

**LEXIGRAF - ΑΝΑΠΤΥΞΗ ΠΛΑΤΦΟΡΜΑΣ ΛΟΓΙΣΜΙΚΟΥ
ΓΙΑ ΑΥΤΟΜΑΤΟΠΟΙΗΜΕΝΟ ΠΟΛΥΓΛΩΣΣΙΚΟ
ΛΕΞΙΚΟΓΡΑΦΙΚΟ DESKTOP PUBLISHING**

Γιάννης Χατζόπουλος

ΠΕΡΙΛΗΨΗ

Η πλατφόρμα λογισμικού LexiGraf αναπτύχθηκε υπό τα MS WINDOWS για να αυτοματοποιήσει τη σύνταξη πολύγλωσσων και δίγλωσσων έντυπων λεξικών ορολογίας. Μπορεί να υποστηρίξει 2-8 γλώσσες ταυτοχρόνως (single byte σετ χαρακτήρων για την 16bit έκδοση, UNICODE υπό ανάπτυξη για την 32bit). Λειτουργεί ως ενοποιημένο περιβάλλον διαχείρισης πολυγλωσσικής ορολογίας και αυτοματοποιημένου desktop publishing. Προσφέρει άμεση μετάβαση από την πρωτογενή συλλογή ορολογίας προς την τυπογραφική πρέσα ή ενδιάμεσα προς πλατφόρμες ηλεκτρονικών λεξικών ορολογίας (CD-ROM). Τα έντυπα πολυγλωσσικά λεξικά ορολογίας που παράγει το LexiGraf υλοποιούνται με την τεχνική του ευρετηρίου όρων. Με συνδυασμούς των διαθέσιμων γλωσσών ανά ζεύγη, αυτομάτως προκύπτουν δίγλωσσα λεξικά ορολογίας. Λειτουργεί σε δίκτυο H/Y, συνδυαστικά με δημοφιλή πακέτα επιτραπέζιων εκδοτικών συστημάτων (πχ. Quark-Χpress) κατά το πρότυπο διασύνδεσης XTAG/ RTF. Μπορεί, όπου το επιτρέπουν οι συνθήκες, να δεχθεί είσοδο λεξιλογικής πληροφορίας από λογισμικό οπτικής αναγνώρισης χαρακτήρων (OCR). Η ανάπτυξη του LexiGraf πραγματοποιήθηκε για να καλυφθούν οι μηχανογραφικές ανάγκες του Γενικού Τμήματος Ξένων Γλωσσών του ΑΠΘ κατά την εκπόνηση ενός τετράγλωσσου λεξικού ορολογίας θετικών επιστημών (Αγγλ/ Γαλλ/ Γερμ/ Ελλην ' 50.000 όρων/ γλώσσα) + 9 δίγλωσσων λεξικών (επιλεγμένοι συνδυασμοί των παραπάνω γλωσσών ανά ζεύγη) + δευτερογενής ηλεκτρονική έκδοση του τετράγλωσσου λεξικού. URL <http://egnatia.ee.auth.gr/~yhat/yiannis/>

**LEXIGRAF - DEVELOPMENT OF AN AUTOMATED MULTILINGUAL
LEXICOGRAPHY DESKTOP PUBLISHING TOOL**

Yiannis Hatzopoulos

SUMMARY

The LexiGraf software base has been developed under MS Windows to assist the creation of multilingual and bilingual terminology dictionaries in print. Any language combination supported by MS Windows™ can be accommodated to a maximum of 8 languages (single byte language sets only for the 16bit version LexiGraf, Unicode 32bit version under development). The system offers direct transition from the terminology collection to the printing press; provides terminology management services, LAN use and excellent quality output to typesetters for subsequent printing press reproduction. Multilingual dictionaries produced by LexiGraf deploy a term reference index technique,

which allows their direct implementation in 2 volumes, regardless of requested number of languages. By combining available languages into pairs bilingual dictionaries are automatically produced, ready for the printing press. Recently introduced features include process integration with OCR products for updating existing terminology resources and connection layers with popular DTP software products (Xtag/RTF protocol). The resulting database files can be utilised for electronic media publishing undertakings. Currently LexiGraf is being used in the Aristotle University - Thessaloniki Greece on a 4 language (Engl/ French/ German/Greek - 50.000 terms/ language) natural sciences dictionary + 9 bilingual ones + electronic versions. URL <http://egnatia.ee.auth.gr/~yhat/yiannis/>

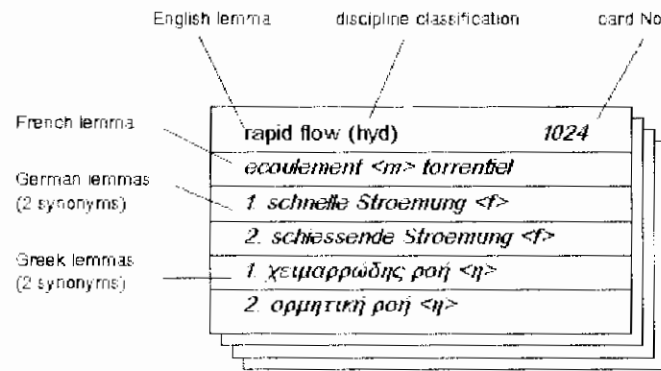
Keywords: Desktop publishing, multilingual lexicography, bilingual lexicography, science terminology, electronic dictionaries

0. ANTI ΕΙΣΑΓΩΓΗΣ

Η Ελληνική γλώσσα ομιλείται από 20.000.000 ανθρώπους σε όλη τη γη. Λόγω του μικρού αυτού αριθμού, η πολυγλωσσική τεκμηρίωση της ελληνικής επιστημονικής ορολογίας είναι ανεπαρκής, παρότι οι ανάγκες επικοινωνίας της επιστημονικής μας κοινότητας σε ένα ανοικτό διεθνές περιβάλλον αυξάνονται καθημερινά. Η πλατφόρμα LexiGraf αναπτύχθηκε για να καλύψει τις ανάγκες μηχανογράφησης κατά την εκπόνηση ενός έντυπου τετράγλωσσου λεξικού ορολογίας θετικών επιστημών (Αγγλ/Γαλλ/Γερμ/Ελλην. 50,000 όρων/γλώσσα) στο Γενικό Τμήμα Ξένων Γλωσσών του Α.Π.Θ. (κκ Στεπανιάν, Χαρσαύλη, Ζάχου, Κωνσταντίνου τηλ. 031.997862 fax 031.998419), μίας σειράς έντυπων δίγλωσσων λεξικών ορολογίας, όπως προκύπτουν από συνδυασμούς των παραπάνω γλωσσών ανά ζεύγη, καθώς και μίας συνοδευτικής έκδοσης CD-ROM του τετράγλωσσου λεξικού.

Το LexiGraf αναπτύχθηκε ως εξειδικευμένο εργαλείο διαχείρισης πολυγλωσσικών βάσεων ορολογίας και αυτοματοποιημένου Desktop publishing, ώστε να πραγματοποιηθεί η μετάβαση από την πρωτογενή συλλογή ορολογίας προς την τυπογραφική πρέσα ή δευτερογενώς προς το πιεστήριο CD-ROM με τα ελάχιστο δυνατό κόστος.

Κατά την ανάπτυξη λήφθηκε μέριμνα, ώστε το λογισμικό να είναι σε θέση να καλύψει μελλοντικές ανάγκες (πρόσθετες γλώσσες, νέες προδιαγραφές σελιδοποίησης, ανοικτή διασύνδεση με άλλα πακέτα λογισμικού, εισαγωγή πληροφορίας από διάφορες πηγές, αξιοποίηση τεχνολογίας δικτύων κκ)



1. ΠΡΩΤΟΓΕΝΗΣ ΣΥΛΛΟΓΗ ΟΡΟΛΟΓΙΑΣ

Το λεξιλογικό υλικό το οποίο κληθήκαμε να επεξεργαστούμε με το LexiGraf ήταν ήδη αποδελτιωμένο σε περίπου 50,000 χειρόγραφες καρτέλες. Παρακάτω διακρίνουμε μία τυπική καρτέλα αποδελτίωσης με αγγλικό λήμμα + συνώνυμα, γνωστικό αντικείμενο, αύξοντο αριθμό, γαλλική μετάφραση + συνώνυμα, γερμανική μετάφραση + συνώνυμα, ελληνική μετάφραση + συνώνυμα. Γραμματικοί προσδιορισμοί συμπεριλαμβάνονταν όπου κρίθηκε αναγκαίο. Η γλώσσα κλειδί ήταν η αγγλική. Σκοπός μας ήταν να αναπτύξουμε ένα περιβάλλον λογισμικού, στο οποίο αφού τροφοδοτηθούν οι διαθέσιμοι ορολογικοί πόροι (*terminological resources*) σε PCs με δακτυλογράφηση (*data-entry*) ή οποιονδήποτε άλλο ενδεδειγμένο τρόπο, να είμαστε σε θέση με το πάτημα ενός κουμπιού να ολοκληρώνουμε αυτομάτως τη διαδοχή των εργασιών: ταξινόμηση/ επεξεργασία/ σελιδοποίηση/ *desktop-publishing*, ώστε να λαμβάνουμε χωρίς πρόσθετη εργασία έτοιμα προς μαζική αναπαραγωγή πολύγλωσσα και δίγλωσσα έντυπα λεξικά. Με μία παρόμοια διαδικασία οδηγούμαστε σε εφαρμογές ηλεκτρονικών λεξικών. Τα γνωστικά αντικείμενα τα οποία περιλαμβάνονταν στην συγκεκριμένη συλλογή πολύγλωσσης ορολογίας ήταν τα ακόλουθα:

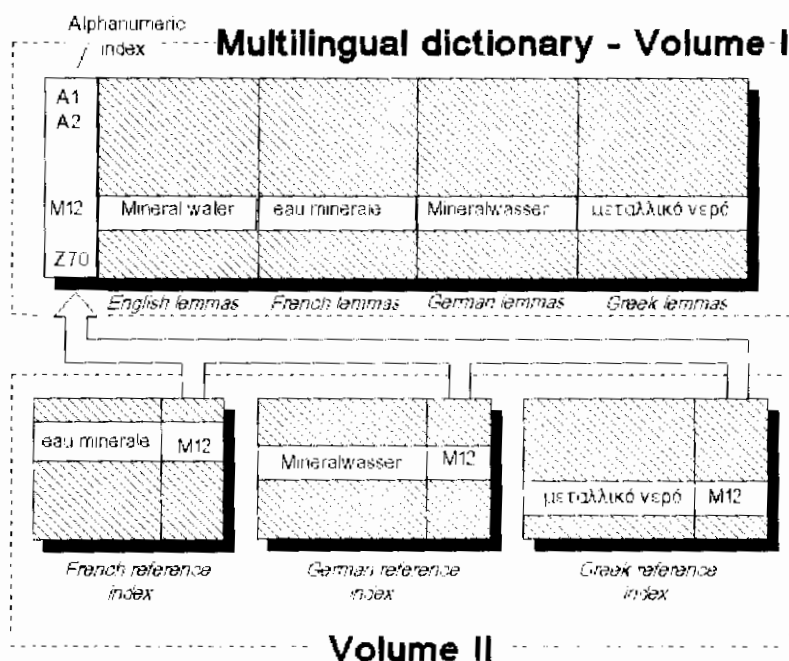
acoustics	advanced mathematics	aerospace engineering	algebra
analytical chemistry	analytical mathematics	astronomy	astrophysics
atomic physics	biochemistry	calculus	chemical engineering
climatology	communications	computer science	control systems
crystallography	electricity	electro-magnetism	electronics engineering
fluid mechanics	food engineering	geochemistry	geodesy
geography	geology	geometry	geophysics
hydrology	inorganic chemistry	mapping	material technology
mechanical engineering	mechanics	metallurgy	meteorology
mineralogy	mining engineering	nuclear physics	nucleonics
oceanology	optics	ordnance	organic chemistry
paleobotany	particle physics	petroleum engineering	petrology
physical chemistry	plasma physics	quantum mechanics	relativity
set theory	solid state physics	spectroscopy	statistical mechanics
statistics	systems engineering	thermodynamics	vectoring

2. ΠΟΛΥΓΛΩΣΣΑ ΛΕΞΙΚΑ ΜΕ ΕΥΡΕΤΗΡΙΑ ΟΡΩΝ

Η τεχνική του ευρετηρίου όρων αποτελεί την πλέον διαδεδομένη δομή σύνταξης έντυπων πολύγλωσσικών λεξικών ορολογίας. Εξαιρετικά δείγματα έχουν κυκλοφορήσει οι εταιρείες HatierTM, LongmanTM, ElsevierTM και LangenscheidtTM. Ανάλογα πολύγλωσσα λεξικά αποτελούνται από δύο διακριτά τμήματα/ τόμους. Στον πρώτο τόμο παρατίθενται οι μεταφράσεις των όρων σε όλες τις γλώσσες του λεξικού, ταξινομημένες ως προς την αντιστοιχία τους στη γλώσσα κλειδί. Κάθε όρος στην γλώσσα κλειδί 'διευθυνσιοδοτείται' μέσα στο πολύγλωσσο λεξικό με έναν ειδικό αλφάριθμο ή αύξοντα αριθμό. Ο δεύτερος τόμος του λεξικού περιέχει χωριστά ευρετήρια όρων για κάθε μία γλώσσα - πλην της γλώσσας κλειδιού - τα οποία παραπέμπουν στις καταχωρήσεις του πρώτου τόμου, χρησιμοποιώντας τον προαναφερθέντα αλφάριθμο. Η μεθοδολογία είναι εντελώς παρόμοια με την ύπαρξη ευρετηρίου όρων σε οποιοδήποτε βιβλίο, και μας επιτρέπει πρόσβαση από και προς οποιδήποτε γλώσσα σε ένα συμπαγές τελικό έντυπο προϊόν.

Τα έντυπα πολύγλωσσα λεξικά ορολογίας, λόγω του τυπικά υψηλού τους κόστους απευθύνονται κυρίως σε επαγγελματίες μεταφραστές και βιβλιοθήκες. Στην περίπτωση μας επιλέχθηκε από την πρώτη στιγμή να μαρφοποιηθεί το διαθέσιμο λεξιλογικό υλικό και να κυκλοφορήσει στην αγορά ως έντυπα πολύγλωσσα λεξικά για τους ακόλουθους λόγους:

Χρησιμοποιώντας τις δυνατότητες των σύγχρονων Η/Υ, προϋπολογίσθηκε ότι το λεξικογραφικό υλικό θα έφτανε στον καταναλωτή σε έντυπη μορφή με μία λιανική τιμή που δεν θα ξεπερνούσε κατά πολύ το κόστος φωτοαντιγραφής του.



Ήταν αναμενόμενο ότι πολλοί από τους ενδιαφερόμενους λαθροαντιγραφείς, λόγω και του μεγάλου μεγέθους της έντυπης έκδοσης (1500 σελ.) δεν θα είχαν σοβαρό λόγο να προτιμήσουν την φωτοαντιγραφή από τη νόμιμη αγορά. Στην περίπτωση αρχικής κυκλοφορίας των λεξικών σε CD-ROM, το κόστος της πειρατικής αναπαραγωγής θα ήταν μόλις το 10-15% της λιανικής τιμής, οπότε θα υπήρχε ισχυρότατο πρόβλημα οπώλειας εσόδων για τον εκδότη. Τα έντυπα λεξικά αρκετά συχνά προτιμώνται από βιβλιοθήκες, έναντι των CD-ROM λόγω του ότι δεν απαιτούν Η/Υ για τη χρήση τους. Άρα είναι προφανές ότι τα λεξικά σε CD-ROM και τα έντυπα πολύγλωσσα λεξικά, είναι δύο διαφορετικά προϊόντα, τα οποία έχουν τη δική τους ξεχωριστή θέση στην αγορά μας. Γι' αυτούς τους λόγους, οι χρήστες του LexiGraf επέλεξαν να προηγηθεί η κυκλοφορία του υλικού τους σε έντυπη μορφή και στη συνέχεια σε ηλεκτρονική.

3. ΔΙΓΛΩΣΣΑ ΛΕΞΙΚΑ

Συνδυάζοντας ο Η/Υ τις διαθέσιμες γλώσσες ανά ζεύγη, μπορεί αυτομάτως να δημιουργήσει δίγλωσσα έντυπα λεξικά ορολογίας. Αν κατά την υλοποίηση ομαδοποιηθούν οι όροι σύμφωνα με τα γνωστικά αντικείμενα στα οποία ανήκουν, μπορούν να υλοποιηθούν έντυπα δίγλωσσα λεξικά συγκεκριμένων επιστημονικών κλάδων μέσα από μία ανομοιογενή συλλογή πολυγλωσσικής ορολογίας. (πχ ελληνο - γαλλικό λεξικό μαθηματικών, ελληνο - αγγλικό φυσικής κοκ). Αυτά τα είδη των λεξικών, λόγω της χαμηλής τους τιμής, της ευκολίας και της ταχύτητας υλοποίησης εμφανίζουν ιδιαίτερα θετικές δυνατότητες απορρόφησης από την αγορά μας.

4. ΗΛΕΚΤΡΟΝΙΚΕΣ ΕΚΔΟΣΕΙΣ

Καθώς η αγορά των ηλεκτρονικών εκδόσεων ωριμάζει σταδιακά και στην Ελλάδα - με παράλληλη εξέλιξη της τεχνολογίας προστασίας κατά της παράνομης αντιγραφής CD-ROM - δημιουργείται αυξανόμενο εμπορικό ενδιαφέρον για την κυκλοφορία CD-ROM πολύγλωσσων λεξικών ορολογίας, απευθυνόμενων στο γενικό επιστημονικό κοινό. Ανταποκρινόμενο στις εξελίξεις, το LexiGraf υποστηρίζει τη μεταφορά των πολύγλωσσων δεδομένων ορολογίας του σε οποιαδήποτε από τις μηχανές (software engines) δημιουργίας ηλεκτρονικών λεξικών που κυκλοφορούν στη διεθνή αγορά. Ταυτοχρόνως, τα ίδια αυτά λεξιλογικά δεδομένα μπορούν να αξιοποιηθούν για τον εμπλουτισμό γνωσιολογικών βάσεων δεδομένων σε εφαρμογές αυτόματης ή επιβηθηθούμενης μετάφρασης (translation memory software), δημιουργώντας πρόσθετα έσοδα για τους ιδιοκτήτες των δεδομένων ορολογίας.

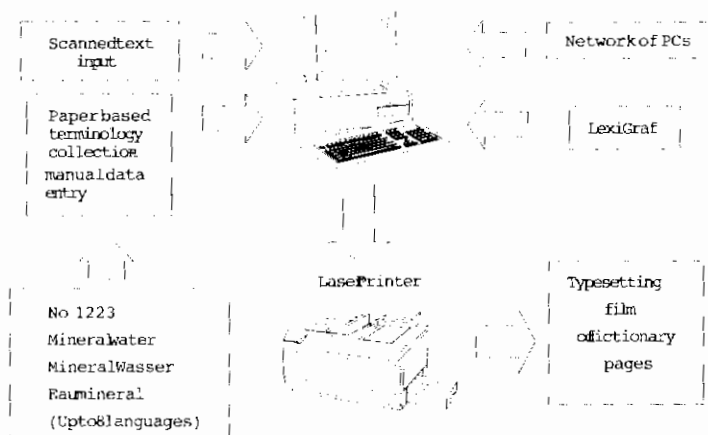
5. Η ΡΟΗ ΕΡΓΑΣΙΩΝ ΣΤΟ LEXIGRAF

Η ροή εργασιών στο LexiGraf χωρίζεται σε τρία διακριτά στάδια:

- Είσοδος πολυγλωσσικής ορολογίας στο σύστημα
- Επεξεργασία
- Έξοδος επεξεργασμένης ορολογίας προς περιφερειακές συσκευές ή προς άλλα πακέτα λογισμικού.

Η ορολογία εισέρχεται στο σύστημα με διάφορους τρόπους, ενδεχομένως χρησιμοποιώντας υφιστάμενες ηλεκτρονικές συλλογές πολύγλωσσας ορολογίας, ομάδες δακτυλογράφων (data entry operators), ή με εισαγωγή πληροφορίας χρησιμοποιώντας μηχανισμούς οπτικής αναγνώρισης χαρακτήρων OCR από ήδη υπάρχοντα έντυπα λεξικά. Η τελευταία αυτή δυνατότητα εξαρτάται καθοριστικά από την ικανότητα του εκάστοτε χρησιμοποιούμενου πακέτου OCR να επεξεργαστεί την

Αφού πραγματοποιηθεί η είσοδος λεξιλογικής πληροφορίας στο LexiGraf, από εκεί και πέρα στο σύστημα το ίδιο επεξεργάζεται, ταξινομεί και μορφοποιεί το υλικό προς πολύγλωσσο ή δίγλωσσο έντυπα λεξικά μέσα σε λίγα λεπτά. Το αποτέλεσμα της επεξεργασίας είναι αρχεία H/Y, κωδικοποιημένα κατά Xtag ή RTF, τα οποία περιέχουν σελιδοποιημένο το υλικό μας, αναγνώσιμο από λογισμικό Desktop Publishing, και στη συνέχεια άμεσα διαθέσιμο για φωτοστοιχειοθεσία σε ειδικές μηχανές φωτοστοιχειοθεσίας Postscript. (Xtag είναι τύπος αρχείων του πακέτου DTP QuarkXpress™, de-facto πρότυπο στο χώρο του DTP, RTF Rich Text Format είναι παρόμοια κωδικοποίηση της Microsoft™ που υποστηρίζεται από πολλά πακέτα λογισμικού, τόσο σε Windows™/PC όσο και σε Macintosh™).



Χάρη σε αυτήν την προσέγγιση μπορούμε να προσθέσουμε οποιοδήποτε στοιχείο μέσα στο κείμενο των λεξικών μας (πχ. εικόνες, σχεδιαγράμματα, τύπους οργανικής χημείας), να αλλάξουμε το στυλ σελιδοποίησης (layout) σε δευτερόλεπτα και να πραγματοποιήσουμε διορθώσεις ταχύτατα. Καθώς τα συστήματα Macintosh™ είναι το στάνταρντ στον εκδοτικό χώρο, τα δεδομένα μπορούν να μεταφερθούν από την πλατφόρμα PC προς την πλατφόρμα Mac με ευκολία χάρη στα πρότυπα Xtag/RTF.

6. ΕΡΓΑΛΕΙΑ ΑΝΑΠΤΥΞΗΣ ΤΟΥ LEXIGRAF

Η έκδοση 2.1 και 2.2 του LexiGraf αναπτύχθηκε με την Turbo Pascal for Windows + Paradox Engine 3.01 της Borland™ (16bit) υπό τα MS-WINDOWS. Οι λειτουργίες διαχείρισης βάσεων δεδομένων παρέχονται από το Paradox με την τεχνική file-sharing. Το σύστημα μπορεί να υποστηρίξει ως 20 χρήστες σε τοπικό δίκτυο. Υπό ανάπτυξη

βρίσκεται η έκδοση LexiGraf 3.0 (32bit) χρησιμοποιώντας την Delphi + Borland InterBase. Είναι μία εφαρμογή client/server, η οποία υποστηρίζει UNICODE και έχει αυξημένες επιδόσεις σε σχέση με την 16bit έκδοση. Θα συνεργάζεται με SGML import filters, και θα προσφέρει μία αυτοματοποιημένη 'μηχανή' δημιουργίας τόσο έντυπων, όσο και ηλεκτρονικών λεξικών.

7. ΛΕΙΤΟΥΡΓΙΚΗ ΠΡΟΣΕΓΓΙΣΗ ΣΤΟ LEXIGRAF

Το LexiGraf υποστηρίζει τρία βασικά επίπεδα πρόσβασης στο δίκτυο: 1. Data entry operator access-level 2. Lexicography administrator access-level 3. Support provision access-level. Κατά την εξέλιξη του συγκεκριμένου project κατέστη προφανές ότι ήταν εξαιρετικά δαπανηρό και δύσκολο να προσληφθούν δακτυλογράφοι (data-entry operators), οι οποίοι θα ήταν σε θέση να χειριστούν όλες τις γλώσσες που υποστήριζε το LexiGraf ταυτοχρόνως. Ας μην ξεχνάμε ότι η πηγαία μας πληροφορία βρισκόταν αποδελτιωμένη σε χειρόγραφες καρτέλες, άρα έπρεπε να βασιστούμε στην ανθρώπινη εργασία για να τροφοδοτήσουμε το σύστημα με ορολογία.

Γι' αυτό το λόγο το LexiGraf χωρίστηκε σε τρία διακριτά τμήματα (modules).

DATA ENTRY MODULE: Αυτόνομο κομμάτι του LexiGraf, το οποίο χειρίζονται οι δακτυλογράφοι-χειριστές (data-entry operators) του συστήματος. Μπορεί να λειτουργήσει αυτόνομα σε κάποιο απομακρυσμένο service bureau, off-line ή on-line σε τοπικό δίκτυο. Επιτρέπει στους δακτυλογράφους (data-entry operators) να δουλεύουν χωριστά την κάθε μία γλώσσα, ανεξάρτητα από τη συνολική εξέλιξη της δουλειάς των συναδέλφων τους. Αν δουλεύουμε το σύστημα off-line, τα λεξικογραφικά δεδομένα ενσωματώνονται με μεταφορές φακέλων/αρχείων σε κάποιο κεντρικό PC (remote lexicography server) σε τακτά χρονικά διαστήματα. Η προσέγγιση βοηθάει στη μείωση του κόστους εισαγωγής πληροφορίας στο σύστημα, διότι ο φόρτος εργασίας μπορεί να καταμεριστεί ισομερώς σε πολλούς operators, οι οποίοι μπορούν να δουλεύουν μία μόνο γλώσσα, σε διαφορετικούς χώρους και ώρες, χωρίς αυτό να δημιουργεί οποιοδήποτε πρόβλημα στην εξέλιξη του project.

TERMINOLOGY IMPORT/EXPORT: Αυτόνομο κομμάτι του LexiGraf, το οποίο αναλαμβάνει τη μεταφορά πληροφορίας από και προς τον 'έξω' κόσμο (πχ. είσοδος από ηλεκτρονικές λίστες ορολογίας, διασύνδεση με λογισμικό OCR, αντίγραφα ασφαλείας (back-up) των δεδομένων, εξαγωγή δεδομένων προς γεννήτριες εφαρμογές ηλεκτρονικών λεξικών κκκ)

ΠΙΝΑΚΑΣ ΚΑΡΤΕΑΔΩΝ			
Νο Κάρτας	Αρ. Γλωσσών	Αλφάριθμος	
Νο Κάρτας	Νο Κάρτας	Νο Κάρτας	Νο Κάρτας
Νο συνωνύμου	Νο συνωνύμου	Νο συνωνύμου	Νο συνωνύμου
Αγγλικό λήμμα	Γαλλικό λήμμα	Γερμαν. λήμμα	Ελλην. λήμμα
Δόκιμος όρος	Δόκιμος όρος	Δόκιμος όρος	Δόκιμος όρος
Αγγλικό πεδίο αφαβήτωσης	Γαλλικό πεδίο αφαβήτωσης	Γερμαν. πεδίο αφαβήτωσης	Ελληνικ. πεδίο αφαβήτωσης
Γνωστικό αντικ.	Γνωστικό αντικ.	Γνωστικό αντικ.	Γνωστικό αντικ.
Αλφάριθμος	Αλφάριθμος	Αλφάριθμος	Αλφάριθμος
Δείκτης κύριου συνωνύμου	Δείκτης κύριου συνωνύμου	Δείκτης κύριου συνωνύμου	Δείκτης κύριου συνωνύμου
Δικαιώματα δικτυακής πρόσβασης	Δικαιώματα δικτυακής πρόσβασης	Δικαιώματα δικτυακής πρόσβασης	Δικαιώματα δικτυακής πρόσβασης
ΠΙΝΑΚΑΣ ΑΓΓΛΙΚΩΝ ΛΗΜΜΑΤΩΝ	ΠΙΝΑΚΑΣ ΓΑΛΛΙΚΩΝ ΛΗΜΜΑΤΩΝ	ΠΙΝΑΚΑΣ ΓΕΡΜΑΝΙΚΩΝ ΛΗΜΜΑΤΩΝ	ΠΙΝΑΚΑΣ ΕΛΛΗΝΙΚΩΝ ΛΗΜΜΑΤΩΝ

ΣΧΕΣΙΑΚΗ ΠΛΗΡΗ ΔΕΔΟΜΕΝΩΝ
ΣΤΟ ΛΕΞΙΓΡΑΦ

LEXICOGRAPHY DTP MODULE: Κομμάτι του LexiGraf, το οποίο ολοκληρώνει την επεξεργασία DTP/σελιδοποίησης των πολύγλωσσων και δίγλωσσων λεξικών μας. Παράγει αυτομάτως λεξικογραφικά αρχεία σελιδοποίησης, κωδικοποιημένα κατά ΧΤΑΓ/RTF, τα οποία μπορούν να διαβαστούν από επαγγελματικά επιτραπέζια εκδοτικά συστήματα (desktop publishing software). Από εκεί δημιουργούνται τα φιλμ φωτοστοιχειοθεσίας, τα οποία χρησιμοποιούνται για την τυπογραφική αναπαραγωγή των λεξικών μας.

Μέσα στα λήμματα μας μπορούμε να συμπεριλάβουμε γραμματικούς προσδιορισμούς, χημικούς τύπους ή παρενθετικούς προσδιορισμούς. Όλη αυτή η πρόσθετη πληροφορία αναγνωρίζεται από το LexiGraf αυτομάτως και δεν επηρεάζει την ορθή ταξινόμηση των λημμάτων σε όλες τις γλώσσες. Επιπλέον, στη φάση του αυτοματοποιημένου DTP, το LexiGraf αυτομάτως αλλάζει τη μορφή (typefacing/ style) των συγκεκριμένων προσδιορισμών, σύμφωνα με τις απαιτήσεις του τμήματος DTP της εκδοτικής εταιρείας που αναλαμβάνει να κυκλοφορήσει τα λεξικά που προκύπτουν (παραμετροποίηση του DTP). Για την ταξινόμηση των λημμάτων, λόγω του μεταβλητού αριθμού γλωσσών που καλείται να χειριστεί το LexiGraf, δε χρησιμοποιήσαμε τις ρουτίνες ταξινόμησης του λειτουργικού συστήματος MS Windows™. Κάθε γλώσσα που υποστηρίζεται διαθέτει ένα εσωτερικό module κανόνων αφαβήτησης/ ταξινόμησης.

Εξετάζοντας το πηγαίο λεξικογραφικό υλικό που κληθήκαμε να επεξεργαστούμε, έγινε αμέσως προφανές ότι υπήρχαν μεμονωμένοι όροι σε κάποιες από τις γλώσσες, οι οποίοι θα μπορούσαν να χαρακτηριστούν ως αδόκιμοι-επεξηγηματικοί (κυρίως στην ελληνική γλώσσα). Ανάλογη πληροφορία έχει θέση σε ένα πολύγλωσσο λεξικό ορολογίας ως ερμηνεία, αλλά όχι σε ένα πχ. ελληνο-αγγλικό λεξικό (οι αδόκιμοι όροι δεν ευρετηριάζονται). Γι' αυτό το λόγο λήφθηκε ειδική μέριμνα ώστε οι αδόκιμοι ή οι επεξηγηματικοί όροι να υπόκεινται σε κατάλληλους χειρισμούς αυτομάτως από το LexiGraf. Παράλληλα, στην περίπτωση που θα διαπιστώσει κάποιο κενό στην απόδοση μίας ερμηνείας σε μία γλώσσα, το λογισμικό μας αυτομάτως αποκρύπτει τη συγκεκριμένη πολύγλωσση ορολογική οντότητα από ένα πολύγλωσσο λεξικό (δε δικαιολογείται η ύπαρξη κενών αποδόσεων σε ένα πολύγλωσσο λεξικό). Η ίδια αυτή πληροφορία θα αξιοποιηθεί κατά τη σύνταξη ενός δίγλωσσου λεξικού, εφόσον το ζεύγος των εμπλεκόμενων γλωσσών είναι συμπληρωμένο, έστω αν οι υπόλοιπες γλώσσες είναι κενές. Ταυτοχρόνως, το LexiGraf μας προειδοποιεί για την ύπαρξη διπλοκαταχωρήσεων στην ορολογία μας, λαμβάνοντας υπόψιν και το γνωστικό αντικείμενο των όρων.

8. ΔΙΑΔΙΚΑΣΙΑ ΔΗΜΙΟΥΡΓΙΑΣ ΕΝΤΥΠΩΝ ΛΕΞΙΚΩΝ

Η δημιουργία έντυπων λεξικών στο LexiGraf εμπεριέχει τα ακόλουθα στάδια:

- Εισαγωγή πολύγλωσσης ορολογίας στο σύστημα.
- Επεξεργασία και ταξινόμηση του υλικού από τους Η/Υ.
- Αποθήκευση των λεξικών (πολύγλωσσων και δίγλωσσων) στο σκληρό δίσκο του Η/Υ με μορφή proprietary μακροεντολών (L-Tags)
- Μετάφραση των L-Tags προς XTAGS/RTF και τροφοδοσία του υλικού σε επαγγελματικά συστήματα DTP (με πλήρη παραμετροποίηση).
- Εκτύπωση των λεξικών μας σε high-end typesetters.

Αφού τροφοδοτηθούν οι βάσεις δεδομένων του LexiGraf με πολυγλωσσική ορολογία, το λογισμικό μας αρρώνει τα περιεχόμενα της γλώσσας κλειδιού για να δημιουργήσει και να καταγράψει τους αλφαριθμούς/ αύξοντες αριθμούς που θα χρησιμοποιηθούν για τη σύνταξη των λεξικών-ευρετηρίων, τα οποία συμπληρώνουν το πολύγλωσσο. Αφού ολοκληρωθεί αυτή η εργασία, ο διαχειριστής λεξικογράφησης (dictionary administrator) μπορεί να επιλέξει αν θα δημιουργήσει κάποιο πολύγλωσσο ή δίγλωσσο λεξικό με βάση τη διατιθέμενη συλλογή ορολογίας. Το αποτέλεσμα της διεργασίας αυτής καταγράφεται στο σκληρό δίσκο του Η/Υ κωδικοποιημένο κατά L-tags. Η πληροφορία αυτή μπορεί να δρομολογηθεί σε κάποιον εκτυπωτή, ώστε να έχουμε σε πολύ σύντομο χρόνο ένα δοκίμιο του υλικού μας, πάνω στο οποίο μπορούμε να εντοπίσουμε τυχόν σφάλματα. Η διόρθωση/ αποκατάσταση των σφαλμάτων μπορεί να γίνει με εύκολο τρόπο από τους χειριστές του συστήματος.

Αν επιθυμούμε να έχουμε την τελική ποιότητα στην οποία θα φτάσει το υλικό στον εκδότη μας, το μόνο που έχουμε να κάνουμε είναι να περάσουμε τα κωδικοποιημένα κατά L-Tags λεξικά μας μέσα από το ειδικό φίλτρο Xtag/RTF του LexiGraf, ώστε να λάβουμε φακέλους αναγνώσιμους από επαγγελματικό λογισμικό DTP. Αν επιλέξουμε να οδηγηθούμε προς εφαρμογή λεξικού σε CD-ROM, θα χρησιμοποιήσουμε τα εργαλεία IMPORT/EXPORT του LexiGraf για να οδηγήσουμε τα δεδομένα μας σε κάποια από τις γεννήτριες εφαρμογές ηλεκτρονικών λεξικών που κυκλοφορούν στην αγορά.

Στην περίπτωση υλοποίησης δίγλωσσου λεξικού, θα επιλέξουμε το ζεύγος των γλωσσών που θέλουμε και στη συνέχεια θα κληθούμε να συμπληρώσουμε μία παραμετρική λίστα γνωστικών αντικειμένων, με βάση την οποία θα συνταχθεί το συγκεκριμένο λεξικό (λεξικό κλάδων). Η εξέλιξη της διαδικασίας ακολουθεί παρόμοια πορεία με αυτήν που περιγράψαμε.

9. ΧΡΗΣΤΙΚΟΤΗΤΑ ΤΟΥ LEXIGRAF

- Μείωση του χρόνου μετάβασης από την πρωτογενή συλλογή ορολογίας προς το τελικό έντυπο ή ηλεκτρονικό προϊόν.
- Ευελιξία στην υλοποίηση πολλών διαφορετικών προϊόντων που σχετίζονται με το χώρο της ορολογίας, ανάλογα με τις ανάγκες της αγοράς την κάθε δεδομένη στιγμή.
- Απαλοιφή των εξόδων DTP και αυτοματοποίηση της λεξικογραφικής επεξεργασίας.
- Μείωση του χρόνου και του κόστους υλοποίησης ενημερωμένων λεξικογραφικών εκδόσεων.
- Μπορεί να χρησιμοποιηθεί για λεξικογραφικά project που εξελίσσονται παράλληλα.
- Δεν απαιτεί εξειδίκευση στη χρήση του.

Γιάννης Χρ. Χατζόπουλος

Διπλ. Ηλεκτρολόγος Μηχανικός

Υψηλάντου 3 Σίνδος 574 00

τηλ. 031-797.532 / fax 031-798.432, κιν. 0977-517.317, Email yhat@egnatia.ee.auth.gr