

ΙΑΤΡΟΛΕΞΗ

Ανάπτυξη Υποδομής Γλωσσικής Τεχνολογίας για το Βιοϊατρικό Τομέα

Neurosoft A.E. --- ΕΑΙΤΥ

ΓΓΕΤ, ΚτΠ, Πρόγραμμα «ΕΠΕΞΕΡΓΑΣΙΑ ΕΙΚΟΝΩΝ, ΗΧΟΥ ΚΑΙ ΓΛΩΣΣΑΣ»

Προϋπολογισμός:	561.240 €
Διάρκεια:	18 μήνες
Επιστημονικός Υπεύθυνος:	Δρ. Αρ. Βαγγελάτος

Αντικείμενο του έργου

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

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

Σε επόμενη φάση (πέρα από το πλαίσιο του παρόντος), είναι δυνατή η επέκταση των αποτελεσμάτων του και η διαμόρφωσή του ώστε να εξυπηρετεί και άλλους σκοπούς (π.χ. υλοποίηση δυνατοτήτων semantic web σε βιοϊατρικούς ιστοχώρους, κλπ).

Αντικείμενο του έργου (παραγόμενα) I

- Μεθοδολογία ανάπτυξης οντολογιών.
- Υπολογιστικοί πόροι και εργαλεία που θα υποστηρίζουν τη μεθοδολογία ανάπτυξης οντολογιών:
 - 1) Το Μορφολογικό Λεξικό της Neurosoft εμπλουτισμένο με βιοϊατρικούς όρους.
 - 2) Μορφοσυντακτικός Σχολιαστής (Morphosyntactic Tagger), ο οποίος θα χαρακτηρίζει μορφοσυντακτικά κάθε λέξη ενός κειμένου.
 - 3) Αναγνωριστής Ονοματικών Φράσεων (Noun-Phrase Chunker), ο οποίος θα ανιχνεύει υποψήφιους πολυλεκτικούς όρους σε κείμενα με χρήση μορφοσυντακτικών κανόνων.
 - 4) Περιβάλλον Ανάπτυξης Οντολογιών βασισμένο στο state-of-the-art σύστημα Protégé (<http://protege.stanford.edu>).
 - 5) Σημασιολογικός Σχολιαστής (Semantic Tagger), ο οποίος θα επισυνάπτει πληροφορία από την οντολογία σε όρους που αναγνωρίζονται σε κείμενα.
 - 6) Μηχανισμός δεικτοδότησης βιοϊατρικών κειμένων βασισμένος.

ΑΝΤΙΚΕΙΜΕΝΟ ΤΟΥ ΈΡΓΟΥ (ΠΑΡΑΓΟΜΕΝΑ) ΙΙ

- Σώμα (corpus) βιοϊατρικών κειμένων.
- Εφαρμογές Ιστού με τις οποίες ο χρήστης θα μπορεί:
 - 1) Να ελέγχει την ορθογραφία μίας λέξης (βιοϊατρικός όρος).
 - 2) Να περιηγείται στην οντολογία πραγματοποιώντας σύνθετες αναζητήσεις για σχέσεις που διέπουν όρους της οντολογίας.
 - 3) Να εισάγει ένα νέο κείμενο και να λαμβάνει ως αποτέλεσμα το κείμενο με μορφοσυντακτικούς και σημασιολογικούς σχολιασμούς των όρων του κειμένου (annotations).
 - 4) Να αναζητά κείμενα με συγκεκριμένους όρους ή συνδυασμό όρων ή σχέσεων που ισχύουν σε όρους.

Ενότητες Εργασίας

1. Σχεδιασμός Σώματος Βιοϊατρικών Κειμένων και Οντολογίας
2. Υλοποίηση Εργαλείων Διαχείρισης Σώματος Κειμένων και Αυτόματης Εξαγωγής Ορολογίας
3. Συλλογή Βιοϊατρικών Κειμένων και Εξαγωγή Ορολογίας
4. Επεξεργασία Βιοϊατρικών Όρων και Δημιουργία Οντολογίας
5. Υλοποίηση Εφαρμογών Ιστού
6. Διάχυση Αποτελεσμάτων

1. Σχεδιασμός Σώματος Βιοϊατρικών Κειμένων και Οντολογίας

Στόχος: Στην ΕΕ1 θα σχεδιαστεί σώμα βιοϊατρικών κειμένων ικανό να αναδείξει την τρέχουσα ορολογία του τομέα και θα οριστεί το μοντέλο οντολογικής αναπαράστασης της ορολογίας.

Έναρξη: T0 **Λήξη:** T0 + 6

Παραδοτέα:

- Σχέδιο Συγκρότησης Σώματος Βιοϊατρικών Κειμένων (τεχνική αναφορά)
- Μοντέλο Αναπαράστασης Οντολογίας (τεχνική αναφορά)

2. Υλοποίηση Εργαλείων Διαχείρισης Σώματος Κειμένων και Αυτόματης Εξαγωγής Ορολογίας

Στόχος: Στην ΕΕ2 θα αναπτυχθούν εργαλεία για τη διαχείριση σώματος κειμένων και την αυτόματη εξαγωγή ορολογίας, τα οποία θα υποστηρίξουν την ανάπτυξη γλωσσικών πόρων (Μορφολογικού Λεξικού και Οντολογίας).

Έναρξη: T0 + 1 Λήξη: T0 + 11

Παραδοτέα:

- Σύστημα Διαχείρισης Σώματος Κειμένων και Εξαγωγής Ορολογίας (λογισμικό)
- Εγχειρίδιο Χρήσης

3. Συλλογή Βιοϊατρικών Κειμένων και Εξαγωγή Ορολογίας

Στόχος: Η ΕΕ3 περιλαμβάνει τη συλλογή και επεξεργασία βιοϊατρικών κειμένων, την εξαγωγή βιοϊατρικών όρων και τον εμπλουτισμό του Μορφολογικού Λεξικού με αυτούς και τέλος τον καθορισμό της ταξινόμιας που θα αποτελέσει τη βάση της οντολογίας.

Έναρξη: T0 + 2 Λήξη: T0 + 9

Παραδοτέα:

- Μονολεκτικοί Όροι Βιοϊατρικής (αρχείο δεδομένων)
- Κανόνες Αναγνώρισης Πολυλεκτικών Όρων Βιοϊατρικής (αρχείο δεδομένων + αναφορά τεκμηρίωσης)
- Πολυλεκτικοί Όροι Βιοϊατρικής (αρχείο δεδομένων)
- Αρχική Ταξινόμια Βιοϊατρικών Όρων (αρχείο δεδομένων)
- Συλλογή βιοϊατρικών κειμένων (αρχείο)
- Τεκμηρίωση συλλογής βιοϊατρικών κειμένων (αναφορά)

4. Επεξεργασία Βιοϊατρικών Όρων και Δημιουργία Οντολογίας

Στόχος: Στην ΕΕ4 θα αναπτυχθεί η οντολογία βιοϊατρικών όρων.

Έναρξη: T0 + 4 Λήξη: T0 + 14

Παραδοτέα:

- Προσαρμοσμένο Protégé (λογισμικό + αναφορά προσαρμογών)
- Αρχική Ταξινόμια εντός Protégé (αρχείο δεδομένων)
- Οντολογία Βιοϊατρικών Όρων (αρχείο δεδομένων)

5. Υλοποίηση Εφαρμογών Ιστού

Στόχος: Στην ΕΕ5 θα αναπτυχθούν εφαρμογές ιστού, οι οποίες θα επιτρέπουν την αξιοποίηση των αποτελεσμάτων του έργου μέσα από το Διαδίκτυο.

Έναρξη: T0 + 11 **Λήξη:** T0 + 18

Παραδοτέα:

- Περιηγητής Οντολογίας
- Μηχανή Αναζήτησης
- Ορθογραφικός Διορθωτής ιστού
- Μορφοσυντακτικός και Σημασιολογικός Σχολιαστής
- Τεκμηρίωση Εφαρμογών ιστού

6. Διάχυση Αποτελεσμάτων

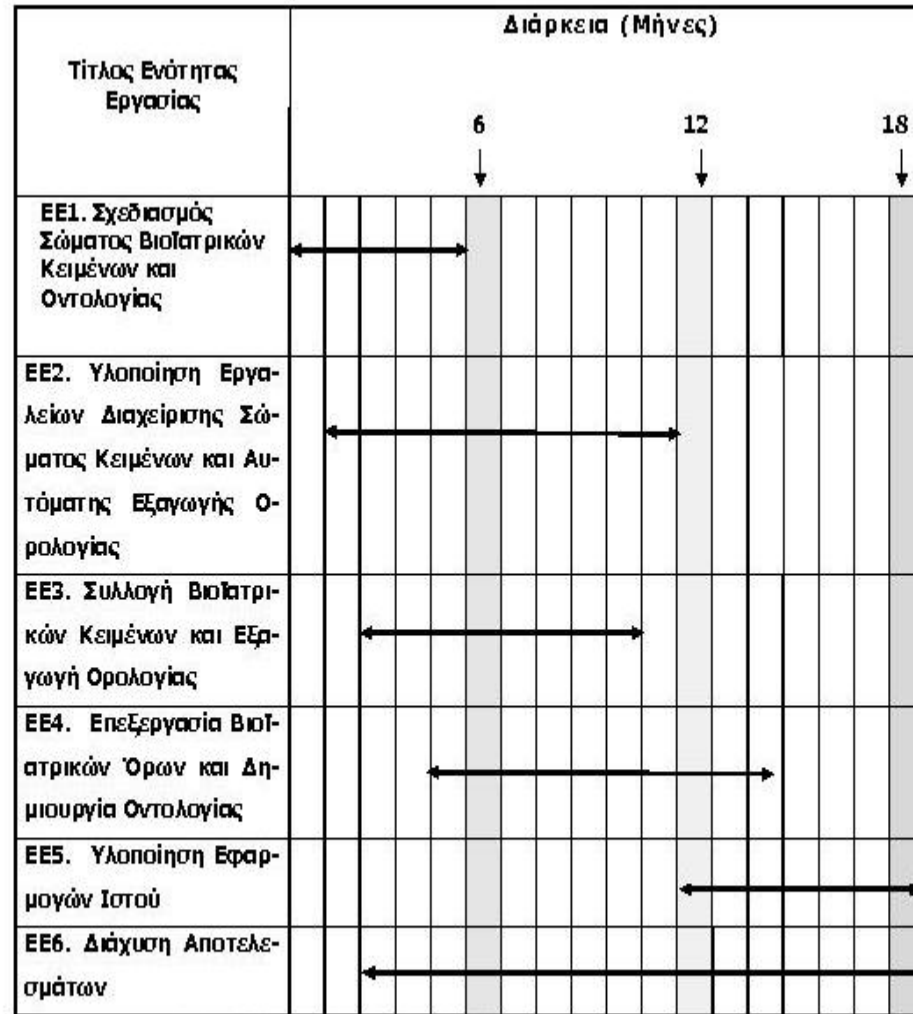
Στόχος: Στην ΕΕ6 θα υλοποιηθεί σειρά ενεργειών και δράσεων δημοσιοποίησης της μεθοδολογίας και των αποτελεσμάτων του προτεινόμενου έργου.

Έναρξη: T0 + 3 **Λήξη:** T0 + 18

Παραδοτέα:

- Υλοποίηση και συντήρηση project site
- Ανακοινώσεις επιστημονικών εργασιών σε συνέδρια / περιοδικά
- Ημερίδα διάχυσης αποτελεσμάτων (πρόγραμμα – υλικό)
- Τελική Αναφορά Έργου

Χρονο- διάγραμμα

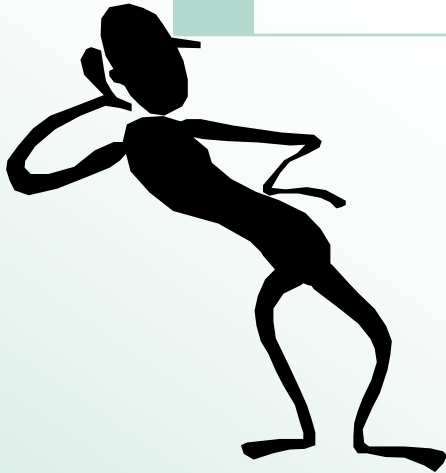


ΥΠΟΜΝΗΜΑ ΠΙΝΑΚΑ / ANNOTATION

Χρησιμοποιείστε τα παρακάτω σύμβολα

←→ Διάρκεια σταδίου Ενότητας Εργασιών

ΙΑΤΡΟΛΕΞΗ



Εν κατακλείδι,

Επί της ουσίας, ένα ιατρο-γλωσσολογικό έργο με
πληροφορική υποστήριξη.

Καλή Επιτυχία