

## 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

#### Π22: Τελική Αναφορά Έργου

Ημερομηνία:	20.12.2007
Έκδοση:	Final
Τύπος:	Εμπιστευτικό

## ΠΕΡΙΕΧΟΜΕΝΑ

<b>1</b>	<b>ΕΙΣΑΓΩΓΗ</b>	<b>3</b>
<b>2</b>	<b>ΥΛΟΠΟΙΗΣΗ ΕΡΓΟΥ</b>	<b>4</b>
<b>2.1</b>	<b>ΠΕΡΙΛΗΨΗ ΠΡΑΓΜΑΤΟΠΟΙΗΘΕΙΣΩΝ ΕΡΓΑΣΙΩΝ</b>	<b>4</b>
<b>2.2</b>	<b>ΑΝΑΛΥΤΙΚΗ ΠΕΡΙΓΡΑΦΗ ΤΩΝ ΕΝΟΤΗΤΩΝ ΕΡΓΑΣΙΩΝ (Ε.Ε.) ΠΟΥ ΥΛΟΠΟΙΗΘΗΚΑΝ</b>	<b>7</b>
2.2.1	ΕΕ1: ΣΧΕΔΙΑΣΜΟΣ ΣΩΜΑΤΟΣ ΒΙΟΪΑΤΡΙΚΩΝ ΚΕΙΜΕΝΩΝ ΚΑΙ ΟΝΤΟΛΟΓΙΑΣ	7
2.2.2	ΕΕ2: ΥΛΟΠΟΙΗΣΗ ΕΡΓΑΛΕΙΩΝ ΔΙΑΧΕΙΡΙΣΗΣ ΣΩΜΑΤΟΣ ΚΕΙΜΕΝΟΥ ΚΑΙ ΑΥΤΟΜΑΤΗΣ ΕΞΑΓΩΓΗΣ ΟΡΟΛΟΓΙΑΣ	12
2.2.3	ΕΕ3: ΣΥΛΛΟΓΗ ΒΙΟΪΑΤΡΙΚΩΝ ΚΕΙΜΕΝΩΝ ΚΑΙ ΕΞΑΓΩΓΗ ΟΡΟΛΟΓΙΑΣ	14
2.2.4	ΕΕ4: ΕΠΕΞΕΡΓΑΣΙΑ ΒΙΟΪΑΤΡΙΚΩΝ ΌΡΩΝ ΚΑΙ ΔΗΜΙΟΥΡΓΙΑ ΟΝΤΟΛΟΓΙΑΣ	20
2.2.5	ΕΕ5: ΥΛΟΠΟΙΗΣΗ ΕΦΑΡΜΟΓΩΝ ΙΣΤΟΥ	24
2.2.6	ΕΕ6: ΔΙΑΧΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ	28
<b>3</b>	<b>ΣΥΝΟΛΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΠΑΡΑΔΟΤΕΑ ΤΟΥ ΕΡΓΟΥ</b>	<b>31</b>
<b>4</b>	<b>ΣΧΟΛΙΑ - ΠΡΟΒΛΗΜΑΤΑ - ΠΑΡΑΤΗΡΗΣΕΙΣ</b>	<b>35</b>
4.1	ΤΕΧΝΟΛΟΓΙΑ / ΤΕΧΝΟΓΝΩΣΙΑ ΠΟΥ ΑΠΟΚΤΗΘΗΚΕ ΣΤΟ ΠΛΑΙΣΙΟ ΤΟΥ ΕΡΓΟΥ	35
4.2	ΣΥΝΕΡΓΑΣΙΑ ΦΟΡΕΩΝ (ΟΙΚΟΝΟΜΙΚΟ ΚΑΙ ΦΥΣΙΚΟ ΑΝΤΙΚΕΙΜΕΝΟ)	35
<b>5</b>	<b>ΔΙΑΧΕΙΡΙΣΗ ΕΡΓΟΥ</b>	<b>37</b>
5.1	ΣΥΝΑΝΤΗΣΕΙΣ	37
5.1	ΠΑΡΑΔΟΤΕΑ	37
<b>6</b>	<b>ΠΑΡΑΡΤΗΜΑΤΑ</b>	<b>38</b>
6.1	ΠΡΑΚΤΙΚΑ ΣΥΝΑΝΤΗΣΕΩΝ	38

## 1 ΕΙΣΑΓΩΓΗ

Η παρούσα αναφορά αντιστοιχεί στο παραδοτέο Π22: «**Τελική Αναφορά Έργου**» του έργου ΙΑΤΡΟΛΕΞΗ. Συνοψίζει ουσιαστικά και εν συντομία, τα όσα έγιναν κατά τη διάρκεια της υλοποίησης του έργου.

Αρχικά περιγράφεται η υλοποίηση του έργου ανά ενότητα εργασίας.

Στη συνέχεια καταγράφονται τα συνολικά αποτελέσματα και τα παραδοτέα του έργου.

Κατόπιν καταγράφονται τα βασικά προβλήματα που ενέσκηψαν και παρατηρήσεις σχετικά με αυτά.

Τέλος, περιγράφεται η διαχείριση του έργου με έμφαση στις συναντήσεις που έγιναν και στα πρακτικά που τηρήθηκαν σε αυτές.

## 2 ΥΛΟΠΟΙΗΣΗ ΕΡΓΟΥ

### 2.1 ΠΕΡΙΛΗΨΗ ΠΡΑΓΜΑΤΟΠΟΙΗΘΕΙΣΩΝ ΕΡΓΑΣΙΩΝ

Το έργο ολοκληρώθηκε σε 20 μήνες από την έναρξή του, εξελισσόμενο σε ικανοποιητικό βαθμό με βάση το χρονοδιάγραμμα.

Πιο αναλυτικά και αναφορικά με τις ενότητες εργασίας που είχαν προδιαγραφεί, έχουμε τα ακόλουθα.

#### ΕΕ1: Σχεδιασμός Σώματος Βιοϊατρικών Κειμένων και Οντολογίας

Σε αυτή την ενότητα εργασίας σχεδιάστηκε ουσιαστικά και σε λεπτομέρεια το συνολικό έργο. Καθορίστηκε και περιγράφηκε το απαιτούμενο σώμα κειμένων και σχεδιάστηκε η μεθοδολογία για την ανάπτυξη της οντολογίας. Τα αποτελέσματα της ενότητας παρουσιάζονται στα δύο παραδοτέα της:

1. Σχέδιο Συγκρότησης Σώματος Βιοϊατρικών Κειμένων (τεχνική αναφορά)
2. Μοντέλο Αναπαράστασης Οντολογίας (τεχνική αναφορά)

#### ΕΕ2: Υλοποίηση Εργαλείων Διαχείρισης Σώματος Κειμένου και Αυτόματης Εξαγωγής Ορολογίας

Σε αυτή την ενότητα αναπτύχθηκαν/προσαρμόστηκαν/εμπλουτίστηκαν τα απαραίτητα εργαλεία για την διαχείριση του σώματος κειμένων και την ανάπτυξη της οντολογίας. Αναλυτικά, αναπτύχθηκαν: α) Αποθήκη Εγγράφων: Πρόκειται για το χώρο αποθήκευσης των κειμένων αλλά και των παραγόμενων από αυτά μεταδεδομένων. β) Προσκομιστής Εγγράφων (Crawler): Διατρέχει συγκεκριμένους δικτυακούς τόπους ή τοπικούς δίσκους και προσκομίζει τα κείμενα προς επεξεργασία. γ) Μετατροπέας Κειμένων (από pdf και html σε txt). Ενώ δεν είχε προδιαγραφεί αρχικά, κρίθηκε απαραίτητη η υλοποίηση του εργαλείου αυτού. δ) Αναγνωριστής Στοιχείων (Tokenizer): Τεμαχίζει ένα κείμενο σε μία σειρά στοιχείων. ε) Μορφοσυντακτικός Σχολιαστής (Morphosyntactic Tagger): Επισυνάπτει μορφοσυντακτικά μεταδεδομένα σε κάθε λέξη του κειμένου. στ) Μορφολογικό Λεξικό. Εμπλουτίστηκε το μορφολογικό λεξικό της Neurosoft με βιοϊατρικούς όρους. ζ) Μηχανισμός Κλίσης Πολυλεκτικών Όρων. Εφαρμογή για την υποστήριξη της κλίσης των όρων αυτών. η) Αναγνωριστής Όρων: Συμβουλευεται το Μορφολογικό Λεξικό, και προσθέτει μεταδεδομένα σε σχέση με τους όρους ενός κειμένου.

Όλα τα παραπάνω ολοκληρώθηκαν κάτω από ένα κοινό “front end” σύστημα. Τέλος, θα πρέπει να σημειωθεί ότι εκτός των παραπάνω, αναπτύχθηκε και ένα σύστημα Συλλογής Λεξιλογικών Συνάψεων (concordancer) που υποστηρίζει τους γλωσσολόγους και τους ιατρούς στην επιμέλεια των όρων. Τα αποτελέσματα αυτά περιγράφηκαν στα δύο παραδοτέα:

3. Σύστημα Διαχείρισης Σώματος Κειμένων και Εξαγωγής Ορολογίας (λογισμικό)

#### 4. Εγχειρίδιο Χρήσης

##### **ΕΕ3: Συλλογή Βιοϊατρικών κειμένων και Εξαγωγή Ορολογίας**

Σε αυτή την ενότητα εργασίας συλλέχθηκε από τον κυβερνοχώρο, το απαραίτητο σώμα βιοϊατρικών κειμένων. Το σώμα αυτό περιέχει 6.276 κείμενα και περίπου 11,5 εκ. λέξεις. Από το σώμα αυτό, εξήχθησαν οι ιατρικοί όροι που μαζί και με τους όρους που συλλέχθηκαν από ιατρικά λεξικά, έφτασαν τους περίπου 9.300. Τους όρους αυτούς επιμελήθηκαν οι ιατροί και οι γλωσσολόγοι. Επιπλέον συνελέγησαν και οι πολυλεκτικοί όροι, που ανέρχονται περίπου σε 5.000. Όλα τα παραπάνω περιγράφονται στα παραδοτέα:

5. Μονολεκτικοί Όροι Βιοϊατρικής (αρχείο δεδομένων)
6. Κανόνες Αναγνώρισης Πολυλεκτικών Όρων Βιοϊατρικής (αρχείο δεδομένων + αναφορά τεκμηρίωσης)
7. Πολυλεκτικοί Όροι Βιοϊατρικής (αρχείο δεδομένων)
8. Αρχική Ταξινόμια Βιοϊατρικών Όρων (αρχείο δεδομένων)
9. Συλλογή βιοϊατρικών κειμένων (αρχείο)
10. Τεκμηρίωση συλλογής βιοϊατρικών κειμένων (αναφορά)

##### **ΕΕ4: Επεξεργασία Βιοϊατρικών Όρων και Δημιουργία Οντολογίας**

Μετά τον ορισμό του Μοντέλου Αναπαράστασης Οντολογίας (ΕΕ1, Παραδοτέο 2), η ΕΕ4 ξεκίνησε υλοποιώντας το μοντέλο αυτό με προσαρμογή/επέκταση του περιβάλλοντος ανάπτυξης οντολογιών Protégé. Στη συνέχεια, η Αρχική Ταξινόμια Βιοϊατρικών Όρων (ΕΕ3, Παραδοτέο 4) εισάχθηκε στο Protégé.

Η ομάδα εξειδικευμένων γιατρών και έμπειρων γλωσσολόγων εξέτασε:

- τους βιοϊατρικούς όρους που συλλέχθηκαν από τα κείμενα,
- το περιβάλλον (context) που βρέθηκαν και κυρίως το περιβάλλον που συνεμφανίστηκαν οι όροι.
- αν οι υποψήφιοι όροι είναι παραλλαγές όρων της οντολογίας ή νέοι όροι,
- αν οι έννοιες που ορίστηκαν στην οντολογία είναι επαρκείς για τον εισαγωγή ενός όρου στην οντολογία ή πρέπει να επεκταθεί η Οντολογία με νέες έννοιες,
- αν η συνεμφάνιση όρων σηματοδοτεί κάποια σχέση που πρέπει να εισαχθεί στην οντολογία.

Επιπλέον στην οντολογία εισάχθηκε επιπλέον πληροφορία σχετικά με τις συσχετίσεις των όρων καθώς και η αγγλική και ελληνική ερμηνεία (για υποσύνολο των όρων).

Τα παραπάνω περιγράφονται στα παραδοτέα:

11. Προσαρμοσμένο Protégé (λογισμικό + αναφορά προσαρμογών)
12. Αρχική Ταξινόμια εντός Protégé (αρχείο δεδομένων)
13. Οντολογία Βιοϊατρικών Όρων (αρχείο δεδομένων)

### ΕΕ5: Υλοποίηση Εφαρμογών Ιστού

Στην ενότητα εργασίας αυτή, αναπτύχθηκαν οι εφαρμογές ιστού που φιλοξενούνται στον ιστοχώρο του έργου (βλ. ΕΕ6). Πιο συγκεκριμένα αναπτύχθηκαν:

- Περιηγητής Οντολογίας, ο οποίος επιτρέπει στο χρήστη την περιήγηση στους όρους της οντολογίας.
- Μηχανή Αναζήτησης στο σώμα βιοϊατρικών κειμένων, τα οποία έχουν ευρετηριαστεί επί τη βάση σημασιολογικών κατηγοριών από την οντολογία.
- Ορθογραφικός Διορθωτής, ο οποίος ελέγχει την ορθογραφία ενός κειμένου ή λέξης χρησιμοποιώντας το Μορφολογικό Λεξικό βιοϊατρικών όρων που υλοποιήθηκε.
- Μορφοσυντακτικός και Σημασιολογικός Σχολιαστής, ο οποίος επισυνάπτει μορφοσυντακτικά (από το Μορφολογικό Λεξικό) και σημασιολογικά (από την Οντολογία) σχόλια σε λέξεις/στοιχεία ενός κειμένου.

Τα παραπάνω περιγράφονται στα παραδοτέα:

14. Περιηγητής Οντολογίας
15. Μηχανή Αναζήτησης
16. Ορθογραφικός διορθωτής ιστού
17. Μορφοσυντακτικός και Σημασιολογικός Σχολιαστής
18. Τεκμηρίωση Εφαρμογών Ιστού

### ΕΕ6: Διάχυση Αποτελεσμάτων

Μετά το 2ο μήνα υλοποίησης, αναπτύχθηκε ο ιστοχώρος του έργου (project site) [www.iatrolexi.gr](http://www.iatrolexi.gr), όπου αναρτήθηκε καθ' όλη τη διάρκεια του έργου όλο το δημοσιοποιήσιμο υλικό και γενικότερα η πορεία του έργου. Το υλικό αυτό μπορεί να το συμβουλευτεί οποιοσδήποτε θέλει να μάθει κάτι για την εξέλιξη του έργου. Επιπλέον στο site αυτό δημιουργήθηκε ένα forum το οποίο χρησιμοποιήθηκε για συζήτηση για την πορεία του έργου, τόσο στα πλαίσια της ομάδας έργου όσο και με κάθε εξωτερικό ενδιαφερόμενο.

Παράλληλα και κατά τη διάρκεια του έργου υπήρξαν μια σειρά από επιστημονικές ανακοινώσεις σε Ελληνικά και Διεθνή συνέδρια σε θέματα που έχουν σχέση με τη διεξαγόμενη έρευνα.

Τέλος στα πλαίσια της διάχυσης των αποτελεσμάτων του έργου, οργανώθηκε σχετική ημερίδα. Στην ημερίδα παρουσιάστηκαν όλες οι σημαντικές εξελίξεις – αποτελέσματα του έργου.

Τα παραπάνω περιγράφονται στα παραδοτέα:

19. Υλοποίηση και συντήρηση project site
20. Ανακοινώσεις επιστημονικών εργασιών σε συνέδρια / περιοδικά.
21. Ημερίδα διάχυσης αποτελεσμάτων (πρόγραμμα – υλικό)
22. Τελική Αναφορά Έργου

## 2.2 ΑΝΑΛΥΤΙΚΗ ΠΕΡΙΓΡΑΦΗ ΤΩΝ ΕΝΟΤΗΤΩΝ ΕΡΓΑΣΙΩΝ (Ε.Ε.) ΠΟΥ ΥΛΟΠΟΙΗΘΗΚΑΝ

Παρακάτω περιγράφονται αναλυτικά οι ενότητες εργασίας του έργου όπως υλοποιήθηκαν.

### 2.2.1 ΕΕ1: Σχεδιασμός Σώματος Βιοϊατρικών Κειμένων και Οντολογίας

#### ΠΕΡΙΓΡΑΦΗ ΕΝΟΤΗΤΑΣ

Σε αυτή την ενότητα εργασίας σχεδιάστηκε ουσιαστικά και σε λεπτομέρεια το συνολικό έργο. Καθορίστηκε και περιγράφηκε το απαιτούμενο σώμα κειμένων και σχεδιάστηκε η μεθοδολογία για την ανάπτυξη της οντολογίας. Πιο συγκεκριμένα, δύο ήταν οι κατευθύνσεις μελέτης της ΕΕ1:

- α) Η εκπόνηση ενός σχεδίου συγκρότησης σώματος βιοϊατρικών κειμένων και
- β) Ο καθορισμός του μοντέλου αναπαράστασης της οντολογίας.

Να σημειωθεί ότι τα παραπάνω καταγράφονται λεπτομερώς στα σχετικά παραδοτέα Π1 και Π2.

#### **A) Σχέδιο συγκρότησης σώματος βιοϊατρικών κειμένων**

Οι φάσεις συλλογής και σχεδιασμού ενός Ελληνικού Σώματος Βιοϊατρικών Κειμένων (ΣΒΚ) στο πλαίσιο του έργου ΙΑΤΡΟΛΕΞΗ παρουσιάζονται στα επόμενα.

Η πρώτη ενέργεια ήταν να εντοπιστούν οι πηγές στο διαδίκτυο και να αξιολογηθούν ως προς το είδος, την πληρότητα, την ποιότητα και την ποικιλία των κειμένων που περιέχουν .

Η δεύτερη ενέργεια ήταν η κατηγοριοποίηση των κειμένων σε ιατρικές θεματικές ενότητες καθώς και ο διαχωρισμός τους με βάση το είδος και τη μορφή τους.

Τέλος, η τρίτη ενέργεια ήταν να σχεδιαστεί ένας τρόπος κωδικοποίησης των κειμένων που θα βοηθήσει στην περαιτέρω επεξεργασία και αξιοποίηση του ΣΒΚ.

#### **Εντοπισμός των πηγών στο διαδίκτυο**

Το έργο ΙΑΤΡΟΛΕΞΗ στόχευσε στο βιοϊατρικό χώρο χωρίς να εστιάζει σε κάποιο συγκεκριμένο πεδίο. Κατά συνέπεια, το ΣΒΚ που καλούμασταν να συγκροτήσουμε θα έπρεπε να περιέχει άρθρα από όλες τις ιατρικές ειδικότητες (οφθαλμολογία, παθολογία, γενική ιατρική, κ.ο.κ.). Προκειμένου δε το ΣΒΚ να είναι αντιπροσωπευτικό, θα έπρεπε να περιλαμβάνει ικανό αριθμό κειμένων από κάθε ειδικότητα καθώς και από κάθε είδος (π.χ.

κείμενα εκπαιδευτικού χαρακτήρα, επιστημονικά κείμενα, επιστημονικά άρθρα, επιστημονικά άρθρα συνεδρίων κτλ.).

Για το λόγο αυτό, έπρεπε να εντοπιστούν πηγές που να περιέχουν ηλεκτρονικά βιοϊατρικά κείμενα που να πληρούν τις παραπάνω προδιαγραφές. Η αναζήτηση στο διαδίκτυο εστιάστηκε σε ιστοτόπους που περιέχουν ελληνικά ιατρικά περιοδικά και σε ιστοτόπους ελληνικών ιατρικών συνεδρίων. Πιο συγκεκριμένα, η αναζήτηση των πηγών έγινε σε δύο επίπεδα. Σε πρώτο επίπεδο μέσω των ιστοτόπων:

- του Ελληνικού Ιατρικού Δικτύου (MedNet Hellas - <http://www.mednet.gr>),
- του Εθνικού Κέντρου Τεκμηρίωσης (<http://antigua.ekt.gr/portal/dt>) και
- της Βιβλιοθήκης του Πανεπιστημίου Μακεδονίας (<http://www.lib.uom.gr/index.php>)

και σε δεύτερο επίπεδο μέσω των μηχανών αναζήτησης:

- του Google (<http://www.google.com>) και
- του Yahoo (<http://www.yahoo.com/>).

Η συγκεκριμένη αναζήτηση μας οδήγησε σε σειρά συμπερασμάτων για την ελληνική πραγματικότητα σε ό,τι αφορά τη συγκέντρωση, κατηγοριοποίηση και δόμηση της πληροφορίας για τα ελληνικά ιατρικά περιοδικά. Η κατάταξη των τίτλων στις λίστες ήταν είτε αλφαβητική είτε θεματική κάτω από αυτοσχέδιες θεματικές κατηγορίες που δημιουργήθηκαν για τις ανάγκες θεματικής περιγραφής της κάθε συγκεκριμένης συλλογής περιοδικών. Οι λίστες περιελάμβαναν, στη συντριπτική τους πλειοψηφία, τίτλους περιοδικών ευρείας κυκλοφορίας (magazines) και, εκτός από τη ζεύξη με το δικτυακό τόπο του περιοδικού, την προαναφερθείσα θεματική κατηγοριοποίηση και, σε ορισμένες περιπτώσεις, μια υποτυπώδη αξιολόγηση των δικτυακών τόπων των περιοδικών, δεν παρείχαν καμιά άλλη πληροφορία. Ένα ποσοστό 10-15% των ζεύξεων δε λειτουργούσε, ένα άλλο περίπου 10% οδηγούσε σε δικτυακούς τόπους που θα μπορούσαν να χαρακτηριστούν καλύτερα ως θεματικές πύλες πρόσβασης ευρύτερου ενδιαφέροντος (π.χ. υγεία) παρά ως ηλεκτρονικά περιοδικά, ενώ ένα μικρότερο ποσοστό οδηγούσε σε ιστοτόπους υπό κατασκευή. Οι ελληνικοί οδηγοί-κατάλογοι του Internet περιελάμβαναν τίτλους περιοδικών χωρίς καμιά διάκριση ως προς:

- Τον τύπο του περιοδικού (ηλεκτρονικές εκδόσεις εντύπων – ηλεκτρονικές μόνον εκδόσεις)
- Το είδος του περιοδικού (περιοδικά ευρείας κυκλοφορίας – ακαδημαϊκά περιοδικά – ενημερωτικά φυλλάδια κτλ.)
- Το ηλεκτρονικό περιεχόμενο (πλήρες κείμενο, μερικό κείμενο, μόνο πίνακες περιεχομένων χωρίς κείμενο, περιλήψεις ή απλώς γενικές πληροφορίες για το περιοδικό)
- Το μορφότυπο του ηλεκτρονικού περιεχομένου (html, pdf, txt, doc, jpeg κ.λπ.)
- Τη γλώσσα και τη χώρα έκδοσης (περιοδικά σε ελληνική γλώσσα που εκδίδονται σε οποιαδήποτε χώρα, περιοδικά σε άλλες γλώσσες που εκδίδονται στην Ελλάδα)
- Την κάλυψη (περιοδικά που διατηρούν μόνον το τρέχον τεύχος, περιοδικά που διατηρούν αρχείο και περιοδικά που διαθέτουν απλώς ένα δείγμα τευχών)
- Την κατάσταση (τρέχοντα περιοδικά – περιοδικά που έχουν αναστείλει την έκδοσή τους)



- Την ενημέρωση (δικτυακοί τόποι περιοδικών που ενημερώνονται ή που δεν έχουν καμία ενημέρωση από τότε που κατασκευάστηκαν)
- Την πρόσβαση (περιοδικά με ελεύθερη ή περιορισμένη πρόσβαση).

Συνολικά εντοπίστηκαν 40 ιστότοποι οι οποίοι περιέχουν υλικό για ιατρικά επιστημονικά περιοδικά και ιατρικά συνέδρια (βλέπε παραδοτέο Π1).

### **Κατηγοριοποίηση Κειμένων σε Θεματικές Ενότητες**

Όπως αναφέρθηκε παραπάνω, τα κείμενα που βρέθηκαν δεν είχαν κάποια συγκεκριμένη κατηγοριοποίηση ως προς τη θεματική ενότητα με αποτέλεσμα το σώμα κειμένων να είναι μη λειτουργικό για κάποιον που θέλει να το χρησιμοποιήσει για συγκεκριμένο σκοπό (π.χ. να αντλήσει πληροφορίες για τον καρδιολογικό τομέα μόνο). Έτσι, για την καλύτερη τεκμηρίωση του σώματος κειμένων, διαχωρίσαμε και κατηγοριοποιήσαμε τα κείμενα με βάση τη θεματική ενότητα στην οποία ανήκουν.

### **Είδη και μορφές κειμένων**

Ως προς τη μορφή, τα κείμενα που βρέθηκαν ήταν περιλήψεις άρθρων, πλήρη άρθρα, κείμενα με πολλά άρθρα σε ένα αρχείο και πρακτικά συνεδρίων. Τα περισσότερα από αυτά περιέχουν εκτός από κείμενο και άλλα στοιχεία, όπως εικόνες, πίνακες, γραφικές παραστάσεις κ.ά. Επιπρόσθετα, ένα μέρος από το πλήθος των κειμένων που συλλέξαμε περιέχει ελληνικά και αγγλικά (μετάφραση) κομμάτια με αποτέλεσμα να μας δίνεται η δυνατότητα (στο μέλλον) για το σχεδιασμό ενός παράλληλου σώματος κειμένων. Η μορφή των κειμένων ήταν είτε HTML είτε PDF.

Ως προς τα είδη των κειμένων, η ομάδα τυποποίησης του TEI (Text Encoding Initiative) έχει δώσει ιδιαίτερη έμφαση στον καθορισμό ενός συνόλου ειδικών επικεφαλίδων (headers). Οι επικεφαλίδες είναι ένας τρόπος τεκμηρίωσης των ηλεκτρονικών κειμένων. Περιγράφουν το ηλεκτρονικό κείμενο και την πηγή στην οποία βρέθηκε (βιβλιογραφικές πληροφορίες, όταν είναι διαθέσιμες), προσδιορίζουν τις μεθόδους και τις συγγραφικές αρχές που συνηγορούν στην αντιγραφή του κειμένου, περιέχουν μη βιβλιογραφικές πληροφορίες που χαρακτηρίζουν το κείμενο και ιστορικό ενημερώσεων και αλλαγών του κειμένου. Στο τμήμα των μη βιβλιογραφικών πληροφοριών μπορεί να περιέχονται και θέματα που αφορούν τη συντακτική και μορφολογική ανάλυση των κειμένων.

Για την παραγωγή των επικεφαλίδων για τα βιοϊατρικά κείμενα που συλλέξαμε, μελετήσαμε με βάση τη βιβλιογραφία, τον τρόπο που ορίζονται οι επικεφαλίδες σε σχέση με τα είδη των κειμένων με σκοπό να αναγνωρίσουμε τα διαφορετικά είδη βιοϊατρικών κειμένων που βρέθηκαν και να τεκμηριώσουμε το καθένα με τις σχετικές επικεφαλίδες.

## Κωδικοποίηση κειμένων

Το σχήμα κωδικοποίησης των κειμένων του ΙΑΤΡΟΛΕΞΗ αποφασίστηκε να βασιστεί στις οδηγίες του TEI Lite, όπως αυτό περιγράφεται αναλυτικά στο κεφάλαιο 4.4.2 του παραδοτέου Π1. Στο ίδιο παραδοτέο περιγράφεται το σχήμα του σώματος κειμένου (corpus) το οποίο αποτελείται από μία βασική επικεφαλίδα (corpus header) ακολουθούμενη από ένα σύνολο κειμένων. Το κάθε κείμενο αποτελείται από μία επικεφαλίδα κειμένου (document header) και το ίδιο το κείμενο (text). Η επικεφαλίδα του σώματος κειμένου τεκμηριώνει το σώμα κειμένου σαν μια ολότητα, ενώ η επικεφαλίδα κειμένου περιέχει μεταδεδομένα για το κάθε κείμενο.

## Β) Μοντέλο Αναπαράστασης Οντολογίας

Στη μελέτη αυτή, καθορίστηκαν τα βήματα για την ανάπτυξη της οντολογίας «ΙΑΤΡΟΛΕΞΗ». Λεπτομερέστερη και αναλυτικότερη περιγραφή μπορεί κανείς να βρει στο παραδοτέο Π2.

### Βήμα 1<sup>ο</sup> : Καθορισμός μοντέλου αναπαράστασης και αρχικής ταξινόμιας

Ο καθορισμός του μοντέλου αναπαράστασης της οντολογίας συνίσταται αφενός στον προσδιορισμό του θεωρητικού πλαισίου και των σημασιολογικών μηχανισμών που θα χρησιμοποιηθούν για την ανάλυση, την περιγραφή και την οργάνωση των εννοιών του βιοϊατρικού τομέα και αφετέρου στον προσδιορισμό των πληροφοριών που θα χρησιμοποιηθούν για την οργάνωση των όρων που κατασημαίνουν κάθε έννοια.

Το μοντέλο αναπαράστασης των όρων περιλαμβάνει ειδικότερα:

- καθορισμό των αρχών με βάση τις οποίες θα δομηθεί η ορολογική εγγραφή στο σύστημα διαχείρισης της οντολογίας,
- καθορισμό των πληροφοριών που θα κωδικοποιούνται για κάθε όρο.

Το μοντέλο αναπαράστασης των εννοιών περιλαμβάνει ειδικότερα:

- τον καθορισμό των βασικών **εννοιών** και την οργάνωσή τους σε **ταξινόμια**,
- τη δήλωση των **ιδιοτήτων** ή **χαρακτηριστικών** κάθε έννοιας, και
- τον εντοπισμό και την περιγραφή των **σχέσεων** κάθε έννοιας με άλλες έννοιες.

Δεδομένου ότι στόχος του ΙΑΤΡΟΛΕΞΗ είναι η δημιουργία μιας γενικής και ανεξάρτητης από εφαρμογές υποδομής για την επεξεργασία της βιοϊατρικής ορολογίας στα ελληνικά, αποφασίστηκε μετά από έρευνα και αξιολόγηση των σχετικών πηγών να χρησιμοποιηθεί ως βασικό μοντέλο αναπαράστασης των βιοϊατρικών εννοιών, με τις ανάλογες προσαρμογές και προσθήκες που θα προκύψουν κατά την επεξεργασία των ορολογικών δεδομένων, το μοντέλο αναπαράστασης και οι ιεραρχίες εννοιών και σχέσεων του Σημασιολογικού Δικτύου

του UMLS, που αποτελεί την πιο εκτεταμένη από άποψη όγκου και ποικιλίας δεδομένων πηγή οντολογικής γνώσης για το θεματικό πεδίο της βιοϊατρικής.

Η επιλογή του συγκεκριμένου μοντέλου στηρίχτηκε ειδικότερα στα εξής κριτήρια:

- Γενικότητα και ευρύτητα
- Ευελιξία και επεκτασιμότητα
- Επεξεργασιμότητα από υπολογιστικά εργαλεία
- Συμβατότητα με διεθνή προϊόντα.

### **Βήμα 2<sup>ο</sup> : Κατάρτιση της μακροδομής της οντολογίας**

Δεύτερο βήμα για την ανάπτυξη της οντολογίας αποτελεί η κατάρτιση της μακροδομής της, δηλαδή του συνόλου των όρων που αποτελούν πραγματώσεις των βιοϊατρικών εννοιών και απαρτίζουν το λεξιλόγιο της. Ακολουθήθηκαν δύο αλληλοσυμπληρούμενες διαδικασίες για τη συγκέντρωση των όρων:

1. **Επεξεργασία δίγλωσσων λεξιλογίων βιοϊατρικών όρων** (αγγλικών - ελληνικών): Στόχος αυτής της εργασίας ήταν να εντοπιστεί το βασικό λεξιλόγιο της ελληνικής βιοϊατρικής ορολογίας που περιλαμβάνεται σε έντυπα και ηλεκτρονικά λεξικά. Πιο συγκεκριμένα, για τον εντοπισμό του βασικού λεξιλογίου έχει σχεδιαστεί η **επεξεργασία** περίπου 14.000 όρων που περιλαμβάνονται στο **ελληνικό MeSH** (έκδοση 1996) και συνεχίζουν να αποτελούν όρους της τρέχουσας έκδοσης του αμερικανικού **MeSH®** (έκδοση 2006) και η **σύγκρισή** τους με το λεξιλόγιο των 20.000 όρων του Λεξικού Βιολογικών και Ιατρικών Όρων της Magenta και των 5.000 βιοϊατρικών όρων που περιλαμβάνονται στο Μορφολογικό Λεξικό της Neurosoft. Ο εντοπισμός του βασικού λεξιλογίου στο πρώτο στάδιο κατάρτισης της μακροδομής αναμένεται ότι θα προσφέρει ορολογικό υλικό για τις περισσότερες από τις σημασιολογικές κατηγορίες της οντολογίας.
2. **Εξαγωγή όρων από ελληνικά βιοϊατρικά κείμενα** : Το ορολογικό υλικό που αντλήθηκε από υπάρχοντα λεξικά στη συνέχεια εμπλουτίστηκε με έννοιες και όρους από ελληνικά βιοϊατρικά κείμενα (που συλλέγησαν για το σκοπό αυτό).

### **Βήμα 3<sup>ο</sup> : Εμπλουτισμός της οντολογίας με νέες έννοιες και νέους όρους**

Η διαδικασία του εμπλουτισμού της οντολογίας έγινε σε τακτά χρονικά διαστήματα και αφορούσε αφενός τη διεύρυνση του πυρήνα της οντολογίας με έννοιες που δεν είχαν προβλεφθεί από το μοντέλο, αφετέρου την προσθήκη στο βασικό λεξιλόγιο όρων που εντοπίζονταν στο διαρκώς εμπλουτιζόμενο με νέο υλικό σώμα κειμένων του ΙΑΤΡΟΛΕΞΗ

## 2.2.2 ΕΕ2: Υλοποίηση Εργαλείων Διαχείρισης Σώματος Κειμένου και Αυτόματης Εξαγωγής Ορολογίας

### ΠΕΡΙΓΡΑΦΗ ΕΝΟΤΗΤΑΣ

Σε αυτή την ενότητα εργασίας, αναπτύχθηκαν/προσαρμόστηκαν/εμπλουτίστηκαν μια σειρά από εργαλεία λογισμικού. Τα εργαλεία αυτά έγιναν για να υποστηρίξουν την ανάπτυξη των γλωσσικών πόρων του έργου, αλλά και την ολοκλήρωση των τελικών στόχων του.

Τα εργαλεία που ολοκληρώθηκαν δεν υλοποιήθηκαν όλα από την αρχή. Με δεδομένο ότι οι δύο φορείς διαθέτουν μακράν εμπειρία στην ανάπτυξη εργαλείων γλωσσικής τεχνολογίας, κάποια αναπτύχθηκαν εκ του μηδενός, κάποια βασίστηκαν πάνω σε ήδη υπάρχοντα προπλάσματα και κάποια εμπλουτίστηκαν. Πιο συγκεκριμένα στα πλαίσια της παρούσας ΕΕ, ολοκληρώθηκαν τα παρακάτω συστημάτων:

- 1) Αποθήκη Εγγράφων (Document Warehouse): είναι ο χώρος αποθήκευσης των κειμένων αλλά και των παραγόμενων από αυτά μεταδεδομένων. Υλοποιήθηκε με βάση την open source βάση δεδομένων MySQL.
- 2) Προσκομιστής Εγγράφων (Crawler): διατρέχει συγκεκριμένους δικτυακούς τόπους ή τοπικούς δίσκους και προσκομίζει τα κείμενα προς επεξεργασία. Υλοποιήθηκε σε Java (JDK 6.0).
- 3) Μετατροπέας Εγγράφων (Document Converter): μετατρέπει τα έγγραφα HTML ή PDF που προσκομίζει ο Crawler σε έγγραφα TXT (δηλ. σε απλό κείμενο). Υλοποιήθηκε σε Java (JDK 6.0). *Το εργαλείο αυτό δεν είχε προβλεφθεί αρχικά μια και είχε γίνει η εκτίμηση ότι θα βρεθούν έγγραφα έτοιμα σε μορφή κειμένου, κάτι που όπως αποδείχθηκε εκ των υστέρων δεν ήταν σωστό.*
- 4) Αναγνωριστής Στοιχείων (Tokenizer): κερματίζει ένα κείμενο σε μία σειρά στοιχείων (tokens: λέξεις, σημεία στίξης, αριθμοί, σύμβολα κτλ.) με τα οποία τροφοδοτούνται οι επόμενες φάσεις επεξεργασίας του κειμένου. Για την υλοποίησή του εξελίχθηκε ένα πρωτόλειο σχετικό σύστημα που είχε αναπτυχθεί στο παρελθόν για ερευνητικούς σκοπούς.
- 5) Μορφοσυντακτικός Σχολιαστής (Morphosyntactic Tagger): επισυνάπτει μορφοσυντακτικά μεταδεδομένα (μέρος του λόγου, γένος, αριθμός πτώση, κλπ.) σε κάθε λέξη του κειμένου (που έχει αναγνωρίσει ο Tokenizer). Αυτό πραγματοποιείται με τη βοήθεια του Μορφολογικού Λεξικού. Υλοποιήθηκε σε Java (JDK 6.0).
- 6) Μορφολογικό Λεξικό. Η Neurosoft A.E. έχει αναπτύξει μορφολογικό λεξικό της Ελληνικής με περίπου 90.000 λήμματα, το οποίο περιέχει και περιορισμένο αριθμό βιοϊατρικών όρων. Το λεξικό αυτό εμπλουτίστηκε με τις άγνωστες λέξεις-όρους (περίπου 9.300) που συλλέχθηκαν στο πλαίσιο της ΕΕ3. Το υπάρχον σύστημα περιγραφής της κλίσης των λημμάτων καλύπτει τον ορισμό της μορφοσυντακτικής πληροφορίας μονολεκτικών όρων.
- 7) Μηχανισμός Κλίσης Πολυλεκτικών Όρων. Στην ΕΕ3 επεκτάθηκε ο μηχανισμός κλίσης μονολεκτικών όρων ώστε να υποστηρίξει τον ορισμό πολυλεκτικών όρων. Για

την υποστήριξη του μοντέλου αυτού, αναπτύχθηκε σχετική εφαρμογή, που παρέχει στο χρήστη τη δυνατότητα εύκολου ορισμού της κλίσης πολυλεκτικών όρων.

- 8) Αναγνωριστής Όρων. Αναπτύχθηκε σχετικό σύστημα το οποίο συμβουλεύεται το Μορφολογικό Λεξικό, καθώς και τους κανόνες που περιγράφουν τη σύνταξη πολυλεκτικών όρων και αναγνωρίζει τους όρους αυτούς σε κείμενα, σε όποια κλιτική μορφή κι αν βρίσκονται.

Σχεδόν όλα τα παραπάνω εργαλεία, μπορούν να χρησιμοποιηθούν αυτόνομα. Επιπρόσθετα, για την καλύτερη και ευκολότερη χρήση τους από τρίτους (ερευνητές ή μη) ολοκληρώθηκαν κάτω από ένα κοινό περιβάλλον που καθοδηγεί το χρήστη σε διαδοχικές ενέργειες/βήματα.

Στο παραδοτέο **Π4: Εγχειρίδιο χρήσης**, περιγράφονται αναλυτικά τα υποσυστήματα αυτά.

Τέλος στο ίδιο παραδοτέο (Π4), περιγράφεται και ένα επιπλέον εργαλείο που αποφασίστηκε να υλοποιηθεί με βάση τις απαιτήσεις των ειδικών (ιατρών – γλωσσολόγων): Ένας **Συλλογέας Λεξιλογικών Συνάψεων** (Concordancer), που όταν χρησιμοποιηθεί με βάση τη συλλογή κειμένων, δίνει πολύ καίρια και σημαντικά στοιχεία στον επιστήμονα σε σχέση με τη λεξικογραφία ενός όρου (<http://www.iatrolexi.gr/tools/webconc>).

## 2.2.3 ΕΕ3: Συλλογή Βιοϊατρικών κειμένων και Εξαγωγή Ορολογίας

### ΠΕΡΙΓΡΑΦΗ ΕΝΟΤΗΤΑΣ

Στην ενότητα εργασίας αυτή συλλέχθηκε από τον κυβερνοχώρο το απαραίτητο σώμα βιοϊατρικών κειμένων. Το σώμα αυτό περιέχει 6.276 κείμενα και περίπου 11,5 εκ. λέξεις. Από το σώμα αυτό, εξήχθησαν οι ιατρικοί όροι που μαζί και με τους όρους που συλλέχθηκαν από ιατρικά λεξικά έφτασαν τους περίπου 9.300. Τους όρους αυτούς επιμελήθηκαν οι ιατροί και οι γλωσσολόγοι. Επιπλέον συλλέχθηκαν και οι πολυλεκτικοί όροι, που ανέρχονται περίπου σε 5.000.

Πιο αναλυτικά σε σχέση με κάθε υποεργασία της ενότητας αυτής:

#### **A. Μονολεκτικοί Όροι Βιοϊατρικής**

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](http://www.google.gr), κτλ).

**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 περιέχονταν στο Μορφολογικό Λεξικό (ΜΛ)<sup>1</sup>.

<sup>1</sup> Δεδομένου ότι το ΜΛ αναγνωρίζει περισσότερους από 1.500.000 κλιτικούς 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.

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. *Χιλή*.

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).

## Β. Κανόνες Αναγνώρισης Πολυλεκτικών Όρων Βιοϊατρικής

Η λεπτομερής καταγραφή και ανάλυση των περίπου 213 διαφορετικών δομών των διλέκτων, τριλέκτων και τετραλέκτων του πεδίου της βιοϊατρικής επέτρεψε την εξαγωγή των περίπου είκοσι βασικών κανόνων σχηματισμού των πολυλεκτικών σύμπλοκων όρων του πεδίου, σύμφωνα με τους οποίους είναι δυνατή η αναζήτηση, ο εντοπισμός και η εξαγωγή τους από τα κείμενα.

Για την περιγραφή των κανόνων αναγνώρισης των πολυλεκτικών σύμπλοκων όρων της βιοϊατρικής χρησιμοποιήθηκε ο πιο διαδομένος φορμαλισμός περιγραφής της σύνταξης τεχνητών γλωσσών που είναι οι **definite-clause γραμματικές** (Pereira and Warren 1980). Στο πλαίσιο αυτού του φορμαλισμού ο κανόνας αναγνώρισης της ονοματικής φράσης (ΟΦ) μπορεί να περιγραφεί ως εξής:

ΟΦ(Γένος, Αριθμός, Πτώση) ←

Άρθ(Γένος, Αριθμός, Πτώση) Ουσ(Γένος, Αριθμός, Πτώση)

Σύμφ  
ωνα

με τον κανόνα αυτό ένα ουσιαστικό μετά από ένα άρθρο συνιστούν μία ονοματική φράση όταν συμφωνούν κατά γένος, αριθμό και πτώση. Επιπλέον, η ονοματική φράση έχει το ίδιο Γένος, Αριθμό και Πτώση με το άρθρο και το ουσιαστικό.



## Γ. Πολυλεκτικοί Όροι Βιοϊατρικής

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

- **Επεξεργασία του αγγλο-ελληνικού λεξιλογίου βιοϊατρικών όρων MeSH** (έκδοση 1996): Ο στόχος αυτής της εργασίας ήταν διττός: 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.
- **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.
- **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 9.300 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>2</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 μπορεί να ανήκει ένας όρος με

<sup>2</sup> Στην παρούσα φάση, το συνολικό μέγεθος του Σώματος Ελληνικών Βιοϊατρικών Κειμένων (ΣΕΒΚ) του ΙΑΤΡΟΛΕΞΗ ανέρχεται σε 11.500.000 λέξεις.



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.

#### Δ. Αρχική Ταξινόμια Βιοϊατρικών Όρων

Όπως ήδη έχει αναφερθεί στο Παραδοτέο Π2 «Μοντέλο Αναπαράστασης της Οντολογίας», στο πλαίσιο του παρόντος έργου η οντολογία αποτελεί την **τυπική περιγραφή της γνώσης του πεδίου της βιοϊατρικής**, για την κατασκευή της οποίας υλοποιήθηκαν μέχρι στιγμής οι εξής υποεργασίες:

1. **Καθορισμός μοντέλου αναπαράστασης οντολογίας:** αφορά αφενός τον προσδιορισμό του θεωρητικού πλαισίου και των σημασιολογικών μηχανισμών που θα χρησιμοποιηθούν για την ανάλυση, την περιγραφή και την οργάνωση των **εννοιών** του βιοϊατρικού τομέα και αφετέρου τον προσδιορισμό των πληροφοριών που θα χρησιμοποιηθούν για την οργάνωση των **όρων** που κατασημαίνουν κάθε έννοια. Για την επιλογή του βασικού μοντέλου αναπαράστασης, εξετάστηκαν μια σειρά από σχετικές οντολογίες – κωδικοποιήσεις (ICD10, MESH, Νόσοι – Διαγνώσεις, UMLS). Κατόπιν μελέτης των μοντέλων αυτών τόσο από την ιατρική όσο και από τη γλωσσική πλευρά τους βασικό μοντέλο αναπαράστασης των βιοϊατρικών εννοιών του ΙΑΤΡΟΛΕΞΗ υιοθετήθηκε το *μοντέλο αναπαράστασης* του Σημασιολογικού Δικτύου του UMLS (<http://umlsinfo.nlm.nih.gov/>), που αποτελεί την πιο εκτεταμένη από άποψη όγκου και ποικιλίας δεδομένων πηγή οντολογικής γνώσης για το θεματικό πεδίο της βιοϊατρικής.
2. **Κατάρτιση αρχικής ταξινόμιας:** για την κατάρτιση της αρχικής ταξινόμιας των βιοϊατρικών όρων ακολουθήθηκε η *από πάνω προς τα κάτω* (top-down) και η *από κάτω προς τα πάνω* (bottom-up) προσέγγιση ανάπτυξης:
  - Στο πλαίσιο της *από πάνω προς τα κάτω* (top-down) προσέγγισης προσαρμόστηκαν στα ελληνικά από τους ειδικούς (ιατρούς γλωσσολόγους) οι κατηγορίες βιοϊατρικών εννοιών που περιλαμβάνονται στις δύο ιεραρχίες του Σημασιολογικού Δικτύου του UMLS.
  - Στο πλαίσιο της *από κάτω προς τα πάνω* (bottom-up) προσέγγισης, η οποία θα εφαρμόζεται σταδιακά μέχρι και την ολοκλήρωση της οντολογίας, η αρχική ταξινόμια εμπλουτίζεται με ειδικότερες κατηγορίες εννοιών, που είτε απαντώνται στο Σώμα Κειμένων του ΙΑΤΡΟΛΕΞΗ είτε αφορούν την ευκρινέστερη διάκριση των σημασιολογικών τύπων του UMLS με βάση τις ιεραρχικές σχέσεις “is\_a” και “part\_of”, δηλαδή των εννοιών που αναφέρονται σε είδη και μέρη, διάκριση που δεν εφαρμόζεται συστηματικά στις

κατηγοριοποιήσεις του Σηματολογικού Δικτύου και της υπάρχουσας αρχικής ταξινόμιας.

### Ε. Συλλογή Βιοϊατρικών κειμένων

Το συγκεκριμένο Σώμα Κειμένων αποτελείται από **11.422.552 λέξεις** οι οποίες περιέχονται σε ελληνικά βιοϊατρικά κείμενα που τυπώθηκαν από το 1996 μέχρι σήμερα και τα οποία βρέθηκαν σε 47 ελληνικές ιστοσελίδες. Οι παραπάνω ιστοσελίδες περιέχουν υλικό για ιατρικά επιστημονικά περιοδικά, ιατρικές εφημερίδες και ιατρικά συνέδρια. Στον πίνακα 1 φαίνονται οι παραπάνω υπερσύνδεσμοι των ιστοτόπων, το πλήθος των κειμένων που βρέθηκαν και η μορφή τους (pdf, htm, txt κ.ο.κ.).

Ως προς τη μορφή, τα κείμενα που βρέθηκαν ήταν περιλήψεις άρθρων, πλήρη άρθρα, κείμενα με πολλά άρθρα σε ένα αρχείο και πρακτικά συνεδρίων. Τα περισσότερα από αυτά περιέχουν εκτός από κείμενο και άλλα στοιχεία, όπως εικόνες, πίνακες, γραφικές παραστάσεις κ.ά. Επιπρόσθετα, ένα μέρος από το πλήθος των κειμένων που συλλέξαμε περιέχει ελληνικά και αγγλικά (μετάφραση) κομμάτια με αποτέλεσμα να μας δίνεται η δυνατότητα (στο μέλλον) για το σχεδιασμό ενός παράλληλου σώματος κειμένων. Συνολικά, βρέθηκαν 6.276 κείμενα από τα οποία, το 69,8% ήταν σε hypertext markup language (.htm) και το 30,2% ήταν σε Portable Document Format (.pdf) μορφή. Το επόμενο στάδιο είναι η μετατροπή των κειμένων από την πραγματική μορφή τους (HTML, PDF) σε text μορφή.

### ΣΤ. Τεκμηρίωση Συλλογής Βιοϊατρικών κειμένων

Όπως αναφέρθηκε και στο παραδοτέο Π1, δεν υπάρχει συγκεκριμένη μεθοδολογία κοινώς αποδεκτή στην ερευνητική κοινότητα, σε σχέση με τη διαδικασία ή τις διαδικασίες που θα πρέπει να ακολουθούνται για τη συλλογή σώματος κειμένων.

Συνεπώς με βάση την γενικότερη θεωρία των σωμάτων κειμένου (corpus) θα πρέπει κανείς να πράξει κατά περίπτωση και αφού απαντήσει σε μια σειρά από ερωτήματα όπως:

- Ποιοι θα είναι οι χρήστες; (π.χ. έρευνα, πηγή για άλλη χρήση, κλπ)
- Ποιος είναι ο σκοπός της συλλογής; (π.χ. βάση για κάποιο λεξικό, έρευνα για στατιστικά μεγέθη, μελέτη κάποιου γλωσσικού φαινομένου, κλπ)
- Πόσο μεγάλη θα πρέπει να είναι η συλλογή; Ποιο είναι ένα ρεαλιστικό μέγεθος;
- Δειγματοληψία ή πλήρης (εξαντλητική) συλλογή;
- Συλλογή ηλεκτρονικών κειμένων ή και έντυπων αντικειμένων (ανάγκη για ψηφιοποίηση).

Επιπλέον το ερευνητικό έργο ΙΑΤΡΟΛΕΞΗ, θέτει το ίδιο μια σειρά από περιορισμούς σε σχέση με τα παραπάνω ερωτήματα. Πρόσθετοι περιορισμοί τέθηκαν από το γενικότερο περιβάλλον του τομέα όπως περιγράφεται στη συνέχεια.

Πιο συγκεκριμένα:

- Οι χρήστες του σώματος βιοϊατρικών κειμένων εκτός των συμμετεχόντων στο έργο, πρόκειται να είναι το σύνολο της ερευνητικής κοινότητας που πιθανώς θα χρειαστεί σχετικές συλλογές για τις έρευνες – μελέτες του.
- Οι βασικοί στόχοι της συλλογής είναι αφενός η εξαγωγή ορολογίας και αφετέρου η παρατήρηση του τρόπου χρήσης (context – περιβάλλον) των βιοϊατρικών όρων.
- Το μέγεθος του σώματος πρέπει να είναι το μέγιστο (ρεαλιστικά) δυνατό, στο πλαίσιο του έργου.

Παράλληλα, η διάρκεια του έργου και ακόμα περισσότερο η διάρκεια της συγκεκριμένης ενότητας εργασίας ( 9 μήνες) έθεσαν έναν ακόμα περιορισμό:

- Η συλλογή θα έπρεπε να προέρχεται από ηλεκτρονικά κείμενα αφού δεν υπήρχαν ούτε χρόνος αλλά ούτε και οι πόροι, για ψηφιοποίηση έντυπου υλικού.

Επιπλέον, μετά από κάποιες προσπάθειες συνεννόησης με εκδοτικούς οίκους ιατρικών βιβλίων και περιοδικών, που απέβησαν άκαρπες, μιας και ήταν σχεδόν αδύνατο να πεισθούν να δώσουν τα κείμενα σε ηλεκτρονική μορφή - έστω και για ερευνητικούς σκοπούς, έγινε κατανοητό ότι η μοναδική πηγή για τη συλλογή του έργου θα έπρεπε να είναι το διαδίκτυο.

Τέλος, κατά την εξέλιξη της διαδικασίας συλλογής, διακριβώθηκε ότι τα βιοϊατρικά κείμενα που υπήρχαν ελεύθερα δημοσιευμένα στο διαδίκτυο ήταν σε μέγεθος τόσα, που να επιτρέπουν την σχεδόν εξαντλητική συλλογή τους.

## 2.2.4 ΕΕ4: Επεξεργασία Βιοϊατρικών Όρων και Δημιουργία Οντολογίας

### ΠΕΡΙΓΡΑΦΗ ΕΝΟΤΗΤΑΣ

Στην ενότητα εργασίας αυτή υλοποιήθηκε η οντολογία που αποτελεί βασικό κορμό για το έργο. Στην αρχή, προσαρμόστηκε το Protégé ώστε να μπορεί να χειριστεί και τα Ελληνικά. Στη συνέχεια η Αρχική Ταξινόμια Βιοϊατρικών Όρων (ΕΕ3, Παραδοτέο 4) εισάχθηκε στο Protégé και αποτέλεσε τη βάση για την ανάπτυξη της τελικής οντολογίας.

Πιο αναλυτικά σε σχέση με κάθε υποεργασία της ενότητας αυτής:

#### **A. Προσαρμογή Protégé**

Στο Protégé και στη φορμαλιστική γλώσσα OWL που τελικά επιλέχθηκε για την ανάπτυξη της οντολογίας των βιοϊατρικών όρων του έργου ΙΑΤΡΟΛΕΞΗ δεν υπάρχει η δυνατότητα ονομασίας των κλάσεων, αλλά και των στιγμιότυπων στα ελληνικά, αφού τα ονόματα αποτελούν αναγνωριστικά της OWL και δεν επιτρέπονται ελληνικοί χαρακτήρες. Υπάρχει όμως η δυνατότητα χρήσης ελληνικών μέσω ετικετών (rdfs:label) και σχολίων-περιγραφών (rdfs:comment) στα ελληνικά που μπορούν να μπουν στις κλάσεις και στα στιγμιότυπα. Για παράδειγμα μια κλάση της οντολογίας του ΙΑΤΡΟΛΕΞΗ στην OWL με ελληνική ετικέτα και ελληνικό σχόλιο έχει ως εξής:

```
<owl:Class rdf:about="#Group">
```

```
<rdfs:comment xml:lang="en">A conceptual entity referring to the classification of individuals according to certain shared characteristics.</rdfs:comment>
```

```
<rdfs:comment xml:lang="el">Μια εννοιολογική οντότητα που αναφέρεται στην ταξινόμηση των ατόμων σύμφωνα με ορισμένα κοινά χαρακτηριστικά.</rdfs:comment>
```

```
<rdfs:label xml:lang="en">Group</rdfs:label>
```

```
<rdfs:label xml:lang="el">Ομάδα</rdfs:label>
```

```
<rdfs:subClassOf>
```

```
<owl:Class rdf:about="#ConceptualEntity"/>
```

```
</rdfs:subClassOf>
```

```
</owl:Class>
```

Βλέπουμε παραπάνω την ετικέτα <rdfs:label xml:lang="el">Ομάδα</rdfs:label>, όπου το "el" δηλώνει ότι είναι στην ελληνική γλώσσα. Κατά αντιστοιχία για το σχόλιο έχουμε <rdfs:comment xml:lang="el"> Μια εννοιολογική οντότητα που αναφέρεται στην ταξινόμηση των ατόμων σύμφωνα με ορισμένα κοινά χαρακτηριστικά.</rdfs:comment>.

Το Protégé μέχρι σήμερα δεν διαθέτει δυνατότητα μεμονωμένης διαχείρισης ετικετών και σχολίων στα ελληνικά. Ωστόσο μπορούμε να προσαρμόσουμε το Protégé, ώστε να

εισάγουμε τους ελληνικούς όρους στην οντολογία του ΙΑΤΡΟΛΕΞΗ, προσθέτοντας ετικέτες και σχόλια στα οποία γράφουμε την ένδειξη “el” στο πεδίο “lang”. Σημειώνουμε ότι η συγκεκριμένη ένδειξη δεν παρέχεται στις επιλογές που διαθέτει τα συγκεκριμένο πεδίο, αλλά μπορεί πολύ απλά να προστεθεί.

Επιπλέον στα πλαίσια του προσαρμοσμένου Protégé και της διευκόλυνσης των ερευνητών αναπτύχθηκε το plug-in του Protégé «IatrolexiTab». Κάνοντας χρήση του IatrolexiTab οι χρήστες μπορούν να πλοηγηθούν στην οντολογία κάνοντας χρήση των ελληνικών ετικετών. Μπορούν επίσης να δουν τα ελληνικά σχόλια (δηλ. την ελληνική περιγραφή του όρου). Επιπλέον μπορούν να αναζητήσουν όρους με βάση την ελληνική ετικέτα.

## **B. Αρχική Ταξινόμια εντός Protégé**

Όπως έχει ήδη τεκμηριωθεί και περιγραφεί και σε προηγούμενα παραδοτέα (βλέπε και Π2, Π8) και με το δεδομένο ότι στόχος του ΙΑΤΡΟΛΕΞΗ είναι η δημιουργία μιας γενικής και ανεξάρτητης από εφαρμογές υποδομής για την επεξεργασία της βιοϊατρικής ορολογίας στα ελληνικά, αποφασίστηκε να χρησιμοποιηθεί ως βασικό μοντέλο αναπαράστασης των βιοϊατρικών εννοιών, με τις ανάλογες προσαρμογές και προσθήκες που θα προκύψουν κατά την επεξεργασία των ορολογικών δεδομένων, το μοντέλο αναπαράστασης και οι ιεραρχίες εννοιών και σχέσεων του Σημασιολογικού Δικτύου του UMLS, που αποτελεί την πιο εκτεταμένη από άποψη όγκου και ποικιλίας δεδομένων πηγή οντολογικής γνώσης για το θεματικό πεδίο της βιοϊατρικής.

Με βάση αυτό το μοντέλο, που αποτελεί το πρώτο επίπεδο της ταξινόμιας και οριστικοποιήθηκε στο Παραδοτέο Π8, η εργασία που εκτελέστηκε στα πλαίσια της υποεργασίας αυτής ήταν:

- α) η εισαγωγή της αρχικής αυτής ταξινόμιας στο Protégé και
- β) η ανάπτυξη της ταξινόμιας προς τα κάτω, καθώς και η επιμέλειά της από τους ειδικούς.

Όσον αφορά στην εισαγωγή της αρχικής ταξινόμιας στο Protégé, δεν ήταν κάτι ιδιαίτερα χρονοβόρο μια και οι κατηγορίες εννοιών (κλάσεις) ήταν 100 στην κατηγορία των οντοτήτων και 35 στην κατηγορία των γεγονότων, σχετικά λίγες δηλαδή σε σχέση με την ανεπτυγμένη ταξινόμια.

Όσον αφορά όμως την ανάπτυξη της ταξινόμιας, αυτή απεδέχθη μια ιδιαίτερα επίπονη διαδικασία, όπου παρόλο που καταβλήθηκε μεγάλη προσπάθεια για να μπορέσει να αυτοματοποιηθεί, τελικά κόστισε ιδιαίτερα πολύ σε ανθρωποχρόνο, κυρίως των ιατρών και των γλωσσολόγων της ομάδας.

Η διαδικασία με την οποία υλοποιήθηκε το 2<sup>ο</sup> βήμα ήταν η παρακάτω:

Αρχικά η αρχική ταξινόμια αναπτύχθηκε με βάση τους μονολεκτικούς και πολυλεκτικούς όρους που έχουν συλλεγεί αντίστοιχα στα παραδοτέα Π5 και Π7. Για όσους από τους όρους αυτούς, προέρχονταν από το MeSH η διαδικασία ήταν σχετικά απλή, μια και το MeSH είναι από μόνο του μια ιεραρχία. Για όσους όμως προέρχονταν από άλλες πηγές (σώμα κειμένων, λεξικά, κτλ.) η διαδικασία ήταν ιδιαίτερα επίπονη και χρονοβόρα.

Μετά την ολοκλήρωση της προσθήκης των όρων στην ιεραρχία, ακολούθησε εκ νέου ένας δεύτερος έλεγχος, στον οποίο βοήθησε πολύ τόσο το ΙΑΤΡΟΛΕΞΙTab (βλ. παραπάνω), όσο και η αυτόνομη εφαρμογή απεικόνισης οντολογιών (βλ. συνέχεια) μέσω της οποίας οι χρήστες μπορούσαν να δουν σε γραφικό περιβάλλον την ιεραρχία που είχε σχηματιστεί και να την ελέγξουν.

Το σύνολο των κόμβων (κλάσεων) που περιλαμβάνει η ταξινόμια στην πλήρως ανεπτυγμένη της μορφή, πλησιάζει τους 16.500.

### Γ. Οντολογία Βιοϊατρικών Όρων

Στην υποεργασία αυτή υλοποιήθηκε η οντολογία βιοϊατρικών όρων. Η μεθοδολογία ανάπτυξης της οντολογίας του έργου ΙΑΤΡΟΛΕΞΗ ήταν η ακόλουθη:

Βήμα 1<sup>ο</sup> : Καθορισμός μοντέλου αναπαράστασης και αρχικής ταξινόμιας

Το βήμα αυτό υλοποιήθηκε στα πλαίσια τόσο του παραδοτέου «Π8: Αρχική Ταξινόμια Βιοϊατρικών όρων» όσο και στο παραδοτέο «Π12: Αρχική ταξινόμια Εντός Protégé».

Βήμα 2ο : Κατάρτιση της μακροδομής της οντολογίας

Δεύτερο βήμα για την ανάπτυξη της οντολογίας αποτελεί η κατάρτιση της μακροδομής της, δηλαδή του συνόλου των όρων που αποτελούν πραγματώσεις των βιοϊατρικών εννοιών και απαρτίζουν το λεξιλόγιο της. Το βήμα αυτό υλοποιήθηκε στα πλαίσια των παραδοτέων «Π5: «Μονολεκτικοί όροι βιοϊατρικής» και Π7: «Πολυλεκτικοί όροι βιοϊατρικής».

Κάθε κόμβος/οντότητα της Οντολογίας αντιστοιχεί σε μια βιοϊατρική έννοια (concept), η οποία πραγματώνεται γλωσσικά μέσω ενός ή περισσότερων (συνώνυμων ή ταυτώσεων) βιοϊατρικών όρων.

Για την αποφυγή της αμφισημίας/πολυσημίας, η οποία είναι εγγενής στην έκφραση των εννοιών μέσω των λέξεων, κάθε έννοια εκπροσωπείται μοναδικά από έναν κωδικό (concept ID). Οι κωδικοί εννοιών της Οντολογίας ΙΑΤΡΟΛΕΞΗ είναι σε πλήρη αντιστοιχία προς τους κωδικούς εννοιών του Σημασιολογικού Δικτύου του UMLS και τους κωδικούς εννοιών του MeSH. Έτσι, οι κόμβοι που είναι κοντά στο ριζικό κόμβο της Οντολογίας φέρουν κωδικούς εννοιών από το Σημασιολογικό Δίκτυο του UMLS (κωδικοί που αρχίζουν από Α ή Β, π.χ. Α1.1.3, Β1.3), από το οποίο γίνεται φανερό ότι η αρχική ταξινόμια (η ραχοκοκαλιά) της Οντολογίας βασίστηκε σε αυτό το Δίκτυο. Από ένα βάθος και κάτω, οι κόμβοι της Οντολογίας φέρουν κωδικούς εννοιών από το MeSH (κωδικοί που αρχίζουν από C, π.χ. C0004714), από το οποίο γίνεται φανερό ότι η αρχική οντολογία επεκτάθηκε με όρους του MeSH.

Αναλυτικά, κάθε κόμβος της Οντολογίας φέρει τις εξής πληροφορίες:

- Τον κωδικό έννοιας (μοναδικός για κάθε κόμβο)
- Ένα ή περισσότερα λεκτικά που αντιστοιχούν στην εκφορά της έννοιας μέσω όρου ή όρων της Αγγλικής (όροι που αντλήθηκαν από το UMLS και το MeSH)
- Ένα ή περισσότερα λεκτικά που αντιστοιχούν στην εκφορά της έννοιας μέσω όρου ή όρων της Ελληνικής (όροι που αντλήθηκαν από το σώμα κειμένων, το ελληνικό MeSH και άλλα βιοϊατρικά γλωσσάρια ή λεξικά)

- Ένα λεκτικό στην Αγγλική που αντιστοιχεί στον ορισμό της έννοιας (από το UMLS και το MeSH)

Οι σχέσεις που αναπαρίστανται μέσω της Οντολογίας είναι:

- Η σχέση υπαγωγής (is-a). Κάθε κόμβος B που υπάγεται σε έναν κόμβο A αναπαριστά μια έννοια που αποτελεί *εξειδίκευση* της έννοιας που αναπαριστά ο κόμβος A. Κατ' επέκταση, η έννοια που αναπαριστά ο κόμβος A αποτελεί *γενίκευση* της έννοιας που αναπαριστά οποιοσδήποτε κόμβος B υπάγεται στον A.
- Η σχέση συνωνυμίας. Όταν ένας κόμβος περιέχει περισσότερα του ενός λεκτικά όρων, είτε στην Ελληνική είτε στην Αγγλική, τότε αυτοί οι όροι είναι συνώνυμοι μεταξύ τους.



## 2.2.5 ΕΕ5: Υλοποίηση Εφαρμογών Ιστού

### ΠΕΡΙΓΡΑΦΗ ΕΝΟΤΗΤΑΣ

Στην ενότητα εργασίας αυτή υλοποιήθηκε οι εφαρμογές ιστού, οι οποίες χρησιμοποιώντας τις υποδομές που έχουν αναπτυχθεί στις προηγούμενες ΕΕ, υλοποιούν και επιδεικνύουν τα χρηστικά αποτελέσματα του έργου.

Αναλυτικά σε σχέση με κάθε υποεργασία της ενότητας αυτής:

#### **A. Περιηγητής Οντολογίας**

Μια από τις εφαρμογές ιστού, είναι ο «Περιηγητής Οντολογίας» μέσω του οποίου ο χρήστης του Web μπορεί να περιηγηθεί εντός της οντολογίας των βιοϊατρικών όρων που αναπτύχθηκε στο ΙΑΤΡΟΛΕΞΗ (βλέπε παραδοτέο Π13. Οντολογία Βιοϊατρικών Όρων). Ο περιηγητής βρίσκεται στο ιστοχώρο του έργου <http://www.iatrolexi.gr/> και είναι άμεσα διαθέσιμος στους χρήστες του διαδικτύου.

Οι δυνατότητες που παρέχει ο Περιηγητής είναι οι εξής:

- Απεικόνιση των όρων της οντολογίας σε δεντρική μορφή και πλοήγηση σε αυτήν.
- Αναζήτηση όρων της οντολογίας και απεικόνιση της θέσης τους στην ιεραρχία της οντολογίας.
- Απεικόνιση της αγγλικής μετάφρασης κάθε όρου.
- Απεικόνιση του μονοπατιού κάθε όρου.
- Απεικόνιση της περιγραφής κάθε όρου.
- Εύρεση του επιλεγμένου όρου στη συλλογή κειμένων (corpus).
- Εύρεση του επιλεγμένου όρου στο διαδίκτυο.

Η εφαρμογή του περιηγητή συνδυάζει ένα σύνολο από τα αποτελέσματα του έργου και παρουσιάζει την εφαρμογή αυτών στην πράξη σε μια ολοκληρωμένη εφαρμογή ιστού. Βασικό κομμάτι του περιηγητή αποτελεί η ίδια η οντολογία των βιοϊατρικών όρων. Τέλος χρησιμοποιείται η μηχανή αναζήτησης (βλέπε παραδοτέο Π15) και η εφαρμογή του «Συλλογέα Λεξιλογικών Συνάψεων» (Concordancer) (βλέπε παραδοτέο Π4) ώστε να δοθεί η δυνατότητα στον χρήστη να βρει σε ποια κείμενα από τη συλλογή κειμένων (corpus) εμφανίζεται ο όρος που έχει επιλέξει και να διαβάσει το κείμενο που ίσως τον ενδιαφέρει.

#### **B. Μηχανή Αναζήτησης**

Κατά τη διάρκεια της υλοποίησης έγινε κατανοητό, ότι το μεγάλο πλεονέκτημα που μπορεί να δώσει η διαδικασία επεξεργασίας που εφαρμόσαμε στο έργο, έγκειται στην ευρετηρίαση (indexing) των κειμένων. Η Ευρετηρίαση κατά τη διάρκεια της προεπεξεργασίας που υφίστανται τα κείμενα. Για την περίπτωση της συλλογής που έχουμε, η προεπεξεργασία



αυτή έχει πραγματοποιηθεί ήδη. Το αποτέλεσμα είναι τα κείμενα να έχουν αποκτήσει ένα πλήθος μεταδεδομένων σε διάφορα επίπεδα (μορφολογικό, συντακτικό και σημασιολογικό). Με τη βοήθεια των μεταδεδομένων αυτών, μπορεί πλέον ο χρήστης να κάνει εξειδικευμένες και «έξυπνες» ερωτήσεις και να λαμβάνει τα σχετικά αποτελέσματα.

Για παράδειγμα:

- Ένα ερώτημα με δύο ή περισσότερες λέξεις ερμηνεύεται από το σύστημα ως "πολυλεκτικός όρος αναζήτησης": θα αναζητηθούν συνάψεις που να περιέχουν όλες αυτές τις λέξεις με την ίδια σειρά που εμφανίζονται στο ερώτημα.
- Εναλλακτικά στη θέση των λέξεων ή συμπληρωματικά μαζί με τις λέξεις μπορούν να αναζητηθούν συμβολικές σταθερές που εκπροσωπούν τιμές μορφοσυντακτικών χαρακτηριστικών των λέξεων. Για παράδειγμα:
  - το ερώτημα ιατρικός ΟΥΣ θα επιστρέψει όλες τις συνάψεις όπου η λέξη *ιατρικός* ακολουθείται από κάποιο ουσιαστικό
  - το ερώτημα ΟΥΣ ΟΥΣ/ΓΕΝ θα επιστρέψει όλες τις συνάψεις όπου ουσιαστικό ακολουθείται από ουσιαστικό σε γενική
  - το ερώτημα ΕΠΘ ασθένεια/ΠΛΗ (συνδυασμένο με λημματοποίηση) θα επιστρέψει όλες τις συνάψεις όπου κάποιο επίθετο ακολουθείται από τύπους του πληθυντικού της λέξης *ασθένεια*.

Μια συμβολική σταθερά συνδυάζεται με μια λέξη ή με άλλες συμβολικές σταθερές μέσω του χαρακτήρα '/', ο οποίος ερμηνεύεται από τη μηχανή αναζήτησης ως και. Για παράδειγμα, με το ερώτημα ασθένεια/ΠΛΗ αναζητούμε συνάψεις που α) περιέχουν τη λέξη *ασθένεια* και β) η λέξη είναι στον πληθυντικό αριθμό, με το ερώτημα ΟΥΣ/ΓΕΝ αναζητούμε συνάψεις λέξεων που α) είναι ουσιαστικά και β) είναι σε γενική πτώση. Οι συμβολικές σταθερές που μπορούν να χρησιμοποιηθούν στα ερωτήματα είναι οι εξής:

- Επιστημονικό Πεδίο/ΙΑΤΡ (βιο/ΙΑΤΡΙκή)
- Μέρος του Λόγου: ΕΠΘ (ΕΠίθετο), ΕΠΡ (ΕΠίρρημα), ΜΤΧ (ΜεΤοΧή), ΟΥΣ (ΟΥΣιαστικό), ΡΗΜ (ΡΗΜα), ΑΡΘ (ΑΡΘρο), ΠΡΘ (ΠΡόΘεση), ΣΝΔ (ΣύΝδεσμος), ΑΝΤ (ΑΝΤωνυμία), ΚΤΗ (ΚΤΗτική Αντωνυμία)
- Γένος: ΑΡΣ (ΑΡΣενικό), ΘΗΛ (ΘΗΛυκό), ΟΥΔ (ΟΥΔέτερο)
- Αριθμός: ΕΝΙ (ΕΝικός), ΠΛΗ (ΠΛΗθυντικός)
- Πτώση: ΟΝΟ (ΟΝΟμαστική), ΓΕΝ (ΓΕΝική), ΑΙΤ (ΑΙΤιατική) ΚΛΗ (ΚΛΗτική)
- Φωνή: ΕΝΡ (ΕΝεΡγητική), ΠΑΘ (ΠΑΘητική)
- Χρόνος: ΕΝΣ (ΕΝεΣτώτας), ΑΟΡ (ΑΟΡΙστος), ΠΡΤ (ΠαΡαΤατικός) ΠΡΚ (ΠαΡαΚείμενος)
- Έγκλιση: ΟΡΙ (ΟΡΙστική), ΥΠΟ (ΥΠΟτακτική), ΠΡΣ (ΠΡοΣτακτική)
- Πρόσωπο: ΠΡ1 (1<sup>ο</sup> Πρόσωπο), ΠΡ2 (2<sup>ο</sup> Πρόσωπο), ΠΡ3 (3<sup>ο</sup> Πρόσωπο)

Τέλος να σημειωθεί ότι κατά την αναζήτηση γίνεται και «**Λημματοποίηση**»: Μπορεί κάποιος να χρησιμοποιήσει αυτή την δυνατότητα αν θέλει να βρει λεξιλογικές συνάψεις ανεξάρτητα από τη μορφολογία των λέξεων του ερωτήματός σας.

Τα αποτελέσματα της αναζήτησης, παρουσιάζονται με την μορφή που δίνει τα αποτελέσματά του και το «google». Με αυτό τον τρόπο χρήστης εξοικειώνεται πιο εύκολα και γρήγορα με το περιβάλλον.

### Γ. Ορθογραφικός διορθωτής ιστού

Με δεδομένη την εμπειρία των δύο φορέων σε εργαλεία ελέγχου της γλώσσας, σε συνδυασμό με τον εμπλουτισμό του μορφολογικού λεξικού που έγινε στο πλαίσιο του έργου, η υλοποίηση ενός ορθογραφικού διορθωτή ιστού δεν ήταν κάτι ιδιαίτερα πολύπλοκο.

Υλοποιήθηκε λοιπόν ένα εργαλείο το οποίο μπορεί να διορθώσει είτε ένα κείμενο είτε απλές λέξεις.

Η λειτουργικότητα του συγκεκριμένου εργαλείου είναι ο έλεγχος της ορθογραφίας βιοϊατρικών όρων ή όρων της κοινής ελληνικής γλώσσας.

Υπάρχει η δυνατότητα διαφορετικών επιπέδων διόρθωσης.

#### Επίπεδο Διόρθωσης

- 1: Φωνητική διόρθωση. Βρίσκει λέξεις που "ακούγονται" το ίδιο.
- 2: Επιφανειακή διόρθωση. Αγνοεί τον τόνο και τα κεφαλαία.
- 3: Απλή διόρθωση. Κατά την αναζήτηση στο λεξικό εκτελεί αντικατάσταση γραμμάτων, "παρεμβολή", κ.α.
- 4: Πλήρης διόρθωση. Συνδυασμός όλων των παραπάνω διορθώσεων.

### Δ. Μορφοσυντακτικός και Σημασιολογικός Σχολιαστής

Μορφοσυντακτικός και Σημασιολογικός Σχολιαστής, είναι ένα λογισμικό που επισυνάπτει μορφοσυντακτικά (από το Μορφολογικό Λεξικό) και σημασιολογικά (από την Οντολογία) σχόλια σε λέξεις/στοιχεία ενός κειμένου. Ο σχολιασμός πραγματοποιείται με τη χρήση επισημειώσεων εκφρασμένων ως XML tags.

Το εργαλείο αυτό είναι ίσως και το πιο ενδιαφέρον, αλλά και αυτό που εκμεταλλεύεται την υποδομή που υλοποιήθηκε στο πλαίσιο του έργου. Στην ουσία διατρέχει το κείμενο εισόδου, χωρίζει (αναγνωρίζει) μία μία τις λέξεις (προαπαθεί με βάση τους κανόνες που έχουν φτιαχτεί να αναγνωρίσει και πολυλεκτικούς όρους, και στη συνέχεια, εισάγει κάποια μεταδεδομένα σε κάθε όρο, είτε μορφολογικά (π.χ. μέρος του λόγου, πτώση, κλπ), είτε σημασιολογικά (αν είναι βιοϊατρικός όρος, κάποια ιδιότητά του, κλπ).

Οπτικά αυτό παρουσιάζεται με διαφορετικά χρώματα σε κάθε αναγνωρισμένο λεκτικό στοιχείο, ενώ επιπρόσθετα ο χρήστης μπορεί να εκμεταλλευτεί και την xml μορφή που παράγεται.

### Ε. Τεκμηρίωση εφαρμογών ιστού

Στην υποεργασία αυτή εκπονήθηκε το εγχειρίδιο χρήσης των εργαλείων που περιγράφονται παραπάνω. Πρόκειται ουσιαστικά για ένα εγχειρίδιο που για κάθε ξεχωριστό εργαλείο περιλαμβάνει οδηγίες για την εγκατάστασή του, τον τρόπο χρήσης του, καθώς και οθόνες από το user interface του.

## 2.2.6 ΕΕ6: Διάχυση Αποτελεσμάτων

### ΠΕΡΙΓΡΑΦΗ ΕΝΟΤΗΤΑΣ

Στην ενότητα εργασίας αυτή υλοποιήθηκαν οι απαραίτητες δράσεις για τη δημοσιότητα και τη διάχυση των αποτελεσμάτων του έργου.

Αναλυτικά σε σχέση με κάθε υποεργασία της ενότητας αυτής:

#### **A. Υλοποίηση και συντήρηση project site**

Ο ιστοχώρος του έργου, υλοποιήθηκε και «ανέβηκε» στο διαδίκτυο ήδη από τον 3<sup>ο</sup> μήνα από την έναρξη του έργου. Αρχικά είχε τη διεύθυνση: [www.iatrolexi.cti.gr](http://www.iatrolexi.cti.gr) και στη συνέχεια κατοχυρώθηκε το όνομα που έχει πλέον και σήμερα: [www.iatrolexi.gr](http://www.iatrolexi.gr).

Η λειτουργικότητα του ιστοχώρου είχε από την αρχή επιλεγεί να έχει κάποια επίπεδα και χρήσεις:

1. **Παρουσίαση του έργου:** υπάρχει η περιγραφή, ο σκοπός καθώς και άλλες λεπτομέρειες σε σχέση με το έργο (ανάδοχος, συνεργαζόμενος φορέας, κλπ).
2. **Παρουσίαση των αποτελεσμάτων του έργου:** δημοσιεύονται τα παραδοτέα αλλά και τα εργαλεία που υλοποιούνται/υλοποιήθηκαν στο πλαίσιο του έργου.
3. **Παρουσίαση επιστημονικών ανακοινώσεων:** δημοσιεύονται οι ανακοινώσεις – επιστημονικές εργασίες που έγιναν στο πλαίσιο του έργου.
4. **Ανακοινώσεις - νέα:** αναρτάται ότι νέο υπάρχει σε σχέση με την εξέλιξη του έργου.
5. **Επικοινωνία:** Υπάρχει η δυνατότητα συζήτησης - επικοινωνίας τόσο για το ευρύ κοινό όσο και σε ιδιωτικό χώρο για τα μέλη της ομάδας του έργου.

Με βάση τα παραπάνω, από την αρχή της λειτουργίας του ο ιστοχώρος λειτούργησε ως σημείο αναφοράς για το έργο τόσο εσωτερικά (εντός της ομάδας έργου) όσο και εξωτερικά (προς κάθε ενδιαφερόμενο).

#### **B. Ανακοινώσεις επιστημονικών εργασιών σε συνέδρια / περιοδικά**

Στο πλαίσιο του έργου έγιναν μια σειρά από επιστημονικές ανακοινώσεις οι οποίες φαίνονται παρακάτω. Επισημαίνεται ότι υπερκαλύφθηκε ο αρχικός στόχος που είχε τεθεί για 6 ανακοινώσεις.

[1] Α. Βαγγελάτος, Γ. Ορφανός, Χ. Τσαλίδης, Χρ. Καλαμαρά: *Ανάπτυξη Οντολογίας Βιοϊατρικών Όρων*, [8ο Πανελλήνιο Επιστημονικό Συνέδριο Management Υπηρεσιών Υγείας και κοινωνικής Φροντίδας \(http://www.mediforce.gr/\)](http://www.mediforce.gr/), Ερέτρια, Οκτώβριος 2006.

- [2] M. Pantazara, E. Mantzari, A. Vagelatos, Ch. Kalamara, A. Iordanidou: *Development of a Greek biomedical corpus*, [11th Panhellenic Conference on Informatics \(PCI 2007\)](http://pci2007.upatras.gr/), (http://pci2007.upatras.gr/), Patras, Greece, May 2007.
- [3] A. Vagelatos, E. Mantzari, M. Pantazara, V. Papapanagiotou, Ch. Kalamara: *Implementation of a Biomedical Ontology for the Greek Language*, [An International Conference on Information Communication Technologies in Health \(ICICTH 2007 - http://www.ineag.gr/ICICTH/\)](http://www.ineag.gr/ICICTH/), Samos, Greece, July 2007.
- [4] Ch. Tsalidis, E. Mantzari, M. Pantazara, Ch. Diolis, A. Vagelatos: *Developing a Greek biomedical corpus for text mining*, [Corpus Linguistics Conference 2007 \(CL 2007\)](http://www.corpus.bham.ac.uk/conference2007/), (http://www.corpus.bham.ac.uk/conference2007/), Birmingham, UK, July 2007.
- [5] A. Vagelatos, E. Mantzari, G. Orphanos, Ch. Tsalidis, Ch. Kalamara, Ch. Diolis: *Implementing the NLP Infrastructure for Greek Biomedical Data Mining*, International Workshop A COMMON NATURAL LANGUAGE PROCESSING PARADIGM FOR BALKAN LANGUAGES, [RANLP 2007](http://www.lml.bas.bg/ranlp2007/) Conference (http://www.lml.bas.bg/ranlp2007/), Borovets, Bulgaria, September 2007.
- [6] Α. Βαγγελάτος, Γ. Ορφανός, Μ. Πανταζάρα, Χ. Διολής, Χρ. Καλαμαρά: *ΙΑΤΡΟΛΕΞΗ: Υποδομή γλωσσικής Επεξεργασίας Βιοϊατρικών Κειμένων*, [9ο Πανελλήνιο Επιστημονικό Συνέδριο Management Υπηρεσιών Υγείας και κοινωνικής Φροντίδας](http://www.mediforce.gr/), (http://www.mediforce.gr/) Κέρκυρα, Οκτώβριος 2007.
- [7] A. Vagelatos, E. Mantzari, G. Orphanos, Ch. Tsalidis, M. Pantazara, Ch. Kalamara, Ch. Diolis: *Biomedical data mining for the Greek language*, [MEDNET 2007](http://www.mednet2007.com/) (http://www.mednet2007.com/), Leipzig, Germany, October 2007.
- [8] Α. Ιορδανίδου, Μ. Πανταζάρα, Ε. Μάντζαρη, Α. Βαγγελάτος, Γ. Ορφανός, Β. Παπαπαναγιώτου: *Ζητήματα αναγνώρισης των πολυλεκτικών σύνθετων όρων στον τομέα της βιοϊατρικής*, [6ο Συνέδριο «Ελληνική Γλώσσα και Ορολογία»](http://www.eleto.gr/gr/Conference06.htm) (http://www.eleto.gr/gr/Conference06.htm), Αθήνα, Νοέμβριος 2007.
- [9] Α. Βαγγελάτος, Α. Ιορδανίδου, Ι. Σταματόπουλος, Γ. Ορφανός, Β. Παπαπαναγιώτου: *Εξόρυξη Γνώσης σε Βιοϊατρικά Κείμενα*, [3ο Πανελλήνιο Συνέδριο για τη Διοίκηση, τα Οικονομικά και τις Πολιτικές Υγείας](http://www.healthcongress.gr/), (http://www.healthcongress.gr/), Αθήνα, Δεκέμβριος 2007.

### Γ. Ημερίδα διάχυσης αποτελεσμάτων

Για τις ανάγκες διάχυσης των αποτελεσμάτων του έργου, διοργανώθηκε ημερίδα. Η Ημερίδα 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 εποικοδομητική συζήτηση και ανταλλαγή απόψεων πάνω σε θέματα κοινού ενδιαφέροντος.

#### **Δ. Τελική Αναφορά Έργου**

Για την καλύτερη τεκμηρίωση του έργου, εκπονήθηκε η παρούσα αναφορά η οποία συνοψίζει όσα έγιναν καθ' όλη τη διάρκεια υλοποίησης.

### 3 ΣΥΝΟΛΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΠΑΡΑΔΟΤΕΑ ΤΟΥ ΕΡΓΟΥ

#### Α) Βασικά Αποτελέσματα του Έργου

Τα αποτελέσματα του έργου μπορούν να χωριστούν/κατηγοριοποιηθούν στις ακόλουθες κατηγορίες:

1. *Τυπικά αποτελέσματα:* Στην κατηγορία αυτή περιλαμβάνονται τα παραδοτέα του έργου. Πιο συγκεκριμένα, μέχρι σήμερα τα παραδοτέα που έχουν ολοκληρωθεί και παραδοθεί είναι τα ακόλουθα:

- Π1. Σχέδιο Συγκρότησης Σώματος Βιοϊατρικών Κειμένων (τεχνική αναφορά).
- Π2. Μοντέλο Αναπαράστασης Οντολογίας (τεχνική αναφορά).
- Π3. Σύστημα Διαχείρισης Σώματος Κειμένων και Εξαγωγής Ορολογίας (λογισμικό).
- Π4. Εγχειρίδιο Χρήσης (του λογισμικού Π3).
- Π5. Μονολεκτικοί Όροι Βιοϊατρικής (αρχείο δεδομένων).
- Π6. Κανόνες Αναγνώρισης Πολυλεκτικών Όρων Βιοϊατρικής (αρχείο δεδομένων + αναφορά τεκμηρίωσης).
- Π7. Πολυλεκτικοί Όροι Βιοϊατρικής (αρχείο δεδομένων).
- Π8. Αρχική Ταξινόμια Βιοϊατρικών Όρων (αρχείο δεδομένων).
- Π9. Συλλογή βιοϊατρικών κειμένων (αρχείο δεδομένων).
- Π10. Τεκμηρίωση συλλογής βιοϊατρικών κειμένων (τεχνική αναφορά).
- Π12. Αρχική Ταξινόμια εντός Protégé (αρχείο δεδομένων).
- Π11. Προσαρμοσμένο Protégé (λογισμικό).
- Π13. Οντολογία Βιοϊατρικών όρων (αρχείο δεδομένων).
- Π14. Περιηγητής Οντολογίας (λογισμικό).
- Π15. Μηχανή Αναζήτησης (λογισμικό).
- Π16. Ορθογραφικός Διορθωτής Ιστού (λογισμικό).
- Π17. Μορφοσυντακτικός και Σημασιολογικός Σχολιαστής (λογισμικό).
- Π18. Τεκμηρίωση Εφαρμογών Ιστού (τεχνική αναφορά).
- Π19. Υλοποίηση και συντήρηση project site.
- Π20. Ανακοινώσεις επιστημονικών εργασιών σε συνέδρια/περιοδικά (τεχνική αναφορά)
- Π21. Ημερίδα Διάχυσης Αποτελεσμάτων.



## Π22. Τελική Αναφορά Έργου (τεχνική αναφορά).

Τα παραδοτέα αυτά έχουν κατατεθεί στην ΓΓΕΤ, μπορεί κανείς όμως να τα βρει και στην ιστοσελίδα του έργου ([www.iatrolexi.gr](http://www.iatrolexi.gr)).

**2. Αυτόνομα Αποτελέσματα:** Στην κατηγορία αυτή περιλαμβάνονται ολοκληρωμένα τμήματα του έργου που νοούνται ως βασικές υποδομές και μπορούν να τύχουν ήδη ερευνητικής εκμετάλλευσης από την ευρύτερη επιστημονική κοινότητα. Σε τίτλους τα ακόλουθα:

- Π3. *Σύστημα Διαχείρισης Σώματος Κειμένων και Εξαγωγής Ορολογίας* (λογισμικό). Το λογισμικό αυτό μπορεί να χρησιμοποιηθεί από όποιον θέλει να κάνει κάτι αντίστοιχο, είτε στο πεδίο της βιοϊατρικής είτε σε άλλο πεδίο.
- Π9. *Συλλογή βιοϊατρικών κειμένων* (αρχείο). Η συλλογή βιοϊατρικών κειμένων, μπορεί να χρησιμοποιηθεί από οποιονδήποτε ερευνητή επιθυμεί να μελετήσει φαινόμενα γλωσσικά ή άλλα, στο συγκεκριμένο πεδίο.
- Π11. *Προσαρμοσμένο Protégé* (λογισμικό). Το εργαλείο Protégé, προσαρμοσμένο ώστε να διαχειρίζεται τα Ελληνικά.
- Π11. Προσαρμοσμένο Protégé (λογισμικό).
- Π13. *Οντολογία Βιοϊατρικών όρων* (αρχείο δεδομένων).
- Π14. *Περιηγητής Οντολογίας* (λογισμικό).
- Π15. *Μηχανή Αναζήτησης* (λογισμικό).
- Π16. *Ορθογραφικός Διορθωτής Ιστού* (λογισμικό).
- Π17. *Μορφοσυντακτικός και Σημασιολογικός Σχολιαστής* (λογισμικό).

**3. Θεωρητικά Αποτελέσματα:** Στην κατηγορία αυτή περιλαμβάνονται μελέτες που έγιναν σε σχέση με την εξέλιξη ή τη συνέχεια του έργου και μπορούν να χρησιμοποιηθούν από κάθε ερευνητή/επιστήμονα που θέλει να προσεγγίσει θεωρητικά τα αντίστοιχα πεδία:

- Π1. *Σχέδιο Συγκρότησης Σώματος Βιοϊατρικών Κειμένων* (τεχνική αναφορά)
- Π2. *Μοντέλο Αναπαράστασης Οντολογίας* (τεχνική αναφορά)
- Π5. *Μονολεκτικοί Όροι Βιοϊατρικής* (αρχείο δεδομένων)
- Π6. *Κανόνες Αναγνώρισης Πολυλεκτικών Όρων Βιοϊατρικής* (αρχείο δεδομένων + αναφορά τεκμηρίωσης)
- Π7. *Πολυλεκτικοί Όροι Βιοϊατρικής* (αρχείο δεδομένων)
- Π8. *Αρχική Ταξινόμια Βιοϊατρικών Όρων* (αρχείο δεδομένων)
- Π13. *Οντολογία Βιοϊατρικών όρων* (αρχείο δεδομένων)
- Π20. *Ανακοινώσεις επιστημονικών εργασιών σε συνέδρια/περιοδικά* (τεχνική αναφορά)

## Β) Πιθανή μελλοντική αξιοποίηση αποτελεσμάτων

Τα αποτελέσματα του έργου, μπορούν να χρησιμοποιηθούν/αξιοποιηθούν με ποικίλους τρόπους:



- *Αυτόνομα αποτελέσματα.* Μπορούν να χρησιμοποιηθούν άμεσα από την ερευνητική κοινότητα για την περαιτέρω προαγωγή της έρευνας.
- *Θεωρητικά αποτελέσματα.* Όπως και τα παραπάνω, δύνανται να χρησιμοποιηθούν ως βάση/αναφορά για οποιαδήποτε σχετική ερευνητική προσπάθεια. Στην κατηγορία αυτή εντάσσονται και οι ερευνητικές εργασίες που έχουν παραχθεί ως σήμερα.
- *Τυπικά Παραδοτέα.* Δεδομένου ότι όλα τα παραδοτέα του έργου, είναι στη διάθεση του ερευνητή/επιστήμονα, μπορούν να αξιοποιηθούν συνολικά από την ερευνητική κοινότητα κατά το δοκούν.

Η σύμπραξη των φορέων που υλοποιεί το έργο, σκέφτεται να αξιοποιήσει άμεσα τους όρους που έχουν συλλεγεί για την παραγωγή ενός εξειδικευμένου ελεγκτή ορθογραφίας για Ελληνικά βιοϊατρικά κείμενα. Η προσπάθεια αυτή χρειάζεται επιπλέον επένδυση σε επίπεδο επεξεργασίας (γλωσσικής) των όρων, ώστε το αποτέλεσμα να είναι αξιόλογο και να γίνει δεκτό από τους χρήστες.

Γενικότερα, ο βασικός στόχος του έργου για την υλοποίηση υποδομής που θα βελτιώσει τη δυνατότητα διαχείρισης βιοϊατρικής πληροφορίας, έχει πληθώρα ωφελούμενων είτε στη μορφή που θα προκύψει στα πλαίσια του ΙΑΤΡΟΛΕΞΗ, είτε με βελτιώσεις προσθήκες που θα μπορεί κάποιος να κάνει με βάση πιο εξειδικευμένη κατεύθυνση.

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 βρει ευκολότερα πιο αξιόπιστη και χρήσιμη πληροφορία για θέματα υγείας που τον απασχολούν. Επιπλέον, φορείς υγείας που παράγουν πληροφορία ιατρικού περιεχομένου (νοσοκομεία, δημόσια διοίκηση, κτλ) 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, κτλ).

### Γ) Κατάλογος δημοσιεύσεων

Στη συνέχεια παρουσιάζεται ο κατάλογος των ερευνητικών εργασιών παρουσιάσεων/δημοσιεύσεων που έχουν πραγματοποιηθεί, ή έχουν γίνει αποδεκτές. Το σχετικό υλικό μπορεί να αναζητηθεί και στο <http://www.iatrolexi.gr/iatrolexi/dhmosieuseis.html>.

ΗΜΕΡΟΜΗΝΙΑ	ΤΙΤΛΟΣ ΑΝΑΚΟΙΝΩΣΗΣ	ΣΥΝΕΔΡΙΟ	Web Site Συνεδρίου
12-15 Δεκεμβρίου 2007	Εξόρυξη Γνώσης σε Βιοϊατρικά Κείμενα	3ο Πανελλήνιο Συνέδριο για τη Διοίκηση, τα Οικονομικά και τις Πολιτικές Υγείας (2007), Αθήνα, Ελλάδα	<a href="http://www.healthcongress.gr/">http://www.healthcongress.gr/</a>
01-03 Νοεμβρίου 2007	Ζητήματα αναγνώρισης των πολυλεκτικών σύνθετων όρων στον τομέα της βιοϊατρικής	6ο Συνέδριο «Ελληνική Γλώσσα και Ορολογία» (2007), Αθήνα, Ελλάδα	<a href="http://www.eleto.gr/gr/Conference06.htm">http://www.eleto.gr/gr/Conference06.htm</a>
07-10 Οκτωβρίου 2007	Biomedical Data Mining for the Greek language	MEDNET 2007, Leipzig, Germany	<a href="http://mednet2007.com/content/">http://mednet2007.com/content/</a>
04 - 07 Οκτωβρίου 2007	ΙΑΤΡΟΛΕΞΗ: Υποδομή γλωσσικής Επεξεργασίας Βιοϊατρικών Κειμένων	9ο Πανελλήνιο Συνέδριο Management Υπηρεσιών Υγείας και Κοινωνικής Φροντίδας(2007), Κέρκυρα, Ελλάδα	<a href="http://www.mediforce.gr/">http://www.mediforce.gr/</a>
27 - 29 Σεπτεμβρίου 2007	Implementing the NLP Infrastructure for Greek Biomedical Data Mining	27-29 Σεπτεμβρίου 2007 Implementing the NLP Infrastructure for Greek Biomedical Data Mining	<a href="http://www.lml.bas.bg/ranlp2007">http://www.lml.bas.bg/ranlp2007</a>
27-30 Ιουλίου 2007	Developing a Greek biomedical corpus for text mining	Corpus Linguistics Conference 2007 (CL 2007), Birmingham, England	<a href="http://www.corpus.bham.ac.uk/conference2007/">http://www.corpus.bham.ac.uk/conference2007/</a>
05-07 Ιουλίου 2007	Implementation of a Biomedical Ontology for the Greek Language	5th ICICTH International Conference on Information Communication Technologies in Health (2007), Samos, Greece	<a href="http://www.ineag.gr/ICICTH/index.html">http://www.ineag.gr/ICICTH/index.html</a>
18-20 Μαΐου 2007	Development of a Greek biomedical corpus	11th Panhellenic Conference on Informatics (PCI 2007), Patras, Greece	<a href="http://pci2007.upatras.gr/">http://pci2007.upatras.gr/</a>
28-30 Σεπτεμβρίου 2006	Ανάπτυξη Οντολογίας Βιοϊατρικών Όρων	8ο Πανελλήνιο Συνέδριο Management Υπηρεσιών Υγείας και Κοινωνικής Φροντίδας, Ερέτρια, Ελλάδα	<a href="http://www.mediforce.gr/">http://www.mediforce.gr/</a>

## 4 ΣΧΟΛΙΑ - ΠΡΟΒΛΗΜΑΤΑ - ΠΑΡΑΤΗΡΗΣΕΙΣ

### 4.1 ΤΕΧΝΟΛΟΓΙΑ / ΤΕΧΝΟΓΝΩΣΙΑ ΠΟΥ ΑΠΟΚΤΗΘΗΚΕ ΣΤΟ ΠΛΑΙΣΙΟ ΤΟΥ ΕΡΓΟΥ

Κατά τη διάρκεια υλοποίησης του έργου, αποκτήθηκε συγκεκριμένη τεχνογνωσία. Η τεχνογνωσία αυτή ανά τομέα μπορεί να περιγραφεί ως εξής:

1. **Οντολογίες:** Μεθοδολογίες/περιβάλλοντα κατασκευής οντολογιών (Protégé, OWL, κτλ).
2. **Σώματα Κειμένων:** Μεθοδολογίες Συλλογής ΣΚ, Εφαρμογές (π.χ. ορολογική έρευνα), Σχολιασμός/Κωδικοποίηση (TEI, CES, Eagle, κλπ), Σώματα Βιοϊατρικών κειμένων (Medpost, Medlex, κτλ)
3. **Ιατρικές οντολογίες/Κωδικοποιήσεις:** UMLS, FMA, Galen, Mesh, SNMED, ICD10, κτλ.
4. **Γλωσσική Τεχνολογία:** Περιγραφή/κωδικοποίηση πολυλεκτικών βιοϊατρικών όρων, Επεξεργασία Βιοϊατρικών όρων
5. **Τεχνικά Ζητήματα:** MySQL, Java, PHP, Apache

Όσον αφορά προϊόντα ή παραδοτέα για τα οποία να έχουν κατοχυρωθεί κάποια licences, πνευματικά δικαιώματα, ή οτιδήποτε άλλο, ΔΕΝ έχει γίνει κάτι τέτοιο.

### 4.2 ΣΥΝΕΡΓΑΣΙΑ ΦΟΡΕΩΝ (ΟΙΚΟΝΟΜΙΚΟ ΚΑΙ ΦΥΣΙΚΟ ΑΝΤΙΚΕΙΜΕΝΟ)

Η συνεργασία των φορέων στην υλοποίηση του έργου, τόσο στο φυσικό όσο και στο οικονομικό αντικείμενο, εξελίχθηκε ομαλά. Οι βασικοί λόγοι για τους οποίους πιστεύουμε ότι έγινε αυτό είναι οι παρακάτω:

- Πολύ σαφώς καθορισμένοι στόχοι από την αρχή του έργου.
- Λίγοι (μόνον δύο) συνεργαζόμενοι φορείς (χωρίς κανέναν «τρίτο» φορέα).
- Εμπειρία στα θέματα που εστίασε το έργο και από τις δύο πλευρές.
- Εμπειρία των υπευθύνων έργου και του επιστημονικού υπευθύνου.
- Εμπειρία των δύο φορέων σε αντίστοιχου τύπου έργα.

Θα πρέπει να τονιστεί ότι στη διάρκεια του έργου, πραγματοποιήθηκαν επιπλέον εργασίες που δεν είχαν προδιαγραφεί, επειδή διαπιστώθηκε ότι ήταν απαραίτητες. Αυτές καταγράφονται παρακάτω:

1. Συλλέκτης Λεξικών Συνάψεων (βλ. Παραδοτέα Π3 και Π4) (<http://www.iatrolexi.gr/iatrolexi/webtools/concordancer/index.html>).

(Ζητήθηκε από τους ειδικούς (γλωσσολόγους και ιατρούς) για την υποβοήθησή τους κατά την επεξεργασία των όρων.)

2. Εφαρμογή Απεικόνισης Οντολογιών (βλ. Παραδοτέο Π12).

(Ζητήθηκε από τους ιατρούς για την καλύτερη (οπτική) παρουσίαση της οντολογίας

– ταξινομίας κατά τη διάρκεια της υλοποίησής της – επιπρόσθετα από το Protégé).

3. Επιπλέον εισαγωγή όρων στην οντολογία, ακόμα και μετά την ολοκλήρωση του σχετικού παραδοτέου (Π5), έως σχεδόν και το τέλος του έργου.

## 5 ΔΙΑΧΕΙΡΙΣΗ ΕΡΓΟΥ

### 5.1 ΣΥΝΑΝΤΗΣΕΙΣ

Κατά τη διάρκεια του έργου πραγματοποιήθηκε σειρά συναντήσεων της ομάδας έργου, στις οποίες συζητούνταν όλα τα θέματα υλοποίησης (χρονοδιάγραμμα, επόμενα βήματα, προβλήματα, κατανομή εργασιών, κλπ).

Οι συναντήσεις γίνονταν πότε στον ένα φορέα (Neurosoft A.E.) και πότε στον άλλο (EAITY). Στις συναντήσεις αυτές τηρούσαμε πρακτικά με βασικό σκοπό την τεκμηρίωση του έργου αλλά και την καταγραφή των αποφάσεων που λαμβάναμε. Τα πρακτικά αυτά ακολουθούν στο παράρτημα.

### 5.1 ΠΑΡΑΔΟΤΕΑ

Στη διάρκεια υλοποίησης του έργου, υπήρχε η υποχρέωση της ολοκλήρωσης και κατάθεσης στην ΓΓΕΤ σειρά παραδοτέων. Τα παραδοτέα αυτά χωρίζονται σε:

- *Τυπικά παραδοτέα υλοποίησης:* Με βάση το Τεχνικό Παράρτημα του έργου είχαν περιγραφεί 22 παραδοτέα. Τα παραδοτέα αυτά ολοκληρώθηκαν, κατατέθηκαν στην ΓΓΕΤ, ενώ έχουν αναρτηθεί και στο site του έργου.
- *Διαχειριστικά παραδοτέα:* Με βάση τους κανόνες υλοποίησης, συμπληρώνονταν και κατατείθονταν στην ΓΓΕΤ οι «τριμηνιαίες εκθέσεις προόδου» καθώς και οι «ετήσια» και «τελική» αναφορά προόδου (τόσο του φυσικού όσο και του οικονομικού αντικειμένου).

## 6 ΠΑΡΑΡΤΗΜΑΤΑ

### 6.1 ΠΡΑΚΤΙΚΑ ΣΥΝΑΝΤΗΣΕΩΝ

9/1/2006

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 1<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** ΕΑΙΤΥ / Εργαστήριο «Ήρων»

---

## ΣΥΝΟΨΗ

Έγινε η πρώτη συνάντηση του έργου «Ιατρολέξη» που εγκρίθηκε από τη ΓΓΕΤ. Το έργο θα υλοποιηθεί από τους συνεργαζόμενους φορείς Neurosoft Α.Ε και ΕΑΙΤΥ.

Στη συνάντηση έγινε παρουσίαση του έργου και συζήτηση για τις διακριτές ενότητες εργασίας.

Επίσης παρουσιάστηκε η καταγραφή που έχει γίνει σε σχέση με διαδικτυακά ιατρικά περιοδικά.

Τέλος παρουσιάστηκε εν τάχει η δουλειά που έχει γίνει μέχρι σήμερα για την ιατρική ορολογία (συλλογή λημμάτων).

### • ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ

Η επόμενη συνάντηση καθορίστηκε για την Πέμπτη 23/2 και ώρα 11.00 στα γραφεία της Neurosoft. Εφόσον (πράγμα μάλλον απίθανο) έχει προχωρήσει η υπογραφή της σύμβασης μέχρι τότε, η συνάντηση θα γίνει πιο νωρίς.

### • ΕΝΕΡΓΕΙΕΣ

Μέχρι την επόμενη συνάντηση αποφασίστηκα να γίνουν:

#### Οι ιατροί:

- Θα πρέπει να συλλέξουν και να προτείνουν μια σειρά από ιατρικά βιβλία τα οποία θα μπορέσουν να αποτελέσουν μια καλή βάση για αποδελτίωση (εφόσον αποφασιστεί ότι χρειάζεται και η αποδελτίωση έντυπου υλικού).
- Να προτείνουν (με την ίδια λογική) κάποια λεξικά (ελληνοαγγλικά - MESH).
- Να δώσουν μια ματιά στη συλλογή των διαδικτυακών ιατρικών περιοδικών για να αξιολογήσουν το επίπεδό τους.

### **Οι πληροφορικοί:**

- Να ψάξουν αν υπάρχουν ηλεκτρονικά λεξικά ιατρικών όρων (Magenda)?
- Να ετοιμάσουν μια παρουσίαση για το Protégé.
- Να δηλώσουν το όνομα [www.iatrolexi.gr](http://www.iatrolexi.gr) που θα χρησιμοποιηθεί για το site του έργου.

### **Οι γλωσσολόγοι:**

- Να σχηματίσουν τελική γνώμη για το τι έχει γίνει μέχρι σήμερα σχετικά με τα λήμματα (είναι κλιμένα? Που βρίσκονται?).

### **Όλοι:**

- Να αναζητηθεί βιβλιογραφία σε σχέση με βιο-ιατρικές οντολογίες.
- Να αναζητηθεί βιβλιογραφία σε σχέση με τη συλλογή σώματος κειμένων (standards?).
- Να αναζητηθούν τυχόν συνεργάτες (εκδοτικοί οίκοι?, Ιατρικοί σύλλογοι?) που θα μπορούσαν να μας διαθέσουν ιατρικά κείμενα που δε βρίσκονται σήμερα στο διαδίκτυο.



16/3/2006

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 2<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** EAITY / Εργαστήριο «Ήρων»

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η δεύτερη συνάντηση του έργου «Ιατρολέξη». Ενημερώθηκαν οι παρευρισκόμενοι ότι την προηγούμενη εβδομάδα, ετοιμάστηκε και κατατέθηκε στη ΓΓΕΤ ο φάκελος του έργου (γι αυτό και χρειάστηκε να αναβληθεί η συνάντησή μας) με τα δικαιολογητικά και το Τεχνικό Παράρτημα που ζητήθηκε για να προχωρήσει η διαδικασία της υπογραφής της σύμβασης.

Στη συνάντηση συζητήθηκαν αρχικά οι γενικές αρχές του έργου. Στη συνέχεια ο Γ.Ο. μας παρουσίασε κάποια πράγματα σε σχέση με τις οντολογίες γενικότερα και τις προσπάθειες που γίνονται διεθνώς.

Σχολιάστηκε το γεγονός ότι υπάρχει μεγάλο εύρος δημοσιευμένων εργασιών στον ερευνητικό χώρο για σχετικά θέματα και κυρίως για τον βιοϊατρικό χώρο και θα πρέπει κάποια στιγμή, αφενός να φτιάξουμε ένα index και αφετέρου να φτιαχτεί μια ηλεκτρονική βιβλιοθήκη που θα φυλάσσονται και θα είναι εύκολο να αναζητηθούν (ίσως στα πλαίσια του site).

Έπειτα έγινε εκτενής συζήτηση για τη μεθοδολογία που θα πρέπει να ακολουθηθεί για την υλοποίηση του έργου.

Τέλος θα πρέπει να σημειωθεί ότι η πρώτη ενότητα εργασίας: **Σχεδιασμός Σώματος Βιοϊατρικών Κειμένων και Οντολογίας**, έχει χρονοδιάγραμμα 6 μήνες και στο πλαίσιο αυτό θα πρέπει να οριστικοποιηθεί ο σχεδιασμός του έργου. Έχει δε δύο βασικά παραδοτέα:

A) Σχέδιο Συγκρότησης Σώματος Βιοϊατρικών Κειμένων (τεχνική Αναφορά)

B) Μοντέλο Αναπαράστασης Οντολογίας (τεχνική αναφορά)

- **ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ**

Η επόμενη συνάντηση καθορίστηκε για την Πέμπτη 30/3 και ώρα 11.00 στα γραφεία της Neurosoft.

- **ΕΝΕΡΓΕΙΕΣ**

Μέχρι την επόμενη συνάντηση αποφασίστηκε να γίνουν:

- Αξιολόγηση και έναρξη συλλογής των ιατρικών κειμένων από το διαδίκτυο
- Διερεύνηση δυνατότητας απόκτησης του τρίτομου βιβλίου Παθολογίας του Harrison
- Ετοιμασία και επίδειξη Protégé με ένα πραγματικό παράδειγμα
- Επίδειξη UMLS ή άλλης βιοϊατρικής οντολογίας
- Θα πάρουμε το DMS4 <http://dsmivtr.org/> (ελληνική μετάφραση της κωδικοποίησης των ψυχιατρικών παθήσεων)
- Να αγοράσουμε κάποια βιβλία (όπως την Ελληνική μετάφραση του MESH, κάποιο λεξικό, κλπ).
- Α συνεχίσουμε την έρευνα στη βιβλιογραφία με σκοπό να προετοιμαστούμε για:
  - Τον τρόπο σχεδιασμού
  - Τις δυνατές χρήσεις της οντολογίας
  - Για το σχεδιασμό του σκελετού της οντολογίας
  - Τον καθορισμό των σχέσεων που θα πρέπει να συμπεριληφθούν

Η βιβλιογραφία που ελέγχεται θα πρέπει με κάποιο τρόπο να διαχέεται και στους υπολοίπους.

Τέλος κάποια πράγματα που δεν υλοποιήθηκαν από την προηγούμενη συνάντηση:

#### Οι ιατροί:

- Να προτείνουν (με την ίδια λογική) κάποια λεξικά (ελληνοαγγλικά - MESH).

#### Οι πληροφορικοί:

- Να ψάξουν αν υπάρχουν ηλεκτρονικά λεξικά ιατρικών όρων (Magenda)?
- Να ετοιμάσουν μια παρουσίαση για το Protégé.

#### Όλοι:

- Να αναζητηθεί βιβλιογραφία σε σχέση με βιο-ιατρικές οντολογίες.
- Να αναζητηθεί βιβλιογραφία σε σχέση με τη συλλογή σώματος κειμένων (standards?).
- Να αναζητηθούν τυχόν συνεργάτες (εκδοτικοί οίκοι?, Ιατρικοί σύλλογοι?) που θα μπορούσαν να μας διαθέσουν ιατρικά κείμενα που δε βρίσκονται σήμερα στο διαδίκτυο.

26/4/2006

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 3<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** EAITY / Εργαστήριο «Ήρων»

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η τρίτη συνάντηση του έργου «Ιατρολέξη».

**(Σήμερα έλαβα ενημέρωση από την ΓΓΕΤ, ότι η απόφαση έναρξης υλοποίησης υπεγράφη στις 26/4 – την ημέρα που κάναμε τη συνάντηση! Συνεπώς το έργο θεωρείται ότι έχει ξεκινήσει!)**

Συζητήθηκε πάλι το ζήτημα του «που βασίζεται κανείς για να ξεκινήσει την υλοποίηση μιας οντολογίας). Επιπλέον διαπιστώθηκε ότι επειδή δεν έχει ξεκινήσει το έργο, δεν υπάρχει η απαιτούμενη αφοσίωση αλλά παρατηρείται δικαίως μια αναβλητική διάθεση (πράγμα που προφανώς με βάση το παραπάνω δεν δικαιολογείται να συνεχίσει να υφίσταται).

Αποφασίστηκε μέχρι την επόμενη συνάντηση να πραγματοποιηθούν πάση θυσία οι ενέργειες που αποφασίστηκαν.

- **ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ**

Η επόμενη συνάντηση καθορίστηκε για την Τετάρτη 17/5 και ώρα 12.00..

- **ΕΝΕΡΓΕΙΕΣ**

Μέχρι την επόμενη συνάντηση αποφασίστηκε να γίνουν (προσθέτω και όσα θέματα δεν ολοκληρώθηκαν από την προηγούμενη συνάντηση):

- Αξιολόγηση και έναρξη συλλογής των ιατρικών κειμένων από το διαδίκτυο
- Διερεύνηση δυνατότητας απόκτησης του τρίτομου βιβλίου Παθολογίας του Harrison
- Ετοιμασία και επίδειξη Protégé με ένα πραγματικό παράδειγμα
- Επίδειξη UMLS ή άλλης βιοϊατρικής οντολογίας
- Θα πάρουμε το DMS4 <http://dsmivtr.org/> (ελληνική μετάφραση της κωδικοποίησης των ψυχιατρικών παθήσεων)
- Να αγοράσουμε κάποια βιβλία (όπως την Ελληνική μετάφραση του MESH, κάποιο λεξικό, κλπ).
- Α συνεχίσουμε την έρευνα στη βιβλιογραφία με σκοπό να προετοιμαστούμε για:

- Τον τρόπο σχεδιασμού
  - Τις δυνατές χρήσεις της οντολογίας
  - Για το σχεδιασμό του σκελετού της οντολογίας
  - Τον καθορισμό των σχέσεων που θα πρέπει να συμπεριληφθούν
- 
- Να αναζητηθεί βιβλιογραφία σε σχέση με βιο-ιατρικές οντολογίες.
  - Να αναζητηθεί βιβλιογραφία σε σχέση με τη συλλογή σώματος κειμένων (standards?).
  - Να αναζητηθούν τυχόν συνεργάτες (εκδοτικοί οίκοι?, Ιατρικοί σύλλογοι?) που θα μπορούσαν να μας διαθέσουν ιατρικά κείμενα που δε βρίσκονται σήμερα στο διαδίκτυο.

17/5/2006

**ΣΥΓΓΡ:** A. Βαγγελάτος

**ΑΝΤΙΚ:** 4<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** EAITY / Εργαστήριο «Ήρων»

---

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η τέταρτη συνάντηση του έργου «Ιατρολέξη».

Αρχικά έγινε ενημέρωση ότι λάβαμε την υπουργική απόφαση που ορίζει και επίσημα την έναρξη του έργου στις 26 Απριλίου 2006. Συνεπώς ξεκινάμε και τυπικά! Για το σκοπό αυτό είπαμε να ορίσουμε τακτικές συναντήσεις κάθε 15 μέρες. Η ημέρα που προτάθηκε είναι η Παρασκευή. Ο τόπος θα ορίζεται κατά περίπτωση. Η επόμενη συνάντηση ορίστηκε για την Παρασκευή 2 Ιουνίου και ώρα 10.00 στο EAITY (Ακταίου 11, Θησείο – 500 μέτρα από τον ομώνυμο σταθμό του τρένου).

Σχετικά με το έργο:

A) Έγινε ενημέρωση ότι έγινε επαφή με τις εκδόσεις Παρισιάνου, όπου καταρχήν υπήρξε επιφυλακτικότητα στο να δοθεί ο Harisson σε ηλεκτρονική μορφή. Τους άφησα κάποιο υλικό και είπαν ότι θα επικοινωνήσουν μαζί μου αφού ενημερωθεί η κα Παρισιάνου. Ενημερωτικά η ελληνική έκδοση του Harisson έχει 50 σελίδες index.

B) Αγοράστηκε (και το περιεργαστήκαμε) το MeSH Βιοιατρικό Λεξικό. Είναι έκδοση του 1991 και περιέχει την μετάφραση του Medical Subject Headings. Δεν περιέχει την δενδρική δομή αλλά ευελπιστούμε ότι μπορούμε να την φτιάξουμε εφόσον το αντιστοιχήσουμε στην αγγλική έκδοση. Την έκδοση αυτή σε δενδρική δομή μπορεί κάποιος να την δει στο: <http://www.nlm.nih.gov/mesh/2006/MeSHtree.A.html>. Τα ζητήματα που θα πρέπει να λυθούν είναι: α) να βρούμε αν υπάρχει η ηλεκτρονική του μορφή – για το σκοπό αυτό θα έρθω σε επικοινωνία με την ΙΑΤΡΟΤΕΚ και β) η μετάφραση είναι του 1991 όπως ειπώθηκε παραπάνω: το αγγλικό MeSH όμως, ανανεώνεται κάθε χρόνο, πράγμα που μπορεί να μας προβληματίσει λίγο. Κατά τα άλλα είναι ίσως μια καλή βάση (με δεδομένο ότι έχει και την ιεραρχική δομή) για να ξεκινήσει κανείς.

Γ) Με δεδομένο ότι έχει ξεκινήσει το έργο και συνεπώς η πρώτη, ενότητα εργασίας: *Σχεδιασμός σώματος Βιοϊατρικών κειμένων και Οντολογίας*, αποφασίστηκε να ξεκινήσουμε να ετοιμάζουμε τα παραδοτέα που είναι δύο τεχνικές αναφορές: 1) Σχέδιο συγκρότησης Σώματος Βιοϊατρικών κειμένων (ITY) και 2) Μοντέλο Αναπαράστασης Οντολογίας (Neuro). Στόχος είναι να φτιάξουμε ένα outline που να το συζητήσουμε την επόμενη φορά και σιγά - σιγά να το γεμίζουμε. Θυμίζω ότι το deadline είναι 6 μήνες από την έναρξη του έργου. Στο πλαίσιο αυτό η Έλενα θα μας στείλει μια σχετική αναφορά που είχε κάνει για το ΕΚΤ και μπορεί να είναι χρήσιμη. Επίσης θα μας στείλει την terminologia anatomica (?).

Δ) Ενημερωθήκαμε ότι το UMLS έχει 30-35 σχέσεις που συνδέουν τις έννοιες. Θα πρέπει να τις δούμε και να βρούμε ποιες από αυτές θα χρησιμοποιήσουμε και μεις.

Ε) Ξεκίνησε η συλλογή κειμένων από τα περιοδικά που είχαμε δει σε παλαιότερη συνάντηση στο Internet.

- **ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ**

Παρασκευή 2/6 και ώρα 10.00.

- **ΕΝΕΡΓΕΙΕΣ**

Μέχρι την επόμενη συνάντηση αποφασίστηκε να γίνουν τα παραπάνω συν κάποια παλαιότερα που δεν έχουν υλοποιηθεί:

- Ετοιμασία και επίδειξη Protégé με ένα πραγματικό παράδειγμα
- Επίδειξη UMLS ή άλλης βιοϊατρικής οντολογίας
- Θα πάρουμε το DMS4 <http://dsmivtr.org/> (ελληνική μετάφραση της κωδικοποίησης των ψυχιατρικών παθήσεων)
- Α συνεχίσουμε την έρευνα στη βιβλιογραφία με σκοπό να προετοιμαστούμε για:
  - Τον τρόπο σχεδιασμού
  - Τις δυνατές χρήσεις της οντολογίας
  - Για το σχεδιασμό του σκελετού της οντολογίας
  - Τον καθορισμό των σχέσεων που θα πρέπει να συμπεριληφθούν
- Να αναζητηθεί βιβλιογραφία σε σχέση με βιο-ιατρικές οντολογίες.
- Να αναζητηθεί βιβλιογραφία σε σχέση με τη συλλογή σώματος κειμένων (standards?).

05/6/2006

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 5<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** EAITY / Εργαστήριο «Ήρων»

---

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η πέμπτη συνάντηση του έργου «Ιατρολέξη».

Αρχικά έγινε ενημέρωση για τη συνάντηση που πραγματοποίησε ο Α.Β. με την κα Παρισιάνου. Η τελευταία, δέχτηκε να συνεργαστεί και συμφώνησε να μας δώσει κείμενα για χρήση στο έργο. Ζήτησε όμως κάποιο συμφωνητικό που θα την διασφαλίζει τυπικά. Συμφωνήσαμε να της στείλουμε ένα συμφωνητικό και να το συζητήσουμε.

Στη συνέχεια ο Γ.Ο. έκανε μια πρόταση για την διαδικασία κατασκευής της οντολογίας η οποία βασίζεται: α) στο semantic network της UMLS και β) στην κατασκευή λημματολογίου με τον τρόπο που θα επιλέξουμε. Στο τελευταίο η πιο πρόσφορη λύση φαίνεται να είναι το Αγγλικό MeSH και με κάποιο τρόπο η γρήγορη μετάφρασή του (είτε με τη χρήση του λεξικού από τις εκδόσεις Παρισιάνου είτε με άλλο τρόπο).

Για την επόμενη συνάντηση στο παραπάνω πλαίσιο, καλό θα ήταν να έχουμε τους σημασιολογικούς τύπους και τις σχέσεις του UMLS τυπωμένους (και ίσως μεταφρασμένους).

- **ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ**

Παρασκευή 22/6 και ώρα 11.00 στα γραφεία του EAITY.

- **ΕΝΕΡΓΕΙΕΣ**

Μέχρι την επόμενη συνάντηση αποφασίστηκε να γίνουν τα παραπάνω συν κάποια παλαιότερα που δεν έχουν υλοποιηθεί:

- Τυπωμένους τους σημασιολογικούς τύπους και τις σχέσεις του UMLS (Έλενα).
- Επίδειξη UMLS (Άρης)
- Να διερευνήσουμε σχετικά με το έργο οντολογία καρδιολογίας που φαίνεται να «τρέχει» στην Ήπειρο.

Δεν έχουν υλοποιηθεί:

- Ετοιμασία και επίδειξη Protégé με ένα πραγματικό παράδειγμα.
- Θα πάρουμε το DSM4 <http://dsmivtr.org/> (ελληνική μετάφραση της κωδικοποίησης των ψυχιατρικών παθήσεων)
- Να συνεχίσουμε την έρευνα στη βιβλιογραφία με σκοπό να προετοιμαστούμε για:
  - Τον τρόπο σχεδιασμού
  - Τις δυνατές χρήσεις της οντολογίας
  - Για το σχεδιασμό του σκελετού της οντολογίας
  - Τον καθορισμό των σχέσεων που θα πρέπει να συμπεριληφθούν



22/6/2006

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 6<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** EAITY / Εργαστήριο «Ήρων»

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η έκτη συνάντηση του έργου «Ιατρολέξη».

Αρχικά έγινε μια επίδειξη της Ιστοσελίδας του έργου που ξεκίνησε να υλοποιεί το ΙΤΥ. Σύντομα θα τη δημοσιεύσουμε για να σχολιαστεί από όλους.

Στη συνέχεια είδαμε πάλι την παρουσίαση του UMLS που είχαμε δει και σε προηγούμενη συνάντηση και συζητήσαμε λίγο τα θέματα του semantic network και πώς μπορούμε αυτά να τα μεταφέρουμε και για το δικό μας έργο.

Επειτα έγινε παρουσίαση του ίδιου του UMLS το οποίο είχε εγκατασταθεί για το σκοπό αυτό σε ένα PC του ΙΤΥ.

Τέλος συζητήσαμε ότι φαίνεται να είναι η καλύτερη τακτική να αρχίσουμε από τη μεταφορά του semantic network του UMLS στα Ελληνικά και στα πλαίσια του δικού μας έργου. Αυτό θα μας βοηθήσει να έχουμε μια συμβατότητα με το ίδιο το UMLS. Οσον αφορά το λεμματολόγιο, φαίνεται το Αγγλικό MeSH να είναι η καλύτερη περίπτωση. Θα πρέπει όμως να βρούμε ένα εύκολο τρόπο να το μεταφράσουμε.

### • ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ

Πέμπτη 6/7 και ώρα 11.00 στα γραφεία του EAITY. Επειδή κάποιοι μπορούν νωρίτερα και κάποιοι αργότερα, βάζουμε κατ' αρχήν αυτή την ώρα και βλέπουμε.

### • ΕΝΕΡΓΕΙΕΣ

Μέχρι την επόμενη συνάντηση αποφασίστηκε να γίνουν τα παραπάνω συν κάποια παλαιότερα που δεν έχουν υλοποιηθεί:

- Να διερευνήσουμε σχετικά με το έργο οντολογία καρδιολογίας που φαίνεται να «τρέχει» στην Ήπειρο.

Δεν έχουν υλοποιηθεί:

- Ετοιμασία και επίδειξη Protégé με ένα πραγματικό παράδειγμα.
- Θα πάρουμε το DSM4 <http://dsmivtr.org/> (ελληνική μετάφραση της κωδικοποίησης των ψυχιατρικών παθήσεων)

- Να συνεχίσουμε την έρευνα στη βιβλιογραφία με σκοπό να προετοιμαστούμε για:
  - Τον τρόπο σχεδιασμού
  - Τις δυνατές χρήσεις της οντολογίας
  - Για το σχεδιασμό του σκελετού της οντολογίας
  - Τον καθορισμό των σχέσεων που θα πρέπει να συμπεριληφθούν

6/7/2006

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 7<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** EAITY / Εργαστήριο «Ήρων»

---

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η έβδομη συνάντηση του έργου «Ιατρολέξη».

Συζητήσαμε τις σχέσεις και τις οντότητες του σημασιολογικού δικτύου του UMLS. Είχαν μεταφραστεί από τους συναδέλφους και ανέλαβαν οι γιατροί να τα δουν (αφού τους σταλούν μέσα στις επόμενες μέρες).

Στη συνέχεια είδαμε το την κωδικοποίηση νόσων – διαγνώσεων (ICD10) που μετά από μια προσπάθεια που έγινε την προηγούμενη εβδομάδα την έχουμε τώρα και στην Αγγλική και στην Ελληνική γλώσσα (σε ηλεκτρονική μορφή). Διαπιστώθηκε ότι το εύρος της είναι μάλλον στενό (λόγω του σκοπού της) για τη δουλειά που την θέλουμε μείς. Έτσι συμφωνήθηκε να παραμείνουμε στο MeSH ως ορολογία και να δούμε με ποιο τρόπο θα μπορέσουμε να έχουμε γρήγορα ΑγγλοΕλληνική μορφή στον υπολογιστή.

Στη συνέχεια έγινε επίδειξη του Protégé και των δυνατοτήτων του, μια και είχαν εισαχθεί οι σχέσεις του UMLS. Φάνηκε ότι κάποια πράγματα το Protégé τα αντιμετωπίζει λίγο διαφορετικά. Επίσης φάνηκε ότι έχει πολλές δυνατότητες, αλλά θα πρέπει εμείς να διερευνήσουμε ποιες και πώς μπορούμε να χρησιμοποιήσουμε. Επίσης κοινή είναι η πεποίθηση ότι θα πρέπει αν πάρουμε όσο περισσότερη πληροφορία από το UMLS μπορούμε με αυτόματο τρόπο.

- **ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ**

Πέμπτη 20/7 και ώρα 11.00 στα γραφεία του EAITY.

- **ΕΝΕΡΓΕΙΕΣ**

Μέχρι την επόμενη συνάντηση αποφασίστηκε να γίνουν τα παραπάνω συν κάποια παλαιότερα που δεν έχουν υλοποιηθεί:

- Να σταλούν οι μεταφράσεις των σχέσεων και των οντοτήτων του UMLS
- Να γίνει επίσκεψη στην ΓΓΕΤ.

- Να διερευνήσουμε σχετικά με το έργο οντολογία καρδιολογίας που φαίνεται να «τρέχει» στην Ήπειρο.
- Θα πάρουμε το DSM4 <http://dsmivtr.org/> (ελληνική μετάφραση της κωδικοποίησης των ψυχιατρικών παθήσεων)
- Να συνεχίσουμε την έρευνα στη βιβλιογραφία με σκοπό να προετοιμαστούμε για:
  - Τον τρόπο σχεδιασμού
  - Τις δυνατές χρήσεις της οντολογίας
  - Για το σχεδιασμό του σκελετού της οντολογίας
  - Τον καθορισμό των σχέσεων που θα πρέπει να συμπεριληφθούν

20/7/2006

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 8<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** EAITY / Εργαστήριο «Ήρων»

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η όγδοη συνάντηση του έργου «Ιατρολέξη».

Αρχικά έγινε ενημέρωση σχετικά με την επίσκεψη των Α. Β. και Γ.Ο. στην ΓΓΕΤ. Συζήτησαν με τον υπεύθυνο από την πλευρά της ΓΓΕΤ κ Πρέκα. Ο τελευταίος τους είπε ότι σχετικά με την προκαταβολή, βλέπει να καθυστερεί μέχρι τον Σεπτέμβριο (αν όχι παραπάνω...). Όσον αφορά το έργο, θα ήθελε τα παραδοτέα να είναι ποιοτικά και πάντα σε σχέση με αυτά που έχουν περιγραφεί στην πρόταση τόσο όσον αφορά στο περιεχόμενο όσο και στο χρόνο παράδοσης. Οι τριμηνιαίες εκθέσεις, θα πρέπει να είναι στην ώρα τους, εκτός της πρώτης που θα δοθεί μαζί με τη δεύτερη. Επίσης όποια τροποποίηση θα θέλαμε να γίνει στο έργο (που δεν πρέπει να είναι πολλές!) θα πρέπει να είναι επαρκώς τεκμηριωμένη.

Στη συνέχεια συζητήθηκε το θέμα της ορολογίας. Όπως είχαμε αποφασίσει από την προηγούμενη συνάντηση, θα υιοθετήσουμε το MeSH.

Τα επόμενα βήματα θα πρέπει να είναι:

1. Έλεγχος Όρων. Οι ιατροί θα αναλάβουν να «δουν» τους όρους του MeSH, με στόχο να ελέγξουν την ορολογία (ορθότητα, τονισμό, κλπ).
2. Έλεγχος ορθογραφίας - κλίσης. Οι γλωσσολόγοι θα αναλάβουν όσες λέξεις δεν έχουν κληθεί αυτόματα να τις ελέγξουν και να τους αναθέσουν τη σωστή κλίση.

Για να οριστικοποιηθούν τα παραπάνω βήματα, αποφασίστηκε η πρώτη εργασία να γίνει άμεσα ώστε μέχρι την επόμενη συνάντηση να έχουμε αποτελέσματα σχετικά με το μέγεθος της δουλειάς.

- **ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ**

Δευτέρα 31/7 και ώρα 12.00 στα γραφεία του EAITY.

- **ΕΝΕΡΓΕΙΕΣ**

Μέχρι την επόμενη συνάντηση αποφασίστηκε να γίνουν τα παραπάνω συν κάποια παλαιότερα που δεν έχουν υλοποιηθεί:

- Να διερευνήσουμε σχετικά με το έργο οντολογία καρδιολογίας που φαίνεται να «τρέχει» στην Ήπειρο.
- Θα πάρουμε το DSM4 <http://dsmivtr.org/> (ελληνική μετάφραση της κωδικοποίησης των ψυχιατρικών παθήσεων)
- Να συνεχίσουμε την έρευνα στη βιβλιογραφία με σκοπό να προετοιμαστούμε για:
  - Τον τρόπο σχεδιασμού
  - Τις δυνατές χρήσεις της οντολογίας
  - Για το σχεδιασμό του σκελετού της οντολογίας
  - Τον καθορισμό των σχέσεων που θα πρέπει να συμπεριληφθούν

31/7/2006

**ΣΥΓΓΡ:** Α. Βαγγελάτος  
**ΑΝΤΙΚ:** 9<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»  
**ΑΝΑΦ:** EAITY / Εργαστήριο «Ήρων»

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η ένατη συνάντηση του έργου «Ιατρολέξη».

Το βασικό θέμα συζήτησης ήταν η επεξεργασία των όρων του MeSH.

Οι όροι του MeSH επεξεργάστηκαν από γλωσσολόγους και τεχνικούς και θα είναι έτοιμα μέχρι το τέλος της εβδομάδας για επιμέλεια από τους γιατρούς. Είναι περίπου 17.000 όροι και θα μοιραστούν στους δύο γιατρούς ισόποσα σε σχέση όμως με τη θεματική κατηγορία που ανήκουν. Για το σκοπό αυτό θα τυπωθούν οι κατηγορίες και οι όροι που περιέχουν και οι ιατροί θα επιλέξουν ποιες είναι ποιο κοντά στην εμπειρία τους.

Η επιμέλεια θα γίνει σε excel με τη λογική ότι όπου υπάρχει λάθος και είναι δυνατόν, θα γράφεται σε νέα στήλη το σωστό. Σε διαφορετική περίπτωση θα καθοριστούν κάποιες σημάνσεις (π.χ. έλεγχος τόνου, έλεγχος ορθογραφίας, άγνωστη λέξη, κλπ.) για περαιτέρω έλεγχο.

Από μια πρώτη προσέγγιση, φαίνεται ότι ένα τρίμηνο είναι αρκετό για την ολοκλήρωση της πρώτης αυτής επεξεργασίας. Για να γίνει καλύτερος παραλληλισμός της εργασίας, θα πρέπει να υπάρχουν ενδιάμεσα παραδοτέα (ανά μηνά?) ώστε να προωθούνται στους γλωσσολόγους για την γλωσσική επιμέλεια.

Στη συνέχεια συζητήσαμε για πιθανές άλλες χρήσεις ή προεκτάσεις του έργου. Κατ' αρχή ενημερώθηκαν όσοι δεν το γνώριζαν ότι στο τεχνικό δελτίο έχουμε μιλήσει για 6 (και όχι 9 όπως λανθασμένα λέχθηκε) ανακοινώσεις που θα προκύψουν στα πλαίσια του έργου. Ήδη έχει σταλεί ένα abstract στο «8<sup>ο</sup> Πανελλήνιο Συνέδριο Management Υπηρεσιών Υγείας» ([www.mediforce.gr](http://www.mediforce.gr)) που περιγράφει την αρχική ιδέα του έργου. Από δω και πέρα θα πρέπει τα μέλη της ομάδας έργου να συζητούν και να αποφασίζουν άλλες προσπάθειες. Πιθανοί στόχοι θα πρέπει να είναι (στον Ελληνικό χώρο) το Πανελλήνιο Ιατρικό Συνέδριο, Γλωσσολογικά Συνέδρια, ή Συνέδρια Πληροφορικής. Όσον αφορά άλλες προεκτάσεις του έργου, μια καλή επέκταση θα ήταν η πλήρης μετάφραση του MeSH 2006, κάτι που θα μας έβαζε στην ίσια γραμμή με κάποιες άλλες χώρες που έχουν μεταφράσει και κυρίως ΣΥΝΤΗΡΟΥΝ την μετάφραση του MeSH. Για παράδειγμα την αντίστοιχη Γερμανική έκδοση

την υποστηρίζει το German Institute of Medical Documentation and Information [http://www.dimdi.de/static/en/klasi/mesh\\_uml/mesh/meshallg.htm](http://www.dimdi.de/static/en/klasi/mesh_uml/mesh/meshallg.htm) (μην αναρωτιέστε - δεν έχουμε τέτοιο στην Ελλάδα!).

Βλέπε και <http://www.nlm.nih.gov/pubs/factsheets/intlmedlars.html>

Τέλος όσοι δεν είχαν γραφτεί στο forum γράφτηκαν και περιμένουμε μεγαλύτερη συμμετοχή μέσω του forum και του site. Στο πλαίσιο αυτό και εφόσον δεν υπάρχουν παρατηρήσεις για το site, ίσως θα πρέπει να το «διαφημίσουμε» λίγο!

#### • ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ

Επειδή η επόμενη συνάντηση είναι πλέον το Σεπτέμβρη και μέχρι τότε ίσως αλλάξουν κάποια πράγματα, βάζουμε ως αρχική ημερομηνία συνάντησης την Τετάρτη 13/9 και ώρα 11.00 στα γραφεία του EAITY και αν κάτι αλλάξει το συζητάμε.

#### • ΕΝΕΡΓΕΙΕΣ

Μέχρι την επόμενη συνάντηση αποφασίστηκε να γίνουν τα παρακάτω:

- Αποστολή του υλικού προς επεξεργασία στους γιατρούς. Οι τελευταίοι θα το σπουδάσουν και θα θέσουν πιθανά προβλήματα, διευκρινίσεις ή άλλο στην επόμενη συνάντηση.
- Να διερευνήσουμε σχετικά με το έργο οντολογία καρδιολογίας που φαίνεται να «τρέχει» στην Ήπειρο.
- Θα πρέπει να αρχίσουν να ετοιμάζονται τα δύο πρώτα παραδοτέα.

1. Σχέδιο Συγκρότησης Σώματος Βιοϊατρικών Κειμένων (τεχνική Αναφορά)

2. Μοντέλο Αναπαράστασης Οντολογίας (τεχνική αναφορά)

- Αυτά θα πρέπει να παραδοθούν στην ΓΓΕΤ στο τέλος Οκτωβρίου. Για το σκοπό αυτό θα πρέπει μέσα στο Σεπτέμβριο να γίνει ένα αρχικό σχεδιάγραμμα του τι θα περιέχουν.
- Θα πάρουμε το DSM4 <http://dsmivtr.org/> (ελληνική μετάφραση της κωδικοποίησης των ψυχιατρικών παθήσεων)
- Να συνεχίσουμε την έρευνα στη βιβλιογραφία με σκοπό να προετοιμαστούμε για:
  - Τον τρόπο σχεδιασμού
  - Τις δυνατές χρήσεις της οντολογίας
  - Για το σχεδιασμό του σκελετού της οντολογίας
  - Τον καθορισμό των σχέσεων που θα πρέπει να συμπεριληφθούν



27/9/2006

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 10<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** ΕΑΙΤΥ / Εργαστήριο «Ήρων»

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η δέκατη συνάντηση του έργου «Ιατρολέξη».

Το βασικό θέμα συζήτησης ήταν η ετοιμασία των δύο πρώτων παραδοτέων:

1. Σχέδιο Συγκρότησης Σώματος Βιοϊατρικών Κειμένων (τεχνική Αναφορά) (ΕΑΙΤΥ)
2. Μοντέλο Αναπαράστασης Οντολογίας (τεχνική αναφορά) (Neuro)

Συζητήθηκαν τα περιεχόμενα που είχαν ήδη ετοιμαστεί από τους δύο φορείς και έγινε κατανομή εργασιών.

Πιο συγκεκριμένα σε σχέση με το πρώτο παραδοτέο, τα κεφάλαια που έχουν σχέση με ιστορικά θέματα και corpus linguistics, καθώς και ζητήματα σχεδιασμού σώματος κειμένων, τα ανέλαβε η Α.Ι. Τα υπόλοιπα θα τα ετοιμάσει το ΙΤΥ που έχει και το κύριο βάρος του ζητήματος της συλλογής του corpus.

Σε σχέση με το δεύτερο παραδοτέο, η Ε.Μ. ανέλαβε τα αρχικά και εισαγωγικά κεφάλαια, ενώ το κομμάτι της μεθοδολογίας ανάπτυξης το ανέλαβε ο Χρ.Ο. Στο θέμα των ιατρικών λεξικών / κψδικοποιήσεων ανέλαβε να συνδράμει ο Α.Β.

Στόχος είναι στην επόμενη συνάντηση (25 Οκτωβρίου) να έχουμε σχεδόν έτοιμα τα παραδοτέα ώστε να τα δούμε σε μια τελική μορφή.

Στη συνέχεια συζητήθηκαν κάποια τεχνικά θέματα σε σχέση με τα εργαλεία που θα πρέπει να υλοποιηθούν. Αρχικά το θέμα της αποθήκευσης και annotation του corpus. Προτάθηκε να δούμε και το σύστημα της IBM UIMA. Επίσης ο Χρ. Τ. ανέλαβε να μας στείλει κάποια σχετικά πράγματα από το έργο Parmenides που υλοποίησε η Neuro.

Σε σχέση με τα λοιπά εργαλεία που θα πρέπει να υλοποιήσει η Neuro, το βασικό είναι ο μηχανισμός κλίσης πολυλεκτικών όρων, μια και τα λοιπά υπάρχουν σε μια κάποια μορφή.

Ανεξάρτητα από την παραπάνω συνάντηση, έγινε συζήτηση με τους γιατρούς σε σχέση με την εξέλιξη του έργου τους. Πράγματι προχωρούν στον έλεγχο της μετάφρασης του MeSH

που τους είχε ανατεθεί. Τα προβλήματα που αντιμετωπίζουν και θέτουν για συζήτηση είναι τα παρακάτω:

Α) κάποιοι όροι δεν μεταφράζονται και συναντώνται σήμερα στα Ελληνικά ιατρικά βιβλία με τη λατινική τους μορφή. Τι θα γίνει με αυτούς;

Β) Στις διπλές ή τριπλές λέξεις που μπορεί να έχει ένας όρος, κρατάμε τη σειρά της αγγλικής, ή όχι. (Εάν όχι, χρειάζεται κόμμα??). Προφανώς η αλλαγή της σειράς προκύπτει μόνο λόγω ανάγκης ταξινόμησης ή κατηγοριοποίησης.

Γ) Κάποιοι όροι που γράφονται με διπλό ll στην αγγλική, το MeSH τους μεταφράζει με ένα «λ» ενώ στα βιβλία συναντώνται με διπλό «λ». Ποια είναι η σωστή γραφή;

Δ) Σε κάποιες περιπτώσεις ένας αγγλικός όρος μεταφράζεται με δύο (ή περισσότερους) Ελληνικούς. Σε αυτήν την περίπτωση τους κρατάμε όλους;

- **ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ**

Επειδή η επόμενη συνάντηση καθορίστηκε για τις 25 Οκτωβρίου. Η ακριβής ώρα και ο τόπος μένει να οριστούν.

- **ΕΝΕΡΓΕΙΕΣ**

Μέχρι την επόμενη συνάντηση αποφασίστηκε να γίνουν τα παρακάτω:

- Ετοιμασία των δύο πρώτων παραδοτέων.
- Θα πάρουμε το DSM4 <http://dsmivtr.org/> (ελληνική μετάφραση της κωδικοποίησης των ψυχιατρικών παθήσεων)

25/10/2006

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 11<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** ΕΑΙΤΥ / Εργαστήριο «Ήρων»

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η εντέκατη συνάντηση του έργου «Ιατρολέξη».

Δύο ήταν τα θέματα της ατζέντας:

α) Η συζήτηση των προβλημάτων που αντιμετωπίζουν οι ιατροί στην επεξεργασία των όρων του MesH και

β) Η ετοιμασία των δύο πρώτων παραδοτέων:

1. Σχέδιο Συγκρότησης Σώματος Βιοϊατρικών Κειμένων (τεχνική Αναφορά)
2. Μοντέλο Αναπαράστασης Οντολογίας (τεχνική αναφορά)

Σε σχέση με το α:

- Παρατηρήθηκαν κάποια σοβαρά λάθη στην απόδοση όρων που αποδείχτηκε όμως ότι είναι από την μεταφορά και επεξεργασία του έντυπου λεξικού (έγιναν κάποιες λάθος αντιστοιχίσεις). Συνεπώς τα λάθη αυτά θα πρέπει να διορθώνονται όπου εντοπίζονται σε συνδυασμό με έλεγχο της έντυπης μορφής.
- Υπάρχουν κάποιοι όροι που δεν μεταφράζονται (δίδεται ο ίδιος όρος - συνήθως στα λατινικά – και για την ελληνική απόδοση). Στις περιπτώσεις που πρόκειται για κύριο όνομα θα κρατάμε τον όρο ως έχει. Στις υπόλοιπες περιπτώσεις θα γίνεται φωνητική μεταγραφή.
- Κάποιοι όροι μεταγράφονται όχι φωνητικά (π.χ. δύο «λ» αντί για ένα, κλπ). Στις περιπτώσεις αυτές θα δίδουμε και τη φωνητική μεταγραφή αλλά και την χρησιμοποιούμενη μορφή.
- Σε σχέση με τη σειρά πολυλεκτικών όρων, θα κρατάμε αυτή που υπάρχει.
- Σε περιπτώσεις που δεν υπάρχουν άρθρα (κυρίως γενικές) ενώ χρειάζονται (π.χ. ελικοβακτηρίδιο του πυλωρού) θα προστίθενται.

Όσον αφορά το β:

Έχουμε καθυστερήσει την ολοκλήρωση του παραδοτέου αλλά βάζουμε ως στόχο την ολοκλήρωσή του μέχρι το τέλος του Νοέμβρη.

Στις 15 Νοέμβρη θα προσθέσουμε όσα κομμάτια έχουν ολοκληρωθεί και θα μοιραστεί για σχολιασμό, ώστε να πετύχουμε τον παραπάνω στόχο.

- **ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ**

Επειδή η επόμενη συνάντηση καθορίστηκε για τις 16 Νοεμβρίου και ώρα 10.00. Ο τόπος μένει να οριστεί.

- **ΕΝΕΡΓΕΙΕΣ**

Μέχρι την επόμενη συνάντηση αποφασίστηκε να γίνουν τα παρακάτω:

- Ετοιμασία των δύο πρώτων παραδοτέων.

23/11/2006

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 12<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** EAITY / Εργαστήριο «Ήρων»

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η δωδέκατη συνάντηση του έργου «Ιατρολέξη».

Δύο ήταν τα θέματα της ατζέντας:

α) Η συζήτηση των προβλημάτων που αντιμετωπίζουν οι ιατροί στην επεξεργασία των όρων του MesH και

β) Η ετοιμασία των δύο πρώτων παραδοτέων:

1. Σχέδιο Συγκρότησης Σώματος Βιοϊατρικών Κειμένων (τεχνική Αναφορά)
2. Μοντέλο Αναπαράστασης Οντολογίας (τεχνική αναφορά)

Σε σχέση με το α), υπάρχει μια καθυστέρηση στην εξέλιξη της επιμέλειας των ιατρικών όρων που έχουν αναλάβει να κάνουν οι ιατροί (λόγω και της δυσκολίας που έχουν κάποιοι όροι εκτός της κοινής ορολογίας). Τέθηκε ως καταληκτική ημερομηνία το τέλος του χρόνου ώστε στη συνέχεια να αναλάβουν οι φιλόλογοι την επιμέλεια.

Σε σχέση με το β) έγιναν παρατηρήσεις στις σχεδόν τελικές εκδόσεις των παραδοτέων που είχαν μοιραστεί σε όλους στις 19/11. Με βάση τις παρατηρήσεις αυτές, θα ετοιμαστεί η επόμενη έκδοση μέσα στις επόμενες 10 ημέρες και θα ξανασταλθεί πάλι σε όλους. Στην επόμενη συνάντηση θα γίνουν οι τελικές παρατηρήσεις και μετά τα παραδοτέα θα αποσταλούν στη ΓΓΕΤ.

Τέλος έγινε και συζήτηση σε σχέση με την ορολογία που χρησιμοποιούμε και τη μεταφορά κάποιων όρων στην Ελληνική. Βασικός στόχος είναι να υπάρχει κοινή ορολογία σε όλη την ομάδα.

### • ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ

Επειδή η επόμενη συνάντηση καθορίστηκε για την Πέμπτη 7 Νοεμβρίου και ώρα 10.00. Ο τόπος μένει να οριστεί.

- **ΕΝΕΡΓΕΙΕΣ**

Μέχρι την επόμενη συνάντηση αποφασίστηκε να γίνουν τα παρακάτω:

- Ετοιμασία των τελικών εκδόσεων των δύο πρώτων παραδοτέων.

17/1/2007

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 13<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** EAITY / Εργαστήριο «Ήρων»

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η 13<sup>η</sup> συνάντηση του έργου «Ιατρολέξη». Το βασικό θέμα συζήτησης ήταν η εξέλιξη – συνέχεια του έργου, με δεδομένο την ολοκλήρωση των δύο πρώτων παραδοτέων (Π1 & Π2).

Κατ' αρχή έγινε μια σύνοψη σε σχέση με την παρούσα κατάσταση:

Με ολοκληρωμένη την ΕΕ1, έχουμε 3 ενότητες εργασίας που τρέχουν παράλληλα:

1. ΕΕ2: Υλοποίηση Εργαλείων Διαχείρισης Σώματος Κειμένων και Αυτόματης Εξαγωγής Ορολογίας (με ημερομηνία ολοκλήρωσης: **τέλος Μαρτίου 2007**).
2. ΕΕ3: Συλλογή Βιοϊατρικών κειμένων και εξαγωγή ορολογίας (με ημερομηνία ολοκλήρωσης: **τέλος Φεβρουαρίου 2007**).
3. ΕΕ4: Επεξεργασία Βιοϊατρικών όρων και Δημιουργία Οντολογίας (με ημερομηνία ολοκλήρωσης: **τέλος Ιουνίου 2007**).

Σε σχέση με αυτά, υπάρχουν τα εξής παραδοτέα:

### ΕΕ2:

- Σύστημα Διαχείρισης Σώματος Κειμένων και Εξαγωγής Ορολογίας (λογισμικό)

στο παραδοτέο αυτό υπάρχουν μια σειρά από υποσυστήματα (καταγράφονται σε σχέση με τον φορέα που είναι υπεύθυνος):

- EAITY: Αποθήκη εγγράφων και προσκομιστής εγγράφων
- Neurosoft: Αναγνωριστής στοιχείων, Μορφοσυντακτικός σχολιαστής, Μορφολογικό λεξικό, Αναγνωριστής όρων
- Παράλληλα οι γλωσσολόγοι θα εξετάσουν το «μηχανισμό κλήσης πολυλεκτικών όρων»
- Εγχειρίδιο χρήσης

### ΕΕ3:

- Μονολεκτικοί όροι Βιοϊατρικής (αρχείο δεδομένων) Neurosoft

- Κανόνες αναγνώρισης πολυλεκτικών όρων Βιοϊατρικής (αρχείο δεδομένων και αναφορά τεκμηρίωσης) Γλωσσολόγοι
- Πολυλεκτικοί όροι Βιοϊατρικής (αρχείο δεδομένων) Neurosoft
- Αρχική Ταξινόμια Βιοϊατρικών Όρων (αρχείο δεδομένων) Neurosoft
- Συλλογή Βιοϊατρικών κειμένων (αρχείο) EAITY
- Τεκμηρίωση συλλογής βιοϊατρικών κειμένων (αναφορά) EAITY

#### ΕΕ4:

- Προσαρμοσμένο Protégé (λογισμικό + αναφορά προσαρμογών) Neurosoft
- Αρχική ταξινόμια εντός Protégé (αρχείο δεδομένων)
- Οντολογία Βιοϊατρικών όρων (αρχείο δεδομένων).

Τα άμεσα βήματα προς την ολοκλήρωση των παραδοτέων (ειδικά της ΕΕ3 που λήγει το τέλος Φλεβάρη) είναι:

- α) Εξαγωγή πολυλεκτικών όρων και ώστε οι γλωσσολόγοι να εργαστούν με σκοπό την μοντελοποίησή τους
- β) Ολοκλήρωση του crawler και τεκμηρίωση της διαδικασίας συλλογής
- γ) Μελέτη του UIMA για την αποθήκευση των εγγράφων του corpus
- δ) ολοκλήρωση των επιμέρους εφαρμογών

#### • ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ

Επειδή η επόμενη συνάντηση καθορίστηκε για την Τετάρτη 31 Ιανουαρίου και ώρα 10.00. Ο τόπος μένει να ορισθεί.



08/02/2007

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 14<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** ΕΑΙΤΥ / Εργαστήριο «Ήρων»

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε σήμερα Τρισκοπέμπτη (!) η 14<sup>η</sup> συνάντηση του έργου «Ιατρολέξη».

Το βασικό θέμα συζήτησης ήταν τα παραδοτέα Π5-Π10 που θα πρέπει να ολοκληρωθούν έως το τέλος του Φεβρουαρίου.

Η τρέχουσα ενότητα Εργασίας που έχει τα άμεσα παραδοτέα είναι η:

4. ΕΕ3: Συλλογή Βιοϊατρικών κειμένων και εξαγωγή ορολογίας (με ημερομηνία ολοκλήρωσης: **τέλος Φεβρουαρίου 2007**).

Σε σχέση με αυτήν, έχουμε να ετοιμάσουμε τα εξής παραδοτέα:

- Μονολεκτικοί όροι Βιοϊατρικής (αρχείο δεδομένων) Neurosoft
- Κανόνες αναγνώρισης πολυλεκτικών όρων Βιοϊατρικής (αρχείο δεδομένων και αναφορά τεκμηρίωσης) Γλωσσολόγοι
- Πολυλεκτικοί όροι Βιοϊατρικής (αρχείο δεδομένων) Neurosoft
- Αρχική Ταξινόμια Βιοϊατρικών Όρων (αρχείο δεδομένων) Neurosoft
- Συλλογή Βιοϊατρικών κειμένων (αρχείο) ΕΑΙΤΥ
- Τεκμηρίωση συλλογής βιοϊατρικών κειμένων (αναφορά) ΕΑΙΤΥ

Συμφωνήσαμε ότι ακόμα και σε αυτά που το παραδοτέο είναι ένα αρχείο, θα πρέπει να υπάρχει ένα 5σέλιδο που να το συνοδεύει.

Επίσης καλό είναι τα παραδοτέα να είναι έτοιμα 10 μέρες πριν το τέλος Φλεβάρη, και να τα διαχύσουμε σε όλους ώστε να γίνουν και πιθανές διορθώσεις.

Το επόμενο θέμα ήταν αυτό των πολυλεκτικών όρων. Οι γλωσσολόγοι έχουν βγάλει μια καταρχήν προσέγγιση για τον τρόπο κωδικοποίησής τους. Πρόβλημα υπάρχει με κάποιους χημικούς σύνθετους όρους, καθώς και με κάποιους που περιέχουν αριθμητικά και άλλα σύμβολα.

Στη συνέχεια έγινε ενημέρωση για την εξέλιξη της επιμέλειας των όρων από τους ιατρούς: Υπάρχει σχετική καθυστέρηση μια και ιδιαίτερα οι όροι χημείας παρουσιάζουν αρκετή δυσκολία.

Τέλος συζητήσαμε για το συνέδριο της ΕΛΕΤΟ που θα γίνει τον Ιούνιο και πιθανή συμμετοχή μας. Ειπώθηκε πως το καλύτερο θέμα που θα μπορούσαμε να στείλουμε έχει να κάνει με τους πολυλεκτικούς όρους που «δουλεύονται» αυτή την περίοδο. Το ακριβές περιεχόμενο θα συζητηθεί – ολοκληρωθεί στην επόμενη συνάντηση.

- **ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ**

Επειδή η επόμενη συνάντηση καθορίστηκε για την Τετάρτη 21 Ιανουαρίου και ώρα 10.00 στο ΕΑΙΤΥ (Δαβάκη 10 και Μυλοπόταμου, Αμπελόκηποι).

02/03/2007

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 15<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** ΕΑΙΤΥ / Εργαστήριο «Ήρων»

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η 15<sup>η</sup> συνάντηση του έργου «Ιατρολέξη» με συμμετοχή και των ιατρών.

Δύο ήταν τα σημαντικά ζητήματα της συνάντησης:

- α) η ολοκλήρωση της επιμέλειας των όρων από τους ιατρούς και
- β) τα παραδοτέα Π5-Π10 που θα πρέπει να ολοκληρωθούν και να αποσταλούν άμεσα.

Αρχικά συζητήθηκε το α) θέμα και έγινε κατανοητό ότι υπάρχουν σημαντικά προβλήματα στην επιμέλεια των όρων τόσο όσον αφορά όρους χημείας όσο και κάποιους διοικητικό-οικονομικούς όρους. Συμφωνήθηκε σε περιπτώσεις που χρειάζεται να δίνονται πάνω από μια εναλλακτικές αποδόσεις.

Όσον αφορά την ολοκλήρωση, συμφωνήθηκε οι ιατροί αν παραδώσουν άμεσα το κομμάτι που έχουν ολοκληρώσει μέχρι στιγμής με τελικό στόχο τέλος Μαρτίου να έχουν ολοκληρωθεί όλοι οι όροι.

Στη συνέχεια συζητήθηκαν οι μη τελικές εκδόσεις των παραδοτέων:

- Μονολεκτικοί όροι Βιοϊατρικής (αρχείο δεδομένων) Neurosoft
- Κανόνες αναγνώρισης πολυλεκτικών όρων Βιοϊατρικής (αρχείο δεδομένων και αναφορά τεκμηρίωσης) Γλωσσολόγοι
- Πολυλεκτικοί όροι Βιοϊατρικής (αρχείο δεδομένων) Neurosoft
- Αρχική Ταξινόμια Βιοϊατρικών Όρων (αρχείο δεδομένων) Neurosoft
- Συλλογή Βιοϊατρικών κειμένων (αρχείο) ΕΑΙΤΥ
- Τεκμηρίωση συλλογής βιοϊατρικών κειμένων (αναφορά) ΕΑΙΤΥ

έγιναν παρατηρήσεις και διορθώσεις και συμφωνήθηκε μέχρι το τέλος της ερχόμενης εβδομάδας να δοθούν οι τελικές εκδόσεις ώστε να σταλούν στη ΓΓΕΤ.

Τέλος συζητήθηκε μια ανακοίνωση που θα σταλεί στο συνέδριο της ΕΛΕΤΟ. Επιλέξαμε το θέμα να είναι οι πολυλεκτικοί όροι και η κωδικοποίησή τους και ανέλαβε η Ε.Μ. να ετοιμάσει μια περίληψη.

- **ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ**

Επειδή και στο τέλος Μαρτίου υπάρχουν 2 παραδοτέα που θα πρέπει να παραδοθούν, καθορίστηκε να υπάρξει μια συνάντηση τεχνικών την Πέμπτη 8 Μαρτίου με ενδεχόμενη συνάντηση όλης της ομάδας την Τετάρτη 14 Μαρτίου.

20/03/2007

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 16<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** EAITY / Εργαστήριο «Ήρων»

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η 16<sup>η</sup> συνάντηση του έργου «Ιατρολέξη» η οποία ήταν αμιγώς τεχνική.

Τι έγινε στο διάστημα που πέρασε:

- Στις αρχές του μήνα παραδόθηκε από τους ιατρούς το πρώτο τμήμα των όρων που έχουν επιμεληθεί. Παρ' όλα αυτά φαίνεται ότι η διαδικασία της επιμέλειας είναι αρκετά πιο επίπονη και χρονοβόρα απ' ότι περιμέναμε.
- Στις 14/3 στάλθηκαν τα παραδοτέα της ΕΕ3.
- Την προηγούμενη εβδομάδα, πραγματοποιήθηκε ακόμα μια τεχνική συνάντηση τα αποτελέσματα της οποίας περιγράφονται και αυτά στο παρών memo.
- Ολοκληρώθηκε και στάλθηκε από την Ε.Μ. το abstract για το συνέδριο της ΕΛΕΤΟ.
- Ήρθε θετική απάντηση στο paper που είχαμε καταθέσει για το συνέδριο που διοργανώνει η ΕΠΥ στις 18-20 Μαΐου 2007 στην Πάτρα.

Το βασικό θέμα συζήτησης ήταν

α) τα παραδοτέα Π3-Π4 που αντιστοιχούν στην ενότητα εργασίας 2 και θα πρέπει να ολοκληρωθούν και να αποσταλούν (τυπικό deadline στις 31/3).

Πιο συγκεκριμένα στο πλαίσιο της ΕΕ2, θα πρέπει να υλοποιηθούν:

- 1) Αποθήκη Εγγράφων (Document Warehouse): είναι ο χώρος αποθήκευσης των κειμένων αλλά και των παραγόμενων από αυτά μεταδεδομένων. Θα υλοποιηθεί πάνω σε ένα Σύστημα Διαχείρισης Σχεσιακών Βάσεων Δεδομένων (ΣΔΣΒΔ - RDBMS) (Ανάπτυξη). EAITY.
- 2) Προσκομιστής Εγγράφων (Crawler): διατρέχει συγκεκριμένους δικτυακούς τόπους ή τοπικούς δίσκους και προσκομίζει τα κείμενα προς επεξεργασία (Ανάπτυξη). EAITY.
- 3) Αναγνωριστής Στοιχείων (Tokenizer): κερματίζει ένα κείμενο σε μία σειρά στοιχείων (tokens: λέξεις, σημεία στίξης, αριθμοί, σύμβολα κτλ.) με τα οποία τροφοδοτούνται οι επόμενες φάσεις επεξεργασίας του κειμένου (Επέκταση). Neurosoft.

- 4) Μορφοσυντακτικός Σχολιαστής (Morphosyntactic Tagger): επισυνάπτει μορφοσυντακτικά μεταδεδομένα (μέρος του λόγου, γένος, αριθμός πτώση, κλπ.) σε κάθε λέξη του κειμένου (που έχει αναγνωρίσει ο Tokenizer). Αυτό γίνεται με τη βοήθεια του Μορφολογικού Λεξικού (Ανάπτυξη). Neurosoft.
- 5) Μορφολογικό Λεξικό. Η Neurosoft έχει αναπτύξει μορφολογικό λεξικό ~90.000 λημμάτων, το οποίο περιέχει και περιορισμένο αριθμό βιοϊατρικών όρων. Το λεξικό αυτό θα εμπλουτισθεί με τις άγνωστες λέξεις-όρους που θα βρεθούν στα βιοϊατρικά κείμενα. Το υπάρχον σύστημα περιγραφής της κλίσης των λημμάτων καλύπτει τον ορισμό της μορφοσυντακτικής πληροφορίας μονολεκτικών όρων (Εμπλουτισμός). Neurosoft.
- 6) Μηχανισμός Κλίσης Πολυλεκτικών Όρων. Το μοντέλο περιγραφής της κλίσης πολυλεκτικών όρων είναι πιο πολύπλοκο από αυτό των μονολεκτικών, διότι η κλίση πολυλεκτικών όρων δεν ακολουθεί μία τυποποιημένη διαδικασία. Υπάρχουν πολυλεκτικοί όροι στους οποίους κλίνεται μόνο η μία συστατική λέξη ενώ οι υπόλοιπες παραμένουν άκλιτες, π.χ. οξειδίο του αζώτου, οξειδία του αζώτου. Σε άλλες περιπτώσεις, όλες οι συστατικές λέξεις ακολουθούν την ίδια κλιτική φόρμα, π.χ. γλουταμινικό νάτριο, γλουταμινικού νατρίου. Θα επεκταθεί ο μηχανισμός κλίσης μονολεκτικών όρων (στοιχείο 5 της παρούσας λίστας) για να υποστηρίξει τον ορισμό πολυλεκτικών όρων. Η εφαρμογή που θα υλοποιήσει το μοντέλο αυτό θα παρέχει στο χρήστη τη δυνατότητα εύκολου ορισμού της κλίσης πολυλεκτικών όρων (Ανάπτυξη). Neurosoft.
- 7) Αναγνωριστής Όρων: συμβουλευεται το Μορφολογικό Λεξικό, όπου περιγράφεται η μορφολογία των μονολεκτικών και πολυλεκτικών όρων, καθώς και κανόνες που περιγράφουν τη σύνταξη πολυλεκτικών όρων και αναγνωρίζει τους όρους αυτούς σε κείμενα, σε όποια κλιτική μορφή κι αν βρίσκονται. Στην περίπτωση των πολυλεκτικών όρων, πρέπει να είναι σε θέση να αναγνωρίζει δομικές παραλλαγές (π.χ. οξειδία του Αζώτου, οξειδία Αζώτου, αζωτούχα οξειδία) και συντμήσεις (π.χ. NOx) (Ανάπτυξη). Neurosoft.

Τα παραδοτέα που θα πρέπει να περιλαμβάνουν τα παραπάνω είναι:

Π3. Σύστημα Διαχείρισης Σώματος Κειμένων και Εξαγωγής Ορολογίας (λογισμικό)

Π4. Εγχειρίδιο Χρήσης

Με βάση τα παραπάνω, αναλύθηκαν ένα - ένα τα παραδοτέα και συμφωνήθηκε ο τρόπος υλοποίησης, που βασικό χαρακτηριστικό του, θα πρέπει να είναι ότι όλα μαζί θα αποτελούν ένα ενιαίο περιβάλλον. Με βάση την πολυπλοκότητα κάποιων εξ αυτών, (6) φαίνεται ότι θα υπάρξει μια μικρή καθυστέρηση στην ολοκλήρωση και αποστολή τους. Παρ' όλα αυτά θα

πρέπει να δοθεί ιδιαίτερη προσοχή, ώστε να ολοκληρωθεί και αυτή η φάση και να επικεντρώσουμε στην φάση ανάπτυξης της ίδιας της οντολογίας.

- **ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ**

Η επόμενη συνάντηση θα καθοριστεί κατόπιν συζήτησης, λίγο μετά το Πάσχα.

27/04/2007

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 17<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** ΕΑΙΤΥ / Εργαστήριο «Ήρων»

---

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η 17<sup>η</sup> συνάντηση του έργου «Ιατρολέξη» η οποία ήταν αμιγώς τεχνική.

Τι έγινε στο διάστημα που πέρασε:

- Σε επικοινωνία που είχα με τους ιατρούς μου ανακοίνωσαν ότι στο αμέσως επόμενο διάστημα θα ολοκληρώσουν την επιμέλεια των όρων που έχουν αναλάβει.
- Ολοκληρώνεται η τριμηνιαία έκθεση και θα σταλεί άμεσα στην ΓΓΕΤ.

Το πρώτο θέμα συζήτησης είναι το πώς προχωράμε. Αναφορικά με τους γλωσσολόγους, έχουν στη διάθεσή τους όρους για επιμέλεια. Αντίστοιχα οι ιατροί, θα πρέπει να αρχίσουν να κατηγοριοποιούν σημασιολογικά τους όρους εκτός MeSH. Αυτό που θα πρέπει να ωριμάσει λίγο είναι ο τρόπος που θα γίνεται αυτό (διαδικασία). Με βάση αυτό όμως θα αρχίσει να χτίζεται και η οντολογία.

Το δεύτερο θέμα συζήτησης ήταν το πώς προχωράμε σε σχέση με τα παραδοτέα Π3 και Π4 τα οποία έχουν καθυστερήσει. Έχει γίνει πρόοδος από την προηγούμενη φορά, αλλά εκτιμάται ότι χρειάζονται ακόμα 15 μέρες για να τελειώσουν και τα τελευταία.

ΜΕ βάση αυτά θα γίνει άμεσα μια ακόμα τεχνική συνάντηση και στη συνέχεια θα κάνουμε μια συνολική συνάντηση της ομάδας έργου.

### • ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ

Η επόμενη συνάντηση θα καθοριστεί για την επόμενη εβδομάδα.



16/05/2007

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 18<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** EAITY / Εργαστήριο «Ήρων»

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η 18<sup>η</sup> συνάντηση του έργου «Ιατρολέξη» στην οποία παρευρέθηκε το μεγαλύτερο μέρος της ομάδας.

Αρχικά έγινε ενημέρωση σε σχέση με το διαρρεύσαν διάστημα: Έγιναν κάποιες τεχνικές συναντήσεις, με στόχο να κλείσει η ΕΕ2 και να παραδοθούν τα σχετικά παραδοτέα. Επίσης έγινε ενημέρωση ότι είναι ο χρόνος για την ετοιμασία της ενδιάμεσης έκθεσης προς τη ΓΓΕΤ.

Στη συνέχεια συζητήθηκε το θέμα της επιμέλειας των όρων από τους γιατρούς. Έγινε κατανοητό ότι η διαδικασία φτάνει στο τέλος της. Υπήρξαν σημαντικές δυσκολίες για την ολοκλήρωσή της μια και κάποιες ομάδες όρων (π.χ. χημικές ενώσεις) έκρυσαν (και κρύβουν) μεγάλη δυσκολία στην απόδοσή τους. Το σημαντικό είναι ότι άμεσα παραδίδεται και τα τελευταία κομμάτια των όρων από τους γιατρούς.

Το επόμενο ζήτημα που συζητήθηκε ήταν η επιμέλεια των όρων από τους γλωσσολόγους: η εργασία αυτή βρίσκεται σε εξέλιξη. Οι γλωσσολόγοι έχουν δύο σύνολα όρων που χρήζουν επεξεργασίας: α) αυτό που τους έχουν παραδώσει οι ιατροί, και β) τις «άγνωστες» λέξεις που προέκυψαν από το corpus.

Κατόπιν συζητήθηκε το θέμα του ποια θα πρέπει να είναι τα επόμενα βήματα. Αυτό που συμφωνήθηκε είναι να γίνει μια προσπάθεια να γίνει κατανοητό, τι μπορούμε να πάρουμε από το UMLS αυτόματα, και στη συνέχεια να προχωρήσουμε. Ένα ζήτημα που θα πρέπει να λυθεί είναι και το ζήτημα του χαρακτηρισμού (κατηγοριοποίησης) των όρων του MESH με βάση τις κατηγορίες που έχουμε ήδη επιλέξει (UMLS). Αυτό είναι μια εργασία που θα πρέπει να αναλάβουν οι ιατροί, αφού οριστικοποιηθεί το προηγούμενο που αναφέρθηκε παραπάνω.

Τέλος, μπήκε στο τραπέζι και η πιθανή συνεργασία με κάποιους φοιτητές του ΕΚΠΑ, οι οποίοι στα πλαίσια της πτυχιακής τους (Αγγλική Φιλολογία) θα μπορούσαν να βοηθήσουν στην συμπλήρωση της ερμηνείας κάποιων όρων.

- **ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ**

Η επόμενη συνάντηση θα γίνει σε περίπου 20 μέρες.

28/06/2007

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 19<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** EAITY / Εργαστήριο «Ήρων»

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η 19<sup>η</sup> συνάντηση του έργου «Ιατρολέξη» στην οποία παρευρέθηκε το μεγαλύτερο μέρος της ομάδας.

Αρχικά έγινε ενημέρωση σε σχέση με το διαρρέυσαν διάστημα: Οριστικοποιήθηκαν και απεστάλησαν τα παραδοτέα της ΕΕ2. Οριστικοποιήθηκαν οι εκθέσεις φυσικού και οικονομικού αντικειμένου. Και έτσι έχουμε μπροστά μας την επόμενη εβδομάδα τον έλεγχο οικονομικού αντικειμένου και στις 20 Ιουλίου την παρουσίαση για την πρόοδο του φυσικού αντικειμένου. Έτσι τη συνέχεια της συνάντησης απασχόλησαν δύο θέματα:

α) η προετοιμασία της παρουσίασης της προόδου του έργου και

β) η εξέλιξη του έργου (ΕΕ4).

Όσον αφορά την παρουσίαση συμφωνήθηκε να αποτελείται από τρία μέρη: μια παρουσίαση του έργου στην αρχή, κάποιες εξειδικευμένες παρουσιάσεις στη συνέχεια και τέλος μια παρουσίαση – επίδειξη του τι έχει φτιαχτεί μέχρι τώρα. Με βάση αυτά θα πρέπει να ετοιμαστούν οι παρακάτω παρουσιάσεις σε power point (στο format που επισυνάπτεται): α) Γενική παρουσίαση (Αρης), β) Corpus (Χρήστος Δ.), γ) Οντολογία (Ελενα + Γιώργος OWL), ε) όροι – πολυλεκτικοί όροι (Μαβίνα), στ) Εργαλεία (Γιώργος). Τέλος μπορεί να γίνει και επίδειξη του site.

Στο πλαίσιο αυτό θα πρέπει να εγκατασταθεί όλο το S/W σε ένα laptop και να υπάρχει και κάπου το site (για την περίπτωση που δεν θα υπάρχει σύνδεση internet). Τέλος αποφασίστηκε την προηγούμενη να γίνει μια πρόβα generale στη Neurosoft.

Όσον αφορά την εξέλιξη του έργου, βρισκόμαστε στην ΕΕ4, η οποία περιλαμβάνει τρία παραδοτέα: Προσαρμοσμένο protégé, Αρχική ταξινομία εντός protégé, και Οντολογία Βιοϊατρικών όρων. Το δεύτερο είναι έτοιμο (απαιτείται απλά μια περιγραφή). Το πρώτο και το τρίτο χρειάζονται κάποια εργασία ακόμα. Στόχος είναι να ολοκληρωθούν τέλος Ιουλίου.

Τέλος συζητήθηκαν οι δύο εργασίες που θα πρέπει να συγγραφούν άμεσα: η μία για την ΕΛΕΤΟ και η δεύτερη για το CL07.

### • ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ

Η επόμενη συνάντηση θα γίνει στις 19 Ιουλίου για την πρόβα της παρουσίασης.

6/08/2007

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 20<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** ΕΑΙΤΥ / Εργαστήριο «Ήρων»

---

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η 20<sup>η</sup> συνάντηση του έργου «Ιατρολέξη», η οποία ήταν μια τεχνική συνάντηση.

Αρχικά συζητήθηκαν οι εκκρεμότητες που υπάρχουν σε σχέση με το έργο:

- α) Έχει καθυστερήσει η συμπλήρωση του τριμηνιαίου δελτίου. Αυτό με δεδομένο ότι ετοιμάσαμε την ετήσια και αναλώσαμε πολύ χρόνο σε αυτή. Βάλαμε ως στόχο να το ολοκληρώσουμε την επόμενη εβδομάδα.
- β) Αρχές Σεπτέμβρη θα πρέπει να μαζέψουμε όλες τις πιθανές ανάγκες τροποποιήσεων (ενδεικτικά: ένταξη Τσαλίδη έστω και άμισθου στο έργο, παράταση χρονοδιαγράμματος, μεταφορά κάποιων κονδυλίων, κ.τ.λ.) και να τις στείλουμε στη ΓΓΕΤ.

Όσον αφορά το φυσικό αντικείμενο του έργου, έχουμε να παραδώσουμε (από το τέλος Ιουνίου) τα παραδοτέα της ΕΕ4 Π11, Π12 και Π13: Προσαρμοσμένο προτέγέ, Αρχική ταξινόμια εντός προτέγέ, και Οντολογία Βιοϊατρικών όρων. Το δεύτερο είναι έτοιμο (απαιτείται απλά μια περιγραφή). Το πρώτο το αναλαμβάνει το ΕΑΙΤΥ και το τρίτο το αναλαμβάνει η Neuro. Στόχος είναι να ολοκληρωθούν εντός του Αυγούστου και να σταλούν αρχές Σεπτέμβρη.

Τέλος συζητήθηκε η εργασία που πέρασε σε workshop του RANLP07 και ανέλαβε να την κοιτάξει και να την παρουσιάσει ο Χ. Τσαλίδης.

- **ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ**

Η επόμενη συνάντηση θα γίνει το Σεπτέμβρη.

10/09/2007

**ΣΥΓΓΡ:** Α. Βαγγελάτος

**ΑΝΤΙΚ:** 21<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»

**ΑΝΑΦ:** EAITY / Εργαστήριο «Ήρων»

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η 21η συνάντηση του έργου «Ιατρολέξη», η οποία ήταν μια τεχνική συνάντηση.

Αρχικά συζητήσαμε τα 3 παραδοτέα του έργου, που εντάσσονται στην ΕΕ4 και έχουν καθυστερήσει να ολοκληρωθούν.

Π11. Προσαρμοσμένο Protégé (λογισμικό + αναφορά προσαρμογών). Έχει ολοκληρωθεί και έχει σταλεί για πιθανά σχόλια. Σχόλια μέχρι στιγμής έχει στείλει η Μ.Π. Περιλαμβάνει την αναφορά και ένα CD.

Π12. Αρχική Ταξινόμια εντός Protégé (αρχείο δεδομένων) Όσον αφορά την αρχική ταξινόμια είναι έτοιμη (από το Π8). Ετοιμάστηκε και η σχετική αναφορά και εστάλη για παρατηρήσεις. Θα περιέχεται μέσα και ένας ανεξάρτητος browser για να μπορεί κανείς να δει τις οντολογίες. ΕΠΕΙΔΗ όμως το κόστος για το πακέτο αυτό είναι δυσανάλογα μεγάλο (δεν ξέρω πως το είχαμε σκεφτεί τότε...) θα πρέπει να αναπτύξουμε με κάποιο τρόπο τι κάναμε... Στη συνάντηση μπήκε το θέμα να μπουν περισσότερα επίπεδα της ιεραρχίας ώστε να δικαιολογηθεί το κόστος. Θα πρέπει να το δούμε.

Π13. Οντολογία Βιοϊατρικών Όρων (αρχείο δεδομένων). Ο Γιώργος ολοκλήρωσε και μας έστειλε σε OWL το MeSH. Όπως μας εξήγησε είναι μια ιεραρχία. Σε κάθε κόμβο θα μπει και το semantic type. Δεν είναι σίγουρο όμως ότι η ιεραρχία μας καλύπτει μια και μιλάμε και για σχέσεις (πέραν της ιεραρχικής). Θα πρέπει να δούμε τι παραπάνω μπορούμε να κάνουμε. Επίσης στο Τεχνικό Παράρτημα λέμε και για φωνητική μεταγραφή και για Ελληνική ερμηνεία. Και σε αυτό θα πρέπει να δούμε τι μπορούμε να κάνουμε. Στο CD με την οντολογία μάλλον θα πρέπει να μπει και εδώ ο browser.

Θα πρέπει τα παραπάνω να ετοιμαστούν σύντομα ώστε να ασχοληθούμε με την τελική φάση του έργου ΕΕ5.

Επίσης, μέσα στις επόμενες μέρες θα πρέπει να ετοιμαστεί και το αίτημα για πιθανές τροποποιήσεις που θέλουμε να κάνουμε. Αυτό θα σταλεί στην ΓΓΕΤ για έγκριση. Βασικό σημείο θα είναι η δίμηνη παράταση ώστε το έργο να τελειώνει στο τέλος του χρόνου (από 26/10 που τελειώνει τώρα).

Τέλος θα πρέπει από τώρα να αρχίσουμε να σκεφτόμαστε για την ημερίδα που θα πρέπει να οργανώσουμε για το έργο.

Εφόσον το αίτημα παράτασης εγκριθεί, η ημερομηνία θα είναι αρχές Δεκέμβρη. Θα πρέπει λοιπόν να αρχίσουμε να σχεδιάζουμε την ατζέντα και τον κατάλογο των προσκεκλημένων!

- **ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ**

Η επόμενη συνάντηση θα γίνει στα μέσα Σεπτέμβρη.

18/10/2007

**ΣΥΓΓΡ:** Α. Βαγγελάτος  
**ΑΝΤΙΚ:** 22<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»  
**ΑΝΑΦ:** EAITY / Εργαστήριο «Ήρων»

---

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η 22<sup>η</sup> συνάντηση του έργου «Ιατρολέξη».

Αρχικά έγινε ενημέρωση ότι έχει ζητηθεί 2μηνη παράταση στο έργο καθώς και κάποιες επιμέρους τροποποιήσεις στο Τ.Π. και περιμένουμε την απάντηση της ΓΓΕΤ

Στη συνέχεια μπήκε το πρώτο βασικό θέμα της συνάντησης το οποίο είναι η ημερίδα που θα διοργανωθεί στο πλαίσιο του έργου. Κατ' αρχή η ημερομηνία προτάθηκε να είναι είτε η Παρασκευή 7 Δεκεμβρίου είτε η Δευτέρα 10 Δεκεμβρίου. Η οριστικοποίηση της ημερίδας θα γίνει μέχρι το τέλος του Οκτωβρίου και με βάση την διαθεσιμότητα των προσκαλεσμένων. Έπειτα συζητήθηκε η δυνατότητα να προσκαλέσουμε 2-3 ανθρώπους εκτός του έργου για να μιλήσουν. Προτάθηκε η κα Αναιάδου (η οποία τελικά δεν μπορεί) καθώς επίσης και οι: κα Φραντζή, Πιπερίδης και κάποιος από τον Δημόκριτο. Σε αυτούς θα γίνουν οι απαραίτητες κρούσεις. Τέλος προτάθηκε ως χώρος το ξενοδοχείο Lycabetus και ανέλαβε να το διερευνήσει ο Χ.Τ. μαζί με τη δυνατότητα παρουσίας κάποιας εταιρείας (όπως π.χ. της ΗΡ).

Κατόπιν έγινε ενημέρωση για δύο θέματα: α) την συνάντηση που είχε την προηγούμενη ημέρα ο Α.Β. στο εργαστήριο του κ Γεωργιάδη στο ΕΚΠΑ για το θέμα της συνεργασίας και β) για την επερχόμενη συνάντηση της επομένης με τους φοιτητές της κας Σοφianού.

Τέλος συζητήθηκαν τα τεχνικά θέματα:

α) Απομένει να δούμε το θέμα του Παραδοτέου Π13 (Οντολογίας). ΘΑ το κοιτάξει καταρχήν ο Γιώργος.

β) Το Π14 (περιγηγητής οντολογίας) ξεκινάει να το υλοποιεί ο Γιάννης Σ.

γ) Το Π16 (Ορθογραφικός διορθωτής ιστού) θα το δει η NeuroLingo.

Με αυτό το σκεπτικό, απομένουν τα Π15 (Μηχανή αναζήτησης) και Π17 (Μορφοσυντακτικός και Σημασιολογικός Σχολιαστής).

- **ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ**

Η επόμενη συνάντηση θα γίνει αρχές Νοέμβρη

9/1/2008

**ΣΥΓΓΡ:** Α. Βαγγελάτος  
**ΑΝΤΙΚ:** 23<sup>η</sup> Συνάντηση έργου «Ιατρολέξη»  
**ΑΝΑΦ:** ΕΑΙΤΥ / Εργαστήριο «Ήρων»

---

## ΣΥΝΟΨΗ

Πραγματοποιήθηκε η 23η συνάντηση του έργου «Ιατρολέξη».

Είδαμε ότι έχουν οριστικοποιηθεί τα παραδοτέα Π13 και Π14.

Τα Π15, Π16 και Π17, ανέλαβε να τα προχωρήσει ο Γ.Ο.

Το Π18 (τεκμηρίωση Εφαρμογών Ιστού), θα το συντονίσει ο Α.Β όπως και το Π22 (Τελική Αναφορά Έργου). Τα Π19, Π20 και Π21 έχουν και αυτά ολοκληρωθεί.

Μπήκε ο στόχος αυτά να ολοκληρωθούν με το τέλος του Ιανουαρίου.

Στη συνέχεια θα πρέπει να δούμε το ζήτημα της τριμηνιαίας έκθεσης και της τελικής έκθεσης η οποία θα πρέπει να ολοκληρωθεί έως τα μέσα του Φεβρουαρίου

## ΕΠΟΜΕΝΗ ΣΥΝΑΝΤΗΣΗ

Η επόμενη συνάντηση θα γίνει εφόσον χρειαστεί στο τέλος Ιανουαρίου.