

**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**

**Π7: Πολυλεκτικοί Όροι Βιοϊατρικής**

**(στο παραδοτέο περιλαμβάνεται οπτικός δίσκος που περιέχει τους όρους)**

<b>Ημερομηνία:</b>	27.02.2007
<b>Έκδοση:</b>	Final
<b>Τύπος:</b>	Εμπιστευτικό

## ΠΕΡΙΕΧΟΜΕΝΑ

1. ΕΙΣΑΓΩΓΗ.....	3
2. ΣΥΓΚΕΝΤΡΩΣΗ ΠΟΛΥΛΕΚΤΙΚΩΝ ΌΡΩΝ - ΠΗΓΕΣ .....	4
3. ΔΙΕΡΓΑΣΙΕΣ ΕΠΕΞΕΡΓΑΣΙΑΣ ΤΩΝ ΠΟΛΥΛΕΚΤΙΚΩΝ ΌΡΩΝ.....	6
4. ΣΥΣΤΑΣΗ ΤΟΥ ΟΡΟΛΟΓΙΟΥ ΤΩΝ ΠΟΛΥΛΕΚΤΙΚΩΝ ΌΡΩΝ.....	7
5. ΟΔΗΓΙΕΣ ΧΡΗΣΗΣ ΤΟΥ ΟΠΤΙΚΟΥ ΔΙΣΚΟΥ.....	10

## 1. ΕΙΣΑΓΩΓΗ

Ένα από τα βήματα για την ανάπτυξη της οντολογίας αποτελεί η κατάρτιση της **μακροδομής** της, δηλαδή του συνόλου των όρων που αποτελούν πραγματώσεις των βιοϊατρικών εννοιών και θα απαρτίζουν το λεξιλόγιο της. Στόχος της ΕΕ3, στην οποία εντάσσεται και το παρόν παραδοτέο, αποτελεί η επεξεργασία των συλλεχθέντων βιοϊατρικών κειμένων και άλλων λεξικογραφικών πηγών στο πεδίο της βιοϊατρικής προκειμένου να εξαχθούν οι όροι που θα αποτελέσουν τη βάση πάνω στην οποία θα χτιστεί η οντολογία του έργου.

Στο πλαίσιο της ΕΕ3, το παρόν παραδοτέο **Π7: «Πολυλεκτικοί Όροι Βιοϊατρικής»** μαζί με το παραδοτέο **Π5: «Μονολεκτικοί όροι Βιοϊατρικής»** συγκροτούν τον αρχικό κατάλογο των όρων της ορολογίας.

Το Π7: «Πολυλεκτικοί Όροι Βιοϊατρικής» περιλαμβάνει:

1. αρχείο δεδομένων με το σύνολο των πολυλεκτικών όρων που έχουν συγκεντρωθεί,
2. σύντομη τεκμηρίωση για το περιεχόμενο των όρων του αρχείου δεδομένων.

Στην τεκμηρίωση που ακολουθεί περιγράφονται:

- οι τρόποι και οι πηγές συγκέντρωσης των πολυλεκτικών όρων,
- οι διεργασίες επεξεργασίας των όρων,
- στοιχεία σχετικά με τη σημασιολογική σύσταση της υπάρχουσας συλλογής των πολυλεκτικών όρων
- οδηγίες χρήσης του οπτικού δίσκου που περιέχει τους πολυλεκτικούς όρους.

## 2. ΣΥΓΚΕΝΤΡΩΣΗ ΠΟΛΥΛΕΚΤΙΚΩΝ ΟΡΩΝ - ΠΗΓΕΣ

Για τη συγκέντρωση των πολυλεκτικών όρων ακολουθήθηκαν οι ακόλουθες τρεις παράλληλες και συμπληρωματικές διαδικασίες:

1. **Επεξεργασία του αγγλο-ελληνικού λεξιλογίου βιοϊατρικών όρων MeSH** (έκδοση 1996): Ο στόχος αυτής της εργασίας ήταν διπτός: αφενός να εντοπιστεί το βασικό λεξιλόγιο της βιοϊατρικής ορολογίας που εκφράζεται μέσω των πολυλεκτικών όρων και το οποίο με την κατάλληλη επεξεργασία θα εμπλουτίσει τη μακροδομή της οντολογίας ΙΑΤΡΟΛΕΞΗ και, αφετέρου να συγκεντρωθεί ένα ικανό και επαρκές σώμα πολυλεκτικών όρων, το οποίο θα χρησιμεύσει ως βάση για τη **διατύπωση μιας γραμματικής σχηματισμού και αναγνώρισης πολυλεκτικών σύμπλοκων όρων** στον τομέα της βιοϊατρικής.
2. **Χειρωνακτικός εντοπισμός πολυλεκτικών όρων** σε ένα υποσύνολο ~500.000 λέξεων του Σώματος Ελληνικών Βιοϊατρικών Κειμένων (ΣΕΒΚ) ΙΑΤΡΟΛΕΞΗ. Στόχος αυτής της εργασίας ήταν να εμπλουτιστεί ο κατάλογος των δομών μορφοσυντακτικής συμπεριφοράς των όρων του MeSH. Το σώμα κειμένων επιβεβαίωσε ως υπαρκτό ένα μεγάλο ποσοστό από αυτές τις δομές, ενώ ταυτόχρονα επιβεβαίωσε τα ποσοστά συχνότητας εμφάνισης των διαφορετικών δομών, όπως αυτά είχαν καταγραφεί στο υποσύνολο των πολυλεκτικών όρων του MeSH που μελετήθηκε.
3. **Εξαγωγή πολυλεκτικών όρων από το Σώμα Ελληνικών Βιοϊατρικών Κειμένων (ΣΕΒΚ) ΙΑΤΡΟΛΕΞΗ**: Η εξαγωγή πολυλεκτικών όρων από τα κείμενα στηρίχθηκε στις εξής δύο υποεργασίες:
  - κωδικοποίηση στο Μορφολογικό Λεξικό (ΜΛ) της Neurosoft περίπου 7.250 μονολεκτικών όρων αλλά και λέξεων που αποτελούν πιθανά συστατικά πολυλεκτικών όρων, έτσι όπως προέκυψαν από την επεξεργασία του λεξιλογίου των μονολεκτικών όρων του **MeSH** και των **άγνωστων λέξεων με υψηλή συχνότητα από το ΣΕΒΚ**<sup>1</sup>.
  - περιγραφή κανόνων αναγνώρισης πολυλεκτικών όρων. Οι διαφορετικές δομές με τις οποίες μπορούν να εμφανίζονται οι πολυλεκτικοί όροι βιοϊατρικής διατυπώθηκαν ως κατηγορήματα Prolog και τροφοδότησαν τον αναγνωριστή όρων. Η αναλυτική περιγραφή των κανόνων αναγνώρισης όρων παρουσιάζεται στο παραδοτέο Π6: «*Κανόνες Αναγνώρισης Πολυλεκτικών Όρων Βιοϊατρικής*».

Σε πρώτο στάδιο, για την αυτόματη αναγνώριση των πολυλεκτικών όρων έγινε επεξεργασία των κειμένων, το αποτέλεσμα της οποίας ήταν η επιλογή, η οργάνωση σε καταλόγους και ταξινόμηση των λέξεων ή/και συνδυασμών λέξεων που πληρούν την ελάχιστη συνθήκη της επιτρεπτής γραμματικής κατηγορίας ή δομής στις οποίες

<sup>1</sup> Στην παρούσα φάση, το συνολικό μέγεθος του Σώματος Ελληνικών Βιοϊατρικών Κειμένων (ΣΕΒΚ) του ΙΑΤΡΟΛΕΞΗ ανέρχεται σε 11.500.000 λέξεις.

μπορεί να ανήκει ένας όρος με βάση τους κανόνες αναγνώρισης. Τα δεδομένα αυτά ποσοτικοποιήθηκαν με βάση στατιστικές μετρήσεις (συχνότητα εμφάνισης και/ή συνεμφάνισης).

Σε δεύτερο στάδιο, έγινε η επεξεργασία των καταλόγων που προέκυψαν, η αξιολόγηση των δεδομένων με βάση στατιστικά και γραμματικά κριτήρια και η πρόκριση των λέξεων ή των συνδυασμών λέξεων που παρουσίαζαν μεγάλη πιθανότητα να είναι όροι.

Οι υποψήφιοι πολυλεκτικοί όροι που εξήχθησαν με τις παραπάνω διαδικασίες από το ΣΕΒΚ ανέρχονται περίπου στις 5.000.

### 3. ΔΙΕΡΓΑΣΙΕΣ ΕΠΕΞΕΡΓΑΣΙΑΣ ΤΩΝ ΠΟΛΥΛΕΚΤΙΚΩΝ ΟΡΩΝ

Στους ~5.000 πολυλεκτικούς όρους που καταγράφηκαν από τις παραπάνω διαδικασίες εφαρμόστηκαν οι ακόλουθες μετατροπές και διεργασίες κανονικοποίησης:

1. Έλεγχος – επιμέλεια από τους ιατρούς και τους γλωσσολόγους ώστε να διορθωθούν λάθη στην απόδοση αλλά και στη μορφή των όρων. Για το σκοπό αυτό, χρησιμοποιήθηκαν από τους ειδικούς τόσο έντυπα Αγγλοελληνικά βιοϊατρικά λεξικά (Μιχαηλίδη, Σταφυλίδη, κ.α.) όσο και ηλεκτρονικές πηγές (Λεξικά Magenta και Ergobyte) αλλά και το ίδιο το διαδίκτυο μέσω γνωστών μηχανών αναζήτησης ([www.google.gr](http://www.google.gr), κλπ).
2. Μετατροπή του πληθυντικού αριθμού, με τον οποίο απαντά συχνά η κεφαλή του πολυλεκτικού όρου στο MeSH, σε ενικό αριθμό, π.χ. *αντισώματα ηπατίτιδας Β* → *αντίσωμα ηπατίτιδας Β* κτλ.
3. Αλλαγή της σειράς της κεφαλής και των συμπληρωμάτων που λειτουργούν ως προσδιοριστικά, καθώς επίσης και των παραλλαγών του όρου με τα κόμματα, π.χ. *ινωδογόνα μη φυσιολογικά* → *μη φυσιολογικά ινωδογόνα*, *αδενοϊών πρωίμες πρωτεΐνες* → *πρωίμες πρωτεΐνες αδενοϊών*, *πρωτεΐνες αυγού, διαιτητικές* → *διαιτητικές πρωτεΐνες αυγού*, κτλ. Εκατοντάδες τέτοιων περιπτώσεων παρουσιάστηκαν στον κατάλογο των όρων του MeSH (για προφανείς ορολογικούς και λεξικογραφικούς λόγους, π.χ. ευκολότερη ταξινόμηση βασισμένη στις λέξεις-κεφαλές). Αυτές έπρεπε να αλλάξουν για να μπορεί το συγκεκριμένο υλικό να εφαρμοστεί στα κείμενα του ΣΕΒΚ.
4. Αντίθετα με ό,τι συνέβη με την επεξεργασία των μονολεκτικών όρων, όπου αφαιρέθηκαν από τον κατάλογο όροι που απαντούν με ξένο όνομα, στον κατάλογο των πολυλεκτικών διατηρήθηκαν όροι οι οποίοι περιλαμβάνουν στα συστατικά τους κάποιο ξένο όνομα, π.χ. *σύνδρομο Caplan*, *ρίκνωση Dupuytren*, *όργανο του Corti*, *φαινόμενο Shwartzman* κτλ.

Στους πολυλεκτικούς όρους που προέκυψαν με βάση τις παραπάνω διεργασίες, καθώς και στους όρους που εξήχθησαν από τα κείμενα αποδόθηκε πλήρης μορφολογική πληροφορία (π.χ. γραμματική κατηγορία και κλίση με βάση το μηχανισμό κλίσης πολυλεκτικών όρων, που αναπτύχθηκε στην ΕΕ2).

## 4. ΣΥΣΤΑΣΗ ΤΟΥ ΟΡΟΛΟΓΙΟΥ ΤΩΝ ΠΟΛΥΛΕΚΤΙΚΩΝ ΟΡΩΝ

Στον πίνακα που ακολουθεί παρουσιάζεται μια πρώτη εκτίμηση σχετικά με τη σημασιολογική σύσταση του αρχικού καταλόγου που συγκροτήθηκε από τους ~5.000 όρους.

Σημασιολογική κατηγορία	Παράδειγμα
Rickettsia ή Χλαμύδια	χλαμύδιο τραχώματος
Αμινοξύ, Πεπτιδίο ή Πρωτεΐνη	αμιδάση πενικιλίνης
Ανατομική Ανωμαλία	εντερικός βλεννογόνος
Ανόργανη Χημική Ουσία	οξειδίο αζώτου
Ανοσοποιητικός Παράγοντας	αντίσωμα ηπατίτιδας Β
Ασθένεια ή Σύνδρομο	στοματική καντινίαση
Ασπόνδυλο	μύγα τσε-τσε
Βακτηρίδιο	αερόβια βακτηριαίμια
Βιοϊατρικό Επάγγελμα ή Απασχόληση	χειρουργική στόματος
Βιοϊατρικό ή Οδοντικό υλικό	σύνθετη ρητίνη
Βιολογικά Ενεργή Ουσία	εκχύλισμα πλακούντα
Γενετική Λειτουργία	αναδιπλασιασμός DNA
Γονίδιο ή Γονιδίωμα	γονίδια ABL
Δείκτης, Αντιδραστήριο ή Διαγνωστική ενίσχυση	πίνακες Spin
Διαγνωστική Διαδικασία	γαστρεντερική ενδοσκόπηση
Δραστηριότητα	παθητική ευθανασία
Δραστηριότητα Υγειονομικής Περιθαλψης	ασφάλιση ψυχιατρικής περίθαλψης
Εικοσανοειδές	θρομβοξάνη Β2
Εκ Γενετής Ανωμαλία	καρυότυπος ΧΥΥ
Εκπαιδευτική Δραστηριότητα	αγωγή υγείας
Εμβρυϊκή Δομή	χοριακή λάχνη
Επάγγελμα ή Τομέας Απασχόλησης	αναπτυξιακή βιολογία
Επαγγελματική Δραστηριότητα	διοίκηση ιατρείου
Επαγγελματική Ομάδα	ιατρικοί γραμματείς
Επικίνδυνη ή Δηλητηριώδης Ουσία	διοξειδίο αζώτου
Επίκτητη Ανωμαλία	κίρσοι οισοφάγου
Εργαστηριακή Διαδικασία	άζωτο ουρίας αίματος
Εργαστηριακό ή Πειραματικό Αποτέλεσμα	λόγος CD4-CD8
Ερευνητική Δραστηριότητα	ανάστροφη εκμάθηση
Εύρημα	αιμορραγία μήτρας
Ηλικιακή Ομάδα	πρόωρο βρέφος
Θεραπευτική ή Προληπτική διαδικασία	κατ' οίκον θεραπεία
Θέση ή Περιοχή Σώματος	βυθός οφθαλμού
Θηλαστικό	ποντικοί Wistar
Ιατρική Συσκευή	νάρθηκας σύγκλεισης
Ιδέα ή Έννοια	αιτία θανάτου
Ιδιότητες Οργανισμών	θερμοκρασία σώματος

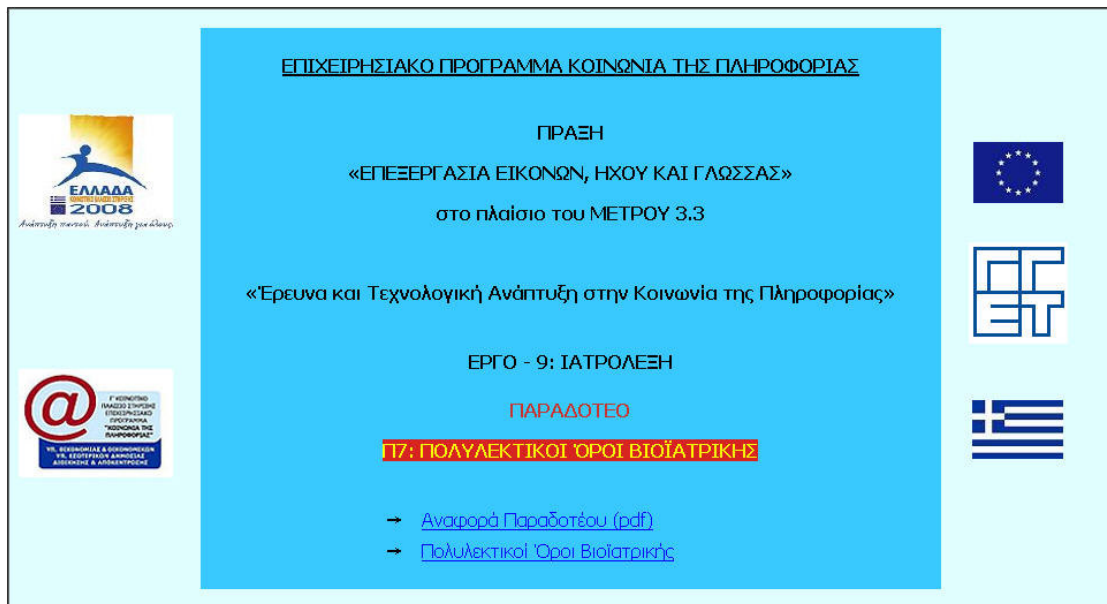
Ιός	ιός ηπατίτιδας Ε
Ιστός	λεμφικός ιστός
Κατασκευασμένο Αντικείμενο	αίθουσα τοκετού
Κοινωνική Συμπεριφορά	ανταγωνιστική συμπεριφορά
Κυβερνητική ή Κανονιστική δραστηριότητα	γονική άδεια
Κυτταρική ή Μοριακή Δυσλειτουργία	βαλεριανή εκφύλιση
Κύτταρο	λεμφοκύτταρα Β
Λειτουργία Κυττάρων	κυτταρική διαίρεση
Λειτουργία Οργάνων ή Ιστού	μυϊκή συστολή
Λειτουργική Χημική Ουσία	υποχλωριώδες νάτριο
Λιπίδιο	λινολεϊκό οξύ
Μεμονωμένη Συμπεριφορά	αντιδραστική αναστολή
Μέρος Σώματος, Όργανο, ή τμήμα οργάνων	όργανο του Corti
Μοριακή Λειτουργία	επαναλαμβανόμενες αλληλουχίες αγκυρίνης
Μύκητας	κρυπτόκοκκος Neoformans
Νεοπλασματική Διεργασία	καρκίνωμα in situ
Νουκλεϊνικό Οξύ, Νουκλεοζίτης ή Νουκλεοτίδα	πρόδρομοι του RNA
Οικογενειακή Ομάδα	διζυγωτικά δίδυμα
Ομάδα Πληθυσμού	αγροτικός πληθυσμός
Οργανική Λειτουργία	εγκυμοσύνη στην εφηβεία
Οργανική Χημική Ουσία	αρσανιλικό οξύ
Οργανοφωσφορική Ένωση	μονοφωσφορική θειαμίνη
Οργάνωση Ιατρικής Περίθαλψης	τράπεζα αίματος
Ορμόνη	έκτοπες ορμόνες
Παθολογική Λειτουργία	συρίγγιο οισοφάγου
Πειραματικό Μοντέλο Ασθένειας	πειραματικός διαβήτης σακχαρώδης
Περιβαλλοντικές Συνέπειες σε Ανθρώπους	ρύπανση νερού
Πνευματική Διεργασία	ανάπτυξη της προσωπικότητας
Πνευματική ή Συμπεριφοριστική Δυσλειτουργία	διαταραχή μάθησης
Πνευματικό Προϊόν	ψυχαναλυτική θεωρία
Ποιοτική Έννοια	αισθητική οδοντιατρική
Ποσοτική Έννοια	άθροισμα των LOD
Σημάδι ή Σύμπτωμα	μυϊκή δυσκαμψία
Στεροειδές	βιταμίνη D
Στοιχείο, Ιόν, ή Ισότοπο	ισότοπο αζώτου
Σύστημα Σώματος	παρασυμπαθητικό νευρικό σύστημα
Σωματική Ουσία	κηλίδα αίματος
Σωματικός Χώρος ή Σύνδεση	άρθρωση καρπού
Τμήμα Κυττάρων	χρωμόσωμα Υ
Τραυματισμός ή Δηλητηρίαση	κάκωση νωτιαίου μυελού
Υδατάνθρακας	υγρομυκίνη Β
Φαινόμενο ή Διεργασία	μητρική αποστέρωση
Φαινόμενο ή Διεργασία προκαλούμενη από άνθρωπο	αισθητηριακή αποστέρωση



Φαρμακολογική Ουσία	εμβόλιο BCG
Φυσικό Φαινόμενο ή Διεργασία	ιοντίζουσα ακτινοβολία
Φυσιολογική Λειτουργία	ανάπτυξη εμβρύου
Φυτό	τοξικό φυτό
Χημική Ουσία που αντιμετωπίζεται δομικά	ενώσεις βορίου
Χρονική Έννοια	δεύτερο στάδιο τοκετού
Χωρική Έννοια	διάταξη πρωτεϊνών

## 5. ΟΔΗΓΙΕΣ ΧΡΗΣΗΣ ΤΟΥ ΟΠΤΙΚΟΥ ΔΙΣΚΟΥ

Στο παρόν παραδοτέο περιλαμβάνεται οπτικός δίσκος (CD) ο οποίος περιέχει τους μονολεκτικούς όρους βιοϊατρικής, καθώς και την παρούσα τεκμηρίωση. Μόλις εισαχθεί ο οπτικός δίσκος στη μονάδα οπτικού δίσκου του υπολογιστή θα ανοίξει αυτόματα ο φυλλομετρητής με την παρουσίαση του παραδοτέου (εικόνα 1).



Εικόνα 1: Στιγμιότυπο παρουσίασης του παραδοτέου.

Στην πρώτη σελίδα παρουσιάζονται στοιχεία του έργου και του συγκεκριμένου παραδοτέου καθώς και δύο υπερσύνδεσμοι (links). Ο πρώτος οδηγεί στην αναφορά του παραδοτέου (παρούσα αναφορά) σε μορφή PDF και ο δεύτερος στα περιεχόμενα της λίστας των πολυλεκτικών βιοϊατρικών όρων.