### 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»

### ЕРГО - 9: IATPOAEEH

## ΠΑΡΑΔΟΤΕΟ

Π16: Ορθογραφικός Διορθωτής Ιστού

(στο παραδοτέο περιλαμβάνεται οπτικός δίσκος που περιέχει το λογισμικό)

Ημερομηνία:	15.12.2007
Έκδοση:	Final
Τύπος:	Εμπιστευτικό





#### ΠΕΡΙΕΧΟΜΕΝΑ

1	ΕΙΣΑΓΩΓΗ	3
2	ΟΡΘΟΓΡΑΦΙΚΟΣ ΔΙΟΡΘΩΤΗΣ ΙΣΤΟΥ	4
2.1	Τεχνικά Χαρακτηριστικά - Εγκατάσταση 5	
2.2	ΟΔΗΓΙΕΣ ΕΓΚΑΤΑΣΤΑΣΗΣ ΤΟΥ ΟΡΘΟΓΡΑΦΙΚΟΥ ΔΙΟΡΘΩΤΗ ΤΟΥ ΕΡΓΟΥ ΙΑΤΡΟΛΕΞΗ5	
2.3	ΕΓΧΕΙΡΙΔΙΟ ΧΡΗΣΗΣ 7	
3	ΟΔΗΓΙΕΣ ΧΡΗΣΗΣ ΤΟΥ ΟΠΤΙΚΟΥ ΔΙΣΚΟΥ	11
4	ΣΧΕΤΙΚΟ ΛΟΓΙΣΜΙΚΟ	12





## 1 ΕΙΣΑΓΩΓΗ

Το παρών παραδοτέο αποτελεί το Π16: «Ορθογραφικός Διορθωτής Ιστού» του έργου ΙΑΤΡΟΛΕΞΗ. Πρόκειται για ένα λογισμικό που υλοποιήθηκε στο πλαίσιο της Ενότητας Εργασίας 5, με τίτλο: «Υλοποίηση Εφαρμογών Ιστού», του έργου ΙΑΤΡΟΛΕΞΗ.

Το εργαλείο αυτό επιτρέπει στο χρήστη να ελέγξει (σε επίπεδο ορθογραφίας) ένα κείμενό του.

Η τεκμηρίωση του εργαλείου αυτού, καθώς και των λοιπών που αναπτύχθηκαν στο πλαίσιο της ΕΕ5, περιγράφεται στο σχετικό παραδοτέο **Π18: Τεκμηρίωση Εφαρμογών Ιστού.** 

Επισυνάπτεται οπτικός δίσκος (CD) στον οποίο περιέχεται το λογισμικό που αναφέρθηκε παραπάνω.





### 2 ΟΡΘΟΓΡΑΦΙΚΟΣ ΔΙΟΡΘΩΤΗΣ ΙΣΤΟΥ

Στο πλαίσιο του έργου και της ενότητας εργασίας: «ΕΕ5. Υλοποίηση Εφαρμογών Ιστού» αναπτύχθηκαν μια σειρά από εφαρμογές ιστού ώστε να είναι δυνατή η διάχυση των αποτελεσμάτων στο ευρύ κοινό. Μια από αυτές τις εφαρμογές είναι ο «ορθογραφικός διορθωτής ιστού» μέσω του οποίου ο χρήστης μπορεί να ελέγξει την ορθογραφία ενός κειμένου. Ο ορθογραφικός διορθωτή βρίσκεται στο ιστοχώρο του έργου <u>http://www.iatrolexi.gr/</u> και είναι άμεσα διαθέσιμος στους χρήστες του διαδικτύου.

🔋 IATPOAEIH - Mozilla Firefox				
Eile Edit View History Bookmarks Iools Help				
<ul> <li>-</li> <li>-</li></ul>	🛛 • 🗼 • 🥑 🛞 🏠 🗋 http://www.iatrolexi.gr/iatrolexi/ergaleia.html			
🌮 Getting Started 🔂 Lates	t Headlines			
ΙΑΤΡΟΛΕΞΗ Ανάπτυξη Υποδομής Γλωσσικής Τεχνολογίας για το Βιοϊατρικό Τομέα				
Εισαγωγή		Εργαλεία		
Περιγραφή				
Αποτελέσματα 🕨	ΕΝΟΤΗΤΑ ΕΡΓΑΣΙΑΣ	τιτλος	ПЕРІГРАФН	
Νέα Χώρος Συζητήσεων	ΕΕ5. Υλοποίηση Εφαρμογών Ιστού	Περιηγητής Οντολογίας	Με το εργαλείο αυτό ο χρήστης μπορεί να πλοηγηθεί στην οντολογία των βιοίατρικών όρων του ΙΑΤΡΟΛΕΣΗ και να αναζητήσει βιοϊατρικούς όρους μέσα σε αυτήν.	
Εταίροι Ομάδα Έργου Σύνδεσμοι	ΕΕ5. Υλοποίηση Εφαρμογών Ιστού	Μηχανή Αναζήτησης	Με το εργαλείο αυτό ο χρήστης μπορεί να κάνει αναζήτηση στο σώμα βιοίταρικών κειμένων, τα οποία έχουν ευρετηριαστεί επί τη βάσων, σμασιολογικών κατηγοριών από την οντολογία, αλλά και σε κείμενα του Παγκόσμιου Ιστού.	
	ΕΕ5. Υλοποίηση Εφαρμογών Ιστού	Ορθογραφικός Διορθωτής Ιστού	Με το εργαλείο αυτό ο χρήστης μπορεί να ελέγξει την ορθογραφία μονολεκτικού βιοίατρικού όρου ή όρου της κοινής ελληνικής γλώσσας.	
	ΕΕ5. Υλοποίηση Εφαρμογών Ιστού	Μορφοσυντακτικός και Σημασιολογικός Σχολιαστής	Με το εργαλείο αυτό ο χρήστης μπορεί να εισάγει ένα κείμενο και να επισυνάπτει μορφοσυντακτικά (από το Μορφολογικό Λεξικό) και σημασιολογικά (από την Οντολογία) σχόλια σε λέξεις/στοιχεία σε αυτό.	
2482 Επισκέπτες από 13-08-2006 ₩3C #TML	ΕΕ2. Υλοποίηση Εργαλείων Διαχείρισης Σώματος Κειμένων και Αυτόματης Εξαγωγής Ορολογίας	Συμφραζόμενα όρων (concordances)	Το εργαλείο αυτό επιπρέπει την αναζήτηση όρων στη συλλογή κειμένων (corpus) και επιστρέφει τον όρο (είτε στη μορφή που δόθηκε, είτε σε κάθε κλιτικό τύπο εφόσον επιλεγεί η «ληματοποίηση») μαζί με τα συμαραζόμενά του ( μια γραμμή) από το κείμενο που βρέθηκε. Ο χρήστης μπορεί στην συνέχεια να επιλέξει και να δει το κείμενο, όπου σημειώνεται που ακριβώς έχει βρεθεί ο όρος.	
Αρχική Σελίδα Επικοινωνία Δικαιώματα Πνευματικής Ιδιοκτησίας Χάρτης Πλοήγησης Σχεδίαση & Ανάπτυξη: Εργαστήριο ΗΡΩΝ - ΕΑΙΤΥ Για την ανάπτυξη αυτής της Δικτυακής Πύλης χρησιμοποιήθηκε ανοιχτό λογισμικό υπό την άδεια GNU/GPL.				

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:

• Φωνητική διόρθωση. Βρίσκει λέξεις που "ακούγονται" το ίδιο.





- Επιφανειακή διόρθωση. Αγνοεί τον τόνο και τα κεφαλαία.
- Απλή διόρθωση. Κατά την αναζήτηση στο λεξικό εκτελεί αντικατάσταση γραμμάτων, "παρεμβολή", κ.α.
- Πλήρης διόρθωση. Συνδυασμός όλων των παραπάνω διορθώσεων.

#### 2.1 ΤΕΧΝΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ - ΕΓΚΑΤΑΣΤΑΣΗ

Ο ορθογραφικός διορθωτής ιστού του ΙΑΤΡΟΛΕΞΗ όσον αφορά το web κομμάτι έχει υλοποιηθεί κάνοντας χρήση της τεχνολογίας JavaServer Pages (JSP) <u>http://java.sun.com/products/jsp/</u> για την ανάπτυξη εφαρμογών ιστού. Η γλώσσα προγραμματισμού που έχει χρησιμοποιηθεί για το application κομμάτι της εφαρμογής είναι η Java και συγκεκριμένα το πακέτο JDK 6. Για την ανάπτυξη των web σελίδων της εφαρμογής έχει γίνει χρήση της HTML, D-HTML και Javascript.

Αναλυτικά η ανάπτυξη και τα τεχνικά χαρακτηριστικά του ορθογραφικού διορθωτή του έργου μπορείτε να τα βρείτε στο αντίστοιχο παραδοτέο.

Για την διόρθωση των όρων έχουν χρησιμοποιηθεί τα μορφοσυντακτικά λεξικά που αναπτύχθηκαν στα πλαίσια του έργου.

## 2.2 ΟΔΗΓΙΕΣ ΕΓΚΑΤΑΣΤΑΣΗΣ ΤΟΥ ΟΡΘΟΓΡΑΦΙΚΟΥ ΔΙΟΡΘΩΤΗ ΤΟΥ ΕΡΓΟΥ ΙΑΤΡΟΛΕΞΗ

Η εφαρμογή ιστού του ορθογραφικού διορθωτή παρέχεται ως ένα σύνολο από αρχεία που αποτελούν τμήματα των εφαρμογών ιστού του έργου ΙΑΤΡΟΛΕΞΗ, το οποίο βρίσκεται στον οπτικό δίσκο που συνοδεύει την παρούσα αναφορά μέσα στον φάκελο με τίτλο '.../software/web\_tools/'. Τα αρχεία που σχετίζονται αποκλειστικά μόνο με τις web σελίδες της εφαρμογή του ορθογραφικού διορθωτή βρίσκονται στον φάκελο με τίτλο '.../software/web\_tools/webapps/speller/'.

Οι εφαρμογές ιστού προϋποθέτουν την ύπαρξη ενός εξυπηρετητή ιστού (web server), ο οποίος θα δίνει στους χρήστες του διαδικτύου την δυνατότητα πρόσβασης σε αυτές. Εμείς προτείνουμε τον Tomcat 6, ο οποίος αποτελεί εφαρμογή ανοικτού κώδικα και είναι διαθέσιμος στο ευρύ κοινό. Το αρχείο εγκατάστασης του Tomcat είναι το 'apache-tomcat-6.0.14.exe' το οποίο βρίσκεται στον οπτικό δίσκο που συνοδεύει την παρούσα αναφορά





μέσα στον φάκελο με τίτλο '…/software/tomcat/'. Πέρα από το υλικό που παρέχουμε για τον Tomcat, επιπλέον υλικό μπορεί να βρει κανείς στη σελίδα <u>http://tomcat.apache.org/</u>.

Για να τρέξουν εφαρμογές ιστού, θα πρέπει να είναι εγκατεστημένο το περιβάλλον Java Runtime Environment 1.6.0 (το αρχείο jre-6u1-windows-i586-p το οποίο εμπεριέχετε στον οπτικό δίσκο που σας έχει δοθεί μέσα στον φάκελο με τίτλο '.../software/jre/' ή μπορείτε να το κατεβάσετε από τη διεύθυνση <u>http://java.sun.com/javase/downloads/index.jsp</u>).

Αφού γίνει εγκατάσταση του Tomcat, στη συνέχεια ο φάκελος με τίτλο ".../software/web\_tools/webapps/iatrolexi/" του οπτικού δίσκου της παρούσας αναφοράς θα πρέπει να αντιγραφεί μαζί με τα περιεχόμενά του στο φάκελο 'webapps' στη θέση του δίσκου που έχει γίνει η εγκατάσταση του Tomcat και βρίσκονται οι σελίδες των εφαρμογών που 'σηκώνει' ο Tomcat. Πριν ξεκινήσει ο Tomcat θα πρέπει να αυξηθεί η μνήμη που χρησιμοποιεί η Java Virtual Machine του. Αυτό γίνεται μέσω της επιλογής 'Configure Tomcat' που υπάρχει στο μενού του Tomcat και ανοίγει τη φόρμα με τίτλο 'Apache Tomcat Properties'. Στο tab της Java βάζουμε την τιμή 128MB στη θέση 'Initial Memory pool' και την τιμή 512MB στη θέση 'Maximum Memory pool'. Η αύξηση της μνήμης είναι απαραίτητη για να διαχειριστούμε το owl αρχείο της οντολογίας.

Επιπλέον θα πρέπει να δημιουργηθεί στον ηλεκτρονικό υπολογιστή που έχει γίνει εγκατάσταση του Tomcat ένας φάκελος στο τοπικό του δίσκο και μέσα σε αυτόν να μπουν τα περιεχόμενα του φακέλου του οπτικού δίσκου της αναφοράς με τίτλο '.../software/web\_tools/data/'. Σε αυτόν τον φάκελο υπάρχουν τα λεξικά, η συλλογή κειμένων, το αρχείο της οντολογίας και διάφορα άλλα βοηθητικά αρχεία που χρησιμοποιούνται από τις εφαρμογές ιστού και είναι απαραίτητα για τη λειτουργία τους. Στη συνέχεια θα πρέπει να αλλάξουν τα paths μέσα στο αρχείο '.../webapps/iatrolexi/WEB-INF/ paths.properties' στη θέση του δίσκου που έχει γίνει η εγκατάσταση του Tomcat, ώστε αυτά να συμφωνούν με τον νέο φάκελο που δημιουργήσατε στον τοπικό δίσκο.

Αφού γίνουν όλα τα παραπάνω με την έναρξη ή επανεκκίνηση του Tomcat οι εφαρμογές ιστού θα είναι πλέον διαθέσιμες στους χρήστες από τη διεύθυνση '...../iatrolexi/'. Η εφαρμογή τρέχει μέσω του web browser Mozilla 2 ή του Internet Explorer 6. Σε περίπτωση που υπάρχει πρόβλημα με τα ελληνικά θα πρέπει να αλλάξει το encoding του εξυπηρετητή σε 'UTF-8'. Αυτό γίνεται στο configuration αρχείο '.../conf/server.xml' του Tomcat προσθέτοντας τη παράμετρο 'URIEncoding="UTF-8" ' στον εξυπηρετητή των HTTP σελίδων. Για παράδειγμα





<Connector port="8080" protocol="HTTP/1.1"

connectionTimeout="20000" enableLookups="true"

redirectPort="8443"

URIEncoding="UTF-8" />

#### 2.3 ΕΓΧΕΙΡΙΔΙΟ ΧΡΗΣΗΣ

Ο χρήστης μπαίνοντας στη σελίδα του «Ορθογραφικού Διορθωτή Ιστού» του ΙΑΤΡΟΛΕΞΗ βλέπει τη παρακάτω σελίδα:

🏉 ІАТРО	LEIH - Ελεγκτής Ορθογραφίας - Windows Internet Explorer	_ 🗆 🛛
00	👻 🔊 http://www.iatrolexi.gr/iatrolexi/webtools/speller/index.jsp 🔽 🗲 🗙 Live Search	P -
氨 SnagIt		
🚖 🏟	🧭 ΙΑΤΡΟΛΕΞΗ - Ελεγκτής Ορθογραφίας	🕶 🎯 Tools 👻 🎽
	Γληκτρολογήστε ή αντιγράψτε και επικολλήστε κείμενο για ορθογραφικό έλεγχο	
		A 1000/

Ο χρήστης έχει τη δυνατότητα να πληκτρολογήσει ή να αντιγράψει και επικολλήσει ένα κείμενο στον ειδικό χώρο και στην συνέχεια να ελέγξει την ορθογραφία του κειμένου πατώντας το κουμπί με τίτλο Έλεγχος Ορθογραφίας. Αν στο κείμενο δεν υπάρχουν ορθογραφικά λάθη εμφανίζεται το παρακάτω μήνυμα:





🟉 ІАТРО/	ΕΞΗ - Ελεγκτής Ορθογραφίας - Windows Internet Explorer	_ 🗆 🗙
$\odot$	🔹 🙋 http://www.iatrolexi.gr/iatrolexi/webtools/speller/index.jsp 💽 🐓 🗙 Live Search	<b>P</b>
氨 SnagIt		
🚖 🏟	💽 ΙΑΤΡΟΛΕΞΗ - Ελεγκτής Ορθογραφίας	🕶 🎯 Tools 👻 🎽
	Κληκτρολογήστε ή αντιγράψτε και επικολλήστε κείμενο για ορθογραφικό έλεγχο         Ο χρήστης έχει τη δυνατότητα να πληκτρολογήσει ή να αντιγράψει και επικολλήσει ένα κείμενο στον ειδικό χώρο και στην συνέχεια να ελέγξει την ορθογραφία του κειμένου πατώντας το κουμπί με τίτλο 'Ελεγχος Ορθογραφίας         Windows Internet Explorer       Μίπαι ανα ορθογραφια λάθη.         Οκ       Ο         Έλεγχος Ορθογραφίας       Καθαρισμός       Οδηγίες Χρήσης	
	💶 🔤 😜 Internet	€ 100% •

Αν υπάρχουν λανθασμένες λέξεις, τότε αυτές εμφανίζονται με κόκκινο. Ο χρήστης έχει την δυνατότητα να διορθώσει μια λέξη τοποθετώντας το ποντίκι πάνω σε μια λανθασμένη λέξη και επιλέγοντας μια λέξη από τη λίστα με τις προτεινόμενες λέξεις του ορθογράφου. Στην παρακάτω εικόνα δίνεται ένα παράδειγμα:





🖉 ІАТРО	\EIH - Ελεγκτής Ορθογραφίας - Windows Internet Explorer	
$\Theta$	<ul> <li>Inter://www.iatrolexi.gr/iatrolexi/webtools/speller/index.jsp</li> <li>Ity Karch</li> </ul>	P -
氨 SnagIt		
*	🧭 ΙΑΤΡΟΛΕΞΗ - Ελεγκτής Ορθογραφίας	🕶 🕥 Tools 👻 🎽
	Ο χρήστης έχει τη δυνατότιτα να πληκτρολογήσει ή να αντιγράψει και επικολλήσει ένα κείμενο στον ειδικό χώρο και στην συνέχια να ελέγξει την ορθογραφία του κειμένου πατώντας το κουμπί με τι συνέχεια συνέχεια συνέχεια συνέχεια συνέχει συνόχια συνέχει συνόχια συνέχει συνέχει συνέχει συνέχει	
	Επεξεργασία Οδηγίες Χρήσης	~
	😜 Internet	🔩 100% 🔻 .

Πατώντας το κουμπί με τίτλο 'Επεξεργασία' το κείμενο επανέρχεται στην αρχική του κατάσταση. Πατώντας το κουμπί με τίτλο 'Καθαρισμός', καθαρίζει η φόρμα και επανέρχεται στην αρχική της κατάσταση. Τέλος υπάρχουν οι οδηγίες χρήσης, οι οποίες εμφανίζονται σε ξεχωριστή σελίδα πατώντας το link με τίτλο 'Οδηγίες Χρήσης'. Η σελίδα με τις οδηγίες χρήσης παρουσιάζεται παρακάτω:





🏉 Οδηγίες Χρήσης Ελεγκτή Ορθογραφίας - Windows Internet Explorer 🗧 🗖 🗙				
00	Ittp://www.iatrolexi.gr/iatrolexi/webtools/speller/help.html	<b>P</b> -		
氨 SnagIt				
🚖 🎄	🍘 Οδηγίες Χρήσης Ελεγκτή Ορθογραφίας	Tools 🔹 👋		
	<b>Ελεγκτής Ορθογραφίας</b> Στόχος του Ελεγκτή Ορθογραφίας είναι να σας βοηθήσει να εντοπίσετε και να διορθώσετε ορθογραφικά λάθη μέσα σε ένα κείμενο.			
	Οδηγίες Χρήσης			
Πληκτρολογήστε ή αντιγράψτε και επικολλήστε μέσα στο πλαίσιο κειμένου το κείμενο που θέλετε να ελέγξετε και πατήστε το κουμπί "Ορθογραφικός Έλεγχος". Αν όλες οι λέξεις του κειμένου σας είναι γνωστές στο ορθογραφικό λεξικό, θα εμφανιστεί ένα παράθυρο διαλόγου με το μήνυμα "Δε βρέθηκαν ορθογραφικά λάθη.". Αν μία ή περισσότερες λέξεις δεν είναι γνωστές στο ορθογραφικό λεξικό και μπορούν να εντοπιστούν πιθανές εναλλακτικές τους, το κείμενό σας τίθεται σε κατάσταση ορθογραφικής διόρθωσης, όπου οι άγνωστες λέξεις φενείναι γνωστές στο ορθογραφικό λεξικό και μπορούν να εντοπιστούν πιθανές εναλλακτικές τους, το κείμενό σας τίθεται σε κατάσταση ορθογραφικής διόρθωσης, όπου οι άγνωστες λέξεις εμφανίζονται σε κόκκινο φόντο. Μεταφέροντας το δείκτη του ποντικιού πάνω σε μια λέξη με κόκκινο φόντο, θα εμφανιστεί ένα μενού με πιθανές - ορθογραφικά στότες- εναλλακτικές. Επιλέγοντας μια εναλλακτική από το μενού, αντικαθιστάτε την άγνωστη λέξη με αυτή, η οποία εμφανίζεται πλέον σε λευκό φόντο. Στην κατάσταση ορθογραφικής διόρθωσης δεν μπορείτε να επεξεργαστείτε το κείμενο παρά μόνο να αντικαταστήσετε τις λέξεις που βρίσκονται σε κόκκινο φόντο με κάποια από αυτές που προτείνει το ορθογραφικό λεξικό. Για να μεταβείτε σε κατάσταση <i>επεξεργασίας</i> , πατήστε το κουμπί "Επεξεργασία".				
		~		
Pone	Therefore the second se	076 -		

Πιο αναλυτική τεκμηρίωση της εφαρμογής, περιλαμβάνεται στο σχετικό παραδοτέο: Π18. Τεκμηρίωση εφαρμογών Ιστού.





## 3 ΟΔΗΓΙΕΣ ΧΡΗΣΗΣ ΤΟΥ ΟΠΤΙΚΟΥ ΔΙΣΚΟΥ

Στο παρόν παραδοτέο περιλαμβάνεται οπτικός δίσκος (CD) ο οποίος περιέχει το λογισμικό που υλοποιεί τον ορθογραφικό διορθωτή ιστού καθώς και την παρούσα αναφορά. Μόλις εισαχθεί ο οπτικός δίσκος στη μονάδα οπτικού δίσκου του υπολογιστή θα ανοίξει αυτόματα ο φυλλομετρητής με την παρουσίαση του παραδοτέου.



Στην πρώτη σελίδα παρουσιάζονται στοιχεία του έργου και του συγκεκριμένου παραδοτέου καθώς και δύο υπερσύνδεσμοι (links). Ο πρώτος οδηγεί στην αναφορά του παραδοτέου (παρούσα αναφορά) σε μορφή PDF και ο δεύτερος στο φακέλου του οπτικού δίσκου που περιέχει το λογισμικό του παραδοτέου.





## 4 ΣΧΕΤΙΚΟ ΛΟΓΙΣΜΙΚΟ

Apache Software Foundation: <u>http://www.apache.org/</u>

The source for Java developers: <u>http://java.sun.com/</u>

Open Source Initiative: <u>http://www.opensource.org/</u>

Tomcat: <u>http://tomcat.apache.org/</u>.



