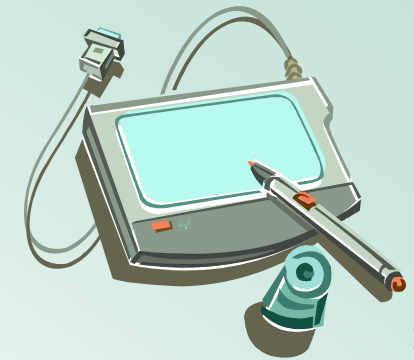
το οποίο αφενός θα οδηγήσει την όλη διαδικασία εξαγωγής ορολογία και ανάπτυξης της οντολογίας και αφετέρου θα είναι διαθέσιμο στους χρήστες/ερευνητές για ανάκτηση/εξόρυξη πληροφορίας.



Εφαρμογές Ιστού

Με τις εφαρμογές ιστού ο εξειδικευμένος και μη χρήστης του διαδικτύου θα μπορεί:

- Να ελέγχει την ορθογραφία μίας λέξης (βιοϊατρικός όρος ή μέρος βιοϊατρικού όρου).
- Να περιηγείται στην οντολογία πραγματοποιώντας σύνθετες αναζητήσεις για σχέσεις που διέπουν όρους της οντολογίας.
- Να εισάγει ένα νέο κείμενο και να λαμβάνει ως αποτέλεσμα το κείμενο με μορφοσυντακτικούς και σημασιολογικούς σχολιασμούς των όρων του κειμένου (annotations).



Εφαρμογές Ιστού

Να αναζητά κείμενα με συγκεκριμένους όρους ή συνδυασμό όρων ή σχέσεων που ισχύουν σε όρους.

Η αναζήτηση θα μπορεί να γίνει

- α) στο σώμα κειμένων που θα συλλεχθεί στα πλαίσια του έργου και
- β) σε κείμενα του Παγκόσμιου Ιστού μέσω ειδικού meta-search engine που θα αναπτυχθεί για το σκοπό αυτό και το οποίο θα χρησιμοποιεί το Google (<http://www.google.com>).

Για τις παραπάνω λειτουργίες, θα αναπτυχθούν και αντίστοιχες υπηρεσίες ιστού (web services), με τη βοήθεια των οποίων ο δικτυακός τόπος θα προσφέρει τη δυνατότητα αξιοποίησης των αποτελεσμάτων από λογισμικό τρίτων στον ερευνητικό χώρο.



ΙΑΤΡΟΛΕΞΗ - Ανάπτυξη Υποδομής Γλωσσικής Τεχνολογίας για το Βιοϊατρικό Τομέα

[Εισαγωγή](#)[Περιγραφή](#)[Αποτελέσματα](#)[Νέα](#)[Χώρος Συζητήσεων](#)[Εταίροι](#)[Σύνδεσμοι](#)

199 Επισκέπτες
από
13-08-2006

Επιχειρησιακό Πρόγραμμα «Κοινωνία της Πληροφορίας»

ΜΕΤΡΟ 3.3

Πρόγραμμα

«ΕΠΕΞΕΡΓΑΣΙΑ ΕΙΚΟΝΩΝ, ΗΧΟΥ ΚΑΙ ΓΛΩΣΣΑΣ»

Τίτλος έργου : **ΙΑΤΡΟΛΕΞΗ: Ανάπτυξη Υποδομής Γλωσσικής Τεχνολογίας για το Βιοϊατρικό Τομέα**

Κωδικός έργου : 9

Ανάδοχος: Neurosoft A.E.

Ερευνητικός Φορέας: EAITY

Θεματικός Τομέας: Εξαγωγή, εξόρυξη και διαχείριση μονογλωσσικών και πολυγλωσσικών δεδομένων, καθώς και πολυμεσικών μεταδεδομένων για ποικιλία εφαρμογών

Αρμόδια Διεύθυνση ΓΓΕΤ: Αυτοτελές Τμήμα Διαχείρισης Διαρθρωτικών Προγραμμάτων (ΔΙΑ)





ΙΑΤΡΟΛΕΞΗ - Ανάπτυξη Υποδομής Γλωσσικής Τεχνολογίας για το Βιοϊατρικό Τομέα



Εισαγωγή

Περιγραφή

Αποτελέσματα

Νέα

Χώρος Συζητήσεων

Εταίροι

Σύνδεσμοι

 199 Επισκέπτες
από
13-08-2006

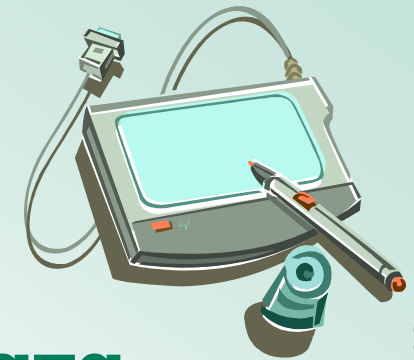
Χώρος Συζητήσεων

Συχνές Ερωτήσεις
 Αναζήτηση
 Αναπάντητα
 Προφίλ
 Αποσύνδεση [vagelat]

Η τελευταία επίσκεψή σας ήταν στις Παρ Σεπ 15, 2006 11:11 am

Χώρος Συζητήσεων

Χώρος Συζητήσεων	Θέματα	Δημοσιεύσεις	Τελευταία Δημοσίευση
ΙΑΤΡΟΛΕΞΗ			
Γενικά Θέματα Στο χώρο αυτό θα συζητούνται γενικά θέματα.	3	4	Παρ Σεπ 15, 2006 9:40 am vagelat →
Πρώτο Παραδοτέο Στο χώρο αυτό θα συζητούνται τα θέματα που αφορούν το 1ο παραδοτέο του έργου.	1	1	Δευ Αύγ 14, 2006 12:12 pm vagelat →
Σχετική Βιβλιογραφία Στο χώρο αυτό θα συζητούνται τα θέματα που αφορούν στην υπάρχουσα βιβλιογραφία.	0	0	Δεν υπάρχουν δημοσιεύσεις
Σώμα Ιατρικών Κειμένων Στο χώρο αυτό θα συζητούνται τα θέματα που αφορούν στην εύρεση και κατηγοριοποίηση Σωμάτων Ιατρικών Κειμένων.	0	0	Δεν υπάρχουν δημοσιεύσεις
UMLS Στο χώρο αυτό θα συζητούνται τα θέματα που αφορούν το UMLS.	0	0	Δεν υπάρχουν δημοσιεύσεις



Αναμενόμενα Αποτελέσματα

Ο βασικός στόχος του έργου για την υλοποίηση υποδομής που θα βελτιώσει την δυνατότητα διαχείρισης βιοϊατρικής πληροφορίας, είναι σίγουρο ότι έχει πληθώρα ωφελουμένων είτε στη μορφή που θα προκύψει στα πλαίσια της παρούσας πρότασης, είτε με βελτιώσεις προσθήκες που θα μπορεί κάποιος να κάνει με βάση πιο εξειδικευμένη κατεύθυνση.



Αναμενόμενα Αποτελέσματα

- Πιο συγκεκριμένα, η επιστημονική κοινότητα του κλάδου της βιοϊατρικής, θα αποκτήσει μια σειρά εργαλείων που θα υποβοηθήσουν σε μεγάλο βαθμό το έργο της: από την καλύτερη αναζήτηση σε ιατρικές πηγές, έως την αυτόματη δεικτοδότηση, χαρακτηρισμό ή άλλη επεξεργασία των κειμένων.
- Παράλληλα ο απλός χρήστης μέσω μιας εξειδικευμένης πύλης αναζήτησης ιατρικών θεμάτων, θα μπορεί να εντοπίσει ευκολότερα πιο αξιόπιστη και χρήσιμη πληροφορία για θέματα υγείας που τον απασχολούν.
- Επιπλέον φορείς υγείας που παράγουν πληροφορία ιατρικού περιεχομένου (νοσοκομεία, δημόσια διοίκηση, κλπ) θα μπορούν να επεξεργαστούν τα δεδομένα τους με πληθώρα τρόπων με στόχο την ικανοποίηση των σκοπών λειτουργίας τους (π.χ. μελέτες συσχέτισης νόσων – ιατρικών εξετάσεων για τα νοσοκομεία, κλπ).

www.iatrolexi.cti.gr

Ανάπτυξη Οντολογίας Βιοϊατρικών Όρων



Εν κατακλείδι,

Με το πέρας του έργου ευελπιστούμε να διαθέσουμε σε κάθε ενδιαφερόμενο μια υποδομή, η οποία είτε στη μορφή που θα παραχθεί, είτε με τις απαραίτητες αλλαγές/προσθήκες, θα βελτιώνει τις δυνατότητες διαχείρισης της βιοϊατρικής πληροφορίας που υπάρχει σε ηλεκτρονική μορφή.

ΕΥΧΑΡΙΣΤΩ