23 Όταν το LSJ γνώρισε τη Βίκυ

Σπύρος Δόικας

ΠΕΡΙΛΗΨΗ

Η παρούσα εργασία αφορά στην **υλοποίηση σε μορφή MediaWiki** (όπως μπορεί κανείς να τη δει στη διεύθυνση http://lsj.translatum.gr) του αρχαιοελληνικού λεξικού ***Liddell, Scott, Jones (LSJ)***. Από την αρχική μορφή xml έγινε επεξεργασία και μετατροπή (με χρήση κανονικών εκφράσεων) σε αρχείο κατάλληλο για χρήση σε MediaWiki. Χαρακτηριστικά στοιχεία της υλοποίησης είναι **α)** η **μεταγραφή των λημμάτων σε διάφορες μορφές και μεταγραμματισμούς** (πολυτονικό με βραχύ/μακρό, πολυτονικό χωρίς βραχύ/μακρό, μονοτονικό, κεφαλαία, λατινικοί χαρακτήρες με τόνους, λατινικοί χαρακτήρες χωρίς τόνους, greeklish, Beta Code)· **β) δυνατότητα αναζήτησης τύπου «αυτόματης συμπλήρωσης»** χωρίς διάκριση πεζών/κεφαλαίων και χωρίς διάκριση διακριτικών σημείων τόσο για ελληνικούς όσο και για λατινικούς χαρακτήρες· **γ) δημιουργία στυλ css** για την τροποποίηση της εμφάνισης· **δ) συλλογή αρχαιοελληνικών αποφθεγμάτων** και δημιουργία επέκτασης για την εμφάνισή τους σε τυχαία σειρά· **ε) παραμετροποίηση του MediaWiki** για λεξικογραφικό έργο αυτής της μορφής **στ) δημιουργία προτύπου εισαγωγής** του αρχείου csv **με ενσωμάτωση των λειτουργιών Semantic Mediawiki** για τη δημιουργία ευρετηρίων βάσει μορφής και μεταγραμματισμού.

Wikifying the LSJ

Spiros Doikas

ABSTRACT

This paper relates the **implementation** of the **Ancient Greek to English dictionary *Liddell, Scott, Jones (LSJ)*** in **MediaWiki** format (http://lsj.translatum.gr). The original **xml** file was processed and converted (using regular expressions) to a file which was appropriate for use in MediaWiki. The main features of this implementation were: **a) transcription of headwords in various forms and transliterations** (polytonic with vrachy/macron, polytonic without vrachy/macron, monotonic, all caps, Latin characters with accents, Latin characters without accents, greeklish, Beta Code); **b) case-insensitive and diacritics-insensitive autocomplete search suggestions** for Greek and Latin characters; **c) css styles** to modify the look and feel; **d) collection of ancient Greek quotes** and development of a MediaWiki **random quote extension**; **e) fine-tuning MediaWiki** in a way that is appropriate for a lexicographic work of this nature; **f)** creation of an **import template** that supports **Semantic Mediawiki functionality**; and **g) creation of indexes** for each form and transliteration.

0 Εισαγωγή

Η **Βίκυ**, ή για την ακρίβεια το λογισμικό τύπου «βίκι» με το όνομα **MediaWiki**, έκανε την εμφάνισή του στον μάταιο τούτο κόσμο στις 8 Οκτωβρίου 2003 με την έκδοση MediaWiki 1.1[1]. Το **LSJ** (για την ακρίβεια το λεξικό αρχαίων ελληνικών προς αγγλικά των **Henry George Liddell**, **Robert Scott** και **Henry Stuart Jones**), πήρε σάρκα και οστά, στην πρώτη του μορφή, το 1843[2] από τις εκδόσεις της Οξφόρδης. Το έργο βασίστηκε στο *Handwörterbuch der griechischen Sprache* του γερμανού λεξικογράφου **Franz Passow**[3] (ημερομηνία πρώτης δημοσίευσης: 1819), το οποίο με τη σειρά του βασίστηκε στο *Kritisches griechisch-deutsches Handwörterbuch* του **Johann Gottlob Schneider**[4] (ημερομηνία πρώτης δημοσίευσης: 1797).

Όπως είναι σαφές από τα παραπάνω, 130 ολόκληρα έτη χωρίζουν τη Βίκυ από το LSJ. Σε άλλες περιπτώσεις μια τόσο μεγάλη διαφορά ηλικίας θα λειτουργούσε αποτρεπτικά. Εν προκειμένω, δεν υπήρξε εμπόδιο στο να γνωριστούν, και όχι μόνο αυτό, αλλά και να ερωτευτούν, *σφόδρα*. Καρπός μάλιστα του έρωτα αυτού, μετά από πολυάριθμες και γεμάτες πάθος συνευρέσεις που έλαβαν χώρα μεταξύ του περίπλοκου γεννητικού κώδικα php της Βίκυς και του σχοινοτενούς αρχαιοελληνικού λημματολογίου του LSJ, υπήρξε ένα χαριτωμένο μωρό με ευχέρεια, από γεννησιμιού του, και στα δύο –τόσο ετερόμορφα– (γον)ιδιώματα των γονιών του. Το μωρό αυτό βαπτίστηκε ανήμερα του Αγίου Βαλεντίνου, 14 Φεβρουαρίου 2013, με παπά και με κουμπάρο, *Translatum LSJ* στον Ιερό Ναό του Οσίου Συμεών του Μεταφραστή, με διεύθυνση IP 74.200.70.96.

Για να γίνουν πραγματικότητα τα παραπάνω, μια σειρά από εργασίες έλαβαν χώρα και μια σειρά προβλημάτων έπρεπε να λυθούν. Κατ' αρχάς καταστρώθηκε ένα σχέδιο με τα ζητούμενα του έργου και στη συνέχεια διερευνήθηκαν οι τρόποι υλοποίησής τους. Οι εργασίες ταξινομήθηκαν εδώ, χάριν ευκολίας, ανάλογα με τη φύση τους σε εργασίες παραμετροποίησης και διαμόρφωσης διακομιστή, εργασίες παραμετροποίησης του MediaWiki και εγκατάστασης επεκτάσεων, εργασίες ανάπτυξης ειδικών επεκτάσεων, εργασίες για την προετοιμασία του λημματολογίου και τη μετατροπή του από xml σε html και εργασίες μετατροπής των λημμάτων σε διάφορες μορφές και μεταγραμματισμούς. Ας πάρουμε όμως τα πράγματα από την αρχή.

---

1 http://en.wikipedia.org/wiki/MediaWiki_version_history

2 http://en.wikipedia.org/wiki/A_Greek%E2%80%93English_Lexicon

3 http://de.wikipedia.org/wiki/Handw%C3%B6rterbuch_der_griechischen_Sprache

4 http://en.wikipedia.org/wiki/Johann_Gottlob_Theaenus_Schneider

1 Εργασίες στο λημματολόγιο

1.1 Προεπεξεργασία

Το αρχικό λημματολόγιο είχε μορφή xml. Για παράδειγμα το λήμμα «ανεπίληστος»:

```
<div2 id="crossa)nepi/lhstos" orig ¯ id="n8400" key="a)nepi/lhstos" type="main"
opt="n"><span class=head>ἀνεπίληστος</span><b>ἀνεπί-ληστος</b>, <itype lang="greek"
opt="n">ον</itype>, <span class=sense><p>   <b>A</b> <i>not to be
forgotten,</i> <span class=bibl><author>Aristaenet.</author>2.13</span>,
<author>Hsch.</author> s.v. <ref type="cross" target="crossa)lastoi=s"
lang="greek">ἀλαστοῖς</ref>. <pos opt="n">Adv.</pos> <span class=foreign>-τως</span>
Sch.<span class=bibl><author>Od.</author>14.174</span>.</span></div2>
```

Έφτασε σε τελική μορφή να είναι:

```
ον, <span class="sense"><p>   <span class="bld">A</span> <b
class="b2">not to be forgotten</b>, <span class="bibl">Aristaenet.2.13</span>, Hsch.
s.v. [[ἀλαστοῖς]]. Adv. <b class="b3">-τως</b> Sch.<span
class="bibl">Od.14.174</span>.</span>
```

Μια σειρά εργασίες εκτελέστηκαν στο λημματολόγιο έτσι ώστε να έρθει σε μορφή html που να του επιτρέπει την εισαγωγή του στο MediaWiki.

1.2 Μετατροπή σε δίστηλο

Η πρώτη εργασία είχε να κάνει με την μετατροπή του σε δίστηλο όπου η πρώτη στήλη είναι το λήμμα και η δεύτερη η ερμηνεία. Γι' αυτό τον σκοπό εκτελέστηκε μια σειρά εντολών εύρεσης/αντικατάστασης έτσι ώστε να προστεθεί ένας χαρακτήρας στηλοθέτη μεταξύ λήμματος και ερμηνείας. Αυτές οι εντολές είχαν τη μορφή κανονικών εκφράσεων (έτσι ώστε να καλύψουν όλες τις δυνατές περιπτώσεις) και τα προγράμματα που χρησιμοποιήθηκαν ήταν κατά κύριο λόγο το **EmEditor**[5] και δευτερεύοντα το **Notepad++**[6].

Ο λόγος για την επιλογή του EmEditor είναι κατ' αρχάς η δυνατότητα που προσφέρει για εργασία με πολύ μεγάλα αρχεία. Εν προκειμένω, το αρχείο ήταν 80 MB και το άνοιγμά του γινόταν σε κλάσμα δευτερολέπτου.

Σε δεύτερη φάση, έγινε μεταφορά του δίστηλου αρχείου σε Excel 2010. Ο λόγος που έγινε η επιλογή του Excel 2010 είναι ότι σε αντίθεση με το Excel 2003 για παράδειγμα, υποστηρίζει έως 1.048.576 σειρές και έως 32.767 χαρακτήρες ανά κελί[7]. Στην περίπτωσή μας το λημματολόγιο αριθμούσε περί τους 120.000 όρους.

---

5 http://www.emeditor.com/

6 http://notepad-plus-plus.org/

7 Βλέπε: http://office.microsoft.com/en-us/excel-help/excel-specifications-and-limits-HP010342495.aspx. Ωστόσο το όριο χαρακτήρων στο κελί αποδείχτηκε τελικά μικρό (η μεγαλύτερη ερμηνεία έφτανε τους 100.000 χαρακτήρες), περικόπτοντας κάποιες ερμηνείες, με αποτέλεσμα να χρειαστεί να γίνει επιπλέον επεξεργασία 200 περίπου όρων κατά την τελική μετατροπή σε αρχείο csv.

1.3 Προεπεξεργασία στήλης όρων

Σε τρίτη φάση, έγινε περαιτέρω επεξεργασία στα λήμματα (στην πρώτη στήλη δηλαδή): **α)** εντοπίστηκαν τα λήμματα τα οποία δεν είχαν στηλοθέτη και ως εκ τούτου δεν χωρίστηκαν σε δύο στήλες και έγινε μη αυτόματος χωρισμός τους· **β)** αφαιρέθηκαν σύμβολα στο τέλος των λημμάτων όπως αστερίσκοι, άνω τελείες, παύλες, κ.τ.λ. **γ)** εντοπίστηκαν περιπτώσεις με κεφαλαία όπου τα διακριτικά δεν ήταν συνδυασμένα με τα γράμματα, και έγινε αντικατάσταση με την συνδυασμένη μορφή διακριτικού και γράμματος (δηλαδή αντί να είναι ένας χαρακτήρας ήταν δύο, όπου π.χ. η δασεία και η οξεία ήταν ένας χαρακτήρας και το γράμμα ένας διαφορετικός) **δ)** εντοπίστηκαν όλα τα λήμματα που επαναλαμβάνονταν με εννοιολογικές υποδιαιρέσεις σε A, B, C, D, στη συνέχεια ομαδοποιήθηκαν, μεταφέρθηκαν σε νέα αρχεία, έγινε περαιτέρω ομαδοποίηση στα λήμματα ανάλογα με τον αριθμό υποδιαιρέσεων και, χρησιμοποιώντας ειδικά γραμμένες μακροεντολές μέσω Excel (μία για κάθε περίπτωση πολλαπλότητας), έγινε συνένωση των ερμηνειών για κάθε λήμμα με πολλαπλές υποδιαιρέσεις. Έτσι, για παράδειγμα, αντί να έχει κανείς τρία ξεχωριστά λήμματα για το «αρκτικός», έχει μόνο ένα στον οποίο θα βρει και τις τρεις νοηματικές υποδιαιρέσεις του[8]. Ακολουθεί η μακροεντολή για τις περιπτώσεις με δύο επαναλήψεις:

```
Public Sub Transpose_2()
Dim rowNum As Long, lr As Long
lr = Range("D" & Rows.Count).End(xlUp).Row
    For rowNum = 2 To lr Step 2
        Range("D" & rowNum).Offset(-1, 1).Value = Range("D" & rowNum).Value
        Range("D" & rowNum).ClearContents
    Next rowNum
End Sub
```

Και το αποτέλεσμα της μακροεντολής (πριν και μετά την εκτέλεση):

	A	B	C	D	E	F	G	H	I	J
1	28	a)/atos1	ἄατος	ἄᾱτος, <gram	Grp opt="n"><gram type="var" opt="n">contr.</grar					
2	29	a)/atos2	ἄατος	ἄᾱτος, <itype lang="greek" opt="n">ον</itype>, <span class=sense						
3	90	a)/bios1	ἄβιος	ἄβιος, <itype lang="greek" opt="n">ον</itype> (A), <span class=se						
4	91	a)/bios2	ἄβιος	ἄβιος, (B), <itype lang="greek" opt="n">ον</itype>, (<itype lang="g						

[8] Παράβαλε για παράδειγμα το «ἀρκτικός» από το Translatum LSJ και το «ἀρκτικός» από το TLG LSJ (http://www.tlg.uci.edu/lsj/) ή το Perseus LSJ. Επιπλέον, το TLG LSJ αν και πραγματοποιεί ανευρέσεις σε άτονες λέξεις, όταν βάζει κανείς έστω και έναν τόνο δεν εμφανίζει αποτελέσματα αυτόματης συμπλήρωσης.

	A	B	C	D	E	F	G	H	I	J
1	28	a)/atos1	ἄατος	ἄᾱτος, <itype lang="greek" opt="n">ον</itype>, <span c						
2	29	a)/atos2	ἄατος							
3	90	a)/bios1	ἄβιος	ἄβιος, (Β), <itype lang="greek" opt="n">ον</itype>, (<it						
4	91	a)/bios2	ἄβιος							

Βλέπουμε στη δεύτερη εικόνα τα περιεχόμενα από το κελί 2D και 4D της πρώτης, μεταφέρθηκαν στις θέσεις 1E και 3E της δεύτερης. Στη συνέχεια έγινε ταξινόμηση βάσει της στήλης D, έτσι ώστε να διαχωριστούν και μετά να διαγραφούν οι σειρές με κενά στην D.

	A	B	C	D	E	F	G	H	I	J
1	28	a)/atos	ἄατος	ἄᾱτος, <itype lang="greek" opt="n">ον</itype>, <span cl						
2	90	a)/bios	ἄβιος	ἄβιος, (Β), <itype lang="greek" opt="n">ον</itype>, (<ity						

Κατόπιν, έγινε συγχώνευση[9] των στηλών D και E προσθέτοντας ως διαχωριστικό το σύμβολο αλλαγής γραμμής στην xhtml (
).

1.4 Προεπεξεργασία στήλης ερμηνειών

Το MediaWiki εξ ορισμού υποστηρίζει κάποιες βασικές ετικέτες html[10]. Δεν αρκούν όμως για τη σωστή απεικόνιση του πηγαίου κώδικα xml. Κάτι τέτοιο θα ήταν δυνατόν με τη χρήση της παραμέτρου $wgRawHtml στο αρχείο LocalSettings.php, πράγμα που όμως θα ενείχε αφενός σοβαρά προβλήματα ασφαλείας (καθώς οποιοσδήποτε χρήστης με δυνατότητα επεξεργασίας του λήμματος θα μπορούσε να εκτελέσει κακόβουλο κώδικα), αφετέρου θα προσέδιδε ιδιαίτερα πολύπλοκη δομή στον κώδικα δυσκολεύοντας την «αναγνωσιμότητά» του και την επεξεργασία του από τους χρήστες.

Έτσι, χρησιμοποιήθηκε μια πλειάδα κανονικών εκφράσεων (γύρω στις 40) για τη μετατροπή του κώδικα xml σε κώδικα html κατάλληλου για βικισελίδες. Παρακάτω ακολουθούν κάποια παραδείγματα (μέσω **EmEditor**).

Για μετατροπή παραπομπών σε εσωτερικές παραπομπές τύπου MediaWiki:

Εύρεση: (<ref type="cross" target=")([^<]+)(lang="greek">)([^<]+)(</ref>)

Αντικατάσταση: [[\4]]

Για δημιουργία στυλ με έντονη γραφή σε υπολήμματα με ελληνικούς χαρακτήρες:

Εύρεση: ()([^<]+)()

Αντικατάσταση: \2

---

[9] Χρησιμοποιήθηκε η μακροεντολή **Merge column data** του προσθέτου για Excel **ASAP-Utilities** (http://www.asap-utilities.com/)

[10] http://meta.wikimedia.org/wiki/Help:HTML_in_wikitext

Ωστόσο, δεν ήταν δυνατή η αποφυγή κάποιων σφαλμάτων σε επίπεδο html (όπως κάποιες ενθυλακωμένες ετικέτες «span» που εμφανίζονταν στη σελίδα html μαζί με το κείμενο) τα οποία επιλύθηκαν με χρήση της παραμέτρου $wgUseTidy[11] στο LocalSettings.php.

2 Μετατροπή των λημμάτων σε διάφορες μορφές και μεταγραμματισμούς

Για την ευκολία του μελετητή έγινε μετατροπή των λημμάτων σε διάφορες μορφές και μεταγραμματισμούς: α) **πολυτονικό με βραχύ/μακρό** (Full diacritics) β) **πολυτονικό χωρίς βραχύ/μακρό** (Medium diacritics) γ) **μονοτονικό** (Low diacritics) δ) **κεφαλαία** (Capitals) ε) **λατινικοί χαρακτήρες με τόνους** (Transliteration A) στ) **λατινικοί χαρακτήρες χωρίς τόνους** (Transliteration B) ζ) **greeklish** (Transliteration C) η) **Beta Code** (Beta Code). Συνολικά 8 διαφορετικές μορφές. Η πρώτη μορφή (πολυτονικό με βραχύ/μακρό) ήταν η αρχική μορφή του πηγαίου λημματολογίου. Για τις υπόλοιπες ακολουθήθηκε μια σειρά από διαφορετικούς τρόπους μετατροπής. Για τα Medium diacritics δημιουργήθηκε μια μακροεντολή που αφαιρούσε το βραχύ και το μακρόν από τα αντίστοιχα γράμματα (διατηρώντας το όποιο άλλο διακριτικό υπήρχε). Για το Low diacritics χρησιμοποιήθηκε μια μακροεντολή μετατροπής σε μονοτονικό από τις μακροεντολές του Translatum[12], καθώς και η αντίστοιχη εφαρμογή του Translatum σε php[13]. Η μετατροπή σε κεφαλαία έγινε μέσω της λειτουργίας του MS Word (Shift+F3). Οι μεταγραμματισμοί Transliteration A και Transliteration B έγιναν μέσω του διαδικτυακού εργαλείου transliterate.com. Η μεταγραφή σε greeklish έγινε με κείμενο εισόδου το Low diacritics χρησιμοποιώντας την εφαρμογή του ΙΕΛ *All Greek to me*[14].

2.1 Beta Code

Για τη μετατροπή σε Beta Code ακολουθήθηκε το πρότυπο του Perseus (με μετατροπή σε πεζά αντί για κεφαλαία όπως γίνεται στο TLG) και χρησιμοποιήθηκε εφαρμογή σε php που αναπτύχθηκε στο translatum ειδικά γι' αυτόν τον σκοπό. Κατ' αρχάς, ανεξάρτητα από το αν ο μεταγραμματισμός γίνεται όλα κεφαλαία ή όλα πεζά, οι χαρακτήρες που θα είναι κεφαλαία στα αρχαία ελληνικά καθορίζονται από ένα αστερίσκο που προηγείται του γράμματος. Για παράδειγμα η λέξη **Πλάτων** μεταγράφεται ως ***PLA/TWN** ή ***pla/twn**. Το σύμβολο της πλαγίας (/) μετά το άλφα συμβολίζει την οξεία.

---

11 http://www.mediawiki.org/wiki/Manual:$wgUseTidy

12 http://www.translatum.gr/forum/index.php?topic=319749.0

13 http://www.translatum.gr/forum/index.php?topic=159473.0

14 http://www.ilsp.gr/el/services-products/products/item/2-langtech/30-all-greek-to-me

Εδώ ίσως αξίζει να αναφέρουμε ότι υπάρχουν αρκετές ασάφειες σχετικά με το πρότυπο Beta Code και διαφορετικές εκδοχές μεταγραμματισμού τόσο για ορισμένους χαρακτήρες όσο και για ορισμένους συνδυασμούς διακριτικών σε χαρακτήρες.

Για παράδειγμα για τα σύμβολα «σ» και «ς» υπάρχουν οι εξής εναλλακτικές. α) μετατροπή και των δύο σε **S**· β) μετατροπή του μεσαίου σίγμα σε **S** και του τελικού σε **J**· γ) μετατροπή του μεσαίου σίγμα σε **S** και του τελικού σε **S2**· δ) μετατροπή του μεσαίου σίγμα σε **S1** και του τελικού σε **S2**.

Άλλο ένα θέμα το οποίο δεν είναι σαφές στο πρότυπο είναι η σειρά αναγραφής των διακριτικών κατά τον μεταγραμματισμό χαρακτήρων με διαλυτικά. Σύμφωνα με το πρότυπο[15], ισχύουν τα ακόλουθα.

Οι χαρακτήρες στα πεζά γράμματα εισάγονται με την ακόλουθη σειρά: 1) γράμμα, 2) πνεύμα, 3) τόνος, 4) υπογεγραμμένη. Π.χ. το **W(=|** αντιστοιχεί στο **ᾧ** (πεζό ωμέγα με δασεία, περισπωμένη και υπογεγραμμένη).

Οι χαρακτήρες στα κεφαλαία γράμματα εισάγονται με την ακόλουθη σειρά: 1) αστερίσκος 2) πνεύμα, 3) τόνος, 4) γράμμα, 5) υπογεγραμμένη. Π.χ. το ***(=W|** αντιστοιχεί στο **ὯΙ** (κεφαλαίο ωμέγα με δασεία, περισπωμένη και υπογεγραμμένη).

Από τα παραπάνω βλέπουμε ότι ενώ στα πεζά το γράμμα γράφεται πάντα πρώτο, στα κεφαλαία γράφεται ορισμένες φορές ακόμη και τέταρτο.

Ωστόσο, στις περιπτώσεις με διαλυτικά και τόνο, για παράδειγμα το **ΰ** (ύψιλον με διαλυτικά και οξεία) μπορεί να μεταγραμματιστεί **U+/** ή **U/+**.

Μια ακόμη αδυναμία του προτύπου είναι ότι ορισμένοι σπάνιοι χαρακτήρες μεταγραμματίζονται με πιο αδιαφανή τρόπο, χρησιμοποιώντας το σύμβολο τοις εκατό (%) και έναν διψήφιο αριθμό. Π.χ. το μακρό μεταγράφεται ως **%26** και το βραχύ ως **%27**.

3 Εργασίες στο MediaWiki

Η έκδοση του MediaWiki που χρησιμοποιήθηκε ήταν η 1.16.5. Οι επεκτάσεις που χρησιμοποιήθηκαν ήταν: SemanticMediaWiki (SMW), Semantic Drilldown, Data Transfer, Parser Functions, MWSearch καθώς και επεκτάσεις που αναπτύχθηκαν ειδικά για το έργο για τη δημιουργία των ευρετηρίων (Normalizer) και την εμφάνιση των αποφθεγμάτων.

[15] http://www.tlg.uci.edu/encoding/

3.1 Αρχείο LocalSettings.php

Το αρχείο LocalSettings.php[16] είναι το κέντρο ελέγχου του MediaWiki μέσω του οποίου είναι δυνατή η παραμετροποίησή του και η διασύνδεσή του με επεκτάσεις που εμπλουτίζουν ή τροποποιούν τη λειτουργικότητά του. Κάποιες από τις παραμέτρους είναι ζωτικής σημασίας για ένα λεξικογραφικό έργο. Για παράδειγμα, από προεπιλογή, το MediaWiki μετατρέπει αυτόματα σε κεφαλαίο το πρώτο γράμμα ενός λήμματος. Κάτι τέτοιο δεν είναι θεμιτό σε ένα λεξικογραφικό έργο όπου η εκδοχή με κεφάλαιο μπορεί να είναι κάτι τελείως διαφορετικό από την εκδοχή με πεζό, π.χ. Ίρις (θεότητα) και ίρις (ίριδα του ματιού). Στη Βικιπαίδεια, τέτοιες περιπτώσεις καλύπτονται από τη λειτουργικότητα των σελίδων αποσαφήνισης η οποία παραπέμπει σε περαιτέρω σελίδες ανάλογα με το θεματικό πεδίο ή τη σημασία, π.χ. «Ίρις (μυθολογία)», «Ίρις (οπτική)», «Ίρις (μετεωρολογία)», κ.τ.λ. Αυτή η προσέγγιση αν και είναι αποτελεσματική για ένα εγκυκλοπαιδικό έργο, δεν θεωρείται κατάλληλη για ένα λεξικογραφικό έργο. Γι' αυτό και δεν χρησιμοποιείται στο Βικιλεξικό. Αντιθέτως, το πρώτο γράμμα των λημμάτων *δεν* μετατρέπεται αυτόματα σε κεφαλαίο. Αυτό επιτυγχάνεται με τη προσθήκη μιας παραμέτρου σχετικής με τα κεφαλαία που επιδέχεται ως τιμή το true ή το false στο εν λόγω αρχείο ρυθμίσεων[17]:

```
$wgCapitalLinks = false;
```

Η εγκατάσταση επεκτάσεων στο MediaWiki γίνεται συνήθως προσθέτοντας τον φάκελο της επέκτασης στον κατάλογο *extensions* και προσθέτοντας τουλάχιστον μία γραμμή στο αρχείο LocalSettings.php για να δηλώσουμε την παρουσία (ή και τον τρόπο λειτουργίας της επέκτασης). Για παράδειγμα η παρακάτω γραμμή διασυνδέει το MediaWiki με την επέκταση SMW Η δεύτερη γραμμή είναι απαραίτητη για την ενεργοποίηση των σημασιολογικών λειτουργιών του SMW στην οποία απαιτείται και ο καθορισμός του ονόματος τομέα του ιστότοπου.

```
require_once("extensions/SemanticMediaWiki/SemanticMediaWiki.php" );
enableSemantics( $wgSitename .'.translatum.gr' );
```

3.2 Εισαγωγή δεδομένων

Η εισαγωγή των δεδομένων έγινε σε δέσμες των 10.000 καταχωρίσεων (μεγαλύτερος αριθμός είχε ως αποτέλεσμα τη μη εισαγωγή τους) μέσω της επέκτασης Data Transfer[18]. Είχε δύο βασικά σκέλη, την εισαγωγή των καθαυτό λημμάτων και την εισαγωγή των

[16] http://www.mediawiki.org/wiki/Manual:LocalSettings.php

[17] http://www.mediawiki.org/wiki/Manual:$wgCapitalLinks

[18] http://www.mediawiki.org/wiki/Extension:Data_Transfer

ανακατευθύνσεων. Το σύνολο των εγγραφών ήταν γύρω στις 240.000. Ένα από τα προβλήματα που ανέκυψαν ήταν η μη εμφάνιση των υπόλοιπων χαρακτήρων στο πεδίο Beta Code όταν περιείχε τον χαρακτήρα της κατακόρυφης ράβδου (|). Το πρόβλημα λύθηκε με την αντικατάσταση του εν λόγω χαρακτήρα στο αρχείο εισαγωγής με την αντίστοιχη αριθμητική οντότητα html (**|**).

Το MediaWiki εκτελεί τις εκκρεμείς εργασίες (και ως τέτοιες εκλαμβάνονται οι εγγραφές που εισάγονται μέσω της επέκτασης Data Transfer) βάσει μιας δέσμης ενεργειών συντήρησης με την ονομασία runJobs.php[19]. Η προεπιλεγμένη ταχύτητα εκτέλεσης είναι μία εργασία ανά σελιδοπροβολή, ωστόσο, αυτό μπορεί να ρυθμιστεί μέσω του LocalSettings.php ή με απευθείας εκτέλεση από τη γραμμή εντολών. Η δεύτερη επιλογή ήταν και αυτή που προτιμήθηκε με χρησιμοποίηση της εντολής:
cd /home/site/public_html/w/wiki/maintenance/ php runJobs.php --maxjobs 10000
Η παράμετρος *--maxjobs* στο τέλος ορίζει τον μέγιστο αριθμό εργασιών προς εκτέλεση.

4 Semantic Mediawiki

Το Semantic MediaWiki (SMW)[20] είναι μια επέκταση για το MediaWiki η οποία επιτρέπει την ενσωμάτωση σημασιολογικών δεδομένων (μεταδεδομένων που περιγράφουν τα δεδομένα) σε βικισελίδες. Στη συνέχεια αυτά τα δεδομένα μπορούν να χρησιμοποιηθούν σε σημασιολογικές αναζητήσεις, κατηγοριοποιήσεις σελίδων και για άλλες χρήσεις. Για το SMW έχουν αναπτυχθεί συμπληρωματικές επεκτάσεις για συγκεκριμένες λειτουργίες. Για παράδειγμα στο παρόν έργο χρησιμοποιήθηκε η επέκταση Semantic Drilldown[21] για τη δημιουργία των ευρετηρίων.

4.1 Ευρετήρια

Τα ευρετήρια[22] βασίστηκαν στις μορφές και τους μεταγραμματισμούς που δημιουργήθηκαν για κάθε λήμμα και όπως αυτά ορίστηκαν με τον κώδικα του προτύπου εισαγωγής. Αυτός ο κώδικας διασυνδέει τις διαφορετικές μορφές και μεταγραμματισμούς με την επέκταση SMW, έτσι ώστε να είναι δυνατή η περαιτέρω αξιοποίησή τους για τη δημιουργία ξεχωριστών ευρετηρίων ανά μορφή/μεταγραμματισμό. Ακολουθεί δείγμα ευρετηρίου σε Beta Code.

---

19 http://www.mediawiki.org/wiki/Manual:RunJobs.php

20 http://www.mediawiki.org/wiki/Extension:Semantic_MediaWiki

21 http://www.mediawiki.org/wiki/Extension:Semantic_Drilldown

22 http://lsj.translatum.gr/wiki/Index:Contents

- ba=
- ba/bac
- babai/
- babaia/c
- ba/baka
- baba/kinos
- ba/bakoi
- baba/kths
- ba/balon
- baba/zw
- babe/lios
- babh/r
- ba/bion
- babi/zw
- babra/zw
- ba/brhkes
- babukw/s
- *babulw/n
- babu/ras
- babu/rtas
- ba/cis
- ba=con
- bada/s
- baddi/n
- ba/dhn
- ... further results

4.2 Πρότυπο εισαγωγής

Το πρότυπο εισαγωγής που χρησιμοποιήθηκε είναι το παρακάτω:

```
[[NormalizedLowDiacritics::{{#normalize:{{{Low diacritics|}}}}}|]]
[[NormalizedCapitals::{{#normalize:{{{Capitals|}}}}}|]]
[[NormalizedTransliterationA::{{#normalize:{{{Transliteration A|}}}}}|]]
[[NormalizedTransliterationB::{{#normalize:{{{Transliteration B|}}}}}|]]
[[NormalizedTransliterationC::{{#normalize:{{{Transliteration C|}}}}}|]]
[[NormalizedBetaCode::{{#normalize:{{{Beta Code|}}}}}|]] {| class="tabl1"
|-
| Full diacritics: <b class="b1">[[Full diacritics::{{{Full diacritics|}}}]]
</b>
| Medium diacritics: <b class="b1">[[Medium diacritics::{{{Medium
diacritics|}}}]]</b>
| Low diacritics: <b class="b1">[[Low diacritics::{{{Low diacritics|}}}]]</b>
| Capitals: <b class="b1">[[Capitals::{{{Capitals|}}}]]</b>
|-
| Transliteration A: <b class="b1">[[Transliteration A::{{{Transliteration
A|}}}]]</b>
| Transliteration B: <b class="b1">[[Transliteration B::{{{Transliteration
B|}}}]]</b>
| Transliteration C: <b class="b1">[[Transliteration C::{{{Transliteration
C|}}}]]</b>
| Beta Code: <b class="b1">[[Beta Code::{{{Beta Code|}}}]]</b>
|}
{{{Definition|}}}
[[Category:LSJ]][[Category:Ancient Greek to English Dictionary]]</includeonly>
```

Στην ουσία πρόκειται για ανάμειξη τριών στοιχείων: κώδικα html, βικικώδικα και php. Οι εντολές **#normalize** παραπέμπουν σε ειδική συνάρτηση κανονικοποίησης[23] για την αφαίρεση των διακριτικών έτσι ώστε να είναι δυνατή η σωστή ταξινόμηση στα διαφορετικά ευρετήρια. Τα στοιχεία τύπου **<b class="b1">** ορίζουν το στυλ εμφάνισης (css) του εκάστοτε στοιχείου. Οι ονομασίες μεταγραμματισμών / μορφών όπως «Beta Code» αντιστοιχούν στα πεδία του αρχείου εισαγωγής csv. Στο τέλος, τα στοιχεία τύπου

[23] Π.χ. μια γραμμή: $norm = preg_replace('/[ΕεΈέἐἑἒἓἔἕἘἙἚἛἜἝὲέῈΈ]/u', 'Ε', $norm);

[[Category:LSJ]] προσθέτουν τις κατηγορίες στις οποίες υπάγεται το κάθε λήμμα. Παρακάτω, δείγμα από το αρχείο εισαγωγής όπως φαίνεται στο EmEditor σε μορφή csv και πινακοποιημένη προβολή. Στη γραμμή 1 βρίσκεται η ονομασία του προτύπου (**LSJ1**) ακολουθούμενη από το όνομα του εκάστοτε πεδίου προς εισαγωγή (π.χ. **LSJ1[Capitals]**).

	Title,	LSJ1[Full d	LSJ1[Med	LSJ1[Low	LSJ1[Capita	LSJ1[Transli	LSJ1[Transl	LSJ1[Transli	LSJ1[Beta	CLSJ1[Definition]↓
2	νωσάμενος,	νωσάμενος,	νωσάμενο	νωσάμενος	ΝΩΣΑΜΕΝΟΣ,	nōsámenos,	nōsamenos,	nosamenos,	nwsa/menos,	"νώσασθαι, <span cl
3	νῶσις,	νῶσις,	νῶσις,	νώσις,	ΝΩΣΙΣ,	nō^sis,	nōsis,	nosis,	nw=sis,	"<span class=""sens
4	νωταγωγέω,	νωτᾰγωγέω,	νωταγωγέ	νωταγωγέω	ΝΩΤΑΓΩΓΕΩ,	nōtagōgéō,	nōtagōgeō,	notagogeo,	nwtagwge/w,	"<span class=""sens
5	νωταγωγός,	νωτᾰγωγός,	νωταγωγό	νωταγωγός	ΝΩΤΑΓΩΓΟΣ,	nōtagōgós,	nōtagōgos,	notagogos,	nwtagwgo/s,	"όν, <span class=""
6	νωταῖος,	νωταῖος,	νωταῖος,	νωταίος,	ΝΩΤΑΙΟΣ,	nōtaîos,	nōtaios,	notaios,	nwtai=os,	"α, ον, poet. <span
7	νωτάκμων,	νωτάκμων,	νωτάκμων	νωτάκμων,	ΝΩΤΑΚΜΩΝ,	nōtákmōn,	nōtakmōn,	notakmon,	nwta/kmwn,	"ονος, ὁ, ἡ, <span
8	νωτάρης,	νωτάρης,	νωτάρης,	νωτάρης,	ΝΩΤΑΡΗΣ,	nōtárēs,	nōtarēs,	notaris,	nwta/rhs,	"ες, (αἴρω) <span c
9	νωτεύς,	νωτεύς,	νωτεύς,	νωτεύς,	ΝΩΤΕΥΣ,	nōteús,	nōteus,	noteys,	nwteu/s,	"έως, ὁ, <span clas
10	νωτηγός,	νωτηγός,	νωτηγός,	νωτηγός,	ΝΩΤΗΓΟΣ,	nōtēgós,	nōtēgos,	notigos,	nwthgo/s,	"όν<b class=""b3"">
11	νωτιαῖος,	νωτιαῖος,	νωτιαῖος	νωτιαίος,	ΝΩΤΙΑΙΟΣ,	nōtiaîos,	nōtiaios,	notiaios,	nwtiai=os,	"α, ον, <span class
12	νωτιάς,	νωτιάς,	νωτιάς,	νωτιάς,	ΝΩΤΙΑΣ,	nōtiás,	nōtias,	notias,	nwtia/s,	"άδος, ἡ, fem. Adj.
13	νωτιδανός,	νωτῐδᾰνός,	νωτιδανό	νωτιδανός	ΝΩΤΙΔΑΝΟΣ,	nōtidanós,	nōtidanos,	notidanos,	nwtidano/s,	"ὁ, a kind of <b cl
14	νωτίζω,	νωτίζω,	νωτίζω,	νωτίζω,	ΝΩΤΙΖΩ,	nōtízō,	nōtizō,	notizo,	nwti/zw,	"only in aor. exc.
15	νώτιος,	νώτιος,	νώτιος,	νώτιος,	ΝΩΤΙΟΣ,	nṓtios,	nōtios,	notios,	nw/tios,	"ον, collat. form c
16	νώτισμα,	νώτισμα,	νώτισμα,	νώτισμα,	ΝΩΤΙΣΜΑ,	nṓtisma,	nōtisma,	notisma,	nw/tisma,	"ατος, τό, (<b clas
17	νωτοβατέω,	νωτοβᾰτέω,	νωτοβατέ	νωτοβατέω	ΝΩΤΟΒΑΤΕΩ,	nōtobatéō,	nōtobateō,	notovateo,	nwtobate/w,	"<span class=""sens
18	νωτόγραπτος	νωτόγραπτος	νωτόγραπ	νωτόγραπτ	ΝΩΤΟΓΡΑΠΤΟΣ	nōtógraptos,	nōtograptos	notograptos,	nwto/grapto	"ον, <span class=""
19	νωτοκοπέω,	νωτοκοπέω,	νωτοκοπέ	νωτοκοπέω	ΝΩΤΟΚΟΠΕΩ,	nōtokopéō,	nōtokopeō,	notokopeo,	nwtokope/w,	"<span class=""sens
20	νῶτον,	νῶτον,	νῶτον,	νώτον,	ΝΩΤΟΝ,	nō^ton,	nōton,	noton,	nw=ton,	"τό, or νῶτος, ὁ, p
21	νωτοπλήξ,	νωτοπλήξ,	νωτοπλήξ	νωτοπλήξ,	ΝΩΤΟΠΛΗΞ,	nōtoplḗx,	nōtoplēx,	notopliks,	nwtoplh/c,	"ῆγος, ὁ, ἡ, <span
22	νωτοστροφέω	νωτοστροφέω	νωτοστρο	νωτοστροφ	ΝΩΤΟΣΤΡΟΦΕΩ	nōtostrophéō	nōtostrophe	notostrofeo,	nwtostrofe/	"<span class=""sens
23	νωτοφορέω,	νωτοφορέω,	νωτοφορέ	νωτοφορέω	ΝΩΤΟΦΟΡΕΩ,	nōtophoréō,	nōtophoreō,	notoforeo,	nwtofore/w,	"<span class=""sens

5 Δημιουργία στυλ css

Οι καταχωρίσεις css καθορίζουν τον τρόπο που θα εμφανιστεί η html των λημμάτων. Ο σκοπός ήταν διττός: αφενός ο οπτικός διαχωρισμός των εννοιών και η (διαφορετικού χρώματος) έμφαση σε όρους και αποδόσεις και αφετέρου η σήμανση με τρόπο ιδανικό για τις μηχανές αναζήτησης (π.χ. με στυλ που ενσωματώνει το στοιχείο **<b>**).

6 Εμφάνιση αποφθεγμάτων

Η χρήση μιας επέκτασης που δημιουργήθηκε ειδικά για την εμφάνιση αρχαιοελληνικών αποφθεγμάτων σε τυχαία σειρά είχε ως σκοπό αφενός να διανθίσει το λεξικογραφικό έργο με ενδιαφέρουσες φράσεις, αφετέρου να λειτουργήσει ως έναυσμα για περαιτέρω διερεύνηση αρχαιοελληνικών λέξεων. Για την επίτευξη του δεύτερου σκέλους, έγινε σε δεύτερη φάση μια βελτίωση της επέκτασης έτσι ώστε να μπορεί να αναλύει (parse) βικικώδικα (ήδη μπορούσε να αναλύει βασική html όπως πλάγια και έντονη γραφή).

Μέρος της συνάρτησης για την ανάλυση των αποφθεγμάτων όπως ήταν αρχικά:

```
    if( $row ) {
        list( $quote, $attribution ) = explode( "\n", $row->quote_text );
        return '<div id="trrandomquote"><div id="trquote">'. $quote .'</div>
<span>'. $attribution .'</span></div>';
```

Και όπως έγινε μετά την τροποποίηση:

```
    if( $row ) {
        list( $quote, $attribution ) = explode( "\n", $row->quote_text );

    $text = '<div id="trrandomquote"><div id="trquote">'.
    $quote .'</div><span>'. $attribution .'</span></div>';

    global $wgParser;
    $parser = clone $wgParser;
    $quote = $parser->parse( $text, Title::newFromText('Sample'), new
ParserOptions() )->getText();

    return $quote;
```

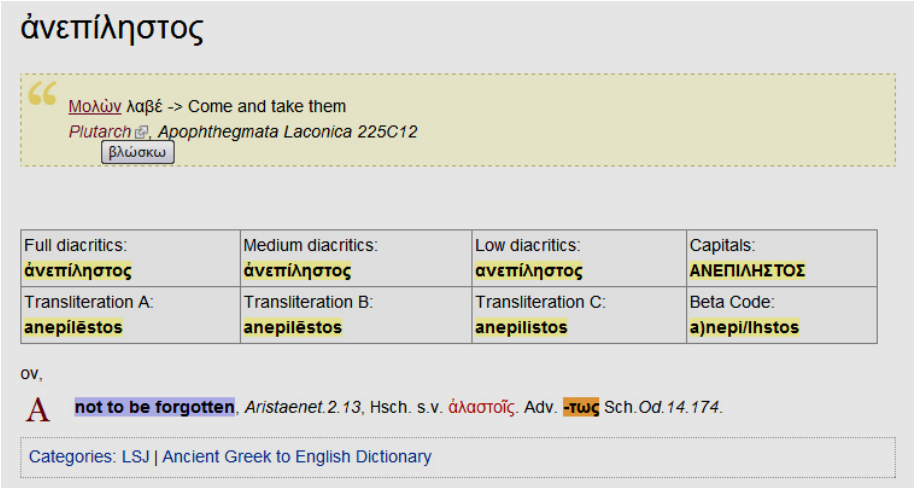
Για παράδειγμα, όταν μια λέξη βρίσκεται ανάμεσα από διπλές αγκύλες στον βικικώδικα με τη μορφή «**[[Λέξη]]**» τότε αυτή μετατρέπεται αυτόματα σε εσωτερικό σύνδεσμο προς το ομώνυμο λήμμα. Η παραπάνω περίπτωση μας καλύπτει όταν η λέξη της παραπομπής είναι ακριβώς η ίδια, αν όμως έχει διαφορετική μορφολογία τότε χρησιμοποιείται διαφορετική βικισύνταξη. Π.χ. αν έχουμε τη φράση «Το νόημα της λέξης» και θέλουμε το «λέξης» να παραπέμπει στο λήμμα «λέξη», τότε η σύνταξη που θα χρησιμοποιηθεί θα είναι της μορφής «**Το νόημα της [[λέξη|λέξης]]**». Στην παρακάτω εικόνα βλέπουμε το λήμμα για το οποία είχα μιλήσει στην αρχή, πώς εμφανίζεται στο πρόγραμμα περιήγησης. Εν προκειμένω, φαίνεται το πλαίσιο «βλώσκω» (σύνδεσμος δηλαδή προς το ομώνυμο λήμμα) καθώς το ποντίκι ήταν πάνω από το «Μολών». Επίσης, βλέπουμε τα τρία είδη επισήμανσης που χρησιμοποιήθηκαν: **κίτρινο** για τις μορφές / μεταγραμματισμούς, **γαλάζιο** για τις αγγλικές ερμηνείες και **πορτοκαλί** για τα ελληνικά υπολήμματα.

Ως εκ τούτου έγινε επιλογή και επεξεργασία μιας σειράς αρχαιοελληνικών αποφθεγμάτων με τις αντίστοιχες μεταφράσεις τους στα αγγλικά και στη συνέχεια εισαγωγή τους στον πίνακα (της βάσης δεδομένων) για την επέκταση των αποφθεγμάτων. Η εισαγωγή/ανανέωση των αποφθεγμάτων γίνεται εκτελώντας από τη γραμμή εντολών μιας δέσμης ενεργειών εισαγωγής μέσω σύνδεσης SSH[24] (ασφαλούς κελύφους). Τα προγράμματα που χρησιμοποιήθηκαν για αυτόν τον σκοπό ήταν το PuTTY[25] και το WinSCP[26] (το δεύτερο διαθέτει και γραφικό περιβάλλον).

[24] http://el.wikipedia.org/wiki/SSH

[25] http://www.putty.org/

[26] http://winscp.net



7 Εργασίες στον διακομιστή

Για να μπορέσει να λειτουργήσει το MediaWiki απαιτείται ένας διακομιστής που να διαθέτει PHP και MySQL, συνήθεις δυνατότητες σε σύγχρονους διακομιστές (ο διακομιστής που χρησιμοποιήθηκε είχε Apache 2.2.24, php 5.3.21, MySQL 5.1.69). Για τη δυνατότητα αυτόματης συμπλήρωσης απαιτείται υποστήριξη Java που σε διακομιστές με cPanel[27] γίνεται με την ενεργοποίηση και παραμετροποίηση του Tomcat[28] από το *WHM* > *Software* > *EasyApache* και αναδόμηση του Apache με επιλογή του πλαισίου για το Tomcat στην 4η οθόνη παραμετροποίησης με τίτλο *Short Options List*.

Στη συνέχεια πρέπει να εγκατασταθεί το Lucene[29] στον διακομιστή (και η σχετική επέκταση στο MediaWiki) και να δημιουργηθεί το αρχείο που θα ελέγχει τη δημιουργία των ευρετηρίων. Να σημειώσουμε ότι το Lucene (αν και δύσκολο να υλοποιηθεί καθώς απαιτεί ριζική πρόσβαση και δικαιώματα σε διακομιστή καθώς και ειδικές τεχνικές γνώσεις) είναι η πλέον ενδεδειγμένη λύση αναζήτησης η οποία έχει υιοθετηθεί και από το Ίδρυμα Wikimedia για τους βικιτόπους του.

Εδώ υπάρχουν δύο δυνατότητες, μία είναι η ρύθμιση για πλήρη αναδημιουργία ευρετηρίων του Lucene και η άλλη η επαυξητική αναδημιουργία (μόνο για τα νέα λήμματα). Αμφότερες μπορούν να αυτοματοποιηθούν μέσω μιας εργασίας χρονοπρογραμματισμού τύπου cron[30].

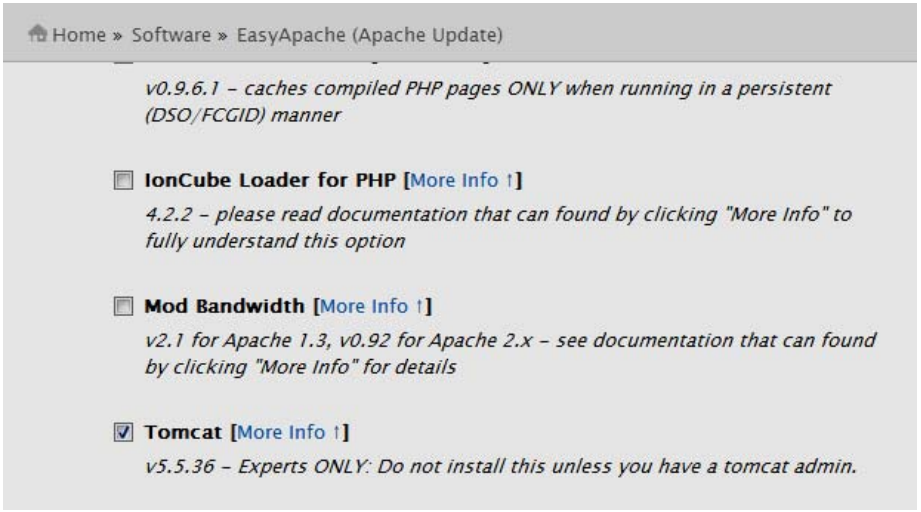
27 http://en.wikipedia.org/wiki/CPanel

28 http://en.wikipedia.org/wiki/Apache_Tomcat

29 http://www.mediawiki.org/wiki/Extension:Lucene-search

30 https://en.wikipedia.org/wiki/Cron

Η επαυξητική αναδημιουργία χρησιμοποιείται για τους ιστότοπους του Ιδρύματος WikiMedia (με αναδημιουργία κάθε 1 ώρα). Είναι δικαιολογημένη η χρήση του καθώς δημιουργείται μεγάλος αριθμός λημμάτων από τους χρήστες, ήταν όμως πολύ δυσκολότερη στην υλοποίηση και μη αναγκαία στην περίπτωσή μας, έτσι προτιμήθηκε η πλήρης αναδημιουργία.



8 Αναζήτηση αυτόματης συμπλήρωσης

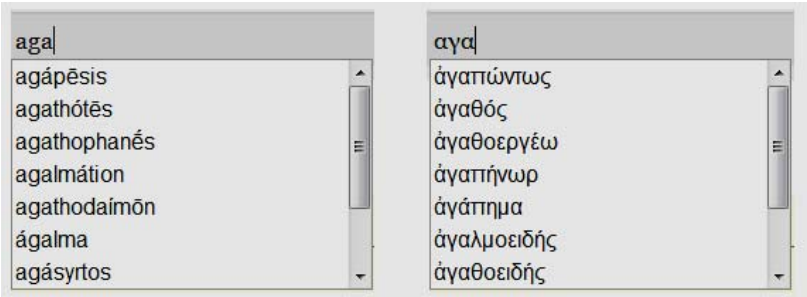
Ένα από τα βασικά ζητούμενα κατά τη φάση προγραμματισμού του έργου ήταν η υλοποίηση δυνατότητας αναζήτησης τύπου «αυτόματης συμπλήρωσης» ή «προτάσεων αναζήτησης» («autocomplete» ή «search suggestions») *χωρίς διάκριση πεζών/κεφαλαίων* και *χωρίς διάκριση διακριτικών σημείων* τόσο για ελληνικούς όσο και για λατινικούς χαρακτήρες (μεταγραμματισμένων ελληνικών λέξεων).

Το παραπάνω ζητούμενο διεκπεραιώνει αυτόματα το Lucene (δημιουργώντας ευρετήρια με κανονικοποιημένες εκδοχές των λημμάτων, δηλαδή χωρίς διακριτικά). Ωστόσο, για να ισχύει το ίδιο για τους λατινικούς χαρακτήρες έπρεπε να υπάρχουν λήμματα γραμμένα με αυτούς. Η λύση ήταν απλή: δημιουργία σελίδων ανακατεύθυνσης[31] για κάθε λήμμα με βάση τον μεταγραμματισμό «Transliteration A». Το MediaWiki «αντιλαμβάνεται» τις σελίδες ανακατεύθυνσης (για τον σκοπό της αναζήτησης) ως κανονικές σελίδες. Έτσι, το ευρετήριο εμπλουτίζεται και με τις εκδοχές μεταγραμματισμένων λημμάτων.

Να σημειώσουμε εδώ ότι η βασική μορφή της αυτόματης συμπλήρωσης στο MediaWiki κάνει ακόμη και διάκριση πεζών/κεφαλαίων (εννοείται ότι επίσης κάνει διάκριση διακριτικών).

[31] http://en.wikipedia.org/wiki/Wikipedia:Redirect

Για παράδειγμα, αν ψάχναμε το λήμμα **Ἶρις** και γράφαμε **ίρις**, δεν θα εμφανιζόταν στην αυτόματη συμπλήρωση. Θα έπρεπε να ξεκινήσουμε με το τονούμενο κεφαλαίο. Αυτό το ζήτημα διορθώνεται με την επέκταση TitleKey[32], ωστόσο, δεν υπάρχει κάποια άλλη επέκταση που να εξασφαλίζει μη διάκριση διακριτικών χωρίς να χρειαστεί εγκατάσταση εφαρμογής και στον διακομιστή[33]. Στην εικόνα δείγμα λειτουργίας της αυτόματης συμπλήρωσης με λατινικούς ή ελληνικούς χαρακτήρες.



9 Επίλογος

Η υλοποίηση ενός έργου όπως η μετατροπή ενός πολύπλοκου λεξικού όπως το LSJ σε μορφή MediaWiki με τις δεδομένες προδιαγραφές, δεν υπήρξε σε καμία περίπτωση μια εύκολη υπόθεση. Παρουσιάστηκαν πολυάριθμα προβλήματα σε κάθε επίπεδο του έργου, ελάχιστα από τα οποία σχολιάστηκαν εδώ. Λόγω της τεχνικής φύσης του εγχειρήματος, προσπάθησα να βοηθήσω τους φιλομαθείς αναγνώστες παρέχοντας συνδέσμους στις υποσημειώσεις.

Εν κατακλείδι, όλα τα ζητούμενα του έργου εκπληρώθηκαν και δομήθηκε μια πλατφόρμα που θα μπορεί να εμπλουτιστεί περαιτέρω στο μέλλον, τόσο μέσω αυτόματης εισαγωγής νέου υλικού, όσο και με ανατροφοδότηση των χρηστών.

Σπύρος Δόικας
Ιδρυτής της μεταφραστικής πύλης www.translatum.gr
Επικοινωνία: www.translatum.gr/contact.htm

[32] https://www.mediawiki.org/wiki/Extension:TitleKey

[33] Εκτός από το Lucene πάντως υπάρχει και το Sphinx (https://www.mediawiki.org/wiki/Sphinx).