ePublishing

Ανοικτή Εκπαίδευση: το περιοδικό για την Ανοικτή και εξ Αποστάσεως Εκπαίδευση και την Εκπαιδευτική Τεχνολογία

Vol 15, No 1 (2019)

Οι Λεξικές Βάσεις Δεδομένων ως Εκπαιδευτικό Εργαλείο Εκμάθησης Ορολογίας στην Υπολογιστική Γλωσσολογία

Δήμητρα Καρβουνιάρη, Ευάγγελος Χρήστος Παπακίτσος

doi: 10.12681/jode.18945

To cite this article:

https://epublishing.ekt.gr | **e-Publisher:** EKT | **Downloaded at:** 26/03/2026 15:02:12

Οι Λεξικές Βάσεις Δεδομένων ως Εκπαιδευτικό Εργαλείο Εκμάθησης Ορολογίας στην Υπολογιστική Γλωσσολογία

The Lexical Databases as an Educational Tool for Learning Terminology in Computational Linguistics

Δήμητρα Καρβουνιάρη
Εκπαιδευτικός ηλεκτρολογίας
Πάντειο Πανεπιστήμιο
Κοινωνικών και Πολιτικών Επιστημών
karvoun_dimi@yahoo.gr

Ευάγγελος Χρήστος Παπακίτσος
Επιστημονικός συνεργάτης
Πανεπιστήμιο Δυτικής Αττικής
papakitsev@sch.gr

Summary
Lexical databases are powerful and easily expandable learning tools in the field of terminology learning. Their usage can provide students with learning experiences in a friendly way, providing access to a variety of sources and information in authentic linguistic environments. This paper describes the design and implementation of a lexical database as a software application of the technological field of Computational Lexicography. The original purpose of creating such a tool was to facilitate the learning of technical terminology of the scientific field of Computational Linguistics, in which the phenomenon of polysemy often occurs, namely that the same term has at least two different meanings. This phenomenon is due both to the translation of English terms into the Greek language and to the combination of different disciplines (Informatics, Mathematics and Linguistics) in the same field of applications.

Keywords
lexicon, lexical database, Computational Lexicography, terminology learning

Περίληψη
Οι λεξικές βάσεις δεδομένων αποτελούν ένα ισχυρό και εύκολα επεκτάσιμο εκπαιδευτικό εργαλείο στον τομέα της εκμάθησης ορολογίας. Η χρήση τους μπορεί να προσφέρει στους σπουδαστές εμπειρίες εκμάθησης με φιλικό τρόπο, παρέχοντας πρόσβαση σε ποικιλία πηγών και πληροφοριών σε αυθεντικά γλωσσικά περιβάλλοντα. Στην παρούσα εργασία περιγράφεται η σχεδίαση και υλοποίηση μιας λεξικής βάσης δεδομένων, ως εφαρμογής λογισμικού του τεχνολογικού πεδίου της Ηλεκτρονικής Λεξικογραφίας. Ο αρχικός σκοπός της δημιουργίας ενός τέτοιου εργαλείου ήταν η διευκόλυνση της εκμάθησης της τεχνικής ορολογίας του επιστημονικού πεδίου της Υπολογιστικής Γλωσσολογίας, στην οποία εμφανίζεται συχνά το φαινόμενο της

πολυσημίας, δηλαδή ο ίδιος όρος να έχει δύο τουλάχιστον διαφορετικές σημασίες. Το φαινόμενο αυτό οφείλεται τόσο στη μετάφραση αγγλικών όρων στην ελληνική γλώσσα όσο και στο συνδυασμό διαφορετικών επιστημών (Πληροφορικής, Μαθηματικών και Γλωσσολογίας) στο ίδιο πεδίο εφαρμογών.

Λέξεις-κλειδιά
ηλεκτρονικό λεξικό, λεξική βάση δεδομένων, Ηλεκτρονική Λεξικογραφία, εκμάθηση ορολογίας

Εισαγωγή

Η χρήση της σύγχρονης τεχνολογίας στον τομέα της διδασκαλίας έχει αυξηθεί κατακόρυφα την τελευταία δεκαετία, σε ποικίλους γνωστικούς τομείς, ενώ η ολοκληρωμένη χρήση ηλεκτρονικών υπολογιστών αποτελεί ήδη συμπληρωματική ή και εναλλακτική λύση έναντι των διδακτικών εντύπων (Αλιβίζος, 2015. Lai, 2006). Σε αυτό το πλαίσιο, η χρήση πολυμέσων συμβάλλει στη βιωματική μάθηση, εμπλουτίζει τις γλωσσικές δεξιότητες και ενισχύει τις στρατηγικές αυτοδιδασκαλίας και την αυτοπεποίθηση των μαθητών (Taylor, 2006). Παράλληλα, επιτρέπει στους μαθητές να προχωρήσουν ενεργά με τον δικό τους ρυθμό, χωρίς αυτός να διαταράσσεται από χρονικούς περιορισμούς ή από την ανυπομονησία που είναι κοινή στα παραδοσιακά περιβάλλοντα μάθησης (Hansen-Smith, 2001). Για παράδειγμα, οι Chun & Plass (1996) διερεύνησαν τον τρόπο με τον οποίο η ικανότητα ανάγνωσης κειμένων μπορεί να ενισχυθεί με μια εφαρμογή πολυμέσων για την εκμάθηση δεύτερης/ξένης γλώσσας. Επιπλέον, διερεύνησαν τη σχέση μεταξύ της κατανόησης και εμπλουτισμού του λεξιλογίου με τη βελτίωση της ικανότητας ανάγνωσης. Τα αποτελέσματα της μελέτης τους έδειξαν ότι οι οπτικοακουστικές δυνατότητες των πολυμεσικών εφαρμογών βοηθούν περισσότερο στη συνολική κατανόηση του αντικειμένου.

Η ικανότητα ανάγνωσης και κατανόησης κειμένων προϋποθέτει από τους μαθητές να κατανοήσουν το λεξιλόγιο που χρησιμοποιείται. Η διαδικασία αυτή περιλαμβάνει μια ποικιλία ψηφιακών και μη μεθόδων, ανάμεσα στις οποίες είναι και η ανεύρεση λέξεων σε ένα λεξικό, για τις οποίες απαιτείται η διερεύνηση των επιπτώσεων τους (Nikolova, 2002). Έτσι, αρκετοί ερευνητές εξέτασαν την αποτελεσματικότητα των ψηφιακών πόρων ως εργαλείων βελτίωσης της ικανότητας ανάγνωσης, της κατανόησης κειμένου και της απόκτησης λεξιλογίου (Laufer & Shumueli, 1997. Watanabe, 1997. Hulstijn, 1992). Ως αποτέλεσμα των ερευνών αυτών, το ερώτημα μετατοπίστηκε από το εάν τα ψηφιακά μέσα μπορούν να διδάξουν λεξιλόγιο στο ποια είδη είναι αποτελεσματικότερα στη διαδικασία αυτή. Ακολούθως, αρκετές μελέτες έχουν ήδη αξιολογήσει τα γενικά οφέλη των σχετικών εφαρμογών λογισμικού στην κατανόηση λεξιλογίου (Hamouda, 2013. Goudarzi, 2012. Behjat, 2011. Ali & Haafar, 2010. Al-Qqaili & Taha, 2007. Yoshii, 2006. Lubliner, 2005. Nation, 2001. Bell & LeBlanc 2000. Lyman-Hayer & Davis, 1996), ενώ επιπλέον η χρήση τέτοιων μεθόδων αξιοποιείται και στην εξ αποστάσεως εκπαίδευση για όλες τις ηλικίες μαθητών (Μανούσου κ.ά., 2017: 18, 23). Άλλη έρευνα επικεντρώθηκε στην ανάπτυξη τεχνολογικού λεξιλογίου τονίζοντας την αυξημένη αποτελεσματικότητα των ειδικά σχεδιασμένων ηλεκτρονικών λεξικών. Υπό το πρίσμα αυτό, τα ψηφιακά λεξικά (βλ. 2. Ψηφιακά Λεξικά) μπορούν να θεωρηθούν ως πολύτιμα εκπαιδευτικά εργαλεία για τη βελτίωση της γνώσης των φοιτητών σε μια ποικιλία από

λεξικά αντικείμενα, σε αυθεντικά γλωσσικά περιβάλλοντα, παρέχοντας στους σπουδαστές φιλικές εμπειρίες εκμάθησης (Bataineh, 2014).

Ψηφιακά Λεξικά

Η εφαρμογή της Πληροφορικής σε όλα τα επιστημονικά πεδία, εκτός από την παροχή ισχυρών εργαλείων επεξεργασίας, δημιούργησε έναν καταιγισμό νέων επιστημονικών και τεχνικών όρων. Ένα λεξικό τεχνικών όρων (μηχανολογίας, ηλεκτρολογίας και ηλεκτρονικών) μπορεί να περιέχει περίπου 13.800 λήμματα (Χιονίδης, 1966), χωρίς μεγάλη συχνότητα ανανέωσης. Αντίστοιχα, ένα λεξικό επιστημονικών όρων (Φυσικής) μπορεί να περιέχει περίπου 9.200 λήμματα (Παπαγιαννακόπουλος, 1979), προσφέροντας ικανοποιητική κάλυψη του αντικειμένου. Ένα λεξικό Πληροφορικής με 10.400 λήμματα (Γαρίδης & Δεληγιαννάκης, 1993) χρειάζεται τακτική ενημέρωση προκειμένου να καλύπτει τη ζήτηση και να συμπεριλαμβάνει τους διαρκώς δημιουργούμενους νέους όρους, αφού σε μια τετραετία (1989-1993) πραγματοποιήθηκαν όπως φαίνεται έξι (6) εκδόσεις. Οι νέοι όροι, ως «εισαγόμενοι» από την Αγγλική γλώσσα (κατά κανόνα), εκτός από ερμηνευτικά σχόλια χρειάζονται και απόδοση στην Ελληνική γλώσσα, προσπάθεια στην οποία καταγίνονται οργανισμοί όπως ο Ελληνικός Οργανισμός Τυποποίησης (ΕΛΟΤ) και η Ελληνική Εταιρεία Ορολογίας (ΕΛΕΤΟ), με διοργάνωση σχετικών συνεδρίων (βλ. www.eleto.gr). Προκειμένου να ανταποκριθεί ένα λεξικό με όρους Πληροφορικής, ανεξαρτήτως του επιστημονικού πεδίου που υποστηρίζει, στην ανάγκη συχνότατης ενημέρωσης, η «παραδοσιακή» μορφή του ως καταλόγου λημμάτων, είτε σε έντυπη είτε σε ηλεκτρονική μορφή (e-book), δεν επαρκεί. Αντίθετα, χρειάζεται προς τούτο η υλοποίησή του ως βάσης δεδομένων (database). Μία τέτοια *λεξική βάση δεδομένων* (lexical database) διευκολύνει τόσο την ανάγκη συχνής ενημέρωσης όσο και τη δημιουργία πολύγλωσσων λεξικών.

Τα πολύγλωσσα λεξικά ελληνικού ενδιαφέροντος έχουν μια ιστορία περίπου 250 ετών (Καβαλλιώτης, 1770. Μοσχοπολίτης, 1802. Ουκούτας, 1797), με στόχο την εκμάθηση της ελληνικής γλώσσας και τη γραπτή αποτύπωση της βλαχικής γλώσσας στο βαλκανικό χώρο (Μπέης & Δασούλας, 2017: 53-54). Σήμερα, υπάρχει πληθώρα τέτοιων ψηφιακών λεξικών της ελληνικής γλώσσας, με χαρακτηριστικά αυτά του Ινστιτούτου Επεξεργασίας του Λόγου (ΙΕΛ: http://www.xanthi.ilsp.gr/dictionaries/) και της Πύλης για την Ελληνική Γλώσσα (http://www.greek-language.gr/greekLang/modern_greek/tools/lexica/triantafyllides/).

Αντίστοιχη ανάγκη λεξικών πολλαπλών ερμηνειών-«εισόδων» υπάρχει και σε απόλυτα σύγχρονες εφαρμογές, όπως αυτές της Υπολογιστικής Γλωσσολογίας (Computational Linguistics) - Γλωσσικής Τεχνολογίας (Linguistic Computing/Engineering) - Επεξεργασίας Φυσικής Γλώσσας (Natural Language Processing). Ήδη το πλήθος των παραπάνω όρων, όπου συνήθως χρησιμοποιείται διαφορετικός όρος από Πληροφορικούς και Γλωσσολόγους (με κάποια διαφορά στόχων), δίνει μια πρώτη ιδέα για το μέγεθος του προβλήματος. Με το γνωστικό αντικείμενο της επεξεργασίας φυσικής γλώσσας ασχολούνται συνήθως (αλλά όχι αποκλειστικά) σε μεταπτυχιακό επίπεδο είτε Πληροφορικοί που αρχικά ειδικεύονται στη Γλωσσολογία είτε Γλωσσολόγοι που αρχικά ειδικεύονται σε θέματα Πληροφορικής. Για τους σκοπούς της διδασκαλίας αυτού του γνωστικού αντικειμένου, η σχετική αγγλική ορολογία αποδίδεται στην Ελληνική γλώσσα και ως εδώ μία «κλασικού τύπου» δίγλωσση λεξική βάση δεδομένων φαίνεται επαρκής. Όμως, η σύμπραξη τριών ουσιαστικά επιστημονικών πεδίων στο ίδιο αντικείμενο

(Πληροφορικής, Μαθηματικών και Γλωσσολογίας) δημιουργεί το φαινόμενο της απόδοσης αγγλικών όρων στην Ελληνική γλώσσα με λέξεις που η κάθε μία έχει διαφορετική σημασία στην Πληροφορική/Μαθηματικά και διαφορετική στην Γλωσσολογία (*πολυσημία*). Μια τέτοια χαρακτηριστική περίπτωση είναι η λέξη-όρος *γράφημα*, όπου στην πρώτη περίπτωση (Πληροφορική/Μαθηματικά) δηλώνει ένα διάγραμμα, ενώ στη δεύτερη περίπτωση (Γλωσσολογία) δηλώνει το σύμβολο ενός φθόγγου («γράμμα»). Γενικότερα, η διαφοροποίηση στην ορολογία μεταξύ των Επιστημών έχει παρατηρηθεί εδώ και αρκετό καιρό (Γιαννουλοπούλου, 2003). Ταυτόχρονα, ο ίδιος όρος (*γράφημα*) αποδίδεται διαφορετικά στην Αγγλική γλώσσα όταν αφορά την Πληροφορική/Μαθηματικά (graph) και διαφορετικά όταν αφορά τη Γλωσσολογία (grapheme). Έτσι στην πραγματικότητα δημιουργείται η ανάγκη υλοποίησης ενός ψηφιακού λεξικού τριών εισόδων (αγγλικός όρος, ελληνικός όρος Πληροφορικής/Μαθηματικών, ελληνικός όρος Γλωσσολογίας), όπου αυτό είναι απαραίτητο.

Με αποκλειστικό σκοπό λοιπόν να διευκολυνθεί η εκμάθηση ορολογίας στο γνωστικό αντικείμενο της Υπολογιστικής Γλωσσολογίας (κυρίως από φοιτητές), αναπτύχθηκε μια μοναδική στο είδος της *λεξική βάση δεδομένων* τριών εισόδων (Καρβουνιάρη, 2015), που θα παρουσιαστεί στη συνέχεια (βλ. 4. Ανάπτυξη Εργαλείου Εκμάθησης Ορολογίας). Ο στόχος της σχεδίασης ήταν να αντιμετωπιστούν αποτελεσματικά τα παραπάνω φαινόμενα πολυσημίας και ακριβούς απόδοσης των αγγλικών όρων, σε σχέση με το ανάλογο γνωστικό πεδίο Πληροφορικής/Μαθηματικών ή Γλωσσολογίας. Της περιγραφής της λεξικής βάσης δεδομένων τριών εισόδων προηγείται μια γενικότερη παρουσίαση των λεξικών βάσεων δεδομένων (βλ. 3. Λεξικές Βάσεις Δεδομένων).

Λεξικές Βάσεις Δεδομένων

Γενικά, μια βάση δεδομένων (database) αποτελείται από το σύνολο των αρχείων στα οποία αποθηκεύονται τα δεδομένα ενός *πληροφοριακού συστήματος*. Τα αρχεία αυτά δεν είναι ανεξάρτητα μεταξύ τους αλλά συνενωμένα. Διακρίνονται όμως από τις έννοιες της *ενοποίησης* και του *μερισμού*. Με τον όρο *ενοποίηση δεδομένων* εννοούμε ότι μέσα στη βάση εξαλείφεται κάθε άσκοπη επανάληψη δεδομένων ενώ με την έννοια *μερισμός δεδομένων* εννοούμε τη δυνατότητα ταυτόχρονης προσπέλασης στα αρχεία της βάσης από πολλούς χρήστες (Silberschatz et al., 2011. Elmasri & Navathe, 2008). Για τη δημιουργία μιας βάσης δεδομένων απαιτείται η λογική σχεδίαση της δομής της. Προκειμένου να επιτευχθεί αυτός ο στόχος παρουσιάζονται ορισμένες έννοιες με ιδιαίτερη σημασία (Καρβουνιάρη, 2015: 18-19):

- Η *οντότητα* (entity) αναφέρεται σε κάθε αντικείμενο, πρόσωπο, γεγονός, κατάσταση ή και αφηρημένη έννοια που προσδιορίζεται από την ανεξάρτητη ύπαρξή του. Είναι δηλαδή κάτι υπαρκτό και διακριτό, για το οποίο ενδιαφερόμαστε να συλλέξουμε πληροφορίες.
- Προκειμένου οι οντότητες να γίνουν περισσότερο διακριτές και προσδιορισμένες, τις τεκμηριώνουμε με συγκεκριμένες ιδιότητες. *Ιδιότητες* (attributes) είναι τα συστατικά στοιχεία που περιγράφουν μια οντότητα. Οι ιδιότητες αποτελούν τον κύριο τρόπο διαφοροποίησης των στοιχείων τα οποία ανήκουν στην οντότητα.
- Οι οντότητες συνδέονται μεταξύ τους με σχέσεις ή διαφορετικά συσχετίζονται. Η *συσχέτιση* (relationship) είναι ένα είδος σύνδεσης μεταξύ των διαφορετικών

οντοτήτων που αναπαριστά μια αντίστοιχη σχέση των αντικειμένων στον πραγματικό κόσμο.

Οι παραπάνω έννοιες χρησιμοποιούνται για τη δόμηση μιας βάσης δεδομένων, σύμφωνα με κάποιο από τα πέντε επικρατέστερα υποδείγματα (*μοντέλα*) σχεδίασης (Παπακίτσος, 2000. Date, 1990): το *ιεραρχικό* (hierarchical), το *δικτυωτό* (network), το *σχεσιακό* (relational), το αντικειμενοστρεφές (object-oriented) και το *Οντοτήτων-Συσχετίσεων* (entity- relationship). Από αυτά τα υποδείγματα, το τελευταίο (*Οντοτήτων-Συσχετίσεων* ή Ο/Σ) προέκυψε αρχικά από τις ανάγκες επεξεργασίας γλωσσικών δεδομένων και προβλημάτων (Παπακίτσος, 2014. Papakitsos et al., 2002. Grigoriadou et al., 2002), συνεπώς κρίνεται καταλληλότερο για τη σχεδίαση λεξικών βάσεων δεδομένων (Παπακίτσος, 2014. Ντάφλου, 2013).

Οι λεξικές βάσεις δεδομένων είναι μια κατηγορία λεξικών όπου είναι δυνατή είτε η αποθήκευση τους σε νέα μέσα ηλεκτρονικής μορφής όπως σε σκληρούς ή οπτικούς δίσκους (CD/DVD), είτε η ανάρτησή τους στο διαδίκτυο και η δημιουργία on-line ψηφιακών λεξικών. Τα λεξικά αυτά ακολουθούν τις αρχές των έντυπων λεξικών για την αναζήτηση μιας λέξης αλλά προσφέρουν επίσης και πολλές άλλες δυνατότητες. Τα βασικά πλεονεκτήματα μιας λεξικής βάσης δεδομένων, έναντι ενός έντυπου, μπορούν να συνοψιστούν ως εξής (Καρβουνιάρη, 2015: 36):

- η επικαιροποίησή τους μπορεί να γίνει εύκολα και γρήγορα,
- η πρόσβαση στο υλικό στηρίζεται στο φιλικό προς τον χρήστη περιβάλλον της διεπαφής,
- υπάρχει δυνατότητα αποθήκευσης πολλών παράλληλων πηγών,
- υπάρχει δυνατότητα χρήσης υπερσυνδέσεων με εξωτερικές πηγές πληροφοριών (εικόνες, ήχους κ.τ.λ.),
- είναι εύκολη η ενσωμάτωση ποικίλων δραστηριοτήτων, ασκήσεων εκμάθησης, κατανόησης κ.τ.λ.,
- η χρήση μέσων όπως ιδιαίτερης γραμματοσειράς, υπογράμμισης κ.τ.λ. μπορεί να χρησιμοποιηθεί για έμφαση σε συγκεκριμένο μέρος της πληροφορίας,
- υποστηρίζεται η δυνατότητα αντιγραφής και μεταφοράς μέρους των πληροφοριών που εμφανίζονται σε νέο έγγραφο ή μέσο,
- ομοίως, η δυνατότητα άμεσης εκτύπωσης,
- και η δυνατότητα μεταφοράς του λεξικού με χρήση ελάχιστου χώρου, κάτι που θα ήταν ιδιαίτερα δύσκολο για ένα ογκώδες έντυπο λεξικό.

  Παράλληλα με τα πλεονεκτήματα μπορούν να αναφερθούν και κάποια μειονεκτήματα, τα οποία παρότι είναι ελάχιστα είναι υπαρκτά. Αυτά είναι τα ακόλουθα (Καρβουνιάρη, 2015: 37):
- ο χρόνος επηρεάζει την συσκευή αποθήκευσης και την αντοχή της,
- το κόστος αγοράς και συντήρησης ενός ηλεκτρονικού υπολογιστή (Η/Υ) και των αποθηκευτικών μέσων που συνδέονται σε αυτόν,
- η εξάρτηση του Η/Υ από ηλεκτρική παροχή ρεύματος,
- η τεχνοφοβία, που μπορεί να αποδειχθεί τροχοπέδη στη μάθηση, ιδιαίτερα για άτομα μεγαλύτερης ηλικίας (eBook, 2009: 7-8).

Τέλος αξίζει να σημειωθεί ότι οι λεξικές βάσεις δεδομένων προσαρμόζονται σε διάφορα πληροφοριακά συστήματα επεξεργασίας φυσικής γλώσσας, ανάλογα με τον τρόπο που σχεδιάζεται η δομή τους, προκειμένου να υποστηρίξουν μια συγκεκριμένη εφαρμογή (Litta & Passarotti, 2017. Bamman & Crane, 2009). Για τη συγκεκριμένη δραστηριότητα

ανάπτυξης υπάρχει και το ISO-24613:2008 (Francopoulo, 2013). Σε αυτή την περίπτωση έχουμε ηλεκτρονικά λεξικά που δημιουργούνται με διάφορα πρότυπα, όπως είναι το WordNet (Princeton University, 2010. Fellbaum, 2005), το FrameNet (Goddard, 2011. Heine & Narrog, 2009), το VerbNet και το Lexicon-grammar (Μαρκαντωνάτου & Φωτοπούλου, 2007), η παρουσίαση των οποίων όμως δεν σχετίζεται άμεσα με την παρούσα εργασία.

Ανάπτυξη Εργαλείου Εκμάθησης Ορολογίας

Η ανάπτυξη ενός εργαλείου εκμάθησης ορολογίας με τη μορφή λεξικής βάσης δεδομένων, όπως αυτό που περιγράφεται παρακάτω, έχει ως στόχο τη διευκόλυνση της εκμάθησης της τεχνικής ορολογίας ενός γνωστικού πεδίου (εδώ της Επεξεργασίας Φυσικής Γλώσσας) και την αντιμετώπιση του συχνού φαινομένου της πολυσημίας. Επιπλέον, η χρήση του εργαλείου έχει στόχο να προσφέρει στους σπουδαστές εμπειρίες εκμάθησης με φιλικό τρόπο, παρέχοντας τους πρόσβαση σε ποικιλία πηγών και πληροφοριών σε αυθεντικά γλωσσικά περιβάλλοντα (Καρβουνιάρη, 2015).

Η διαδικασία ανάπτυξης αποτελείται από δύο μέρη: το σχεδιασμό της δομής της (βλ. 4.1 Σχεδιασμός Δομής) και τη δημιουργία της διεπαφής (βλ. 4.2 Διεπαφή Λεξικού). Το πρώτο μέρος αφορά κυρίως τον χειριστή της λεξικής βάσης δεδομένων (operator), δηλαδή τον τεχνικό που ενημερώνει τα δεδομένα και επεμβαίνει στη δομή της, ενώ το δεύτερο μέρος αφορά κυρίως τον χρήστη (user) της εφαρμογής λογισμικού (Παπακίτσος, 2014: 72-76). Ο κυρίαρχος τρόπος υλοποίησης της λεξικής βάσης δεδομένων είναι μέσω ενός εργαλείου σχεσιακής βάσης δεδομένων (Magidow, 2015). Στη συγκεκριμένη εργασία χρησιμοποιήθηκε η MsAccess 2010 (Καρβουνιάρη, 2015).

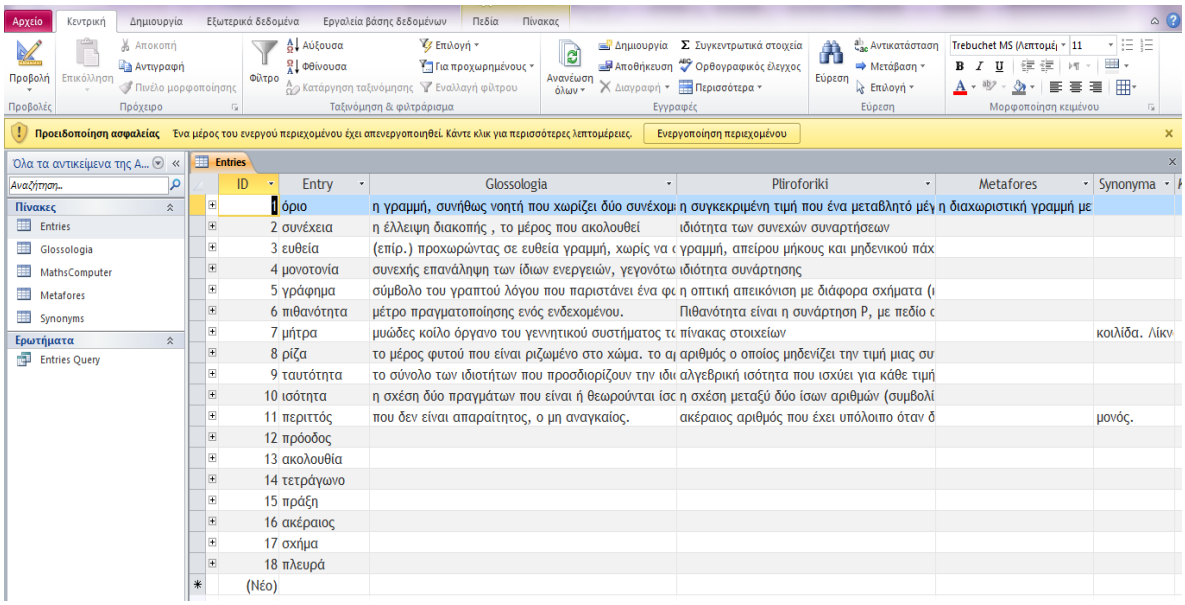
Σχεδιασμός Δομής

Ο τρόπος δόμησης μιας λεξικής βάσης δεδομένων αποτελεί μια σύνθετη και χρονοβόρα διαδικασία που απαιτεί ιδιαίτερη προσοχή για αποφυγή λαθών, επαναλήψεων, ατελειών κ.τ.λ. Για τον λόγο αυτό κρίνεται χρήσιμος ο ορισμός των εννοιών *μακροδομή* και *μικροδομή* ενός λεξικού (Καρβουνιάρη, 2015: 38-39):

- Η μακροδομή περιλαμβάνει το πλήθος όλων των λημμάτων που εμπεριέχονται στο λεξικό καθώς και τη σειρά που θα εμφανίζονται. Έτσι ένα λεξικό με πλούσια μακροδομή θεωρείται ότι περιέχει μεγάλο αριθμό λημμάτων, ενώ ένα λεξικό με φτωχή μακροδομή το αντίθετο.
- Για τον ορισμό του λήμματος απαιτείται η χρησιμοποίηση άλλων λέξεων, πληροφοριών, παραδειγμάτων, γραμματικών στοιχείων κ.τ.λ που θα προσδιορίσουν τη σημασία της. Η επιλογή των πληροφοριών γίνεται βάσει συγκεκριμένων κριτηρίων που διέπουν ολόκληρο το λεξικό και αποτελούν το αντικείμενο της μικροδομής.

  Συμπερασματικά, θα μπορούσε να αναφερθεί ότι η μακροδομή ασχολείται με τον αριθμό και τη σειρά των λημμάτων ενώ η μικροδομή με τη μορφή και το περιεχόμενο επεξηγήσεων των λημμάτων. Στις βασικές αρχές που καθορίζουν την συγγραφή ενός λεξικού κυρίαρχο ρόλο παίζουν οι εξής δύο κανόνες:
- Το λεξικό πρέπει να αποτελεί ένα *κλειστό* κύκλωμα. Δηλαδή όλες οι λέξεις που εμφανίζονται στη μικροδομή πρέπει να εμπεριέχονται και στη μακροδομή.
- Όλες οι λέξεις που χρησιμοποιούνται για τον ορισμό ενός λήμματος πρέπει να είναι πιο συχνές απ’ αυτές που ορίζουν.

Κάθε λήμμα χαρακτηρίζεται από ένα σύνολο πληροφοριών που συντελούν στην επεξήγηση του ορισμού του. Σε αυτές περιλαμβάνονται η φωνητική μεταγραφή του, ο κώδικας κλίσης του, οι διάφορες σημασίες του (κυριολεκτικές-μεταφορικές), ο ορισμός, παραδείγματα και χρήσεις μέσα σε προτάσεις, συντακτικές πληροφορίες, η ετυμολογία του, συνώνυμα, αντώνυμα κ.τ.λ.



Εικόνα 1: Απεικόνιση της λεξικής βάσης δεδομένων σε MsAccess 2010 (Καρβουνιάρη, 2015)

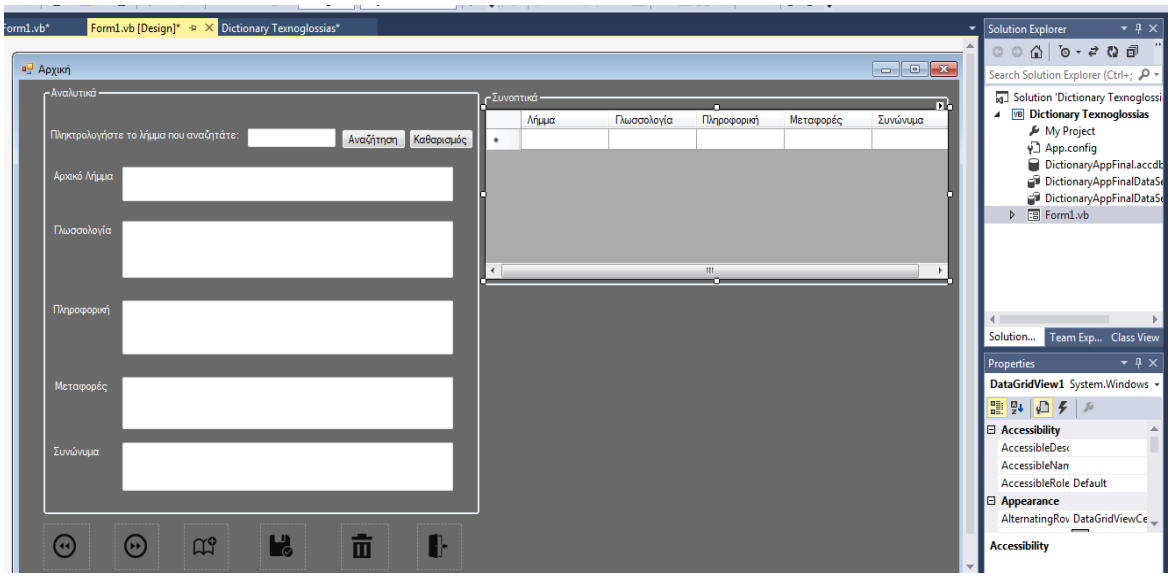
Με βάση τα προηγούμενα, η δομή μιας λεξικής βάσης δεδομένων (Εικ. 1) που δύναται να υποστηρίζει πολλαπλές εφαρμογές ορολογίας (ελληνική-ελληνική, αγγλική-ελληνική) αποτελείται τουλάχιστον από τα παρακάτω στοιχεία (μικροδομή), στα αντίστοιχα *πεδία* (στήλες):

- Το λήμμα (οντότητα) στην ελληνική ορολογία (Entry).
- Έναν αριθμό (ιδιότητα) που χαρακτηρίζει κάθε εγγραφή-λήμμα μοναδικά (*κλειδί*) και το διαφοροποιεί από τα υπόλοιπα (ID).
- Την ερμηνεία του λήμματος (ιδιότητα), όπως θα τη χρησιμοποιούσε ένας γλωσσολόγος (Glossologia) με την αγγλική μετάφραση του όρου.
- Την ερμηνεία του λήμματος (ιδιότητα), όπως θα τη χρησιμοποιούσε ένας πληροφορικός (Pliroforiki) με την αγγλική μετάφραση του όρου.
- Την ύπαρξη μεταφορών (ιδιότητα), στην οποία θα αποθηκεύονται όλες οι μεταφορικές ερμηνείες ενός λήμματος με τη μορφή κειμένου (Metafores).
- Την ύπαρξη συνώνυμων (συσχέτιση), στην οποία θα αποθηκευτούν όλα τα συνώνυμα ενός λήμματος με τη μορφή κειμένου (Synonyma).

  Η διαφορά της λεξικής βάσης δεδομένων από ένα κοινό λεξικό (σε έντυπη ή σε ηλεκτρονική μορφή) είναι η δυνατότητα επέκτασης είτε της μικροδομής (όπως π.χ. συμπεριλαμβάνοντας και άλλες γλώσσες με την προσθήκη αντίστοιχου πεδίου-στήλης) είτε της μακροδομής μέσω του εύκολου εμπλουτισμού του λεξικού με νέα λήμματα. Η τελευταία αυτή δυνατότητα υλοποιείται από τη *διεπαφή* (interface) του λεξικού.

Διεπαφή Λεξικού

Η σχεδίαση και υλοποίηση της διεπαφής δημιουργεί ένα ολοκληρωμένο πλέον ψηφιακό εργαλείο λεξικού γλωσσικής τεχνολογίας, μέσω της οποίας θα αλληλεπιδρά ο χρήστης (user) με την εφαρμογή (Εικ. 2). Στη συγκεκριμένη εργασία, η διεπαφή υλοποιήθηκε στο περιβάλλον Visual Studio Ultimate 2013 (Καρβουνιάρη, 2015: 46).



Εικόνα 2: Μορφή του Παραθύρου Διεπαφής (Καρβουνιάρη, 2015)

Οι τυπικές δυνατότητες μιας τέτοιας διεπαφής περιλαμβάνουν:

- Αναζήτηση λήμματος, όπου εμφανίζονται όλες οι σχετικές πληροφορίες (φόρμα στο αριστερό τμήμα του παραθύρου διεπαφής).
- Εμφάνιση όλου του περιεχομένου της λεξικής βάσης δεδομένων (πίνακας στο κεντρικό τμήμα του παραθύρου διεπαφής).
- Πλοήγηση σε προηγούμενα και επόμενα λήμματα (δύο πρώτα από αριστερά εικονίδια-πλήκτρα της γραμμής κάτω αριστερά του παραθύρου διεπαφής).
- Προσθήκη νέου λήμματος (τρίτο από αριστερά εικονίδιο-πλήκτρο της γραμμής κάτω αριστερά του παραθύρου διεπαφής).
- Αποθήκευση των αλλαγών και ενημέρωση της βάσης δεδομένων (τέταρτο από αριστερά εικονίδιο-πλήκτρο της γραμμής κάτω αριστερά του παραθύρου διεπαφής).
- Διαγραφή επιλεγμένου λήμματος (πέμπτο από αριστερά εικονίδιο-πλήκτρο της γραμμής κάτω αριστερά του παραθύρου διεπαφής).
- Τερματισμός της εφαρμογής (έκτο από αριστερά εικονίδιο-πλήκτρο της γραμμής κάτω αριστερά του παραθύρου διεπαφής).
- Αλφαβητική ταξινόμηση της βάσης δεδομένων μετά την προσθήκη νέου λήμματος (επιλογές μενού στο δεξιό τμήμα του παραθύρου διεπαφής).

  Άλλες δυνατότητες επέμβασης που είναι τυπικές σε μια εφαρμογή *ηλεκτρονικής λεξικογραφίας*, όπως π.χ. οι ταξινομήσεις κατά εννοιολογική σημασία ή με βάση τις καταλήξεις, μπορούν να υλοποιηθούν μέσω του μενού στο δεξιό τμήμα του παραθύρου διεπαφής.

Συμπεράσματα - Επίλογος

Στην παρούσα εργασία περιγράφηκε η σχεδίαση και υλοποίηση μιας λεξικής βάσης δεδομένων, ως εφαρμογής λογισμικού του τεχνολογικού πεδίου της Ηλεκτρονικής Λεξικογραφίας. Ο αρχικός σκοπός της δημιουργίας ενός τέτοιου εργαλείου ήταν η διευκόλυνση της εκμάθησης της τεχνικής ορολογίας του επιστημονικού πεδίου της Επεξεργασίας Φυσικής Γλώσσας, στην οποία εμφανίζεται συχνά το φαινόμενο της πολυσημίας, δηλαδή ο ίδιος όρος να έχει δύο τουλάχιστον διαφορετικές σημασίες. Το φαινόμενο αυτό οφείλεται τόσο στη μετάφραση αγγλικών όρων στην ελληνική γλώσσα όσο και στο συνδυασμό διαφορετικών επιστημών (Πληροφορικής, Μαθηματικών και Γλωσσολογίας) στο ίδιο πεδίο εφαρμογών.

Ιδιαίτερος προβληματισμός σημειώθηκε κατά την φάση της σχεδίασης, καθώς αυτή μπορούσε να πραγματοποιηθεί με διάφορους δυνατούς τρόπους όπως η χρήση δύο τύπων διεπαφών, όπου ανάλογα με την ιδιότητα του χρήστη, αυτός θα μπορούσε να βλέπει μόνο την ερμηνεία του λήμματος που τον ενδιαφέρει. Δηλαδή, ο Γλωσσολόγος να βλέπει μόνο τη Γλωσσολογική ερμηνεία και ο Πληροφορικός μόνο την ερμηνεία του λήμματος κατά τις αρχές της Πληροφορικής. Επιπλέον, ήταν (και παραμένει) δυνατή η χρήση υπερσυνδέσεων που θα προσφέρουν ή θα παραπέμπουν σε περισσότερες πληροφορίες, μια δυνατότητα που διευρύνει εξαιρετικά την εκπαιδευτική αξία ενός ηλεκτρονικού λεξικού. Τελικά, επικράτησε ο σχεδιασμός που προσφέρει ταυτόχρονη παρουσίαση των ερμηνειών λόγω της πειραματικής φύσης της παρούσας εργασίας. Στο ψηφιακό λεξικό που σχεδιάστηκε έγινε πρόβλεψη για αλφαβητική ταξινόμηση των λημμάτων. Ωστόσο το εργαλείο αυτό θα μπορούσε να διευρυνθεί δίνοντας τη δυνατότητα στον χρήστη να εφαρμόζει επιπλέον τρόπους ηλεκτρονικής καταγραφής, όπως για παράδειγμα η ετυμολογική ή η αντίστροφη καταγραφή, ακολουθώντας τους βασικούς κανόνες της Ηλεκτρονικής Λεξικογραφίας. Τέλος, η δυνατότητα ανάρτησής του σε έναν ηλεκτρονικό ιστότοπο παρέχει την ευκολία της εξ αποστάσεως χρήσης του από κάθε ενδιαφερόμενο.

Συμπερασματικά, οι λεξικές βάσεις δεδομένων αποτελούν ένα ισχυρό και εύκολα επεκτάσιμο εκπαιδευτικό εργαλείο στον τομέα της εκμάθησης ορολογίας. Η χρήση τους μπορεί να προσφέρει στους σπουδαστές εμπειρίες εκμάθησης με φιλικό τρόπο, παρέχοντας πρόσβαση σε ποικιλία πηγών και πληροφοριών σε αυθεντικά γλωσσικά περιβάλλοντα, ανεξαρτήτως του γνωστικού αντικειμένου.

Βιβλιογραφία

Ali, M., & Haafar, M. (2010). *A comparative study of model and traditional reading comprehension*. Kuala Lumpur: University of Malaya.

Al-Qqaili, M., & Taha, S. (2007). *Jordanian EFL undergraduate students' critical reading skills*. MA thesis, Yarmouk University, Irbid, Jordan.

Bamman, D., & Crane, G. (2009). Computational Linguistics and Classical Lexicography. *Digital Humanities Quarterly*, *3*(1), http://www.digitalhumanities.org/dhq/vol/3/1/000033/000033.html.

Bataineh, A.M. (2014). The Effect of Electronic Dictionaries and Hypermedia Annotations on English Major Students' Reading Comprehension and Vocabulary Learning. *International Journal of Linguistics*, *6*(4), 102-115.

Behjat, F. (2011). Teacher correction or word processors: Which is a better option for the improvement of EFL students' writing skill? *Journal of Language Teaching and Research*, *2*(6), 1430-1434.

Bell, L., & LeBlanc, B. (2000). The language of glosses in L2 reading computer: Learners' preferences. *Hispania*, *83*(2), 274-285.

Chun, M., & Plass, L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, *80*(2), 183-198.

Date, C.J. (1990). *An Introduction to Database Systems* (Vol. I, 5th edition). Addison-Wesley.

eBook (2009). *Πρόταση για τον Σχεδιασμό και την Εισαγωγή του «Ηλεκτρονικού Βιβλίου» στην Εκπαίδευση*. Αθήνα: Παιδαγωγικό Ινστιτούτο.

Elmasri, R., & Navathe, S.B. (2008). *Θεμελιώδεις Αρχές Συστημάτων Βάσεων Δεδομένων* (5η έκδοση). Αθήνα: Δίαυλος.

Fellbaum, C. (2005). WordNet and wordnets. In K. Brown et al. (eds.), *Encyclopedia of Language and Linguistics* (2nd edition, pp. 665-670). Oxford: Elsevier.

Francopoulo, G. (ed.) (2013). *LMF: Lexical Markup Framework*. Hoboken, NJ: Wiley-ISTE.

Goddard, C. (2011). *Semantic Analysis: A Practical Introduction*. Oxford: Oxford University Press.

Goudarzi, Z. (2012). The effect of input enhancement of collocations in reading on collocation learning and retention of EFL learners. *International Education Studies*, *6*(4), 63-71.

Grigoriadou, M., Papakitsos, E., & Philokyprou, G. (2002). Semantic Modeling in Morpheme-based Lexica for Greek. In Proceedings of the 5th Systems Science Congress "Res-Systemica" (Special Issue: www.res-systemica.org), Crete.

Hamouda, A. (2013). A Study of Dictionary Use by Saudi EFL Students at Qassim University. *Study in English Language Teaching*, *1*(1), 227-257.

Hansen-Smith, E. (2001). Computer-assisted language learning. In R. Carter, & D. Nunan (eds.), *The Cambridge Guide to Teaching English to Speakers of other Languages*. Cambridge: CUP.

Heine, B., & Narrog, H. (2009). *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press.

Hulstijn, H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P.J.L. Arnaud, & H. Behoint (eds.), *Vocabulary and Applied Linguistics* (pp. 113-125). London: MacMillan.

Lai, C. (2006). *The advantages and disadvantages of computer technology in second language acquisition*. Unpublished doctoral thesis, Prairie View A&M University.

Laufer, B., & Shumueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC Journal*, *28*(1), 89-113.

Litta, E., & Passarotti, M. (eds.) (2017). *Proceedings of the Workshop on Resources and Tools for Derivational Morphology* (DeriMo). Milano, Italy, 5-7 October 2017.

Lubliner, S. (2005). The effect of comprehensive vocabulary instruction on title I students' meta-cognitive word-learning skills and reading comprehension. *Journal of Literacy Research*, *4*(3), 70-83.

Lyman-Hayer, M., & Davies, J. (1996). The Case for Computer-Mediated Reading: Une Vie de Boy. *The French Review*, *69*(5), 775-792.

Magidow, A. (2015). A relational database model and prototype for storing diverse discrete linguistic data. *Journal for Language Technology and Computational Linguistics*, *30*(1), 27-45.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: CUP.

Nikolova, R. (2002). Effect of students' participation in authoring of multimedia materials on student acquisition of vocabulary. *Language Learning & Technology*, *6*(1), 100-122.

Papakitsos, E., Gregoriadou, M., & Philokyprou, G. (2002). Modelling a morpheme-based Lexicon for Modern Greek. *Literary and Linguistic Computing*, *17*(4), 475-490.

Princeton University (2010). *About WordNet*. Princeton, NJ: Princeton University. Available at: https://wordnet.princeton.edu/.

Silberschatz, A., Korth, H.F., & Sudarshan, S. (2011). *Η Πλήρης Θεωρία των Βάσεων Δεδομένων* (6η έκδοση). Αθήνα: Γκιούρδας, Μ.

Taylor, M. (2006). The effects of CALL versus traditional L1 glosses on L2 reading comprehension. *CALICO Journal*, *23*(2), 309-318.

Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *SSLA*, *19*, 287-307.

Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology*, *10*(3), 85-101.

Αλιβίζος, Σ. (2015). Διδασκαλία ως φαινόμενο χωροχρονικής αποπλαισίωσης και ο ρόλος των νέων και ψηφιακών Μέσων για το σχεδιασμό της. *Ανοικτή Εκπαίδευση: το περιοδικό για την Ανοικτή και εξ Αποστάσεως Εκπαίδευση και την Εκπαιδευτική Τεχνολογία*, *11*, 8-19.

Γαρίδης, Π.Κ., & Δεληγιαννάκης, Ε.Ν. (1993). *Σύγχρονο Λεξικό Πληροφορικής* (6η έκδοση). Αθήνα: Εκδόσεις Δίαυλος.

Γιαννουλοπούλου, Γ. (2003). Θετικές και Ανθρωπιστικές Επιστήμες: Υπάρχουν Διαφορές στην Επιστημονική Ορολογία; ΕΛΕΤΟ – 4ο Συνέδριο «Ελληνική Γλώσσα και Ορολογία», Αθήνα, 30 Οκτωβρίου - 1 Νοεμβρίου 2003.

Καβαλλιώτης, Α. (1770). *Πρωτοπειρία παρά του σοφολογιωτάτου και αιδεσιμωτάτου Διδασκάλου Ιεροκήρυκος, και Πρωτοπαπά Κυρίου Θεοδώρου Αναστασίου Καβαλλιώτου του Μοσχοπολίτου ξυντεθείσα, και νυν τύποις εκδοθείσα δαπάνη του εντιμοτάτου, και χρισημωτάτου Κυρίου Γεωργίου Τρίκουπα, του και Κοσμήσκη επιλεγόμενου εκ πατρίδος Μοσχοπόλεως*. Βενετία.

Καρβουνιάρη, Δ. (2015). *Ψηφιακό Λεξικό Όρων Γλωσσικής Τεχνολογίας: Φάση Προετοιμασίας*. Διπλωματική εργασία, Διεπιστημονικό Διαπανεπιστημιακό ΠΜΣ «Τεχνογλωσσία», ΕΚΠΑ & ΕΜΠ.

Μανούσου, Ε., Κοντογεωργάκου, Β., Γεωργιάδη, Ε., & Κόκκαλη, Α. (2017). Παιδαγωγικό υλικό στην σχολική εξ αποστάσεως εκπαίδευση. Μελέτη Περίπτωσης: το εξ αποστάσεως εκπαιδευτικό Κέντρο της Βικτώρια στην Αυστραλία. *Ανοικτή Εκπαίδευση: το περιοδικό για την Ανοικτή και εξ Αποστάσεως Εκπαίδευση και την Εκπαιδευτική Τεχνολογία*, *13*, 5-25.

Μαρκαντωνάτου, Σ., & Φωτοπούλου, Α. (2007). *Το Εργαλείο «Έκφραση»*. Εισήγηση του 8ου Διεθνούς Συνεδρίου Ελληνικής Γλωσσολογίας (ICGL8). Πανεπιστήμιο Ιωαννίνων, 30/8-2/9 2007.

Μοσχοπολίτης, Δ. (1802 [1794]). *Εισαγωγική Διδασκαλία περιέχουσα Λεξικόν Τετράγλωσσον των τεσσάρων κοινών διαλέκτων, ήτοι της απλής Ρωμαϊκής, της εν Μοισία Βλαχικής, της Βουλγαρικής και της Αλβανιτικής*. Βενετία.

Μπέης, Σ., & Δασούλας, Φ. (2017). Πρόταση για το σύστημα γραφής της βλαχικής γλώσσας. *Γλωσσολογία/Glossologia*, *25*, 51-69.

Ντάφλου, Ο. (2013). *Αυτόματος Μορφολογικός Αναλυτής της Νέας Ελληνικής σε Γλώσσα Προγραμματισμού Visual C: Βελτίωση Αλγορίθμων*. Διπλωματική Εργασία: Διεπιστημονικό Διαπανεπιστημιακό ΠΜΣ «Τεχνογλωσσία», ΕΚΠΑ & ΕΜΠ.

Ουκούτας, Κ. (1797). *Νέα Παιδαγωγία ήτοι Αλφαβητάριον εύκολον του μαθείν τα νέα παιδία τα ρωμανο-βλάχικα Γράμματα εις κοινήν χρήσιν των Ρωμανο-Βλάχων, νυν πρώτον συνετέθη και εδιορθώθη παρά του αιδαισιμοτάτου εν ιερεύσιν κ. κ. Κωνσταντίνου του Ουκούτα, Μοσχοπολίτου, Χαρτοφύλακος και Πρωτοπαπά εν τη Ποσνάνια της μεσημβρινής Προυσίας, και δι' αυτού χάριν εδόθη εις τύπον δια καύχημα του Γένους*. Εν Βιέννη: Τύποις Μαρκίδων Πούλιου.

Παπαγιαννακόπουλος, Π. (1979). *Αγγλο-Ελληνικόν Λεξικόν Φυσικών Όρων*. Αθήνα: Μ. Καρδαμίτσας.

Παπακίτσος, Ε.Χ. (2000). *Συμβολή στη Μορφολογική Επεξεργασία της Νέας Ελληνικής: Λειτουργική Αποσύνθεση – Καρτεσιανό Ηλεκτρονικό Λεξικό*. Διδακτορική διατριβή, ΕΚΠΑ.

Παπακίτσος, Ε.Χ. (2014). *Γλωσσική Τεχνολογία Λογισμικού: ΙΙ. Πραγμάτωση*. Αθήνα: ΕΒΕ.

Χιονίδης, Χ. (1966). *Αγγλο-Ελληνικόν Λεξικόν Τεχνικών Όρων* (2η έκδοση). Αθήνα: Ατλαντίς.