

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

Π5: ΜΟΝΟΛΕΚΤΙΚΟΙ ΟΡΟΙ ΒΙΟΪΑΤΡΙΚΗΣ

(στο παραδοτέο περιλαμβάνεται οπτικός δίσκος που περιέχει τους όρους)

Ημερομηνία:	27.02.2007
Έκδοση:	Final
Τύπος:	Εμπιστευτικό

ΠΕΡΙΕΧΟΜΕΝΑ

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

1. ΕΙΣΑΓΩΓΗ

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 για την εξαγωγή της ορολογίας που θα αποτελέσει τη βάση πάνω στην οποία θα χτιστεί η οντολογία του ΙΑΤΡΟΛΕΞΗ. Έτσι, 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 λεξιλόγιό της.

Το παρόν παραδοτέο **Π5 «Μονολεκτικοί Όροι Βιοϊατρικής»** περιλαμβάνει:

1. αρχείο δεδομένων (στον επισυναπτόμενο οπτικό δίσκο) με το σύνολο των μονολεκτικών όρων που έχουν συγκεντρωθεί,
2. την παρούσα σύντομη τεκμηρίωση για το περιεχόμενο των όρων του αρχείου δεδομένων.

Πιο συγκεκριμένα, στην τεκμηρίωση που ακολουθεί περιγράφονται:

- οι τρόποι και οι πηγές συγκέντρωσης των όρων,
- οι διεργασίες επεξεργασίας των όρων,
- στοιχεία σχετικά με τη σύσταση της συλλογής των όρων.
- οδηγίες χρήστης του οπτικού δίσκου που περιέχει το αρχείο των όρων.

2. ΣΥΓΚΕΝΤΡΩΣΗ ΜΟΝΟΛΕΚΤΙΚΩΝ ΟΡΩΝ - ΠΗΓΕΣ

Για τη συγκέντρωση των μονολεκτικών όρων με βάση τις απαιτήσεις του έργου ΙΑΤΡΟΛΕΞΗ, ακολουθήθηκαν οι ακόλουθες δύο αλληλοσυμπληρούμενες διαδικασίες:

Επεξεργασία δίγλωσσων λεξιλογίων βιοϊατρικών όρων (αγγλικών - ελληνικών): Στόχος αυτής της εργασίας ήταν να εντοπιστεί το βασικό λεξιλόγιο της βιοϊατρικής ορολογίας που περιλαμβάνεται σε έντυπα και ηλεκτρονικά λεξικά της ελληνικής, και το οποίο με την κατάλληλη επεξεργασία θα μπορούσε να αποτελέσει τη βάση της οντολογίας ΙΑΤΡΟΛΕΞΗ. Πιο συγκεκριμένα, για τον εντοπισμό του βασικού λεξιλογίου έγινε **επεξεργασία** περίπου 7.300 μονολεκτικών όρων που περιλαμβάνονται στο **ελληνικό MeSH** (έκδοση 1996) και συνεχίζουν να αποτελούν όρους της τρέχουσας έκδοσης του αμερικανικού **MeSH®** (έκδοση 2006). Για την επεξεργασία των όρων χρησιμοποιήθηκαν από τους ειδικούς (ιατρούς γλωσσολόγους) 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, κλπ).

Εξαγωγή όρων από ελληνικά βιοϊατρικά κείμενα: Παράλληλα με την επεξεργασία του λεξιλογίου του MeSH έγινε εντοπισμός και εξαγωγή των άγνωστων λέξεων από το Σώμα Ελληνικών Βιοϊατρικών Κειμένων (ΣΕΒΚ) του ΙΑΤΡΟΛΕΞΗ, το συνολικό μέγεθος της τρέχουσας μορφής του οποίου ανέρχεται σε περίπου 11.500.000 λέξεις. Ως άγνωστες λέξεις θεωρήθηκαν όσοι λεκτικοί τύποι δεν περιείχονταν στο Μορφολογικό Λεξικό (ΜΛ)¹.

Στο ΣΕΒΚ περιέχονται συνολικά 262.000 μοναδικό μονολεκτικοί τύποι, ενώ στο MeSH 16.500. Οι μονολεκτικοί τύποι που είναι κοινό στο ΣΕΒΚ και στο MeSH είναι 9.500. Από τους αυτούς, οι 1.163 ήταν άγνωστοι στο ΜΛ.

Οι άγνωστοι μονολεκτικοί τύποι του MeSH είναι 6.173 (επί συνόλου 16.500 μοναδικών μονολεκτικών τύπων), ενώ οι άγνωστοι μονολεκτικοί τύποι του ΣΕΒΚ ήταν 31.400 (επί συνόλου 262.000 μοναδικών μονολεκτικών τύπων). Πιο αναλυτικά, οι άγνωστοι τύποι του ΣΕΒΚ, με συχνότητα > 1 είναι 12.700 (εκ των οποίων με συχνότητα > 2 είναι 8.641 και με συχνότητα > 3 είναι 5.264) ενώ με συχνότητα 1 είναι 18.700. Αξίζει να σημειώσουμε ότι οι μονολεκτικοί τύποι του ΣΕΒΚ με συχνότητα 1 (δηλ. 18.700 μοναδικό μονολεκτικοί τύποι) δεν είναι όλοι χρήσιμοι και αξιοποιήσιμοι αφού μεταξύ των άλλων περιέχουν και αρκετά λάθη: λέξεις κολλημένες μεταξύ τους, λέξεις που τους λείπουν γράμματα, λέξεις κολλημένες με αριθμούς και άλλα σύμβολα κ.ο.κ. Συνεπώς, κρίθηκε καλό να μην τους συμπεριλάβουμε στην πρώτη φάση επεξεργασίας του λεξιλογίου, αλλά αργότερα.

Οι άγνωστες λέξεις που εντοπίστηκαν στο ΣΕΒΚ ανήκουν σε τρεις κατηγορίες:

1. όροι,

¹ Δεδομένου ότι το ΜΛ αναγνωρίζει περισσότερους από 1.500.000 κλητικούς τύπους ελληνικών λέξεων του γενικού λεξιλογίου, υπήρχε μεγάλη πιθανότητα οι άγνωστες λέξεις που εντοπίστηκαν σε κείμενα έντασης βιοϊατρικών όρων να είναι μονολεκτικοί ή μέρη πολυλεκτικών όρων.

2. κύρια ονόματα και

3. εσφαλμένοι τύποι (κομμένες λέξεις, διάφορα λάθη κτλ.).

Σε πρώτη φάση, απομακρύνθηκαν με αυτόματο τρόπο από τον κατάλογο των άγνωστων λέξεων όσες είχαν αρχικό γράμμα κεφαλαίο (τοπωνύμια, ονόματα προσώπων, ιδρυμάτων, νοσοκομείων κτλ.). Σε δεύτερη φάση, το υλικό εξετάστηκε προσεκτικά από τους ειδικούς (ιατρούς και γλωσσολόγους) έτσι ώστε να απομακρυνθούν με χειρωνακτικό τρόπο και οι εσφαλμένοι τύποι.

Το ορολογικό υλικό που αντλήθηκε από τις παραπάνω πηγές εμπλουτίστηκε με μορφολογική πληροφορία και ενσωματώθηκε στο Μορφολογικό Λεξικό (ΜΛ) της Neurosoft Α.Ε.

Ακολούθησε μορφοσυντακτικός σχολιασμός του Σώματος Κειμένων του έργου με δύο στόχους:

1. τον **εντοπισμό των άγνωστων λέξεων**, δηλαδή των λέξεων των οποίων η μορφή δεν είναι αναγνωρίσιμη από το ΜΛ, και την **ενσωμάτωσή τους στο ΜΛ** μετά από σχετική επεξεργασία και

2. τον **αποκλεισμό** από το ορολόγιο του ΙΑΤΡΟΛΕΞΗ όρων που εντοπίστηκαν στο MeSH αλλά δεν εμφανίστηκαν καθόλου στα κείμενα βιοϊατρικής του ΙΑΤΡΟΛΕΞΗ και του Google.

3. ΔΙΕΡΓΑΣΙΕΣ ΕΠΕΞΕΡΓΑΣΙΑΣ ΤΩΝ ΟΡΩΝ

Μια σειρά από μετατροπές και διεργασίες κανονικοποίησης εφαρμόστηκαν στους μονολεκτικούς όρους που αντλήθηκαν με τους τρόπους που αναφέρθηκαν στα προηγούμενα, όπως περιγράφεται στη συνέχεια:

1. Έλεγχος – επιμέλεια από τους ιατρούς και τους γλωσσολόγους ώστε να διορθωθούν λάθη στην απόδοση αλλά και στη μορφή των όρων.
2. Μετατροπή του πληθυντικού αριθμού, με τον οποίο απαντά συχνά ο μονολεκτικός όρος π.χ. στο MeSH, σε ενικό αριθμό, π.χ. *τοξίνες* → *τοξίνη*, *τριχοστοματίδες* → *τριχοστοματίδα*, *μυξομύκητες* → *μυξομύκητα* κτλ.
3. Αφαίρεση των όρων που απαντούν στο MeSH με ξένο όνομα, το οποίο δεν έχει αποδοθεί στα ελληνικά, π.χ. *Coxiella*, *biomphalaria*, *bulinus*. Ως ξένοι θεωρήθηκαν και οι όροι λατινικής προέλευσης. Ωστόσο, στην κατηγορία αυτή δεν περιλήφθηκαν και παρέμειναν στον κατάλογο τα αρκτικόλεξα και τα ακρώνυμα, π.χ. DNA, RNA.
4. Διατήρηση του κεφαλαίου αρχικού γράμματος μόνο στις περιπτώσεις των κυρίων ονομάτων, π.χ. **Χ**ιλή.

Στους μονολεκτικούς όρους που προέκυψαν από τις παραπάνω διεργασίες, καθώς και στους όρους που εξήχθησαν από τα κείμενα αποδόθηκε πλήρης μορφολογική πληροφορία (π.χ. γραμματική κατηγορία και κλίση).

4. ΣΥΣΤΑΣΗ ΤΟΥ ΟΡΟΛΟΓΙΟΥ ΤΩΝ ΜΟΝΟΛΕΚΤΙΚΩΝ ΟΡΩΝ

Οι διαδικασίες της συγκέντρωσης όρων από λεξιλόγια και λεξικά και κείμενα βιοϊατρικής και της επεξεργασίας τους από ειδικούς γιατρούς και γλωσσολόγους οδήγησαν στην κατάρτιση ενός αρχικού καταλόγου που περιλαμβάνει **7.242** μονολεκτικούς όρους. Από αυτούς 4.576 είναι ουσιαστικά, 2.542 είναι επίθετα, 55 είναι ρήματα, 62 είναι μετοχές και 7 είναι επιρρήματα.

Στον πίνακα που ακολουθεί παρουσιάζεται μια πρώτη περιγραφή σχετικά με το πώς κατανέμονται με βάση την αρχική ταξινόμια του ΙΑΤΡΟΛΕΞΗ οι όροι του αρχικού ορολογίου².

Σηματολογική κατηγορία	Παράδειγμα
Rickettsia ή Χλαμύδια	ερλίχια
Αμινοξύ, Πεπτιδίο ή Πρωτεΐνη	ωκυτοκίνη
Αμφίβιο	τρίτωνας
Ανατομική Ανωμαλία	σκολίωση
Ανατομική Δομή	φρύδι
Ανόργανη Χημική Ουσία	ηλεκτρολύτες
Ανοσοποιητικός Παράγοντας	φυματίνη
Αντιβιοτικό	μονοβακτάμη
Ασθένεια ή σύνδρομο	ωτοσκλήρυνση
Ασπόνδυλο	αράχνη
Βακτηρίδιο	ακτινομύκητας
Βιοϊατρικό Επάγγελμα ή Απασχόληση	ωτορινολαρυγγολογία
Βιοϊατρικό ή Οδοντικό Υλικό	υπόθετο
Βιολογικά Ενεργή Ουσία	ψευδάργυρος
Γενετική Λειτουργία	μετάλλαξη
Γεωγραφική Περιοχή	Χιλή
Γονίδιο ή Γονιδίωμα	ψευδογονίδιο
Δείκτης, Αντιδραστήριο ή Διαγνωστική Ενίσχυση	υδρόθειο
Διαγνωστική Διαδικασία	αγγειογραφία
Δραστηριότητα	ταφή
Δραστηριότητα Μηχανών	ψύξη

² Το συνημμένο αρχείο των μονολεκτικών όρων δεν περιλαμβάνει τις λέξεις με τους εντονότυπους χαρακτήρες στον πίνακα, οι οποίες δηλώνουν:

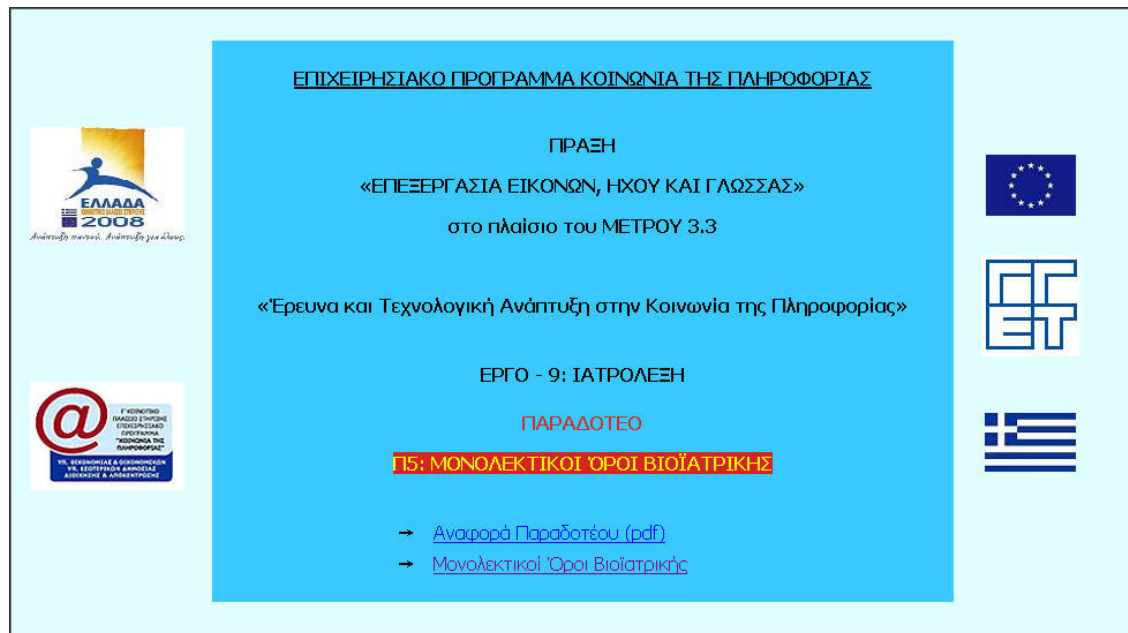
- διάφορες ανθρώπινες δραστηριότητες (π.χ. υποτροφία, πιστοποίηση, ταφή, βασανισμός κτλ.),
- ποιοτικά χαρακτηριστικά (π.χ. ωχρότητα, σκληρότητα κτλ.),
- ομάδες ανθρώπων (π.χ. φαρμακοποιοί, φυλακισμένοι κτλ.),
- μη επιστημονικές ονομασίες μερών του σώματος (π.χ. φρύδι, ώμος κτλ.),
- κοινές ονομασίες ειδών ζώων (π.χ. χρυσόψαρο, χήνα, χελώνα, αράχνη κτλ.),
- κοινές ονομασίες ειδών φυτών (π.χ. ευκάλυπτος),
- τοπωνύμια (π.χ. Χιλή),
- διάφορες επαγγελματικές ειδικότητες (π.χ. ωκεανογραφία, στατιστική κτλ.)
- λέξεις του κοινού λεξιλογίου (π.χ. ψωμί, τροφή, χειρονομία κτλ.).

Δραστηριότητα Υγειονομικής Περιθαλψης	τηλεϊατρική
Εικοσανοειδές	ριοππροσίλη
Εκ Γενετής Ανωμαλία	ψευδοερμαφροδιτισμός
Εκπαιδευτική Δραστηριότητα	υποτροφία
Εμβρυϊκή Δομή	γαστρίδιο
Επάγγελμα ή Τομέας Απασχόλησης	ωκεανογραφία
Επαγγελματική Δραστηριότητα	μικροφωτογράφιση
Επαγγελματική Ομάδα	φαρμακοποιοί
Επικίνδυνη ή Δηλητηριώδης Ουσία	χλώριο
Επίκτητη Ανωμαλία	κίρσοι
Εργαστηριακή Διαδικασία	χρωματομετρία
Εργαστηριακό ή Πειραματικό Αποτέλεσμα	υπεργαμασφαιριναιμία
Ερευνητική Δραστηριότητα	στατιστική
Ερπετό	χελώνα
Εύρημα	φλεγμονή
Ζώο	χορδωτά
Ηλικιακή Ομάδα	νεογνά
Θεραπευτική ή Προληπτική Διαδικασία	ωθηκεκτομή
Θέση ή Περιοχή Σώματος	ώμος
Θηλαστικό	άλογο
Ιατρική Συσκευή	νάρθηκας
Ιδιότητες Οργανισμών	γονότυπος
Ιός	ψευδομοναδοφάγος
Ιστός	χόνδρος
Καθημερινή ή Ψυχαγωγική Δραστηριότητα	αθλοπαιδιές
Κατασκευασμένο Αντικείμενο	χειρουργείο
Κοινωνική Συμπεριφορά	χειρονομία
Κυβερνητική ή Ρυθμιστική Δραστηριότητα	πιστοποίηση
Κυτταρική ή Μοριακή Δυσλειτουργία	αιμόλυση
Κύτταρο	ωκύτταρο
Λειτουργία Κυττάρων	χημειοταξία
Λειτουργία Οργάνων ή Ιστού	ωθυλακιορρηξία
Λειτουργική Χημική Ουσία με Λειτουργικά Χαρακτηριστικά	οξειδωτικό
Λιπίδιο	χλωροβουτανόλη
Μεμονωμένη Συμπεριφορά	δακτυλολειχία
Μέρος Σώματος, Όργανο, ή Τμήμα Οργάνων	ωθήκη
Μοριακή Λειτουργία	φωτοφωσφορυλίωση
Μύκητας	ασπέργιλλος
Νεοπλασματική Διεργασία	χοριοκαρκίνωμα
Νουκλεϊνικό οξύ, Νουκλεοζίτης ή Νουκλεοτίδα	ψευδοουριδίνη
Οικογενειακή Ομάδα	τετράδυμα
Ομάδα Πληθυσμού	φυλακισμένοι
Οργανική Λειτουργία	αφόδευση
Οργανική Χημική Ουσία	ωχρατοξίνη
Οργανισμός	ομοζυγώτης

Οργανοφωσφορική Ένωση	φωσφορουλοχολίνη
Ουσία	κοπριά
Παθολογική Λειτουργία	ψευδοκύηση
Περιβαλλοντικές Συνέπειες σε Ανθρώπους	φθορίωση
Πνευματική Διεργασία	κιναίσθηση
Πνευματική ή Συμπεριφοριστική Δυσλειτουργία	ψευδαίσθηση
Πνευματικό Προϊόν	συνταγολόγιο
Ποιοτική Έννοια	σκληρότητα
Ποσοτική Έννοια	επιπολασμός
Πτηνό	χήνα
Σημάδι ή Σύμπτωμα	ωχρότητα
Στεροειδές	χοληστερόλη
Στοιχείο, Ιόν ή Ισότοπο	άζωτο
Σύστημα Σώματος	σκελετός
Σωματική Ουσία	ιδρώτας
Σωματικός Χώρος ή Σύνδεση	σύναψη
Τμήμα Κυττάρων	ριβόσωμα
Τραυματισμός ή Δηλητηρίαση	έγκαυμα
Τροφή	ψωμί
Υδατάνθρακας	ψυχοσίνη
Φαινόμενο ή Διεργασία	υποτροπή
Φαινόμενο ή Διεργασία Προκαλούμενη από Άνθρωπο	βασανισμός
Φαρμακολογική Ουσία	αναισθητικό
Φύκι	πλαγκτόν
Φυσικό Φαινόμενο ή Διεργασία	απαμίνωση
Φυσιολογική Λειτουργία	φωτοσύνθεση
Φυτό	ευκάλυπτος
Χημική Ουσία	οξύ
Χημική Ουσία που Αντιμετωπίζεται Δομικά	άλας
Χρονική Έννοια	τελοφάση
Ψάρι	χρυσόψαρο

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 βιοϊατρικών όρων.