Απόκτηση φορητών προσωπικών υπολογιστών από τους πρωτοετείς φοιτητές στην Τριτοβάθμια Εκπαίδευση, για το έτος 2009–2010

ΟΠΣ 217114

Τίτλος Παραδοτέου:
Μελέτη Εργαλείων Ορθογραφικής Διόρθωσης και Θησαυρών

Εθνικό Δίκτυο Έρευνας και Τεχνολογίας Α.Ε.
Greek Research and Technology Network S.A.

Επιτελική Σύνοψη

Το παρόν έγγραφο παρουσιάζει τις προδιαγραφές για την ανάπτυξη και τεχνική υποστήριξη, για διάστημα ενός έτους, Πληροφοριακού Συστήματος (ΠΣ), που θα λειτουργεί on-line μέσω του διαδικτύου, διαχείρισης του εμπλουτισμού του λεξικού ορθογραφικής διόρθωσης ελληνικών με το Hunspell και Θησαυρού συνωνύμων της ελληνικής γλώσσας, με άδεια EUPL.

Σε λογισμικά με άδεια ανοιχτού κώδικα χρησιμοποιείται ευρέως ο ορθογραφικός διορθωτής Hunspell. Τέτοια λογισμικά είναι το OpenOffice.org, FireFox, Thunderbird, Chrome, Opera κ.ά.. Επιπλέον, το αποκεντρωμένο μοντέλο συνεργασίας χρηστών και το πλαίσιο της συνιδιοκτησίας του παραγόμενου προϊόντος, που το Κίνημα Ανοιχτού Λογισμικού εισήγαγε, έχει αποδειχτεί ότι συμβάλλουν στη δημιουργία μιας αφοσιωμένης κοινότητας χρηστών που μπορούν να χρησιμοποιούν και επεκτείνουν το ΠΣ σε σταθερή βάση. Η υποστήριξη του ΠΣ που θα αναπτυχθεί προσανατολίζεται προς αυτή την κατεύθυνση.

Συνολικά, η προσπάθεια αφορά

- στη βελτίωση των επιδόσεων του Hunspell για την ελληνική γλώσσα, κάτι που μεταφράζεται σε εμπλουτισμό τού λεξικού ορθογραφικής διόρθωσης και συνωνύμων με νέες λέξεις αλλά και τη συγγραφή κανόνων
- στην υιοθέτηση ενός συνεργατικού μοντέλου επέκτασης και συντήρησης που εμπλέκει άμεσα και κινητοποιεί τη συμμετοχή χρηστών

Περιεχόμενα

Λίστα Εικόνων

1 Εισαγωγή

Μία από τις δημοφιλέστερες χρήσεις των ηλεκτρονικών υπολογιστών είναι η προετοιμασία και επεξεργασία εγγράφων. Με τη βοήθεια ειδικών εφαρμογών αυτή η εργασία γίνεται εύκολη και ξεκούραστη, όμως η αποδοτικότητα των χρηστών επαφείεται ολοκληρωτικά στη συντακτική τους ικανότητα και τη γνώση της γλώσσας. Γι' αυτό το λόγο αναπτύχθηκαν εργαλεία που αναγνωρίζουν ορθογραφικά λάθη ή λέξεις που δεν ορίζονται στη γλώσσα και διευκολύνουν σημαντικά την προετοιμασία και διόρθωση κειμένων, αυξάνοντας σημαντικά την ταχύτητα παραγωγής ποιοτικών, καθαρών από λάθη εγγράφων. Τα εργαλεία αυτά είναι συνήθως ενσωματωμένα στις εφαρμογές έκδοσης κειμένου και ονομάζονται *ορθογράφοι*.

Το έγγραφο αυτό παρουσιάζει τις προδιαγραφές για ένα Πληροφορικό Σύστημα (ΠΣ) που θα υποστηρίζει τη συντήρηση και την επέκταση ενός ορθογράφου, του Hunspell πιο συγκεκριμένα, ενώ θα χρησιμοποιεί ως βάση και κανάλι επικοινωνίας το διαδίκτυο και θα στηρίζεται σε μια κοινότητα χρηστών για τη συνεχή ανάπτυξή του.

Αρχικά, στο κεφάλαιο 2 δίδεται λεπτομέρης περιγραφή της υπάρχουσας τεχνογνωσίας στο χώρο των ελληνικών ορθογράφων. Η περιγραφή των προσπαθειών ανάπτυξης ορθογράφων περιλαμβάνει ελληνικές πρωτοβουλίες αλλά κυρίως απαριθμεί διεθνή εργαλεία που προτάσσουν υποστήριξη για την ελληνική γλώσσα (ενότητα 2.1). Γίνεται φανερή, δε, η διάκριση των παραπάνω από έργα ανάπτυξης γραφικών διεπαφών για υπάρχοντες ορθογράφους που περιγράφονται στην ενότητα 2.2. Στη συνέχεια περιγράφονται συνεργατικές προσπάθειες ανάπτυξης ορθογράφων παγκοσμίως (ενότητα 2.3). Με τη βοήθεια του διαδικτύου και με κίνητρο τη συνιδιοκτήσια του παραγόμενου προϊόντος, καθώς και την προσθήκη χαρακτηριστικών προς ίδιο όφελος, έκαναν την εμφάνισή τους συνεργατικά μοντέλα ενεργής συμμετοχής χρηστών από όλο τον κόσμο.

Από αυτές τις επισκοπήσεις προκύπτουν με φυσικό τρόπο τα χαρακτηριστικά του διαδικτυακού, συνεργατικού ΠΣ διαχείρισης εμπλουτισμού του λεξικού για την ελληνική γλώσσα, που παρουσιάζεται στο κεφάλαιο 3. Το κεφάλαιο αυτό αποτελεί το συνδετικό κρίκο ανάμεσα στην υπάρχουσα τεχνογνωσία και τις προδιαγραφές του ΠΣ τεκμηριώνοντας το λόγο ανάπτυξης του ΠΣ.

Οι προδιαγραφές του ΠΣ περιγράφονται στο κεφάλαιο 4. Αυτές διακρίνονται σε τεχνικές προδιαγραφές (ενότητα 4.1), που περιγράφουν το τεχνικό πλαίσιο λειτουργίας του ΠΣ, και λειτουργικές προδιαγραφές (ενότητα 4.2), που περιγράφουν τη λειτουργικότητα του συστήματος. Στο κεφάλαιο παρουσιάζονται επίσης οι προδιαγραφές των κανόνων του ορθογράφου Hunspell, στους οποίους βασίζεται η λεξικογραφική διόρθωση (ενότητα 4.3).

Το κεφάλαιο 5 παρουσιάζει τις δυνατότητες του ΠΣ για εξαγωγή προς εγκατάσταση και χρήση του εργαλείου ορθογραφικής διόρθωσης στο περιβάλλον εργασίας συμβατού προγράμματος του υπολογιστή π.χ. *Opera*.

Το κεφάλαιο 6 περιγράφει τις βοηθητικές επεκτάσεις, που το ΠΣ πρέπει να υποστηρίζει, καθώς και τις τεχνικές προδιαγραφές των λειτουργικών συστημάτων, εφαρμογών στα οποία οι επεκτάσεις θα είναι διαθέσιμες.

Τέλος, το κεφάλαιο 7 παρουσιάζει τα παραδοτέα του έργου ανάπτυξης του ΠΣ ενώ το κεφάλαιο 8 αναφέρει την άδεια χρήσης του ΠΣ.

2 Υπάρχουσα Τεχνογνωσία

2.1 Ορθογραφικά Λεξικά της Ελληνικής

2.1.1 ΣΥΜΦΩΝΙΑ

Η ΣΥΜΦΩΝΙΑ [35] είναι ένα προϊόν γλωσσικής τεχνολογίας του Ινστιτούτου Επεξεργασίας του Λόγου. Παρέχει τις λειτουργίες του ορθογραφικού ελέγχου, προσθήκης νέων λέξεων στο λεξικό, γραμματικού χαρακτηρισμού λέξεων και ελέγχου συντακτικής συμφωνίας των λέξεων. Μάλιστα, σύμφωνα με την περιγραφή του πρϊόντος, οι δύο τελευταίες λειτουργίες αποτελούν καινοτομία ενάντι αντίστοιχων προϊόντων.

Το λογισμικό βρίσκεται ενσωματωμένο στο περιβάλλον του Microsoft Word 1997 και 2000. Είναι κατάλληλο για κάθε χρήστη συγγραφής κειμένου και απαλλάσσει τα γράπτα από ορθογραφικά λάθη και λάθη πληκτρολόγησης.

Οι υπηρεσίες που προσφέρει παρουσιάζουν τα εξής χαρακτηριστικά:

- Ορθογραφικός έλεγχος

  Η ΣΥΜΦΩΝΙΑ πραγματοποιεί τον ορθογραφικό έλεγχο σε πραγματικό χρόνο ταυτόχρονα με την εισαγωγή κειμένου από πλευράς του χρήστη και χρησιμοποιεί μια διακριτική γραμμή για να υποδείξει ότι μια λέξη δεν είναι έγκυρη σύμφωνα με τη βάση δεδομένων που διατηρεί. Σε αυτή την περίπτωση ο χρήστης δύναται να πατήσει δεξί κλικ πάνω στη λέξη και αμέσως εμφανίζονται παρεμφερείς προτάσεις διόρθωσης.

  Το λεξικό των 65.000 λημμάτων παράγει όλους τους κλιτούς τύπους, πάνω από 1.600.000 σε αριθμό. Αυτή η ευρεία κάλυψη της ελληνικής γλώσσας έχει ως αποτέλεσμα τη συνύπαρξη ορθογραφικής και μορφολογικής ποικιλίας (π.χ. *κοιτάν, κοιτούν, κοιτούνε*) που είναι σημαντικά χαρακτηριστικά της σύγχρονης ελληνικής. Αυτά τα χαρακτηριστικά πρέπει να λαμβάνει υπόψη ένας ορθογράφος. Η αλληλεπίδραση με το χρήστη γίνεται μέσω της διεπαφής που προσφέρει το Microsoft Word.

- Συντακτικός έλεγχος

  Πλέον του ορθογραφικού ελέγχου, η ΣΥΜΦΩΝΙΑ ελέγχει κάθε λέξη σε σχέση με τις λέξεις που την περιβάλλουν, π.χ. το ουσιαστικό με το επίθετο και το άρθρο, το ρήμα με τις αντωνυμίες. Ο συντακτικός έλεγχος βρίσκει εφαρμογή και χρησιμότητα κατά κόρον στην αποσαφήνιση λέξεων που μοιράζονται το ίδιο άκουσμα αλλά γράφονται διαφορετικά και διακρίνονται στα γραμματικά τους χαρακτηριστικά, όπως οι λέξεις *χαίρεται* και *χαίρετε*.

  Τα προβλήματα αυτά αναφέρονται ως αμφισημίες. Εφόσον η λέξη συγκεντρώνει τα χαρακτηριστικά που ικανοποιούν τους αντίστοιχους κανόνες συντακτικού

τότε γίνεται αποδεκτή διαφορετικά είτε προτείνεται η ίδια λέξη τροποποιημένη, π.χ. σε άλλη πτώση, ή προτείνεται μια άλλη λέξη με τα αναμενόμενα χαρακτηριστικά.

- Προσθήκη νέων λέξεων

  Η λειτουργία αυτή παρέχει τη δυνατότητα στο χρήστη να εμπλουτίσει το λεξικό με νέες λέξεις ώστε να μην τις αντιμετωπίζει ως ξένες και λανθασμένες. Η εισαγωγή των λέξεων είναι διαδραστική και γίνεται με την πλήρη κλίση τους.

  Χαρακτηριστικά, για τη λέξη *πιθανοτικής* ο χρήστης καλείται να εισάγει ότι η λέξη είναι επίθετο γένους θηλυκού, αριθμού ενικού, πτώσης γενικής. Στη συνέχεια το σύστημα επιτρέπει στο χρήστη να επιλέξει τη σωστή κλίση του επιθέτου μέσα από επιλογές που εμφανίζει. Αυτό είναι ιδιαίτερα χρήσιμο για κείμενα που περιλαμβάνουν ειδική ορολογία.

  Παρά τη δυνατότητα που προσφέρει για προσθήκη νέων λέξεων, αυτή αφορά μόνο τη συγκεκριμένη εγκατάσταση λογισμικού και επομένως η ΣΥΜΦΩΝΙΑ ως σύστημα δεν επωφελείται από αυτή τη λειτουργικότητα, δεν επεκτείνεται.

- Γραμματικός χαρακτηρισμός

  Με τη ΣΥΜΦΩΝΙΑ ο χρήστης μπορεί να πάρει όλες τις πληροφορίες για το γραμματικό χαρακτηρισμό μιας λέξης κάνοντας δεξί κλικ πάνω της.

Η ΣΥΜΦΩΝΙΑ είναι ορθογράφος δεύτερης γενιάς. Οι τυπικοί ορθογράφοι, που ανήκουν στην πρώτη γενιά, αγνοούν καίρια χαρακτηριστικά της γλώσσας και ως εκ τούτου μειονεκτούν σε απόδοση και εργονομία.

Οι ορθογράφοι πρώτης γενιάς:

- Αποτυγχάνουν να εντοπίσουν συντακτικά ορθογραφικά λάθη, αφού στηρίζονται στην επεξεργασία λέξης προς λέξη
- Συμπεριφέρονται στις άγνωστες λέξεις ως απλές συμβολοσειρές και αναγκάζουν το χρήστη να επαναλάβει την εισαγωγή της νέας λέξης για κάθε κλιτό τύπο ενός άγνωστου λήμματος.
- Βασίζονται σε λεξικό που υποστηρίζει μόνο τη δημοτική γλώσσα αγνοώντας χιλιάδες λέξης της καθαρεύουσας που ακόμη χρησιμοποιούνται ευρέως.

Σύμφωνα με τον Δρ. Ν. Γλαρό [1], η πρωτοπορία της ΣΥΜΦΩΝΙΑΣ μεταφράζεται σε σημαντικά οφέλη για τους χρήστες:

- Η διόρθωση των κειμένων πραγματοποιείται ταχύτερα, ευκολότερα και αποδοτικότερα.
- Τα κείμενα παρουσιάζουν βελτιωμένη ποιότητα συγγραφής
- Διατηρείται και διαδίδεται η ορθή γραφή της γλώσσας.
- Βελτιώνεται η συνεργασία των διαφόρων εμπλεκόμενων στην παραγωγή εγγράφων.

[1] Δρ. Ηλ/γος Μηχανικός του Ινστιτούτου Επεξεργασίας του Λόγου και πρόσωπο επικοινωνίας για τη ΣΥΜΦΩΝΙΑ

Η ΣΥΜΦΩΝΙΑ παρουσιάζει επίσης έντονη προσαρμοστικότητα σε ειδικά περιβάλλοντα που παράγουν ειδικές κατηγορίες κειμένων. Καταλυτικό ρόλο παίζει η διαδικασία εκμάθησης της ιδιολέκτου και της ορολογίας από αντιπροσωπευτικά κείμενα. Σε μικρό χρονικό διάστημα, το λογισμικό διόρθωσης κειμένων μπορεί να προσαρμοστεί στις νέες απαιτήσεις.

Η ποιότητα του ορθογράφου ΣΥΜΦΩΝΙΑ οδήγησε την εφημερίδα *Καθημερινή* να ζητήσει τη συμβολή της στον τομέα της διόρθωσης κειμένων. Ο πηρύνας του ορθογράφου ΣΥΜΦΩΝΙΑ έχει ενσωματωθεί στο λογισμικό που χρησιμοποιεί η *Καθημερινή* για εκδοτικές εφαρμογές.

Εκτός από τις λειτουργίες που έχουμε αναφέρει η συνεργασία αυτή απέφερε μια νέα λειτουργία, το περιβάλλον συντήρησης ηλεκτρονικού λεξικού το οποίο αναπτύσσεται στην υποενότητα 2.3.2.

Συγκεντρωτικά, η ΣΥΜΦΩΝΙΑ μπορεί να:

- χρησιμοποιηθεί ως ειδική ενότητα λογισμικού ενσωματωμένη σε επεξεργαστή κειμένου της αγοράς
- αναπτυχθεί ως ανεξάρτητη εφαρμογή
- υποστηρίξει οποιαδήποτε υπολογιστική πλατφόρμα
- λειτουργήσει σε μοντέλο πελάτη-εξυπηρετητή ή τοπικά στον υπολογιστή του χρήστη
- λειτουργήσει ως διαδικτυακή υπηρεσία

Τέλος, Η ΣΥΜΦΩΝΙΑ, όπως κάθε άλλο λογισμικό, επιβάλλει ελάχιστες απαιτήσεις συστήματος ώστε να μπορεί να εκτελεστεί. Απαιτεί λοιπόν PC Pentium επεξεργαστή, 32 MB RAM μνήμη, 45 MB αποθηκευτικό χώρο στο δίσκο, Microsoft Windows 95 ή νεώτερο λειτουργικό σύστημα και Microsoft Word 1997 ή 2000 επεξεργαστή κειμένου.

2.1.2 Ελληνικός Ορθογράφος

Ο Ελληνικός Ορθογράφος [2] είναι ένας ηλεκτρονικός ελληνικός ορθογράφος ικανός να ελέγχει και να αφαιρεί σφάλματα από το γραπτό λόγο της κοινής νεοελληνικής, της γραφόμενης κυπριακής και της αθλητικής, οικονομικής και χρηματιστηριακής ορολογίας. Είναι μια επιστημονική εφαρμογή που προκύπτει από ερευνητική σύμπραξη του Αριστοτέλειου Πανεπιστημίου Θεσσαλονίκης και του Πανεπιστημίου Κύπρου.

Η ανάγκη ανάπτυξης του εργαλείου έγκειται στην ανεπάρκεια των υπάρχοντων ηλεκτρονικών ελληνικών ορθογράφων να παρέχουν:

- αναγνώριση και διόρθωση ειδικής ορολογίας
- προσθήκη στοιχείων της γραφόμενης κυπριακής
- ενσωμάτωση ηλεκτρονικών λεξικών

Ο Ελληνικός Ορθογράφος στηρίζει τον εμπλουτισμό του στην ικάνοτητα αφομοίωσης βάσεων από άλλα λεξικά, χαρακτηριστικό που ενισχύει σημαντικά τη δυνατότητα και το ρυθμό επέκτασής του. Η λειτουργία του Ελληνικού Ορθογράφου έχει ως βάση ηλεκτρονικά λεξικά και τοπικές γραμματικές. Τα εν λόγω λεξικά[2] της κοινής νέας

ελληνικής σχεδιάστηκαν και αναπτύσσονται[3] από τη Μονάδα Αυτόματης Επεξεργασίας Φυσικών Γλωσσών του Εργαστηρίου Μετάφρασης και Επεξεργασίας του Λόγου[4]. Η επεξεργασία των λέξεων επιτυγχάνεται με τη χρήση του *GenereFlexion* [20] το οποίο παράγει όλους τους κλιτούς τύπους των καταχωρημένων λέξεων.

Ενώ τα λεξικά θα συνεισφέρουν πλήθος γλωσσικών μονάδων προς αναγνώριση, ο ορθογραφικός έλεγχος θα βασιστεί στο *Λεξικό της Κοινής Νεοελληνικής του Ινστιτούτου Νεοελληνικών Σπουδών*[5] και το *Αντίστροφο Λεξικό της Νέας Ελληνικής*[6].

Τα κριτήρια με τα οποία αυτά επιλέχθηκαν είναι:

- η βιβλιογραφική αναγνώριση
- ο πλούτος γλωσσικών μονάδων που περιέχουν (πάνω από 30.000 λήμματα)
- η πρόσφατη χρονολογία έκδοσης

Για τη γραφόμενη κυπριακή, η αναγνώριση τύπων θα πραγματοποιηθεί από την αποδελτίωση γραπτών του κυπριακού τύπου από τις εφημερίδες *Φιλελεύθερος, Πολίτης, Χαραυγή* και *Σημερινή*.

Λόγω των αποκλίσεων της κοινής νεοελληνικής και της κυπριακής ο ορθογράφος θα χρησιμοποιεί τρία είδη σήμανσης, μία για τους μη αποδεκτούς τύπους είτε για την κοινή νεοελληνική είτε για τη γραφόμενη κυπριακή, μία για τους αποδεκτούς τύπους της κυπριακής ποικιλίας και μία τελευταία για κυπριακά ονόματα, τα οποία διαφέρουν από τα ελληνικά μόνο ως προς τον τονισμό π.χ. *Παπάμιχαηλ* (κυπριακή) - *Παπαμιχαήλ* (κοινή νεοελληνική).

Η αναγνώριση και ο έλεγχος της ειδικής ορολογίας στα πεδία του αθλητισμού, των οικονομικών και του χρηματιστηρίου είναι αποτέλεσμα συγκεντρωτικής καταγραφής και ομαδοποίησης της ίδιας ερευνητικής ομάδας[7]. Η αθλητική ορολογία προέρχεται από δύο κατηγορίες, αθλήματα θερινών και χειμερινών ολυμπιακών αγώνων και υπολογίζεται σε 37.000 γλωσσικές μονάδες ένω οι οικονομικοί και χρηματιστηριακοί όροι φτάνουν τις 8.500 περίπου.

2.1.3 Λεξισκόπιο

Το Λεξισκόπιο [25] είναι ένα προϊόν της Neurolingo και συνδυάζει λειτουργίες σύνθετου γλωσσικού ελέγχου και παροχής πληροφοριών για τα μέρη του λόγου. Παρέχει ενοποιημένα Μορφολογικό Λεξικό, Συλλαβιστή, Ορθογράφο, Λημματοποιητή και Θησαυρό Συνωνύμων - Αντιθέτων σε ένα εργαλείο.

- Μορφολογικό λεξικό

  Το Μορφολογικό λεξικό αποτελεί τη βάση για όλα τα γλωσσικά εργαλεία της Neurolingo. Περιέχει 90.000 γλωσσικές μονάδες, καθεμία από τις οποίες συνδέεται με μορφολογικούς κανόνες (κλιτικά υποδείγματα). Από αυτούς τους κανόνες παράγονται όλοι οι κλιτικοί τύποι, οι οποίοι απαριθμούνται σε 1.200.000.

[2] 18.000 ρήματα, 70.000 απλά ουσιαστικά, 28.000 πολυλεκτικές μονάδες, 50.000 κύρια ονόματα, 16.000 απλά και σύνθετα επιρρήματα, 3.000 τοπωνύμια, 1.000 γραμματικές λέξεις.

[3] Η βιβλιογραφική πηγή περιγράφει την εξέλιξη του ορθογράφου το έτος 2005.

[4] `http://linginfo.frl.auth.gr`, Τμήμα Γαλλικής Γλώσας και Φιλολογίας, Φιλοσοφική Σχολή, Α.Π.Θ.

[5] Ίδρυμα Μανόλη Τριανταφυλλίδη, 1998, 50.000 λήμματα

[6] Άννα Αναστασιάση-Συμεωνίδη, 2002, 180.000 λήμματα

[7] Université de Rennes II, Πρόγραμμα Euradic, στο πλαίσιο του προγράμματος Technolangues (μονόγλωσσες και δίγλωσσες γλωσσολογικές πηγές) / Ειδικό Λεξικό: Παραγωγή πολύγλωσσου λεξικού αθλημάτων.

Το Μορφολογικό λεξικό προσφέρει πλούσια πληροφορία:

- Ορθογραφική, παρουσιάζει ποια γράμματα και με τι σειρά σχηματίζουν ορθές λέξεις
- Συλλαβισμού, παρουσιάζει ποιες συλλαβές αποτελούν τον κλιτικό τύπο
- Μορφηματική, παρουσιάζει ποια μορφήματα (πρόθημα, θέμα, επίθημα, κατάληξη) αποτελούν τον κλιτικό τύπο
- Μορφοσυντακτική, παρουσιάζει ποια λέξη παράγει έναν κλιτικό τύπο και ποια είναι τα μορφοσυντακτικά χαρακτηριστικά του (μέρος του λόγου, γένος, πτώση, αριθμός, πρόσωπο, φωνή, χρόνος)
- Υφολογική, παρουσιάζει τα υφολογικά χαρακτηριστικά του κλιτικού τύπου αν υπάρχουν, π.χ. ο τύπος *παίρνανε* ταξινομείται ως προφορικός (σε σύγκριση με τον τύπο *έδιναν*)
- Ορολογική, παρουσιάζει το ειδικό λεξιλόγιο που ανήκει ένας ειδικός όρος, π.χ. ο τύπος *αβιογένεση* αποτελεί όρο της Βιολογίας.

Εκτός από όρους της κοινής νοελληνικής το Μορφολογικό λεξικό περιλαμβάνει και τύπους ειδικού λεξιλογίου. Συγκεκριμένα περιλαμβάνει 10.000 ελληνικά τοπωνύμια (ελληνικοί νομοί, δήμοι, κοινότητες, επαρχίες, πόλεις, χωριά) και 6.000 όρους βιοϊατρικής.

- Συλλαβιστής

  Ο Συλλαβιστής αναγνωρίζει όλα τα πιθανά σημεία συλλαβισμού μιας λέξης. Η λειτουργία του στηρίζεται σε κανόνες και λεξικό εξαιρέσεων. Μάλιστα ο Συλλαβιστής επιδέχεται εκπαίδευση αφού ένα μέρος των κανόνων που χρησιμοποιεί έχουν παραχθεί αυτόματα βάσει πληροφοριών συλλαβισμού από το Μορφολογικό Λεξικό. Αυτοί οι κανόνες του επιτρέπουν να υποδεικνύει το σωστό συλλαβισμό 24 διαφορετικών συνδυασμών φωνηέντων, που μπορεί να χωρίζονται στο συλλαβισμό ή όχι (*συνίζηση*).

  Το λεξικό εξαιρέσεων προηγείται στην εφαρμογή από τους κανόνες και περιέχει γλωσσικές μονάδες για τις οποίες οι κανόνες αποφαίνονται λανθασμένα. Οι μονάδες, 300 στον αριθμό περίπου, υπάρχουν συλλαβισμένες και σε αρκετές περιπτώσεις οδηγούν σε σημασιολογική ασάφεια, π.χ. *ή-πια έναντι του ή-πι-α*. Στις περίπτωσεις αυτές ο Συλλαβιστής δε διενεργεί το συλλαβισμό φωνηέντων που αλλάζουν το νόημα.

  Ο Συλλαβιστής συλλαβίζει χωρίς σφάλμα κάθε έναν από τους 1.200.000 λεκτικούς τύπους του Μορφολογικού Λεξικού. Για τύπους εκτός αυτού, έχει αναμενόμενο ποσοστό λάθους <0.3%, το οποίο αφορά μόνο λέξεις που περιλαμβάνουν κάποιον από τους 24 προαναφερθέντες συνδυασμούς φωνηέντων.

  Η χρησιμότητα του Συλλαβιστή έγκειται στις γραμμές μικρού μήκους, όπως εφημερίδων, όπου τα συστήματα ηλεκτρονικής στοιχειοθεσίας παράγουν κακό αποτέλεσμα καθώς δημιουργούνται μεγάλα κενά σε κάποιες γραμμές. Συνεπώς εκτυπώσιμος χώρος μένει ανεκμετάλλευτος και χαλάει η αισθητική της σελίδας.

- Ορθογράφος

  Ο ορθογράφος αντλεί γλωσσικές μονάδες από λεξικά, γενικού και ειδικού λεξιλογίου καθώς και ξενόγλωσσα. Βασική πηγή, πάντως αποτελεί το Μορφολογικό Λεξικό της Neurolingo. Ο ορθογράφος ενσωματώνει συνεχώς λεκτικούς τύπους από

δημοσιογραφικά και λογοτεχνικά κείμενα καθώς επίσης και αναφορές από χρήστες.

Στα δευτερεύοντα λεξικά υπάγονται το *Λεξικό Ελληνικών Τοπωνυμιών* που περιλαμβάνει περίπου 10.000 ονόματα καποδιστριακών τοπωνυμιών και το *Αγγλικό Λεξικό* που περιέχει 200.000 αγγλικούς λεκτικούς τύπους. Το τελευταίο είναι χρήσιμο στην εξέταση ελληνοαγγλικών κειμένων σε μονόγλωσσα περιβάλλοντα όπως το *MS Outlook, Excel, Access, Mozilla Thunderbird κλπ* όπως επίσης και σε περιβάλλοντα που δεν παρέχουν λειτουργία ανίχνευσης γλώσσας με αλλαγή του πληκτρολογίου (*MS Word 1998 για Mac OS*).

- Λημματοποιητής

  Ο Λημματοποιητής παραλαμβάνει σαν είσοδο έναν λεκτικό τύπο και υποδεικνύει το λημματικό τύπο που αντιστοιχεί, π.χ. στον τύπο *κατέστη* επιστρέφει το λημματικό τύπο *καθιστώ*. Σε περίπτωση περισσότερων της μίας αντιστοιχιών, ο Λημματοποιητής τις επιστρέφει όλες, π.χ. για τον τύπο *απαντήσεις* επιστρέφει *απαντώ* και *απάντηση*. Η λειτουργία του βασίζεται στο Μορφολογικό Λεξικό της Neurolingo.

  Η αδρή μορφολογία της ελληνικής γλώσσας υπαγορεύει τη χρήση του Λημματοποιητή σε εφαρμογές ευρετηρίων και αναζήτησης, ειδικά στο πλαίσιο του παγκόσμιου ιστού. Η χρήση του συνεπάγεται τον εμπλουτισμό της αναζήτησης με παράγωγα λήμματα με βάση τον υποκείμενο λεκτικό τύπο, π.χ. πέρα από τον όρο *αυτοκίνητο* να πραγματοποιηθεί αναζήτηση και για τους όρους *αυτοκινήτου, αυτοκινήτων*. Συνεπώς, τα ερωτήματα διευρύνονται για να συμπεριλάβουν λεκτικούς τύπους που προέρχονται από το ίδιο λήμμα.

  Ο Λημματοποιητής βρίσκεται ενσωματωμένος στα σύστηματα ευρετηριασμού και αναζήτησης:

  – Apache Lucene,

    Η λειτουργικότητα παρέχεται μέσω απογόνου της Java κλάσης *org.apache.lucene.analysis.Analyzer*.

  – Oracle Text,

    Η λειτουργικότητα παρέχεται μέσω αποθηκευμένων διαδικασιών.

  – Micorsoft Indexing Service / SQL Server Full Text Search,

    Η λειτουργικότητα παρέχεται μέσω υλοποίησης της διεπαφής *iStemmer COM*

- Περιηγητής Θησαυρού

  Ο Περιηγητής Θησαυρού (βλ. εικόνα 2.1[8]) πραγματοποιεί αναζήτηση και παρουσίαση λημμάτων. Επίσης παρέχει πρόσβαση στην ηλεκτρονική έκδοση του Θησαυρού. Αναγνωρίζει λήμματα με βάση όλους τους κλιτικούς τύπους κάθε λέξης ακόμη και σε περίπτωση ανορθογραφίας χάρη στη συνεργασία με το Λημματοποιητή και τον Ορθογράφο.

Το Λεξισκόπιο συνεργάζεται επιτυχημένα με μια πληθώρα επεξεργαστών κειμένου:

- MS Word 2000/2002(XP)/2003 για Windows και X/2004 για Macintosh,

[8] http://www.neurolingo.gr/el/technology/application_tools/thesaurus_browser.jsp

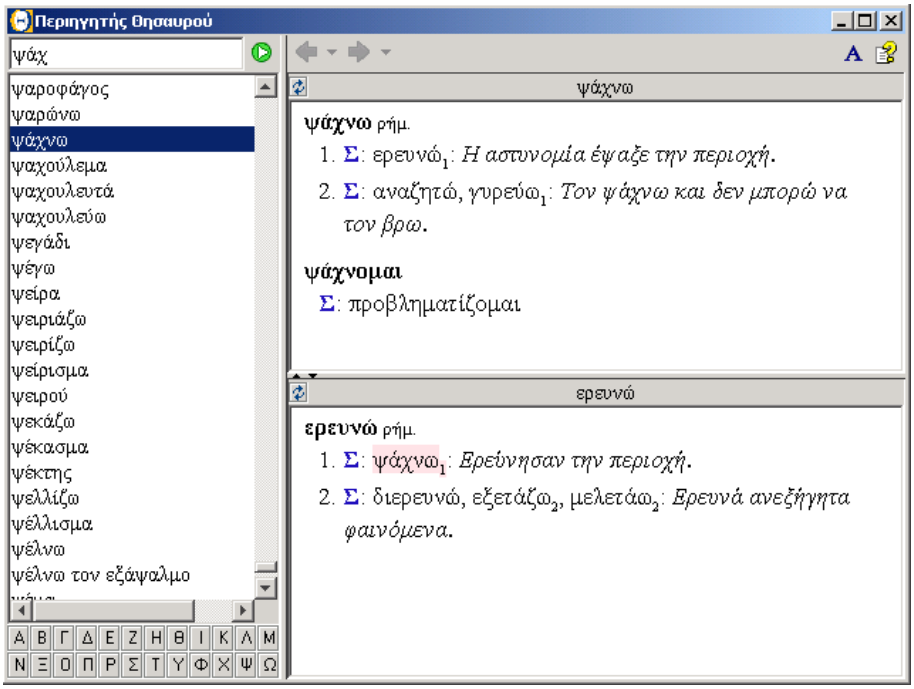

Figure 2.1: Γραφική διεπαφή Περιηγητή Θησαυρού

- OpenOffice Writer 1.1-1.5 & 2.0-2.0.2 για Windows, Linux, Solaris x86 και Solaris Sparc,
- StarOffice Writer 7 & 8 για Windows, Linux, Solaris x86 και Solaris Sparc,
- NeoOffice Writter 1.1-1.5 & 2.0-2.0.3 για Mac OS X,
- QuarkXpress 4.03 για Windows,
- Adobe InDesign 2.0/3.0(CS)/CS2 για Windows και Macintosh,
- Adobe Photoshop 7.0/8.0(CS)/CS2 για Windows και Macintosh,
- Adobe Illustrator CS/CS2 για Windows και Macintosh.

2.1.4 Αυτόματος Πολυτονιστής

Ο Αυτόματος Πολυτονιστής [36] είναι ένα προϊόν λογισμικού της ΜΑΤΖΕΝΤΑ το οποίο αναπτύχθηκε κατεξοχήν για τη μετατροπή μονοτονικών κειμένων σε πολυτονικά όμως παρέχει και λειτουργικότητα ορθογράφησης. Αναφορικά με τον πολυτονισμό, όπως χαρακτηριστικά ονομάζεται αυτή η μετατροπή, ο Αυτόματος Πολυτονιστής υποστηρίζει πολυτονισμό κειμένων γραμμένων στα αρχαία ελληνικά, στην αρχαΐζουσα, στην καθαρεύουσα και στη δημοτική ενώ εφαρμόζει κάθε φορά τη σωστή γραμματική.

Βασίζεται σε έξυπνους αλγόριθμους για την εφαρμογή των κανόνων τονισμού και χρησιμοποιεί πλούσια βάση δεδομένων λέξεων και τύπων. Ο πολυτονιστής κάνει

τη συγγραφή ορθογραφημένων πολυτονικών κειμένων εύκολη ακόμη και για τους αδαείς.

Ο Αυτόματος Πολυτονιστής είναι διασυνδεδεμένος με το Microsoft Word, όπου εκτελούνται όλες οι λειτουργίες. Ο χρήστης μπορεί να επιλέξει ρυθμίσεις πολυτονισμού, να εκτελέσει μετατροπές από και προς άλλα μονοτονικά και πολυτονικά συστήματα και να συγγράψει από την αρχή πολυτονικά κείμενα.

Τα πολυτονικά κείμενα που παράγει ο Αυτόματος Πολυτονιστής είναι μορφής *Unicode* και η επεξεργασία τους μπορεί να γίνει μέσω των εφαρμογών επεξεργασίας κειμένου *Microsoft Word, MATZENTA OfficeSuite Pro Windows* και μέσω επαγγελματικών προϊόντων σελιδοποίησης που υποστηρίζουν *Unicode* κωδικοποίηση π.χ. *inDesign*. Ακόμη και αν δεν υποστηρίζεται η εν λόγω κωδικοποίηση υπάρχει ρύθμιση στον Πολυτονιστή για τη μετατροπή του πολυτονικού κειμένου.

Ο πολυτονισμός με αυτό το εργαλείο είναι μια αυτόματη διαδικασία όπου ο χρήστης χρειάζεται μόνο στις περιπτώσεις που περισσότερες από μία επιλογές είναι σωστές. Ο Πολυτονιστής εμφανίζει πληροφορίες για κάθε επιλογή (εικόνα 2.2).

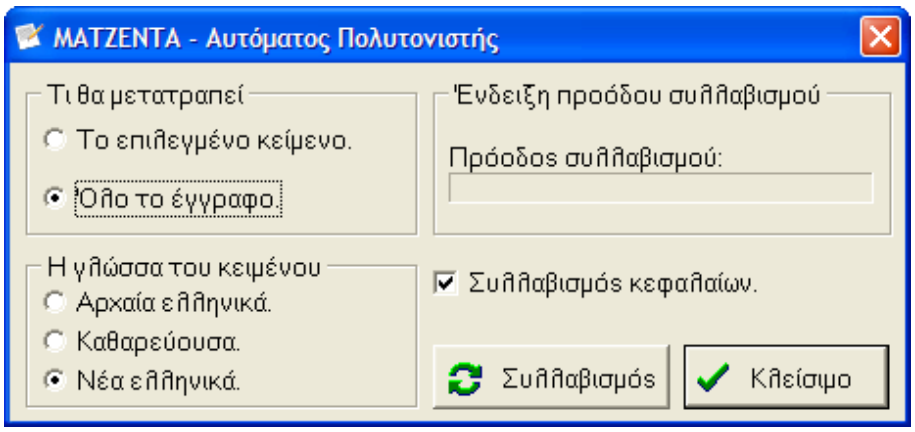

Figure 2.2: Γραφική διεπαφή Πολυτονιστή

Ο Πολυτονιστής υποστηρίζει τις εξής λειτουργίες:

- Αυτόματος πολυτονισμός μονοτονικών κειμένων της νέας ή αρχαίας ελληνικής.
- Συγγραφή πολυτονικών κειμένων
- Διόρθωση κειμένων για την αποδοτικότερη ορθογραφική επιμέλεια των κειμένων. Εμφανίζει κάθε λέξη που παρουσιάζεται στο κείμενο μία μόνο φορά εξοικονομώντας 70-80% χρόνου επιμέλειας.
- Συλλαβισμός πολυτονικού κειμένου, ώστε οι λέξεις να κόβονται στη σωστή συλλαβή όταν βρεθούν στο τέλος της γραμμής και να μην υπάρχουν μεγάλα κενά μεταξύ των λέξεων.
- Εξαγωγή σε άλλα πολυτονικά συστήματα, μιας και πολλές εφαρμογές δεν παρέχουν τη δυνατότητα επεξεργασίας *Unicode* εγγράφων(*QuarkXpress, PageMaker*). Υποστηρίζεται εξαγωγή σε *Macintosh, Πολυτονιστή 1 MATZENTA, πολυτονικό GR-soft* και *HTML* για έκδοση στο Διαδίκτυο.

- Προσθήκη λέξεων στη βάση δεδομένων του Πολυτονιστή
- Τεκμηρίωση μετατροπών, καθώς εμφανίζει πληροφορίες για την κλίση, τη δομή και τον τονισμό που επιλέγει για κάθε λέξη
- Διαρκής επέκταση του Πολυτονιστή μέσω Διαδικτύου. Ο Πολυτονιστής είναι ένα ενεργό σύστημα λογισμικού το οποίο αναβαθμίζεται συνεχώς. Οι νεότερες εκδόσεις του είναι άμεσα διαθέσιμες μέσω Διαδικτύου όμως η συντήρηση και επέκτασή του γίνεται κεντρικά από αφιερωμένο προσωπικό της εταιρίας.
- Συγγραφή μονοτονικού και πολυτονικού κειμένου σε υπολογιστές με λειτουργικό σύστημα Macintosh. Η διαλειτουργικότητα του ορθογραφικού λογισμικού είναι ένα πολύ σημαντικό θέμα στην εποχή της επαναχρησιμοποίησης εργαλείων και κώδικα με πρωτεστάντη το κίνημα ανοιχτού λογισμικού, όπου συσκευές με διαφορετικές προδιαγραφές παρέχουν προγράμματα διαχείρισης εγγράφων.

2.1.5 Ιατρολέξη

Η Ιατρολέξη [34] είναι ένα ερευνητικό έργο το οποίο ολοκληρώθηκε υπό την αιγίδα της *Γενικής Γραμματείας Έρευνας και Τεχνολογίας* και στόχευε στην παροχή υποδομής για εξειδικευμένες εφαρμογές επεξεργασίας φυσικής γλώσσας στον τομέα της Βιοϊατρικής όπως ευρετηρίαση κειμένων, εξαγωγή και ανεύρεση πληροφοριών μέσα από κείμενα, εξόρυξη πληροφορίας και λειτουργία ερωταποκρίσεων. Γι αυτό το σκοπό αναπτύχθηκαν εργαλεία και υποδομή εστιασμένα στο επιστημονικό πεδίο ώστε να υποστηρίξουν βέλτιστα τις ανάγκες διαχείρισης της ψηφιοποιημένης πληροφορίας.

- Μορφολογικό λεξικό

  Το Μορφολογικό λεξικό αποτελεί τη βάση στην οποία τα υπόλοιπα εργαλεία στηρίζονται για τη λειτουργία τους. Περιέχει 100.000 λέξεις της βιοϊατρικής ορολογίας.

- Περιηγητής Οντολογίας

  Επιτρέπει την περιήγηση στην οντολογία[9] των βιοϊατρικών τύπων που αναπτύχθηκε κατά τη διάρκεια του ερευνητικού έργου και την αναζήτηση βιοϊατρικών όρων.

- Ορθογραφικός Διορθωτής Ιστού[10]

  Χρησιμεύει στον έλεγχο της ορθογραφίας βιοϊατρικού όρου ή άλλου όρου της νεοελληνικής γλώσσας.

- Μορφοσυντακτικός και Σημασιολογικός Σχολιαστής[11]

  Χρησιμεύει στην εισαγωγή μορφοσυντακτικών σχολίων, από το Μορφολογικό λεξικό, και συμασιολογικών σχολίων, από την Οντολογία.

- Μορφοσυντακτικός Σχολιαστής[12]

  Παρουσιάζει τη μορφοσυντακτική επεξήγηση για κάθε λέξη ενός κειμένου.

[9] http://www.iatrolexi.gr/iatrolexi/files/Iatrolexi-corpus.pdf
[10] http://www.iatrolexi.gr/iatrolexi/webtools/speller/index.jsp
[11] http://www.iatrolexi.gr/iatrolexi/webtools/annotator/index.jsp
[12] http://www.iatrolexi.gr/iatrolexi/tools/morphoTagger.php

- Συμφραζόμενα όρων[13]

  Χρησιμεύει στην αναζήτηση όρων τους οποίους παρουσιάζει σε όλους τους κλιτικούς τύπους αν επιλεγεί λημματοποίηση μαζί με τα συμφραζόμενα από την ίδια γραμμή κειμένου. Επίσης ο όρος που αναζητήθηκε μπορεί να ιχνηλατηθεί στο κείμενο.

Τα εργαλεία είναι εγκατεστημένα σε εξυπηρετητές στο Διαδίκτυο και ως εκ τούτου είναι διαθέσιμα προς δοκιμή-χρήση διαδικτυακά. Παρόλα αυτά δεν προσφέρεται κάποιος τρόπος επέμβασης σε αυτά με στόχο τη συντήρηση και επέκταση της λειτουργίας τους.

2.1.6 Ispell

Ο Ispell [15, 19, 14] είναι ένας ορθογράφος φτιαγμένος για συστήματα *Unix* και υποστηρίζει τις περισσότερες Δυτικές γλώσσες. Από το 1997 υπάρχει και υποστήριξη για την ελληνική γλώσσα[14]. Παρέχει πληθώρα διεπαφών για την ενσωμάτωση με άλλα συστήματα, συμπεριλαμβανομένης μιας προγραμματιστικής διεπαφής για τη διασύνδεση με συντάκτες κειμένου όπως ο *emacs*.

Ο Ispell προτείνει διορθώσεις με βάση την απόσταση *Damerau-Levenshtein*[15] που πρέπει να είναι το πολύ 1. Δεν στηρίζεται σε κανόνες προφοράς για να προβλέψει πιο δύσκολες διορθώσεις.

Ιστορικά ο Ispell συνδέεται με ένα πρόγραμμα αναπτυγμένο από τον *R. E. Gorin* σε *Assembly* γλώσσα για την οικογένεια υπολογιστικών μηχανημάτων *PDP-10*. Στη συνέχεια αναπτύχθηκε διασύνδεση για τη γλώσσα *C* και ο ορθογράφος συνέχισε να επεκτείνεται. Το γενικευμένο περιγραφικό σύστημα του Ispell, που βασίζεται στα μορφήματα, χρησιμοποείται και από άλλους ορθογράφους.

Ο Ispell λειτουργεί διαβάζοντας ένα αρχείο κειμένου λέξη λέξη μέχρι να συνάντησει κάποια λέξη που δεν υπάρχει στο λεξικό του. Σε αυτή την περίπτωση παρουσιάζει παρεμφερείς έγκυρες λέξεις και δίνει επιπλέον τη δυνατότητα στο χρήστη να προσθέσει την άγνωστη λέξη στο λεξικό.

Ο Ispell κατέχει την πρωτοπορία στη χρήση της προγραμματιστικής διεπαφής, που αρχικά προοριζόταν για τον *emacs*. Εφαρμογές πλέον χρησιμοποιούν αυτή τη δυνατότητα για να παρέχουν την υπηρεσία του ορθογράφου. Ο Ispell διατίθεται με συγκεκριμένη άδεια ανοιχτού λογισμικού.

2.1.7 GNU Aspell

Ο GNU Aspell [3, 5, 4], ή απλά Aspell, είναι ένας ορθογράφος ελεύθερου, ανοιχτού λογισμικού που αναπτύχθηκε για να αντικαταστήσει τον Ispell. Αποτελεί τον προτυποποιημένο ορθογράφο για το σύστημα GNU. Μπορεί να χρησιμοποιηθεί επίσης σε άλλα συστήματα τύπου Unix και σε Windows.

Το λογισμικό διατίθεται με άδεια *GNU Lesser General Public Licence (GNU LGPL)* ενώ η τεκμηρίωση με *GNU Free Documentation Licence (GNU FDL)*. Ο Aspell υποστηρίζει πάνω από 70 γλώσσες μεταξύ των οποίων και η ελληνική.

Η ανωτερότητα του Aspell σε σχέση με τον Ispell βρίσκεται στη δυνατότητα του πρώτου να διορθώσει κείμενα με κωδικόποιηση *UTF-8* χωρίς τη χρήση ειδικών λεξικών. Επίσης ο Aspell παρέχει υποστήριξη για τη χρήση πολλαπλών λεξικών άμεσα και

[13] http://www.iatrolexi.gr/iatrolexi/webtools/concordancer/index.html
[14] http://dmst.aueb.gr/dds/sw/greek/ispell/index.html
[15] http://en.wikipedia.org/wiki/Damerau–Levenshtein_distance

χειρίζεται έξυπνα προσωπικά λεξικά όταν εκτελούνται περισσότερες από μία Aspell διαδικασίες.

Ο Aspell έχει ενσωματωθεί στα ακόλουθα συστήματα λογισμικού:

- *Pidgin*[16],
- *Digsby*[17],
- *Lyx*[18],
- *Notepad++*[19],
- *Opera*[20],
- *gedit*[21],
- *abiword*[22]

Μετρήσεις[23] έχουν δείξει ότι ο Aspell σταθερά προτείνει καλύτερες διορθώσεις από τον Ispell και μάλιστα αποδίδει πολύ καλύτερα από τον ορθογράφο του Microsoft Word 1997 και τους υπόλοιπους ορθογράφους της εποχής που αναπτύχθηκε. Επίσης παρέχει υποστήριξη για έλεγχο αρχείων *(La)TeX* και *HTML* και για άλλες γλώσσες πέραν της αγγλικής στο χρόνο εκτέλεσης.

2.1.8 MySpell

Ο ορθογράφος MySpell [17, 23] συμπεριλαμβανόταν στον *OOo Writer*[24] του *OpenOffice.org*. Από την έκδοση 2.0.2 το *OpenOffice.org* χρησιμοποιεί τον ορθογράφο *Hunspell* (2.1.10).

Ο MySpell αναπτύχθηκε για να ολοκληρώσει διάφορα εργαλεία ορθογράφησης στο *OpenOffice.org*. Είναι γραμμένος σε *C++* και υποστηρίζει συμπίεση μορφημάτων βασισμένος στον *Ispell*. Ο MySpell είναι πολύ αποδοτικός από άποψη χώρου που καταναλώνει καθώς αποθηκεύει βάσεις λέξεων και αναφορές σε μορφήματα με τα οποία αυτές συνδέονται.

Ο MySpell χρησιμοποιείται από τα ακόλουθα προγράμματα:

- *Mozilla Thunderbird 2.0* και παλαιότερες εκδόσεις. Ο *Mozilla Thunderbird 3.0* χρησιμοποιεί το *Hunspell*.
- *Mozilla Firefox 2.0* και παλαιότερες εκδόσεις. Ο *Mozilla Firefox 3.0* χρησιμοποιεί το *Hunspell*.
- *Aegisub 2.0*[25] μπορεί να χρησιμοποιήσει ένα λεξικό με μικρές τροποποιήσεις.

Ο ορθογράφος *Aspell* (2.1.7) και ο επεξεργαστής κειμένου *Vim 7* μπορούν να χρησιμοποιήσουν ένα λεξικό που έχει δημιουργηθεί για το MySpell.

[16] http://en.wikipedia.org/wiki/Pidgin_(software)
[17] http://en.wikipedia.org/wiki/Digsby
[18] http://en.wikipedia.org/wiki/LyX
[19] http://en.wikipedia.org/wiki/Notepad\%2B\%2B
[20] http://en.wikipedia.org/wiki/Opera_(web_browser)
[21] http://en.wikipedia.org/wiki/Gedit
[22] http://en.wikipedia.org/wiki/Abiword
[23] http://aspell.net/test/cur/
[24] http://en.wikipedia.org/wiki/OpenOffice.org_Writer
[25] http://en.wikipedia.org/wiki/Aegisub

2.1.9 MySQL Spell Checker

Ο MySQL spell checker [26] είναι ένας απλοϊκός ορθογράφος. Αντιπαραβάλλει λέξεις σε ένα κείμενο ενώ το λεξικό του είναι αποθηκευμένο σε μια βάση δεδομένων *MySQL*. Με τη βοήθεια της μεθόδου *soundex()* της *PHP* προτείνει διορθώσεις για λέξεις που δεν αναγνωρίζει.

2.1.10 Hunspell

Ο Hunspell [24] είναι ένας ορθογράφος και μορφολογικός αναλυτής ειδικά σχεδιασμένος για γλώσσες που παρουσιάζουν έντονη μορφολογία, πολύπλοκη λεκτική δομή και κωδικοποίηση χαρακτήρων. Η λειτουργία του στηρίζεται σε λεξικά με καταχωρημένους τύπους λέξεων και σε κανόνες συμπεριφοράς του ορθογραφικού διορθωτή, υπό συγκεκριμένες συνθήκες. Όπως παραπέμπει και τ' όνομά του αναπτύχθηκε αρχικά για την Ουγγρική γλώσσα.

Είναι βασισμένος πάνω στον ορθογράφο *MySpell* (2.1.8) και διατηρεί συμβατότητα με τα λεξικά του *MySpell*. Μια κύρια διαφορά μεταξύ των δύο είναι ότι ο τελευταίος χρησιμοποιεί κωδικοποίηση ενός Byte για κάθε χαρακτήρα που αναπαριστά ο υπολογιστής ενώ ο πρώτος *Unicode UTF-8*.

Ο Hunspell είναι ενσωματωμένος στα εξής προϊόντα λογισμικού:

- *OpenOffice.org*, από την έκδοση 2.0.2 και ύστερα
- *Mozilla Firefox*, από την έκδοση 3 και ύστερα
- *Mozilla Thunderbird*, από την έκδοση 3 και ύστερα
- *Mozilla Seamonkey*, από την έκδοση 2 και ύστερα
- *Eclipse*, με τη χρήση *Hunspell4Eclipse*
- *WinShell*, ολοκληρωμένο περιβάλλον χρήσης για *TeX* και *LaTeX* σε *Windows*
- *Yudit*, *Unicode* επεξεργαστής κειμένου σε *Windows*
- *Opera 10+*,
- *Google Chrome*,
- *The Bat!*, πρόγραμμα διαχείρισης ηλεκτρονικού ταχυδρομείου
- *Apple Mac OS X Snow Leopard*
- *Omega T*, εργαλείο μετάφρασης ανοιχτού λογισμικού
- *XTuple*, εφαρμογή προγραμματισμού πόρων
- *XTuple*,
- *XTuple*, εφαρμογή προγραμματισμού πόρων

Επίσης μπορεί να χρησιμοποιηθεί από το *Enchant* (2.2.2) και τον *Emacs*. Ο Hunspell αποτελεί ανοιχτό λογισμικό και διατίθεται με άδειες *GPL, LGPL* και *MPL-trilicense*. Μπορεί να διασυνδεθεί με διάσημες γλώσσες προγραμματισμού όπως είναι η *Java, Perl, Python, .NET, Delphi, Ruby* και *JavaScript*.

Τα χαρακτηριστικά του Hunspell τον κάνουν ιδανικό για τον ορθογραφικό έλεγχο της ελληνικής γλώσσας, που παρουσιάζει πλούσια μορφολογία, σύνθετη κωδικοποίηση και λεκτική δομή. Η ήδη υπάρχουσα βάση των 500.000 λέξεων και παραπάνω αποτελεί επίσης σημαντικό πλεονέκτημα.

Τέλος, ο Hunspell αποτελεί λογισμικό ανοιχτού κώδικα και, όπως φαίνεται και από τα παραδείγματα, χρησιμοποιείται ευρέως σε έργα ανοιχτού λογισμικού, αναδεικνύοντας με αυτό τον τρόπο τη φιλική άδεια χρήσης του.

2.2 Διεπαφές Ορθογραφικών Λεξικών της Ελληνικής

2.2.1 Pspell

Ο ορθογράφος Pspell (Portable Spell Checker Interface Library) [4, 16] δημιουργήθηκε για να παρέχει μια ενιαία διεπαφή προς τα συστήματα ορθογράφησης. Χρησιμοποιήθηκε στον προγραμματισμό όπως στη γλώσσα *C* και διατίθεται με την άδεια *GNU LGPL*.

Ο Pspell είναι ανενεργός από το 2001 και έχει αντικατασταθεί από τον ορθογράφο *Aspell* (2.1.7).

2.2.2 Enchant

Το Enchant [8, 21] είναι ελεύθερο λογισμικό που αναπτύχθηκε ως τμήμα του επεξεργαστή κειμένου *AbiWord* για να παρέχει πρόσβαση σε υπάρχοντες ορθογράφους. Παρουσιάζει μια ενιαία διεπαφή για τη διασύνδεση με κατεξοχήν λειτουργικότητα που παρέχει ένας κοινός ορθογράφος. Έχει τη δυνατότητα να φορτώσει πολλές διεπαφές με ορθογράφους αμέσως.

Αναλυτικά, το Enchant παρέχει διεπαφές για τους εξής ορθογράφους:

- *Aspell* (2.1.7)
- *Pspell* (2.2.1)
- *Ispell* (2.1.6)
- *Hunspell* (2.1.10)
- *MySpell* (2.1.8)
- *Uspell*, υποστηρίζει κυρίως Εβραϊκά και Ανατολικές Ευρωπαϊκές γλώσσες και φιλοξενείται στο αποθετήριο του *AbiWord*
- *Hspell*, υποστηρίζει την Εβραϊκή γλώσσα
- *AppleSpell*, χρησιμοποιείται από το λειτουργικό σύστημα Mac OS X
- *Voikko*[26], υποστηρίζει τη Φινλανδική γλώσσα
- *Zemberek*, υποστηρίζει την τουρκική γλώσσα

Διατίθεται με άδεια *GNU LGPL* με την επιπλέον δήλωση ότι οποιοδήποτε εργαλείο μπορεί να φορτωθεί και να χρησιμοποιηθεί από το Enchant. Με αυτόν τον τρόπο μπορεί να χρησιμοποιεί τους ορθογράφους κάθε πλατφόρμας και ταυτόχρονα να δίνει τη δυνατότητα στους χρήστες να χρησιμοποιήσουν τον ορθογράφο που προτιμούν.

2.2.3 cocoAspell

Ο cocoAspell [22] είναι μια διεπαφή του *Aspell* (2.1.7) ειδικά φτιαγμένη για το λειτουργικό σύστημα Mac OS X. Τα πλεονεκτήματα που παρέχει ο cocoAspell συνοψίζονται σε:

[26] `http://voikko.sourceforge.net/`

- Υποστήριξη διόρθωσης εγγράφων σε επίπεδο συστήματος. Όλες οι εφαρμογές που χρησιμοποιούν τη διεπαφή του συστήματος για διόρθωση εγγράφων έχουν πρόσβαση στον ορθογράφο. Για παράδειγμα, οι εφαρμογές *Mail, OmniWeb, ProjectBuilder, TextEdit* μπορούν να χρησιμοποιήσουν τον ορθογράφο σε διαφορετικές γλώσσες.

- Ένα παράθυρο προτιμήσεων παρέχεται με τον cocoAspell ως διεπαφή για την επιλογή και παραμετροποίηση λεξικών. Πολλές επιλογές είναι διαθέσιμες έτσι ώστε να μπορεί κάθε χρήστης να προσαρμόσει τον ορθογράφο στις ανάγκες του.

2.2.4 KSpell

Ο KSpell [11] παρέχει πρόσβαση στους ορθογράφους *Ispell* (2.1.6), *Aspell* (2.1.7) και *Hspell* κατ' επιλογή. Παρέχει επίσης διεπαφή με λειτουργικότητα εύρεσης, προσθήκης, αντικατάστασης λέξεων. Μπορεί να χρησιμοποιηθεί για τον έλεγχο αρχείων *ASCII*, για την υλοποίηση διόρθωσης στο Διαδίκτυο και για τη διόρθωση αρχείων ειδικής μορφής.

Καθώς το Διαδίκτυο καλύπτει με την πάροδο του χρόνου όλο και περισσότερες ανάγκες χρηστών, αυξανόμενα εκμεταλλευόμαστε τις τεχνολογίες που προσφέρει για επεξεργασία εγγράφων (π.χ. έγγραφα *Google*). Συνεπώς, η υπηρεσία ενός ορθογράφου που λειτουργεί , συντηρείται και επεκτείνεται στο ίδιο περιβάλλον αξιολογείται πολύ σημαντική και αυξάνει την παραγωγικότητα των χρηστών.

2.2.5 Flyspell

Ο Flyspell [27] αποδίδει λειτουργία ορθογράφησης στον *Emacs* τη στιγμή της πληκτρολόγησης. Παρεμβαίνει ελάχιστα στην εφαρμογή ενώ στιγματίζει λανθασμένες λέξεις με την ολοκλήρωσή τους.

Ο Flyspell είναι ανεξάρτητος των γλωσσών γιατί ο χρήστης είναι ελεύθερος να διαλέξει το δικό του λεξικό. Είναι συμβατός με αρχεία τύπου *TeX*. Ο χρήστης μπορεί μάλιστα να επιλέξει την εκκίνηση του ορθογράφου αυτόματα με τη χρήση εντολής στα αρχεία *TeX*.

Προτείνει διορθώσεις για λανθασμένες λέξεις με τη μορφή αναδυόμενων μενού. Πατώντας δεξί κλικ σε μια στιγματισμένη λέξη εμφανίζεται ένα μενού με επιλογές διόρθωσης. Εναλλακτικά, δίνεται η δυνατότητα αγνόησης του λάθους ή πρόσθεσης της λέξης στο λεξικό.

Τέλος, ο Flyspell παρέχει δυνατότητες αυτόματης διόρθωσης. Με την εντολή *M - \t* στον *Emacs* αντικαθιστά τη λανθασμένη λέξη με μια πιθανή διόρθωση. Σε περίπτωση που περισσότερες της μίας διορθώσεων είναι πιθανές, αυτές ταξινομούνται και σε κάθε εκτέλεση της εντολής μια νέα διόρθωση, η επόμενη, επιλέγεται από τη λίστα. Η επιλογή μπορεί να γίνει είτε αλφαβητικά είτε με βάση την πιθανότητα να είναι σωστή η αντικατάσταση.

2.2.6 GNOME-Spell

Ο ορθογράφος GNOME-Spell [12] βασίζεται στον ορθογράφο *Pspell* και παρέχει υπηρεσίες ελέγχου ορθογραφίας σε εφαρμογές που εκτελούνται στο γραφικό περιβάλλον GNOME. Η τελευταία του έκδοση 1.0.7, παρουσιάστηκε τον Οκτώβριο 2006.

2.2.7 GtkSpell

Ο GtkSpell [10] παρέχει υπόδειξη και αντικατάσταση ανορθογραφημένων λέξεων. Με δεξί κλικ πάνω σε μια λανθασμένη λέξη εμφανίζει διάφορες επιλογές αντικατάστασης. Χρησιμοποιεί πλήρως βιβλιοθήκη *Pspell* ή *Aspell*. Η τελευταία του έκδοση 2.0.11 παρουσιάστηκε τον Μάιο 2005.

2.2.8 GaSpell

Ο GaSpell [9] παρουσιάζει γραφική διεπαφή του ASpell προς το χρήστη. Η τελευταία του έκδοση 0.30-4 παρουσιάστηκε τον Ιουλίο 2000.

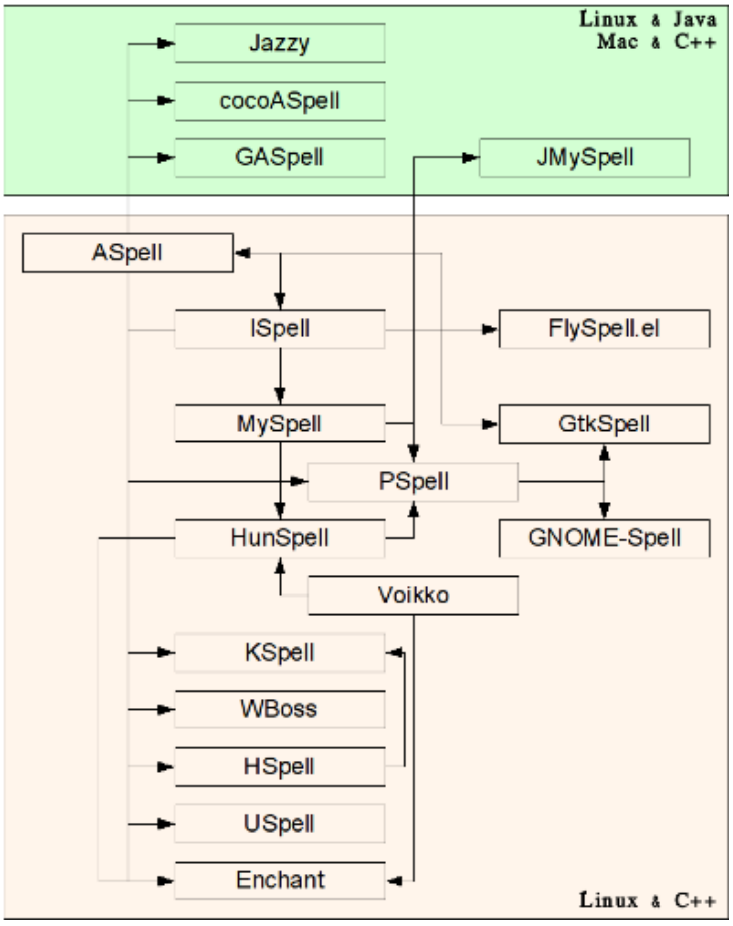

Figure 2.3: Συσχέτιση ορθογράφων

Το γράφημα (βλ. εικόνα 2.3[27]) απεικονίζει την εξέλιξη των ορθογράφων. Ορθογράφοι που παρουσιάζονται μόνο στο διάγραμμα δεν παρέχουν υποστήριξη για την ελληνική γλώσσα. Οι ορθογράφοι συσχετίζονται μεταξύ τους σε κώδικα ή αλγοριθμική λογική.

[27] http://nemertes.lis.upatras.gr/dspace/bitstream/123456789/1813/1/diplwmatiki.pdf

2.3 Συνεργατική Ανάπτυξη Λεξικών μέσω Δικτύου

2.3.1 Συνεργατικές Τεχνολογίες και Ανοιχτό Λογισμικό σε Έργα Επεξεργασίας Φυσικής Γλώσσας

Οι Streiter et al [30] υπογραμμίζουν τη σημασία της συνεργατικής ανάπτυξης εργαλείων επεξεργασίας ηλεκτρονικών κειμένων, όπως ορθογράφοι, ελεγκτές γραμματικής και βοηθοί συλλαβισμού για τις επονομαζόμενες *μη κεντρικές γλώσσες*[28].

Υποστηρίζουν ότι επιτυχημένα στυλ επιστημονικής συνεργασίας όπως συναντάμε στην ανάπτυξη ανοιχτού λογισμικού και στην ανάπτυξη της *Wikipedia* μπορούν να οφελήσουν σημαντικά το σχεδιασμό έργων επεξεργασίας φυσικής γλώσσας ειδικά για τις μη κεντρικές γλώσσες. Οι συγγραφείς δίνουν έμφαση στο ανοιχτό λογισμικό, στη σημασία του αποθετηρίου λογισμικού και προτείνουν οι μη κεντρικές γλώσσες να ενστερνιστούν την προσέγγιση του ανοιχτού λογισμικού για τους πόρους που αναπτύσσουν.

Η επιλογή αυτή παρουσιάζει σημαντικά πλεονεκτήματα:

- συντήρηση και επέκταση κώδικα και δεδομένων-γλωσσικών μονάδων από αφοσιωμένη κοινότητα

  Οι μη κεντρικές γλώσσες δύσκολα τυγχάνουν χρηματοδότησης για εφαρμογές επεξεργασίας φυσικής γλώσσας και στερούνται βασικών εργαλείων. Η εξασφάλιση της συνέχισής τους είναι θεμελιώδης προτεραιότητα.

  Η κοινότητα ανοιχτού λογισμικού είναι μιας πρώτης τάξης ευκαιρία να ξεφύγουν οι μη κεντρικές γλώσσες από την απομόνωση και την αφάνεια χωρίς να εμπλακούν σε άνισο ανταγωνισμό.

  Το ιδεώδες του ανοιχτού λογισμικού βοηθά προς αυτή την κατεύθυνση αφού προσφέρει ισχυρά κίνητρα ενεργής συμμετοχής στη βάση της ελευθερίας και συνιδιοκτησίας του λογισμικού.

  Δεδομένα σε ένα αποθετήριο που μοιράζονται δομή και χαρακτηριστικά με τα υπόλοιπα δεδομένα στο αποθετήριο οφελούνται κατά τη συντήρηση αφού μπορούν να εισπράξουν ενημερώσεις αυτόματα ακόμη και αν δεν προορίζονταν γι αυτά.

- παραγωγή δημοσιεύσεων και ώθηση της επιστημονικής προόδου

  Το κίνημα ανοιχτού λογισμικού υποστηρίζει εκ βάσης τη διαθεσιμότητα κώδικα και δεδομένων. Ως εκ τούτου η διάχυση και η επιβεβαίωση των επιστημονικών αποτελεσμάτων σε συνέδρια και περιοδικά βοηθούνται σημαντικά με ορίζοντα την επιστημονική πρόοδο.

2.3.2 Περιβάλλον Συντήρησης του Ηλεκτρονικού Λεξικού ΣΥΜΦΩΝΙΑ

Αναπτύχθηκε το 2003 εν μέσω συνεργασίας του *ΙΕΛ* με την *Καθημερινή* για χρήση στο περιβάλλον της δεύτερης. Επωμίζεται τη διαρκή ανάπτυξη του συστήματος διόρθωσης κειμένων της Καθημερινής (υποενότητα 2.1.1) συνεργατικά μέσω δικτύου (βλ. εικόνα 2.4[29]).

Το περιβάλλον συντήρησης:

[28] Το χαρακτηριστικό της κεντρικότητας για μια γλώσσα μετράται από την ύπαρξη και το επίπεδο εργαλείων επεξεργασίας φυσικής γλώσσας και έρευνας.

[29] www.ilsp.gr/files/KathimeriniStory_5_7_05.ppt

- υποστηρίζει διαχείριση του ηλεκτρονικού λεξικού για πλατφόρμες *Windows* και *Macintosh*.
- αποτελεί ανεξάρτητη εφαρμογή που εκτελείται σε περιβάλλον *Windows*
- Χρησιμοποιείται μόνο από εξουσιοδοτημένους χρήστες
- Υποστηρίζει την εισαγωγή, μεταβολή και διαγραφή λέξεων
- Υποστηρίζει τη δημιουργία, τροποποίηση και διαγραφή υπολεξικών
- Ενημερώνει το λεξικό από τα προσωπικά λεξικά που διατηρούν οι χρήστες τοπικά στον υπολογιστή τους. Οι προτεινόμενες προσθήκες ελέγχονται από τους διορθωτές της εφημερίδας.
- Η μορφή του λεξικού για *Macintosh* παράγεται αυτόματα μετά την ενημέρωση της αντίστοιχης των Windows.

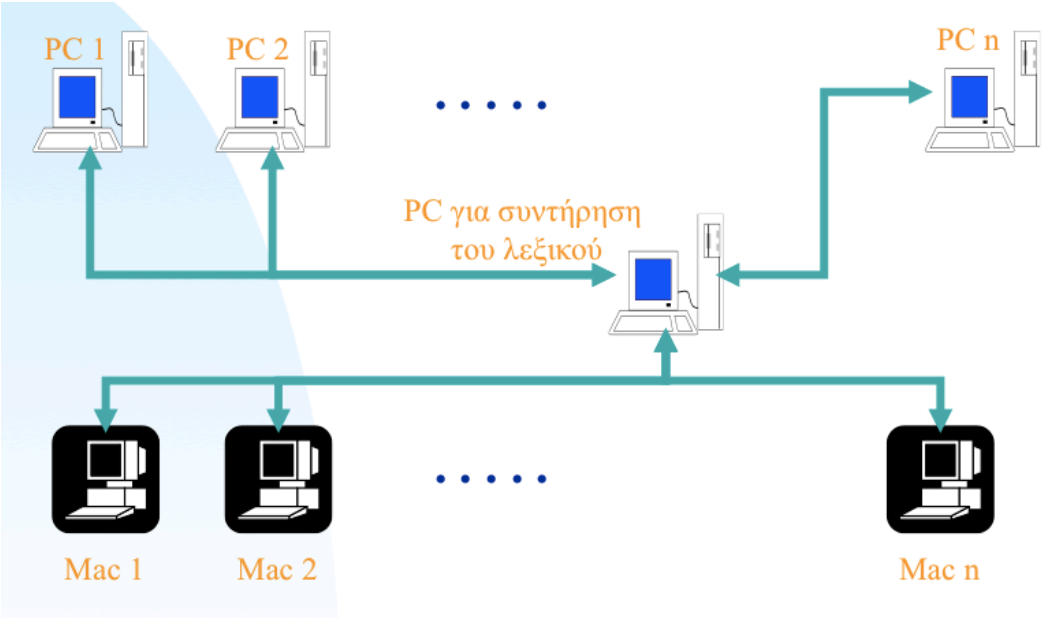

Figure 2.4: Διάγραμμα λειτουργίας περιβάλλοντος συντήρησης

Το περιβάλλον συντήρησης επιτρέπει στους εξουσιοδοτημένους χρήστες να εισάγουν νέες λέξεις στο λεξικό. Οι πρόσθετες λέξεις αποστέλλονται στον υπολογιστικό κόμβο που εκτελείται το περιβάλλον συντήρησης και εισάγονται σε μια ουρά για να ελεγχθεί η εγκυρότητά τους από τους διορθωτές της εφημερίδας. Μετά την επιβεβαίωση εισαγωγής των νέων λέξεων στο λεξικό, το λεξικό επαναμεταγλωττίζεται για να συμπεριλάβει τους νέους γλωσσικούς τύπους και οι υπολογιστές των χρηστών ενημερώνονται αυτόματα με την καινούρια έκδοση του λεξικού.

Ένα κύριο χαρακτηριστικό του συστήματος είναι η ιεράρχηση των χρηστών σε εξουσιοδοτημένους, μη εξουσιοδοτημένους και ελεγκτές. Μόνο οι εξουσιοδοτημένοι χρήστες μπορούν να κάνουν αλλαγές και υποθέτωντας ότι κάθε χρήστης ταυτοποιείται

με ένα συγκεκριμένο υπολογιστή, κάθε αλλαγή συνδέεται με έναν χρήστη. Επιπλέον, οι αλλαγές ελέγχονται ως προς την ορθότητά τους από τους έμπειρους ελεγκτές και με αυτόν τον τρόπο διασφαλίζεται η ακεραιότητα των δεδομένων.

2.3.3 Collaborative Open Lexicon Development

Η προσπάθεια αυτή [29] έχει ως στόχο την οργάνωση ενός λεξικού που τροφοδοτείται από έγγραφα που κυκλοφορούν στο Διαδίκτυο, εκπαιδεύεται μέσω τεχνικών μηχανικής μάθησης, ενημερώνεται από χρήστες και έρχεται να προβάλλει την ανάγκη διαλειτουργικότητας.

Ο πηρύνας της ιδέας αποτελείται από:

- ανοιχτή δομή, βασισμένη στην τεχνολογία *XML*
- ανοιχτά πρωτόκολλα επικοινωνίας, βασισμένα στην τεχνολογία XML ερώτησης απάντησης
- ανοιχτή συμμετοχή, με προσθήκη λέξεων μετά από επιβεβαίωση
- ανοιχτό λογισμικό και δεδομένα, με ελεύθερα εργαλεία λογισμικού και περιεχόμενο λεξικού
- κείμενα ως πηγή εμπλουτισμού του λεξικού, ώστε να αντανακλάται η χρήση και η σημασία των όρων στην καθημερινότητα

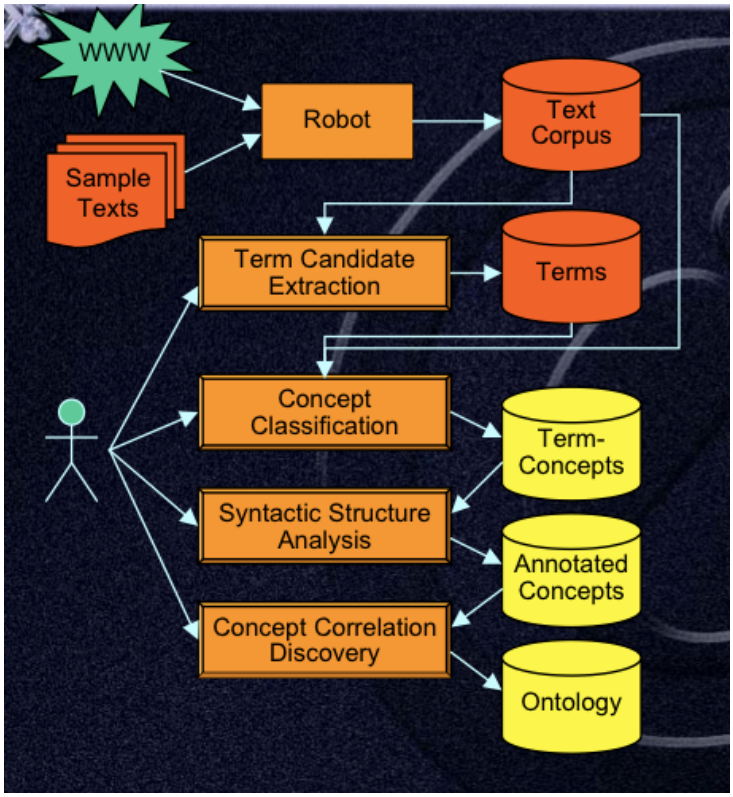

Figure 2.5: Διάγραμμα λειτουργίας ανοιχτού λεξικού

Η λειτουργία του λεξικού, όπως φαίνεται και στην εικόνα 2.5 [29] ξεκινά από τη συλλογή εγγράφων που προέρχονται είτε από το Διαδίκτυο είτε αποτελούν δείγματα προς εκπαίδευση του υποσυστήματος τεχνητής νοημοσύνης. Το υποσύστημα τεχνητής νοημοσύνης πραγματοποιεί, μέσω μηχανικής μάθησης, εξαγωγή υποψήφιων όρων προς εισαγωγή στο λεξικό με τη συμμετοχή χρηστών. Οι χρήστες από το παρόν βήμα συμβάλλουν στη διαδικασία και συμπληρώνουν τις τεχνολογίες τεχνητής νοημοσύνης.

Στη συνέχεια, πραγματοποιείται κατηγοριοποίηση των όρων με βάση το πλαίσιο χρήσης και ακολουθεί συντακτική ανάλυση της δομής της γλωσσικής μονάδας. Οι χρήστες αποδίδουν σημασιολογικές συσχετίσεις ανάμεσα σε όρους με αποτέλεσμα τη δημιουργία, εμπλουτισμό και εκπαίδευση οντολογίας. Η οντολογία διατηρεί κανόνες τους οποίους χρησιμοποιεί για να συσχετίσει σημασιολογικά όρους μεταξύ τους.

Η αρχιτεκτονική του συστήματος (βλ. εικόνα 2.6 [29]) συνδυάζει κεντρικοποιημένες δραστηριότητες, προς εξασφάλιση ακεραιότητας και συνέπειας των δεδομένων και αποκεντρωμένο δίκτυο τοποθεσιών, προς εξασφάλιση ανοιχτής συμμετοχής και ανεμπόδιστης κλωνοποίησης των δεδομένων του συστήματος.

Η αρχιτεκτονική είναι βασισμένη σε εργασίες, οι οποίες παράγονται από τα εργαλεία ανάλυσης κειμένων. Οι συμμετέχοντες αναλαμβάνουν εργασίες μέσω Διαδικτύου, εργάζονται τοπικά στον υπολογιστή τους και τις παραδίδουν όταν τις διεκπεραιώσουν.

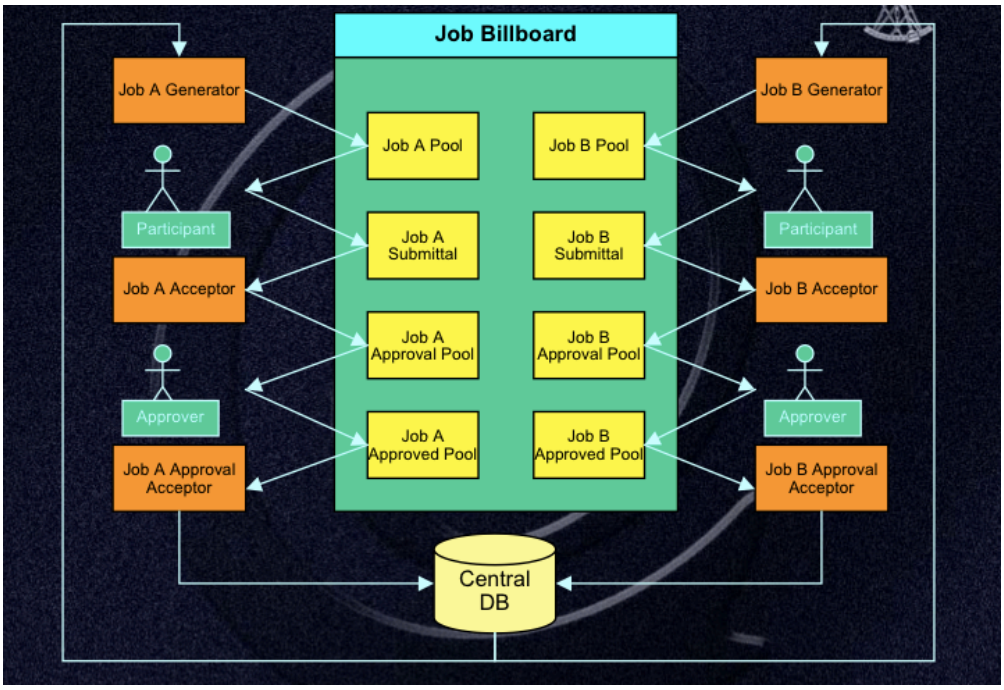

Figure 2.6: Διεκπεραίωση εργασιών ανοιχτού λεξικού

Το πλήρες δίκτυο σύνδεσης και επικοινωνίας (εικόνα 2.7 [29]) περιλαμβάνει υπολογιστικούς κόμβους:

- συμμετεχόντων, οι οποίοι διεκπεραιώνουν εργασίες
- ελεγκτών, οι οποίοι μοιράζονται την ίδια βάση δεδομένων, στενά συγχρονισμένη και παρέχουν υπηρεσία στους γειτονικούς κόμβους συμμετεχόντων
- διαμεσολαβητών (προαιρετικά), οι οποίοι διανέμουν την επικοινωνία μεταξύ ελεγκτών και συμμετεχόντων

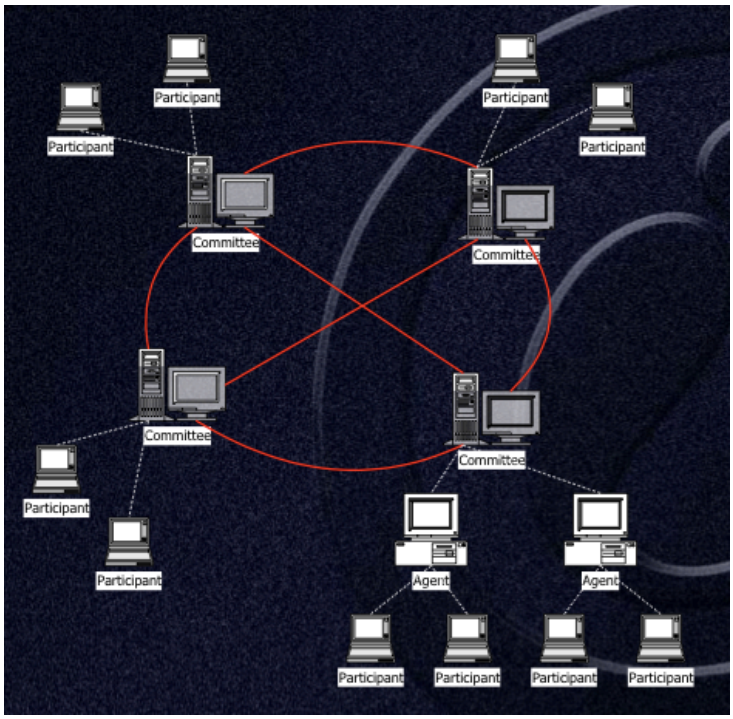

Figure 2.7: Δίκτυο σύνδεσης ανοιχτού λεξικού

2.3.4 Collaborative Dictionary of the English language

Το συνεργατικό λεξικό της αγγλικής γλώσσας [32] έχει τη βάση του στο λεξικό *Webster* του 1913 και έχει συμπληρωθεί με κάποιους ορισμούς από το λεξικό *WordNet*[30]. Κινητήρια δύναμή του πάντως είναι οι εθελοντές ανά τον κόσμο που συνεισφέρουν λέξεις και ο αυστηρός έλεγχος των προτεινόμενων προσθηκών.

Το ηλεκτρονικό λεξικό αποτελεί επίσης το θεμέλιο λίθο για την ανάπτυξη ένος σύγχρονου, εύπεπτου, εγκυκλοπαιδικού λεξικού, το οποίο θα είναι προσβάσιμο ελεύθερα μέσω του Διαδικτύου. Η συντήρηση και επέκταση του λεξικού εναπόκειται, και πάλι,

[30] http://wordnet.princeton.edu/

στην απανταχού κοινότητα χρηστών με στόχο να δημιουργηθεί μια μεγάλη βάση γνώσης διαθέσιμη δωρεάν.

Στο Διαδίκτυο επίσης κυκλοφορούν παράγωγες εκδόσεις του λεξικού σε διαφορετική μορφή ή εμπλουτισμένες με διεπαφή:

- στον κατάλογο *extent96*[31] του *Project Gutenberg*[32]
- η ομάδα ανάπτυξης *DICT*[33]
- το *GCIDE*[34] του έργου *GNU*
- το *ARTFL*[35] έργο του *Πανεπιστημίου του Σικάγο*

2.3.5 SAIKAM Online Dictionary

Το SAIKAM [1] είναι ένα ιαπωνικό - ταϊλανδικό έργο ανάπτυξης λεξικού που αναπτύσσεται μέσω Διαδικτύου με τις συνεργατικές προσπάθειες χρηστών. Η αρχική βάση του λεξικού δημιουργήθηκε από δύο λεξικά μετάφρασης, ιαπωνικών σε αγγλικά και ταϊλανδικών σε αγγλικά με την αυτόματη αναγνώριση όμοιων αγγλικών όρων και το ταίριασμα των αντίστοιχων όρων στα ιαπωνικά και στα ταϊλανδικά. Συνεπώς, η αρχική βάση δεν μπορεί να θεωρηθεί ούτε πλήρης ούτε ακριβής.

Το έργο έχει αναπτυχθεί με τη μορφή ιστοσελίδας που παρέχει μετάφραση όρων από τα ιαπωνικά στα ταϊλανδικά και αντίστροφα ενώ ταυτόχρονα δίνει τη δυνατότητα σε μια αφοσιωμένη, εγγεγραμμένη ομάδα χρηστών, να ενημερώνουν και να βελτιώνουν το περιεχόμενο του λεξικού. Επίσης επιτρέπει στους χρήστες να αναζητήσουν ιαπωνικές προτάσεις από Ιαπωνικά σώματα κειμένων που έχουν συλλεχθεί από το Διαδίκτυο και να συλλέξουν χρήσιμα στατιστικά στοιχεία χρήσης του λεξικού.

Αυτές οι υπηρεσίες είναι προσβάσιμες μέσω προτυποποιημένου λογισμικού περιήγησης Διαδικτύου. Το περιβάλλον του Διαδικτύου φιλοξενεί μια ευρεία γκάμα τεχνολογιών, συνεπώς μια μεγάλη πρόκληση για τις εφαρμογές που δραστηριοποιούνται είναι η συνεργασία. Πρώτης τάξης παράδειγμα αποτελεί το τοπίο των προγραμμάτων περιήγησης διαδικτύου όπου προβλήματα επικοινωνίας του λεξικού με περιηγητές μπορεί να στοιχίσει ακριβά.

Σύμφωνα με προτάσεις που παρείχαν οι χρήστες του SAIKAM δύο νέες υπηρεσίες ενσωματώθηκαν σε αυτό, ένα εργαλείο πελάτη και αυτόματη αναζήτηση στο λεξικό για ιστοσελίδες.

- Εργαλείο πελάτη

  Εκτός από τη διαδικτυακή διεπαφή, το SAIKAM προσφέρει πλέον και ένα εργαλείο πελάτη ώστε να μπορούν οι χρήστες να το εκτελούν τοπικά στον υπολογιστή τους. Συνεπώς, οι χρήστες με εγκατεστημένο λειτουργικό σύστημα *Microsoft Windows* μπορούν να λάβουν το εργαλείο πελάτη από το Διαδίκτυο και να το χρησιμοποιήσουν σαν ηλεκτρονικό λεξικό στον υπολογιστή τους.

  Αφορμή γι' αυτή την υπηρεσία αποτέλεσε η απαίτηση διαθεσιμότητας του λεξικού εκτός διαδικτυακής συνδεσιμότητας. Επίσης το κόστος σύνδεσης στο Διαδίκτυο αλλά και η ταχύτητα μεταφοράς δεδομένων από τον εξυπηρετητή

[31] http://www.gutenberg.org/ebooks/673
[32] http://en.wikipedia.org/wiki/Project_Gutenberg
[33] http://en.wikipedia.org/wiki/DICT
[34] http://en.wikipedia.org/wiki/GCIDE
[35] http://artfl-project.uchicago.edu/

στον υπολογιστή του χρήστη είναι απαγορευτικά[36]. Πολλοί χρήστες σταμάτησαν να ενδιαφέρονται για την επέκταση του λεξικού λόγω του μεγάλου χρόνου αναμονής.

Το περιεχόμενο του λεξικού στην πλευρά του χρήστη δεν είναι εντελώς στατικό. Σε αραιά χρονικά διαστήματα, ο πελάτης συνδέεται στον εξυπηρετητή και:

- λαμβάνει τελευταίες ενημερώσεις από τη βάση δεδομένων
- δεσμεύει λέξεις για να επεξεργαστεί τοπικά
- αφού τελειώσει την επεξεργασία ορισμών όρων καταθέτει τις αλλαγές στη βάση δεδομένων

- Αυτόματη αναζήτηση στο λεξικό για ιστοσελίδες

  Αυτή η υπηρεσία λειτουργεί σαν διαφανής ενδιάμεσος εξυπηρετητής για την περιήγηση σε ιαπωνικές ιστοσελίδες. Ο χρήστης ανοίγει την ιστοσελίδα της αυτόματης αναζήτησης SAIKAM και εισάγει την ιστοσελίδα που επιθυμεί να περιηγηθεί. Το SAIKAM ανακατευθύνει το πρόγραμμα περιήγησης στον ενδιάμεσο εξυπηρετητή περνώντας σαν παράμετρο τη ζητούμενη ιστοσελίδα.

  Ο εξυπηρετητής διαθέτει ένα εργαλείο ανάλυσης κειμένου το οποίο αναγνωρίζει μια λίστα ιαπωνικών λέξεων που περιέχονται στο κείμενο. Αφού λάβει τη σελίδα, αναλύει το κείμενο για γνωστές ιαπωνικές λέξεις και επιστρέφει την αρχική σελίδα με τους ορισμούς των λέξεων συνημμένες στον περιηγητή του χρήστη.

  Η ανάλυση του κειμένου πραγματοποιείται αφού έχουν απαλειφθεί όλοι οι αγγλικοί χαρακτήρες και σύμβολα. Στη συνέχεια το κείμενο τεμαχίζεται σε λίστες λέξεων και το σύστημα σχηματίζει εκφράσεις αναζητώντας μετάφραση για τη μακρύτερη το δυνατόν έκφραση στη βάση δεδομένων.

  Με αυτό τον τρόπο ο χρήστης μπορεί να διαβάσει ιαπωνικές ιστοσελίδες πιο άνετα, απαλλαγμένος από το κόστος να συμβουλεύεται λεξικό για κάθε λέξη που δε γνωρίζει.

Το SAIKAM ως σύστημα λογισμικού είναι δομημένο σύμφωνα με το μοντέλο πελάτη εξυπηρετητή. Η πλευρά του εξυπηρετητή (σύμφωνα και με την εικόνα 2.8 [1]) αποτελείται από:

- *Βάση δεδομένων*

  Μια *SQL* βάση δεδομένων που συντηρεί τους πίνακες του λεξικού, όπως το σύνολο των ιαπωνικών λέξεων του λεξικού, το σύνολο των ταϊλανδικών λέξεων του λεξικού, τους ορισμούς λέξεων και στατιστικά δεδομένα. Επίσης τα προφίλ των χρηστών και τα ημερολόγια πρόσβασης κρατώνται εκεί.

  Όλες οι αλλαγές που πραγματοποιούνται στη βάση δεδομένων είναι χρονολογημένες. Το ίδιο ισχύει και για τις εκδόσεις των λεξικών που κατέχουν οι χρήστες τοπικά στον υπολογιστή τους. Με αυτό τον τρόπο οι ενημερώσεις που λαμβάνουν οι χρήστες ανταποκρίνονται μόνο σε αυτές που έπονται της έκδοσης του λεξικού τους.

[36] εν έτει 2000 στην Ταϊλάνδη

- Διαδικτυακή διεπαφή

  Ένα σύνολο από έγγραφα *HTML, CGI* εφαρμογές σε γλώσσα *Perl* και *Java applets* που παρέχουν υποστήριξη τριών γλωσσών για την περιήγηση στο λεξικό και την ενημέρωσή του.

- Υπηρεσίες προστιθέμενης αξίας

  ’Ενα σύνολο υποσυστημάτων που παρέχουν αυτόματη αναζήτηση λέξεων και υπηρεσίες αναζήτησης.

- Γέφυρα πελάτη

  Μια διαδικασία που περιμένει σύνδεση από εργαλείο πελάτη και εκτελεί συγχρονισμό δεδομένων μεταξύ πελάτη SAIKAM και του εξυπηρετητή.

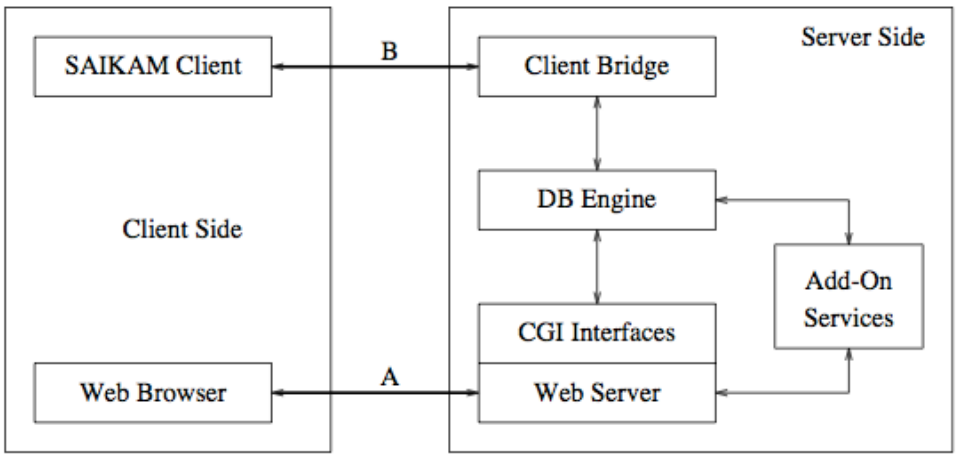

Figure 2.8: Αρχιτεκτονική Συστήματος SAIKAM

Τέλος, η συνεργατική συντήρηση του λεξικού επιτυγχάνεται μέσα από μια διαδικασία που διασφαλίζει την ποιότητα των εργασιών:

- Ο χρήστης εγγράφεται στην ομάδα ανάπτυξης του SAIKAM και το σύστημα του αναθέτει όνομα χρήστη και κωδικό.

- Κατά την αναζήτηση ιαπωνικών λέξεων στον πελάτη SAIKAM, ο χρήστης μπορεί να αποφασίσει να προσθέσει λέξεις στο καλάθι του προς επεξεργασία

- Στη συνέχεια συνδέεται στον εξυπηρετητή χρησιμοποιώντας τα στοιχεία του και ζητά αποκλειστική πρόσβαση εγγραφής για τις λέξεις μέσα στο καλάθι του. Ο εξυπηρετητής παραχωρεί πρόσβαση για όσες λέξεις δεν έχουν δεσμευτεί

από άλλα καλάθια. Αυτές οι λέξεις χαρακτηρίζονται ως επεξεργάσιμες στον πελάτη και δεν επιτρέπονται αλλαγές[37] σε αυτές από τη διαδικτυακή διεπαφή.

- Οι λέξεις που έγιναν αποδεκτές προς επεξεργασία επιστρέφονται στον πελάτη και η σύνδεση με τον εξυπηρετητή μπορεί να τερματιστεί. Ο χρήστης μπορεί να επεξεργαστεί τις λέξεις τοπικά στον υπολογιστή του.

- Ο χρήστης συνδέεται ξανά στον εξυπηρετητή για να καταχωρήσει τις αλλαγές του. Οι αλλαγές αποθηκεύονται και χρονολογούνται. Οι επεξεργασμένες λέξεις αφαιρούνται από το καλάθι του πελάτη.

2.3.6 Lingwo - Collaborative Dictionary

Το Lingwo [28] αποτελείται από μια μικρή οικογένεια μονάδων λογισμικού με στόχο τη λειτουργία ενός διαδικτυακού, συνεργατικού λεξικού. Το λεξικό προσπαθεί να υποστηρίξει σύγχρονες, επαγγελματικές λεξικογραφικές πρακτικές. Εστιάζει στη μετάφραση λεξικών, τα οποία συντηρούνται συνεργατικά από μια ομάδα, όπως ένα *wiki*.

Το Lingwo αναπτύσσεται στην πλατφόρμα διαχείρισης περιεχομένου ανοιχτού λογισμικού *Drupal*.

2.3.7 Wictionary - The Free Dictionary

Το Wictionary [33] είναι ένα συνεργατικό έργο με στόχο την παραγωγή ενός πολύγλωσσου, ελεύθερα διαθέσιμου λεξικού ικανό να περιγράψει όλες τις λέξεις από όλες τις γλώσσες χρησιμοποιώντας ορισμούς και περιγραφές στις αντίστοιχες γλώσσες.

Σχεδιασμένο όπως η *Wikipedia* το Wictionary έχει επεκταθεί και περιλαμβάνει πλέον θησαυρό συνωνύμων, βίβλο εκφράσεων, στατιστικά γλωσσών και μακροσκελή παραθέματα. Το όραμα είναι να δίδεται ουσιαστική πληροφορία για τη σημασία μιας λέξης. Για κάθε λέξη λοιπόν παρέχεται ετυμολογία, προφορά, παραδείγματα, συνώνυμα, αντώνυμα και μεταφράσεις.

Το Wictionary είναι ένα *wiki*, το οποίο σημαίνει ότι υπόκειται σε επεξεργασία και το σύνολο του περιεχομένου του καλύπτεται από διπλή άδεια, *Creative Commons Attribution-ShareAlike 3.0 Unported License* και *GNU Free Documentation License*.

Το Wictionary επιβάλλει αυστηρές συμβάσεις μορφοποίησης και κριτήρια αποδοχής υλικού. Γι' αυτό το λόγο υπάρχουν σαφείς οδηγίες, ενότητα βοήθειας, απομονωμένο περιβάλλον πειραματισμού και ενημερωτική ιστοσελίδα αποκλειστικά για τους χρήστες που συνεισφέρουν στην ανάπτυξη του Wictionary.

Η ενεργής συμμετοχή στο έργο ανάπτυξης του Wictionary προϋποθέτει εγγραφή έτσι ώστε η συνεισφορά να είναι ιχνηλατίσιμη.

2.3.8 Irishionary

Το Irishionary [7] είναι ένα συνεργατικό λεξικό που αναπτύσσεται μέσω διαδικτύου. Ξεκίνησε από μηδενική βάση το 2008 και σήμερα απαριθμεί πάνω από 1.000 μέλη, 5.000 λέξεις και 7.500 μεταφράσεις.

Όταν μια λέξη απουσιάζει από το λεξικό ένας εγγεγραμμένος χρήστης μπορεί να την εισάγει.

[37] Αλλαγές στις λέξεις μπορούν να πραγματοποιηθούν είτε από τη διαδικτυακή διεπαφή είτε από το πρόγραμμα πελάτη. Για να αποφευχθούν συγκρουόμενες αλλαγές ένα επιπλέον χαρακτηριστικό εισήχθη, το οποίο σημειώνει την κατάσταση επεξεργασίας και που αυτή συμβαίνει.

2.3.9 Distributed Medical Dictionary

Το *πανεπιστήμιο του Baylor* αναπτύσσει ένα ιατρικό λεξικό [13]. Εκμεταλλευόμενοι τα δίκτυα υπολογιστών υψηλής ταχύτητας, παρέχουν μια εθνική ιατρική υποδομή πληροφοριών εξαιρετικά σημαντική για το επιστημονικό πεδίο.

Το μοντέλο της αρχιτεκτονικής υποστηρίζει συνεργατική ανάπτυξη μιας αποκεντρωμένης, διαδικτυακής βάσης δεδομένων που φιλοξενεί ιατρική ορολογία. Το σύστημα δίνει έμφαση στην υψηλή διαθεσιμότητα, στο χρόνο απόκρισης, στην υποστήριξη τοπικών διαλέκτων και έλεγχο του λεξιλογίου.

2.3.10 Papillon Project

Ο Serasset [31] περιγράφει το έργο Papillon που έχει ως στόχο τη δημιουργία ενός πολύγλωσσου λεξικού γενικής χρήσης. Επηρεασμένο από τα έργα ανοιχτού λογισμικού επιτρέπει στους χρήστες να συνεργαστούν μέσω του Διαδικτύου.

Οι χρήστες έχουν στη διάθεσή τους αρχική βάση λεξικών ανεπτυγμένη από άλλους και εργαλεία για να επεκτείνουν ή να διορθώσουν το πολύγλωσσο λεξικό. Μια επέκταση του συστήματος αφορά τη δυνατότητα των χρηστών να ορίσουν τις προσωπικές τους όψεις στη βάση δεδομένων.

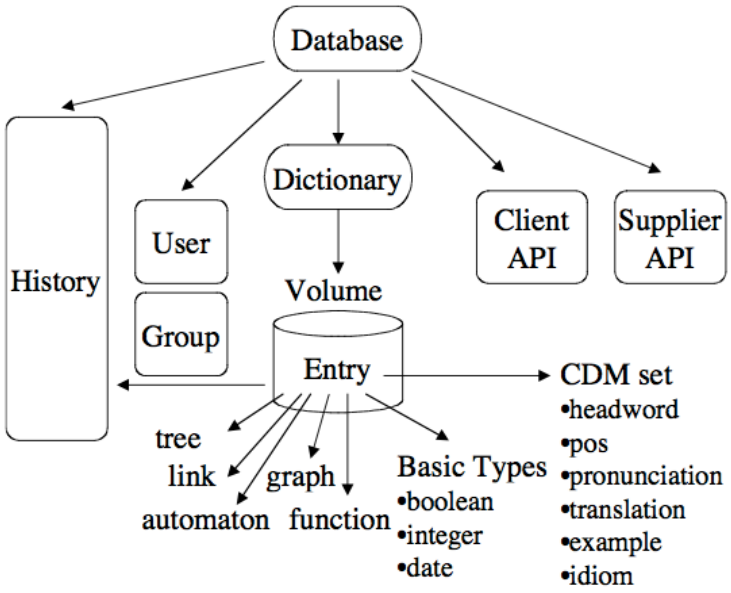

Figure 2.9: Αρχιτεκτονική Συστήματος Papillon

Ο συγγραφέας ορίζει ένα πλήρες πλαίσιο για τη δημιουργία λεξικών. Το πλαίσιο είναι γενικό έτσι ώστε να επιτρέπει προσαρμογή ετερογενών λεξικών. Το πλαίσιο καταλήγει στον ορισμό μιας γλώσσας τύπου *XML* της *Dictionary Markup Language (DML)*.

Το σύνολο των λεκτικών όρων στη βάση δεδομένων μπορεί να περιγραφεί με στοιχεία *DML*. Ολόκληρη η ιεραρχία των αρχείων, στοιχείων και ισιοτήτων *XML* περιγράφεται

με τη χρήση σχημάτων *XML*. Στην εικόνα 2.9 [31] περιγράφεται η οργάνωση των κύριων στοιχείων *DML*.

2.3.11 Longdo

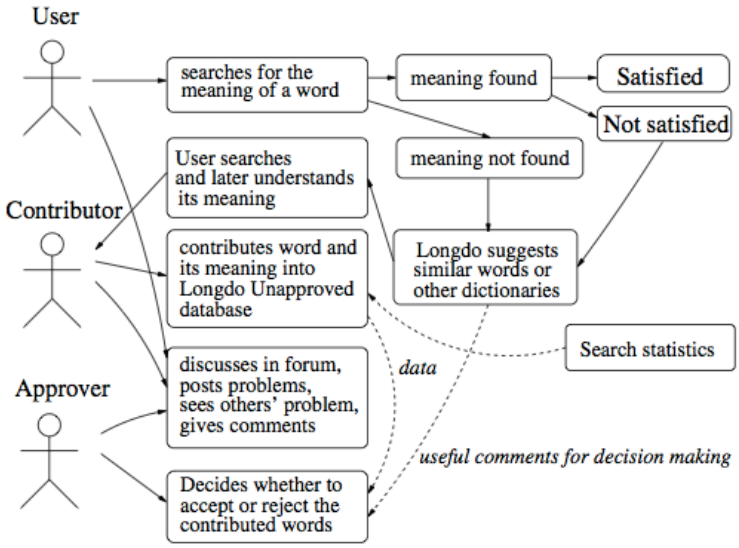

Figure 2.10: Συνεργατικό μοντέλο Longdo

Το συνεργατικό μοντέλο του Longdo [18] (βλ. εικόνα 2.10 [18]) περιλαμβάνει:

- τους *συντηρητές λεξικού*,

  Δημιούργησαν το σύστημα μεταγλώττισης του λεξικού. Η συντήρηση και η επέκταση του λεξικού σε συνεργασία με τους εθελοντές είναι στις αρμοδιότητες τους.

- τους *χρήστες*

  Οι χρηστές αναζητούν λέξεις στο λεξικό μέσω της διαδικτυακής διεπαφής ή λογισμικού πελάτη. Αν μια λέξη δε βρεθεί, τότε το σύστημα εμφανίζει παρεμφερείς εγγραφές. Επίσης εμφανίζει άλλα λεξικά σε περίπτωση που η λέξη είναι έγκυρη αλλά δεν περιλαμβάνεται στο λεξικό. Μετά την κατανόηση της άγνωστης λέξης ένας χρήστης μπορεί να αναλάβει το ρόλο του συνεισφέροντα.

- τους *συνεισφέροντες*

  Οι συνεισφέροντες μπορούν να προτείνουν άγνωστες λέξεις προς εισαγωγή στο λεξικό μέσω μιας φόρμας.

- τους *ελεγκτές*

  Το Longdo περιλαμβάνει δύο βάσεις δεδομένων, την *εγκεκριμένη* και την *ανέλεγκτη*. Οι ελεγκτές περιηγούνται στη βάση δεδομένων με τις ανέλεγκτες εγγραφές και

τις επεξεργάζονται. Οι δεκτές εγγραφές μεταφέρονται στη βάση δεδομένων με τις εγγεκριμένες εγγραφές. Οι συζητήσεις στο φόρουμ του λεξικού λαμβάνονται υπόψη κάποιες φορές. Οι ελεγκτές λαμβάνουν πρόσκληση για να αναλάβουν αυτό το ρόλο δεδομένου της καλής γνώσης της γλώσσας και της διαθεσιμότητας χρόνου τους.

2.3.12 EDICT

Το EDICT [6] ξεκίνησε το 1991 με στόχο να παρέχει ένα δωρεάν διαθέσιμο λεξικό ιαπωνικών σε αγγλικά επεξεργάσιμο από υπολογιστή. Το συνεργατικό μοντέλο στηρίχθηκε στις ανταλλαγές ηλεκτρονικού ταχυδρομείου και αρχείων μέσω Διαδικτύου.

Η αρχική δομή των εγγραφών στο αρχείο EDICT ήταν αρκετά απλή και σύντομα έγινε φανερό ότι χρειαζόταν μια πιο πλούσια δομή για να αναπαραστήσει τις πολυπλοκότητες του ιαπωνικού λεξικού.

Το 1999 μια *XML* έκδοση (*JMdict*)[38] εισήχθη η οποία επέτρεπε ετικέτες και ετεροαναφορές. Επίσης διευκόλυνε τη μετάφραση σε άλλες γλώσσες όπως γαλλικά, γερμανικά, ρώσσικα. Το αρχείο JMdict, το οποίο είναι σε κωδικοποίηση *UTF-8* είναι το κύριο αποτέλεσμα του έργου, ενώ η αρχική μορφή παράγεται ακόμη για συστήματα που στηρίζονται σε αυτή.

Η μορφή EDICT επεκτάθηκε σε EDICT2 και αντανακλά τη δομή των εγγραφών *XML*. Χρησιμοποιείται από διάφορα συστήματα συμπεριλαμβανομένου του *WWWJDIC*[39]. Εκδόσεις επίσης παράγονται στην *XML* μορφή που χρησιμοποιεί η εφαρμογή *dict* της *Apple* και στην *EPWINGJIS X 4081* που χρησιμοποιείται από πολλά συστήματα ιαπωνικών ηλεκτρονικών λεξικών.

Αυτό το έργο θεωρείται πρότυπη αναφορά ιαπωνικού - αγγλικού λεξικού στο Διαδίκτυο και χρησιμοποιείται από τη βάση δεδομένων *Unihan*[40] και πολλά άλλα ιαπωνικά - αγγλικά έργα. Από το 2000 το έργο διαχειρίζεται από το *Electronic Dictionary Research and Development Group(EDRDG)*. Το 2010 η συντήρηση του λεξικού μεταφέρθηκε σε διαδικυακή βάση δεδομένων. Τον Ιούνιο του 2011 το αρχείο του EDICT περιείχε περίπου 155.000 εγγραφές.

[38] http://en.wikipedia.org/wiki/JMdict
[39] http://en.wikipedia.org/wiki/WWWJDIC
[40] http://en.wikipedia.org/wiki/Unihan_Database

3 Ορθογράφοι: Χρήση και Συνεργατική Ανάπτυξη

Η ανάπτυξη εύχρηστων και αποτελεσματικών ορθογράφων οφελεί σημαντικά τους χρήστες, οι οποίοι χρησιμοποιούν τον ορθογράφο είτε σαν εργαλείο τοπικά στον υπολογιστή τους είτε σαν υπηρεσία μέσω του Διαδικτύου. Οι ορθογράφοι αυξάνουν την παραγωγικότητα και την αποδοτικότητα στην προετοιμασία εγγράφων, διευκολύνουν και επιταχύνουν τη συνεργασία και προωθούν την ορθή χρήση της γλώσσας.

Οι ορθογράφοι μπορεί να επωφεληθούν από το συνεργατικό μοντέλο ανάπτυξης και να αποκτήσουν έναν πηρύνα οπαδών, οι οποίοι κινούμενοι από ίδιο συμφέρον φροντίζουν για τη συντήρηση και επέκταση του ορθογράφου σύμφωνα με τις αρχές του ανοιχτού/ελεύθερου λογισμικού. Όσο πληθαίνει η κοινότητα χρηστών, θα προοδεύει ταχύτατα και ο ορθογράφος. Μάλιστα, οι όποιες προσθήκες στον ορθογράφο συνήθως συνοδεύονται από επισήμανση της συνεισφοράς. Η συμμετοχή επιβραβεύεται κατ' αυτό τον τρόπο και κάθε εργασία είναι καταλογίσιμη.

Επιπλέον, το Διαδίκτυο δίνει τη δυνατότητα σε χρήστες από όλο τον κόσμο να συνεισφέρουν στην επέκταση του ορθογράφου, όπου κάθε συνεισφορά αποθηκεύεται σε μια κεντρική βάση δεδομένων και ενσωματώνεται μετά από έλεγχο. Ανά τακτά χρονικά διαστήματα, ο ορθογράφος ανανεώνεται με τις τελευταίες αλλαγές. Είτε προσφέρεται σαν υπηρεσία που εκτελείται σε κάποιο εξυπηρετητή είτε σαν εργαλείο που εκτελείται τοπικά στον υπολογιστή του χρήστη, οι αλλαγές φτάνουν σε όλους τους χρήστες και όλοι είναι κερδισμένοι με αυτό τον τρόπο.

Εν τέλει, το συνεργατικό μοντέλο είναι ιδανικό για την ανάπτυξη ορθογράφου αφού η ανάγκη, η ευκαιρία και η δυνατότητα συμμετοχής διαδίδονται σε όλο τον κόσμο μέσω της διαδικτυακής υποδομής. Το ελεύθερο/ανοιχτό λογισμικό κινητοποιεί τους χρήστες να συνεισφέρουν με σημαία την αναγνωρισιμότητα της προσφοράς τους και το αίσθημα της ιδιοκτησίας ενώ όλη η κοινότητα χρηστών αποταμιεύει τις προσθήκες. Γίνεται φανερό ότι το παρόν μοντέλο ανάπτυξης και επέκτασης του λογισμικού οδηγεί σε ταχύτερη πρόοδο, προϊόν εστιασμένο στις ανάγκες των χρηστών και αυξημένη ικανοποίηση από τη χρήση του.

4 Προδιαγραφές

Με βάση τα πορίσματα της επισκόπησης των ελληνικών ορθογράφων και της συνεργατικής ανάπτυξης σκιαγραφούνται τα χαρακτηριστικά του ΠΣ προς ανάπτυξη, πρώτα οι τεχνικές προδιαγραφές και στη συνέχεια οι λειτουργικές. Τέλος, περιγράφονται οι προδιαγραφές των κανόνων που χρησιμοποιεί ο ορθογράφος για τη διόρθωση κειμένων.

4.1 Τεχνικές προδιαγραφές

Για την αύξηση τής απόδοσης τού έργου στοχεύουμε στη μεγαλύτερη δυνατή εμπλοκή τού κοινού, των ανθρώπων που χρησιμοποιούν κάποιο λογισμικό, γράφουν κείμενα και αντιμετωπίζουν ελλείψεις και λάθη στην ορθογραφική διόρθωση. Για αυτό και ως πλατφόρμα λειτουργίας τού ΠΣ επιλέγεται το Διαδίκτυο. Χαρακτηριστικά τής υποδομής λειτουργίας τού ΠΣ:

- Ανάπτυξη με εργαλεία ανοιχτού λογισμικού τα οποία είναι ευρέως διαδεδομένα και δεν έχουν περιοριστικούς όρους χρήσης και διάθεσης.
- Λειτουργία σε διακομιστές ανοιχτού λογισμικού, βασισμένων σε Linux.
- Περιβάλλον εργασίας στο web (μέσω browser). Όλες οι οθόνες, χειριστών, στελεχών, διαχειριστή θα είναι σε web περιβάλλον.
- Υποστήριξη όλων των γνωστών φυλλομετρητών (όπως Internet Explorer, Firefox, Safari, Chrome κ.ο.κ.).
- Διασφάλιση της ακεραιότητας των δεδομένων και προστασία των ευαίσθητων προσωπικών δεδομένων.

4.2 Λειτουργικές προδιαγραφές

- *Διαχείριση χρηστών.*

  Το ΠΣ παρέχει τη δυνατότητα εγγραφής χρηστών. Ο χρήστης δημιουργεί ένα λογαριασμό, διαμορφώνει το προφίλ με βασικά στοιχεία επικοινωνίας και συνδέεται με όνομα και κωδικό τής επιλογής του.

- *Διαβάθμιση χρηστών.*

  Ένας χρήστης μπορεί να είναι απλός χρήστης ή διαχειριστής τού ΠΣ.

- *Καταγραφή συνεισφοράς.*

  Όταν ένας συνδεδεμένος χρήστης συνεισφέρει με κάποιον τρόπο στο έργο, η συνεισφορά του καταγράφεται στο ημερολόγιο τού ΠΣ. Η συνεισφορά μη συνδεδεμένων χρηστών καταγράφεται με τη διεύθυνση IP με την οποία συνδέθηκαν.

- *Ρύθμιση αποδοχής συνεισφορών.*

  Το ΠΣ δίνει τη δυνατότητα επιλογής από το Διαχειριστή να επιτρέπει ή να αποτρέπει τη συνεισφορά από μη εγγεγραμμένους χρήστες.

- *Αναζήτηση λέξης στο ορθογραφικό λεξικό ή στο Θησαυρό.*

  Ο χρήστης πληκτρολογεί μία λέξη ή ρίζα λέξης και το ΠΣ επιστρέφει τη λίστα των καταχωρημένων λέξεων στο λεξικό ορθογραφίας και στο θησαυρό, που ταιριάζουν με τη ζητούμενη. Έτσι, οι χρήστες μπορούν να βλέπουν εάν μία λέξη ή τύπος της είναι ήδη καταχωρημένη.

- *Προτάσεις καταχώρησης νέων λέξεων στο ορθογραφικό λεξικό.*

  Ο χρήστης πληκτρολογεί μία λέξη και την υποβάλλει για έλεγχο και ενσωμάτωση στη βάση δεδομένων.

  Στην περίπτωση που ο χρήστης υποβάλλει μία λέξη, αυτή καταχωρείται στη βάση δεδομένων σε κατάσταση “προς έγκριση”. Ένας διαχειριστής τού ΠΣ θα μπορέσει στη συνέχεια να την ελέγξει και να την εγκρίνει ή να την απορρίψει ως λανθασμένη επιλογή.

  Ο χρήστης μπορεί να υποβάλλει και δέσμη λέξεων, όχι μόνον μία προς μία. Αυτό γίνεται είτε με γραφή τους τη μία κάτω ή δίπλα από την άλλη, είτε επικολλώντας λέξεις από το Πρόχειρο είτε με την εισαγωγή τού περιεχομένου ενός αρχείου κειμένου.

- *Ανάπτυξη κλιτικού συστήματος λέξεων.*

  Ο χρήστης υποβάλλει ένα θέμα λέξης, επιλέγει το μέρος τού λόγου (ουσιαστικό, επίθετο, ρήμα κ.λπ.) και το ανάλογο κλιτικό σύστημα (κλίνεται όπως...). Το ΠΣ αυτόματα αναπαραγάγει το κλιτικό σύστημα τής λέξης. Τα κλιτικά συστήματα που θα ακολουθηθούν είναι αυτά που καταγράφηκαν από το Ίδρυμα Τριανταφυλλίδη στην έκδοση τού ομώνυμου λεξικού τής νέας ελληνικής.

  Ο χρήστης μπορεί να υποβάλλει δέσμη λέξεων και όχι μόνον μία, όπως παραπάνω. Βέβαια, κάθε λέξη σε μία δέσμη πρέπει να είναι τού ιδίου μέρους τού λόγου και να κλίνεται με τον ίδιο τρόπο.

  Ο χρήστης βλέπει τις λέξεις και το κλιτικό σύστημα τής κάθε μίας για τυχόν λάθος επιλογές και αποφασίζει για καταχώρηση ή για αναίρεση και διορθώσεις. Καταχωρούμενες οι λέξεις που παράγονται αυτόματα από το ΠΣ με την ανάπτυξη τού κλιτικού τους συστήματος, συνδέονται μεταξύ τους ώστε να γνωρίζουμε ποιές είναι τύποι τού ίδιου θέματος λέξης.

  Ένας διαχειριστής τού ΠΣ θα μπορέσει στη συνέχεια να ελέγξει και να εγκρίνει ή να απορρίψει ως λανθασμένη την πρόταση.

- *Προτάσεις καταχώρησης νέων θεμάτων ή συνωνύμων στο Θησαυρό.*

  Ο χρήστης πληκτρολογεί μία λέξη και συνώνυμα για αυτήν. Το ΠΣ ελέγχει εάν υπάρχει η λέξη καταχωρημένη στο Θησαυρό και εάν δεν υπάρχει την προσθέτει. Στη συνέχεια ελέγχει ένα ένα τα προτεινόμενα συνώνυμα και προσθέτει όσα δεν υπάρχουν και τα συνδέει με την προτεινόμενη λέξη.

  Ένας διαχειριστής τού ΠΣ θα μπορέσει στη συνέχεια να την ελέγξει και να την εγκρίνει ή να την απορρίψει ως λανθασμένη επιλογή.

- *Διόρθωση καταχωρημένων λέξεων.*

  Όπως με την προσθήκη νέων λέξεων, οι χρήστες μπορούν να υποβάλλουν προτάσεις για επανέλεγχο ήδη καταχωρημένων λέξεων. Ο χρήστης γράφει μία λέξη και την αναζητά. Το σύστημα την εμφανίζει και ο χρήστης τη χαρακτηρίζει ως αμφισβητούμενη προσθέτοντας και τα σχόλιά του, που πιστεύει ότι είναι το λάθος, ώστε να ελεγχθεί από τους διαχειριστές.

- *Δημοσίευση σε κοινωνικά δίκτυα.*

  Στο προφίλ τού χρήστη προβλέπεται σημαία που δείχνει εάν ο χρήστης επιθυμεί να γίνονται δημοσιεύσεις στη σελίδα του στο Facebook για τη συνεισφορά του. Η αρχική θέση αυτής τής σημαίας είναι ON.

  Όταν μία συνεισφορά γίνεται από εγγεγραμμένο χρήση, και η σχετική σημαία στο προφίλ του είναι ON, το ΠΣ αυτόματα δημιουργεί μία δημοσίευση στη σελίδα τού χρήστη στο Facebook για αυτή τη συνεισφορά. Η μορφή και το μήνυμα τής δημοσίευσης θα συζητηθεί κατά την υλοποίηση.

4.3 Προδιαγραφές κανόνων Hunspell

Οι κανόνες που θα διαμορφωθούν πρέπει να αντιμετωπίζουν τουλάχιστον τα παρακάτω ζητήματα. Θα εκτιμηθεί θετικά η πρόταση επιπλέον περιπτώσεων που πρέπει να αντιμετωπιστούν.

- Δήλωση ελληνικού αλφάβητου.
- Εξίσωση κεφαλαίων με πεζά.
- Εξίσωση δίφθογγων με φωνήεντα ή σύμφωνα.
- Ρύθμιση αναγνώρισης λέξεων με κεφαλαία με και χωρίς τόνο.
- Ρύθμιση αντιμετώπισης εγκλίσεων.
- Ρύθμιση αντιμετώπισης συνθετικών (ανά, κατά, διά κ.λπ.).

5 Εξαγωγή για χρήση

Με βάση τις καταχωρημένες ενεργές λέξεις τού ορθογραφικού λεξικού και του Θησαυρού συνωνύμων, το ΠΣ δημιουργεί αυτόματα το αρχείο γλωσσικής υποστήριξης τής ελληνικής για το Hunspell και την προτείνει στο χρήστη για online λήψη. Ο χρήστης λαμβάνει το αρχείο και το εγκαθιστά στο λογισμικό που χρησιμοποιεί, π.χ. στο OpenOffice.org. Έτσι, ανά πάσα στιγμή μπορεί οποιοσδήποτε να λάβει και να εγκαταστήσει την τελευταία μορφή τής γλωσσικής υποστήριξης για τα ελληνικά.

6 Βοηθητικές επεκτάσεις

Απαιτείται η δημιουργία των απαραίτητων επεκτάσεων για τους φυλλομετρητές και άλλα λογισμικά ώστε οι χρήστες να μπορούν να συνεισφέρουν τη στιγμή κατά την οποία εργάζονται, π.χ. γράφουν κείμενα και βρίσκουν ελλείψεις και λάθη στη βάση δεδομένων τού Hunspell για τα ελληνικά. Οι βοηθητικές επεκτάσεις πρέπει να υποστηρίζουν το σενάριο:

- Ο χρήστης γράφει ένα κείμενο στο OpenOffice.org, σε μία φόρμα στο Firefox ή σε άλλο φυλλομετρητή ή λογισμικό που χρησιμοποιεί το Hunspell.
- Το Hunspell χαρακτηρίζει τη λέξη ως ανορθόγραφη και προστίθεται η χαρακτηριστική κόκκινη υπογράμμιση.
- Ο χρήστης κάνει δεξί κλικ πάνω στη λέξη.
- Εάν η λέξη υπάρχει ήδη στο ορθογραφικό λεξικό ο χρήστης επιλέγει την προτεινόμενη ορθογραφικώς ορθή μορφή της.
- Εάν από τις προτάσεις φανεί ότι η λέξη δεν υπάρχει στο ορθογραφικό λεξικό, ο χρήστης μπορεί να επιλέξει την αποστολή της στο ΠΣ ώστε να ελεγχθεί και να συμπεριληφθεί με την προβλεπόμενη διαδικασία.

Οι βοηθητικές επεκτάσεις πρέπει να λειτουργούν κατ' ελάχιστον στα λογισμικά:

- FireFox 3.x Windows / Linux / MacOS X.
- Chrome Windows / Linux / MacOS X.
- OpenOffice.org 3.x Windows / Linux / MacOS X.

Θα εκτιμηθεί θετικά η συμβατότητα με περισσότερα λογισμικά.

7 Παραδοτέα

Ως παραδοτέα νοούνται τα παρακάτω:

- Το λογισμικό τού ΠΣ το οποίο λειτουργεί βάσει των τεχνικών προδιαγραφών και υλοποιεί τις λειτουργικές προδιαγραφές, όπως αμφότερες περιγράφονται παραπάνω.
- SQL Queries (συμβατές και με MySQL) δημιουργίας των απαραίτητων βάσεων δεδομένων και των πινάκων τους, τους οποίους χρησιμοποιεί το ΠΣ.
- SQL Queries (συμβατές και με MySQL) προσθήκης στις Βάσεις Δεδομένων των λέξεων τού ορθογραφικού λεξικού και του Θησαυρού συνωνύμων, που είναι διαθέσιμες σήμερα για τους χρήστες λογισμικών που χρησιμοποιούν το Hunspell.
- Τις βοηθητικές επεκτάσεις σε εκτελέσιμη και πηγαία μορφή για κάθε λογισμικό και πλατφόρμα που περιγράφεται στο ανάλογο τμήμα.

8 Άδεια χρήσης

Το λογισμικό θα διατίθεται με άδεια EUPL[1].

[1] http://www.osor.eu/eupl

Βιβλιογραφία

[1] V Ampornaramveth. Saikam: An online dictionary development project. In *Proc. of the 4th Intl. Workshop on Academic Information Networks and Systems, February 1998, NACSIS Seminar House, Karuizawa, Japan*, February 1998.

[2] Anna Anastassiadis-Symeonidis, Tita Kyriacopoulou, Georgia Nikolaou, Anna Panayotou-Triantaphyllopoulou, and Vasiliki Foufi. Terminology and automatic spelling correction. In *5th Conference of Greek Language and Terminology, Leukosia, Cyprus*, Nov 2005.

[3] Kevin Atkinson. Aspell. `http://en.wikipedia.org/wiki/GNU_Aspell`.

[4] Kevin Atkinson. Pspell. `http://en.wikipedia.org/wiki/Pspell`.

[5] Kevin Atkinson. Aspell. `http://aspell.net/`, 2004.

[6] Jim Breen. Electronic dictionary. `http://www.csse.monash.edu.au/~jwb/edict_doc.html`, 1991.

[7] Community. Irishionary. `http://www.irishionary.com/`.

[8] Enchant developers. Enchant. `http://en.wikipedia.org/wiki/Enchant_(software)`.

[9] Gaspell developers. Gaspell. `http://freshmeat.net/projects/Gaspell/`.

[10] GtkSpell developers. Gtkspell. `http://gtkspell.sourceforge.net/`.

[11] Kspell developers. Kspell. `http://api.kde.org/3.1-api/classref/kspell/index.html`.

[12] Radek Doulik. Gnome-spell. `http://linux.softpedia.com/get/Text-Editing-Processing/Others/Gnome-Spell-31362.shtml`.

[13] J Fowler, G Buffone, and D Moreau. The architecture of a distributed medical dictionary. *Medinfo MEDINFO*, 8 Pt 1:126–130, 1995.

[14] GNU. Ispell. `http://www.gnu.org/software/ispell/`, 2006.

[15] R. E. Gorin. Ispell. `http://en.wikipedia.org/wiki/Ispell`.

[16] The PHP Group. Pspell. `http://php.net/manual/en/book.pspell.php`, 2001.

[17] Kevin Hendricks. Myspell. `http://en.wikipedia.org/wiki/MySpell`.

[18] Pattara Kiatisevi, Vuthichai Ampornaramveth, Er I. Kovács, and Haruki Ueno. 2003) “searching in the longdo thai online dictionary service. In *Proceedings of the 4 th International Conference on Information Technologies InTech*, pages 17–19, 2003.

[19] Geoff Kuenning. Ispell. `http://lasr.cs.ucla.edu/geoff/ispell.html`, 1996.

[20] T Kyriacopoulou, S Mrabti, and A Yannacopoulou. Le dictionnaire électronique des noms composés en grec moderne. *Lingvisticæ Investigationes*, 25(1):7–28, 2002.

[21] Dom Lachowicz. Enchant. `http://www.abisource.com/projects/enchant/`, 1998.

[22] Anton Leuski. cocoaspell. `http://cocoaspell.leuski.net/`.

[23] mentors.debian.net. Myspell. `http://mentors.debian.net/package/myspell-el-gr`, 2008.

[24] Laszlo Nemeth. Hunspell. `http://hunspell.sourceforge.net/`.

[25] Neurolingo. Λεξισκόπιο. `http://www.neurolingo.gr/online_tools/lexiscope.htm`.

[26] Sam. Mysql spell checker. `http://www.phpclasses.org/package/2597-PHP-Spell-check-texts-with-a-MySQL-table-as-dictionary.html`.

[27] Manuel Serrano. Flyspell. `http://www-sop.inria.fr/members/Manuel.Serrano/flyspell/flyspell.html`.

[28] David Snopek. Lingwo - collaborative dictionary. `http://drupal.org/project/lingwo`.

[29] Virach Sornlertlamvanich. Building a dictionary from www. `203.185.96.228/virach/sites/default/files/paper/buildingWWWdic.pdf`.

[30] Oliver Streiter, Kevin P. Scannell, and Mathias Stuflesser. Implementing nlp projects for noncentral languages: instructions for funding bodies, strategies for developers. *Machine Translation*, 20:267–289, December 2006.

[31] Gilles Sérasset. Frameworks, implementation and open problems for the collaborative building of a multilingual lexical database, 2002.

[32] Volunteers. Collaborative international dictionary of the english language. `http://en.wikipedia.org/wiki/Collaborative_International_Dictionary_of_English`.

[33] Jimmy Wales. Wictionary - the free dictionary. `http://en.wiktionary.org/wiki/Wiktionary:Main_Page`.

[34] Α Ιορδανίδου, Μ Πανταζάρα, Ε Μάντζαρη, Γ Ορφανός, Α Βαγγελάτος, and Β Παπαπαναγιώτου. Ζητήματα αναγνώρισης των πολυλεκτικών σύμπλοκων όρων στον τομέα της βιοϊατρικής. In *6ο Συνέδριο «Ελληνική Γλώσσα και Ορολογία» (2007), Αθήνα, Ελλάδα*, Νοέμβριος 2007.

[35] Ινστιτούτο Επεξεργασίας Λόγου. Συμφωνία. http://www.ilsp.gr/el/services-products/products/item/2-langtech/14-simfonia.

[36] MATZENTA. Αυτόματος Πολυτονιστής. http://www.magenta.gr/index.php/Λογισμικό/Πολυτονιστές/Αυτόματος-Πολυτονιστής-Επαγγελματική-έκδοση.html.