

Θησαυροί και συστήματα ανάκτησης πληροφοριών σε στενή σχέση αλληλεπίδρασης

Χαρά Μπρίντεζη, Διονύσης Γιαννίμπας

Εισαγωγή

Από την πρώτη στιγμή της εμφάνισής τους, οι θησαυροί αξιοποιούνται στο πλαίσιο προηγμένων συστημάτων πληροφόρησης, οι περισσότεροι δε από αυτούς αναπτύσσονται και λειτουργούν ως λεξιλόγιο πρόσβασης σε μεγάλες βάσεις δεδομένων επιστημονικού και τεχνικού περιεχομένου.

Με τη διαρκή αύξηση των ηλεκτρονικών πληροφοριακών πηγών και τη συνεχή εξέλιξη των τεχνολογιών της πληροφορίας, οι εφαρμογές των θησαυρών διευρύνονται συνεχώς σε όλους τους χώρους όπου γίνεται διαχείριση πληροφορίας. Είναι χαρακτηριστικό ότι παρ' όλο που αρχικά δημιουργήθηκαν για χρήση κυρίως στα πλαίσια των βάσεων δεδομένων, έχουν αρχίσει πλέον σε όλο τον κόσμο να εισχωρούν και στους χώρους των βιβλιοθηκών, καθώς είναι φανερό ότι μπορούν να συνεισφέρουν στη λύση των «προβλημάτων» που τους έχει δημιουργήσει η νέα πληροφοριακή πραγματικότητα. Τα προβλήματα αυτά σχετίζονται με το γεγονός ότι οι συλλογές τους περιλαμβάνουν πλέον πέραν του εντύπου και άφθονο ηλεκτρονικό υλικό, το οποίο πρέπει να περιγραφεί με αποδοτικό τρόπο, συμβατό με το σύγχρονο ηλεκτρονικό περιβάλλον. Το περιβάλλον αυτό υπαγορεύει τη χρήση υστεροσυνδυασμένων εργαλείων, όπως είναι οι θησαυροί. Η διαπίστωση αυτή προκύπτει από τα ποσοστά ικανοποίησης των αναγκών των χρηστών από τα υστεροσυνδυασμένα συστήματα, τα οποία υπερτερούν συγκρινόμενα με τα ποσοστά που προέρχονται από τα παραδοσιακά προσυνδυασμένα συστήματα.

Ένα επιπλέον χαρακτηριστικό της εποχής μας είναι ότι ενώ παραδοσιακά ένας θησαυρός κάλυπτε συγκεκριμένο θεματικό πεδίο, οι σημερινές συνθήκες και οι ανάγκες των βιβλιοθηκών, απαιτούν πλέον συχνά τη χρήση τέτοιων λεξιλογίων για την οργάνωση περιεχομένου με ποικίλο εύρος και βάθος θεματικής κάλυψης.

Τα τελευταία χρόνια έχουν ωριμάσει και στην Ελλάδα οι συνθήκες για την αξιοποίηση τέτοιων ισχυρών εννοιολογικών εργαλείων στην επιστήμη της πληροφόρησης. Ο πολύπλευρος εκσυγχρονισμός που συντελείται σε βιβλιοθήκες κάθε τύπου (ακαδημαϊκές και ερευνητικές, λαϊκές και σχολικές) καθιστά ακόμη πιο επιτακτική την επίλυση χρόνιων προβλημάτων που σχετίζονται, μεταξύ άλλων, με την έλλειψη υποδομής στον τομέα της θεματικής οργάνωσης περιεχομένου.

Αισιοδοξία δημιουργεί η ευρύτερη κινητικότητα που παρατηρείται στο συγκεκριμένο πεδίο και έχει εκδηλωθεί με σημαντικά εγχειρήματα που βρίσκονται είτε σε φάση σχεδιασμού είτε σε πλήρη εξέλιξη, όπως ο διαρκώς αναπτυσσόμενος Κατάλογος Θεματικών Επικεφαλίδων της Εθνικής Βιβλιοθήκης, ο Κατάλογος Καθιερωμένων Τύπων του Συλλογικού Καταλόγου των Δημόσιων Βιβλιοθηκών, ο Κατάλογος Ονομάτων και Θεμάτων της Συλλογής των Σχολικών Βιβλιοθηκών, οι όροι που έχει επεξεργαστεί η Βιβλιοθήκη του Τεχνικού Επιμελητηρίου με βάση υλικό από ειδικές επιτροπές ορολογίας, η θεματική οργάνωση των βάσεων δεδομένων του Εθνικού Κέντρου Κοινωνικών Ερευνών και η προσπάθεια της Κοινοπραξίας Ακαδημαϊκών Βιβλιοθηκών για τη δημιουργία αρχείου θεματικών όρων που θα υποστηρίξει τον συλλογικό τους κατάλογο. Γενικότερα, πάντως, εύκολα διαπιστώνει κανείς ότι η χρήση θησαυρών στις ελληνικές βιβλιοθήκες είναι πολύ περιορισμένη. Ως εκ τούτου, το ΕΚΤ προχώρησε στην ανάπτυξη του Θησαυρού Ελληνικών Όρων, του πρώτου θησαυρού γενικού περιεχομένου στην ελληνική γλώσσα (με απόδοση των όρων στην αγγλική).

Ο Θησαυρός εκπονήθηκε στο πλαίσιο του έργου «Υποστήριξη της Εκπαιδευτικής Πράξης από Κινητές Βιβλιοθήκες» (ΕΠΕΑΕΚ II), με σκοπό την υποστήριξη της θεματικής πρόσβασης στο υλικό των Δημόσιων και Δημοτικών Βιβλιοθηκών και των αντίστοιχων Κέντρων Πληροφόρησης, αλλά και γενικότερα σε κάθε είδους περιεχόμενο. Ήταν ένα βήμα που το υπαγόρευε η νέα πραγματικότητα που διαμορφώνεται στον χώρο της πληροφόρησης, δηλαδή, η ανάπτυξη ψηφιακών συλλογών, ο πολλαπλασιασμός και η διασύνδεση των βάσεων δεδομένων, η εφαρμογή διαδικασιών αυτοματοποίησης, η λειτουργία κέντρων πληροφόρησης, η παροχή νέων υπηρεσιών, η κυριαρχία των OPAC.

Η πρώτη αυτή έκδοση διατίθεται δωρεάν στις βιβλιοθήκες με τη μορφή CD-ROM, ως ηλεκτρονική εφαρμογή που αναπτύχθηκε σε συνεργασία με την «iLang Γλωσσική Καινοτομία», εταιρεία-τεχνοβλαστό του Ινστιτούτου Επεξεργασίας του Λόγου. Έχει επίσης προγραμματιστεί η δημιουργία αντίστοιχης εφαρμογής στον διαδικτυακό κόμβο του ΕΚΤ, με την οποία οι χρήστες θα έχουν πρόσβαση στην εκάστοτε ενημερωμένη και εμπλουτισμένη έκδοση του Θησαυρού.

Στην τρέχουσα έκδοση περιλαμβάνονται οι βασικές έννοιες από ένα ευρύ φάσμα γνωστικών πεδίων, με μεγαλύτερη έμφαση σε θέματα ελληνικού και ευρωπαϊκού ενδιαφέροντος, χωρίς πάντως να επιδιώκεται ιδιαίτερη κάλυψη σε βάθος. Αντίθετα, δόθηκε ιδιαίτερη βαρύτητα στην όσο το δυνατόν σωστότερη δόμηση των σχέσεων μεταξύ των εννοιών, αφού από νωρίς κρίθηκε σκόπιμο ότι, για τα πρώτα του βήματα, ο Θησαυρός προέχει να σταθεί καλά στα πόδια του, παρά να διανύσει μεγάλη απόσταση. Μάλιστα, η τεκμηρίωση μιας «καλής πρακτικής» αποτέλεσε προτεραιότητα σε κάθε πτυχή του έργου, καθώς θεωρήθηκε ότι θα είναι το σημαντικότερο μέρος της υποδομής που αφήνει πίσω της η πρώτη έκδοση. Συνεπώς, ο Θησαυρός Ελληνικών Όρων, ως έχει, δεν φιλοδοξεί να καλύψει τις ανάγκες των ακαδημαϊκών ή ερευνητικών βιβλιοθηκών.

Εντούτοις, αποτελεί έναν πυρήνα βασικών εννοιών πάνω στον οποίο μπορεί να βασιστεί η δόμηση συγκεκριμένων θεματικών περιοχών, αξιοποιώντας μάλιστα την υπάρχουσα υποδομή από πλευράς λογισμικού, τεχνολογίας και μεθοδολογίας στη διαχείριση του υλικού. Έτσι, μέσα από συνεργασίες με τους ενδιαφερόμενους φορείς, μπορεί να δρομολογηθεί η περαιτέρω ανάπτυξη και εξειδίκευση του Θησαυρού σε επιλεγμένους τομείς, ώστε σε βάθος χρόνου να είναι σε θέση να εξυπηρετεί πλήρως όσο το δυνατόν περισσότερους από τους πληροφοριακούς οργανισμούς της χώρας.

Η εμπειρία που αποκομίσθηκε κατά τη διαδικασία δημιουργίας του εν λόγω εργαλείου είναι μεγάλη. Επίσης, σε όλα τα στάδια ιδιαίτερη βαρύτητα δόθηκε στην παρακολούθηση των διεθνών εξελίξεων για την εύρεση καλών πρακτικών δόμησης, τεχνολογικής υποστήριξης και αξιοποίησης των θησαυρών. Παράλληλα μελετήθηκαν οι υπάρχουσες οδηγίες συγκρότησης και ανάπτυξης θησαυρών καθώς και οι διαδικασίες αναπροσαρμογής τους σύμφωνα με τις επιταγές του σύγχρονου πληροφοριακού περιβάλλοντος, που θέλει πλέον τους θησαυρούς να είναι ηλεκτρονικά εργαλεία και μάλιστα σε θέση από τη μια να εκμεταλλεύονται και από την άλλη να τροφοδοτούν την τεχνολογία. Σήμερα, όλες οι πτυχές της διαχείρισης των θησαυρών (συγκρότηση, συντήρηση, εμπλουτισμός), ακόμα και η διάθεσή τους προς χρήση, πρέπει να στηρίζονται σε ειδικό λογισμικό και αντίστροφα, πολλά συστήματα ανάκτησης πληροφοριών (βάσεις δεδομένων, μηχανές αναζήτησης, συστήματα διαχείρισης περιεχομένου) στηρίζονται σε θησαυρούς και λειτουργούν με τη βοήθειά τους.

Από την αρχή έγινε αντιληπτή η στενή σχέση αλληλεπίδρασης μεταξύ Συστημάτων Ανάκτησης Πληροφοριών (ΣΑΠ) και θησαυρών. Εντοπίστηκαν χαρακτηριστικά των θησαυρών που εξυπηρετούνται άριστα από τεχνικές που έχουν υιοθετήσει τα ΣΑΠ, όπως και συγκεκριμένες λειτουργίες των ΣΑΠ που διαμορφώνουν, άμεσα ή έμμεσα, συγκεκριμένες μεθόδους δόμησης και παρουσίασης ενός θησαυρού.

Στην ανακοίνωση αυτή γίνεται μια προσπάθεια παρουσίασης των πιο σημαντικών στοιχείων μέσω των οποίων ξεπροβάλλει η προαναφερθείσα σχέση αλληλεπίδρασης. Πρέπει να σημειωθεί

εδώ, ότι στους χρήστες συμπεριλαμβάνονται και οι ευρετηριαστές, για τους οποίους γίνεται ιδιαίτερη μνεία στα σημεία που τους αφορούν περισσότερο.

Επίσης, στις περισσότερες περιπτώσεις τίθεται ως προϋπόθεση τα δεδομένα να είναι ευρετηριασμένα με έναν θησαυρό. Ωστόσο, σε κάποια σημεία οι τεχνικές που παρουσιάζονται μπορούν να εφαρμοστούν και σε συστήματα που χρησιμοποιούν ένα θησαυρό μόνο για αναζητήσεις με βάση τη φυσική γλώσσα.

Αξίζει επίσης να αναφερθεί ότι τα παραδείγματα που αναφέρονται μπορούν να έχουν εφαρμογή τόσο σε μονόγλωσσους όσο και σε πολύγλωσσους θησαυρούς.

Η ανακοίνωση χωρίζεται σε τρεις βασικές ενότητες. Η πρώτη αφορά στους καταλληλότερους τρόπους παρουσίασης ενός ηλεκτρονικού θησαυρού, η δεύτερη στις ενδεδειγμένες τεχνικές πρόσβασης (πλοήγηση, αναζήτηση) και η τρίτη στον τομέα της διαλειτουργικότητας.

1. Παρουσίαση

Δεν είναι τυχαίο ότι στην ηλεκτρονική εποχή μιλάμε πλέον για διεπαφή ή περιβάλλον χρήσης. Η σύγχρονη αντίληψη για την επαφή του χρήστη με το προϊόν ή την υπηρεσία που τον ενδιαφέρει επιβάλλει ενιαίο και ολοκληρωμένο τρόπο αλληλεπίδρασης μέσα από ένα οικείο, λειτουργικό και ευέλικτο περιβάλλον.

Η δομή των θησαυρών, με τις πλούσιες «κάθετες» και «οριζόντιες» διασυνδέσεις μεταξύ των όρων, συνιστούσε ανέκαθεν μια πρόκληση για τη σχεδίαση της παρουσίασης του λεξιλογίου. Οι κλασικές λύσεις στις έντυπες εκδόσεις περιλαμβάνουν κυρίως την *αλφαβητική* και την *ιεραρχική* (ή *συστηματική*) παρουσίαση. Η πρώτη αποτελεί την κύρια πρόσβαση στο λεξιλόγιο, και συνήθως συμπεριλαμβάνει την παρουσίαση του συνόλου της πληροφορίας για κάθε όρο με τη μορφή «δελτίου» (εμφάνιση δηλαδή ολόκληρης της εγγραφής του όρου), ενώ η δεύτερη παρέχει εποπτεία των «κάθετων» σχέσεων κάθε έννοιας (η οποία αντιπροσωπεύεται από τον αντίστοιχο όρο) αναδεικνύοντας έτσι τη θέση της στο σύστημα των εννοιών.

Για αποτελεσματικότερη πρόσβαση στο λεξιλόγιο παρέχονται συνήθως, πέρα από την αλφαβητική παρουσίαση, ειδικά ευρετήρια, τα γνωστά KWIC και KWOC, που χρησιμοποιούνται για τον εντοπισμό περιφραστικών όρων με βάση και τις υπόλοιπες λέξεις εκτός από την αρχική.

Όπως επισημαίνεται και στα σχετικά πρότυπα [7], ένας έντυπος θησαυρός χρειάζεται τόσο τις δύο βασικές μορφές παρουσίασης (αλφαβητική και ιεραρχική) όσο και τα ειδικά αυτά ευρετήρια. Αντίθετα σε έναν ηλεκτρονικό θησαυρό απαραίτητη είναι μόνο η ιεραρχική (πλαισιωμένη από την παρουσίαση του «δελτίου» όρου), ενώ η χρησιμότητα των ευρετηρίων, συμπεριλαμβανομένου και του αλφαβητικού, εξαρτάται από το πόσο προηγμένος είναι ο μηχανισμός αναζήτησης που παρέχει το περιβάλλον χρήσης. (βλ. ενότητα 2: *Πρόσβαση*)

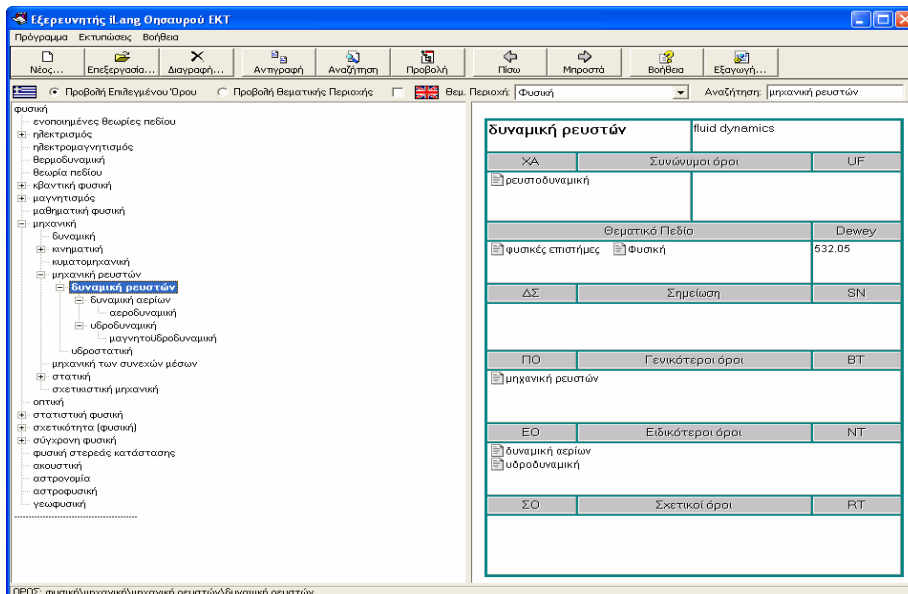
Ευνόητο είναι ότι οι ηλεκτρονικοί θησαυροί οφείλουν να αξιοποιούν τις νέες τεχνολογίες και να αποφεύγουν να μιμούνται άκριτα τον τρόπο παρουσίασης των έντυπων θησαυρών. Η εξοικείωση των χρηστών ηλεκτρονικών υπολογιστών και κυρίως του Διαδικτύου με περιβάλλοντα αναζήτησης όπως αυτά του «Windows Explorer», των μηχανών αναζήτησης ή των θεματικών πυλών έχει δώσει τον πρώτο λόγο στα γραφικά (Graphical User Interfaces), στο υπερκείμενο και στην πλοήγηση (browsing).

Η επιτυχία ενός θησαυρού εξαρτάται άμεσα από την ικανότητά του να «αντιγράφει» τον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου. Στο μυαλό του ανθρώπου αποθηκεύεται ένας τεράστιος και πολύπλοκος ιστός αλληλεξαρτήσεων μέσα στον οποίο καμία λέξη δεν είναι «απομονωμένο νησί». Κάθε σκέψη, λέξη ή εικόνα συνδέεται συνειρμικά με άλλες σχετικές λέξεις και έννοιες μέσω λεπτών σχέσεων. Τα ηλεκτρονικά μέσα έχουν τη δυνατότητα εκμετάλλευσης των σημασιολογικών δικτύων των θησαυρών που διευκολύνουν τις τάσεις *συνειρμών* (associations) του κάθε χρήστη. Επομένως, ο ιδανικότερος τρόπος παρουσίασης ενός όρου που επιλέχτηκε, είναι

η εμφάνισή του μέσα στο γενικότερο εννοιολογικό περιβάλλον του στον Θησαυρό. Παρακάτω, αναφέρονται οι τρόποι με τους οποίους αυτό μπορεί να επιτευχθεί.

Η ιεραρχική παρουσίαση –ιδιαίτερα στην πιο κλασική της έκφραση με τη χρήση εσοχών για την υποδήλωση της υπαγωγής της μίας έννοιας στην άλλη, γνωστή και ως δενδρική δομή– συμπεκνώνει με τον πλέον παραστατικό και άμεσο τρόπο τη «ραχοκοκαλιά» του συστήματος των εννοιών. Στην πλειοψηφία τους οι ερευνητές παραδέχονται ότι για την καλύτερη κατανόηση ενός όρου, στην αρχική εμφάνισή του «δέντρου» του, χρειάζεται να είναι ορατοί όλοι οι γενικότεροι όροι έως τον όρο κορυφής και τουλάχιστον δύο επίπεδα ιεραρχίας κάτω από αυτόν. Το ίδιο προτείνει και το Getty Institute για την παρουσίαση των όρων του «AAT Thesaurus».

Στην Εικόνα 1 φαίνεται η υλοποίηση της παραπάνω παραδοχής στον Θησαυρό Ελληνικών Όρων, μέσω μιας διεπαφής τύπου «Windows Explorer», που προσφέρει επιπλέον ένα **γραφικό** περιβάλλον **οικείο** σε όλους, και γι' αυτό κατάλληλο και για τους λιγότερο πεπειραμένους. Στο αριστερό μέρος της οθόνης εμφανίζεται συνοπτικά το «δέντρο» του επιλεγμένου όρου, από τον όρο κορυφής (Top Term) μέχρι το πρώτο επίπεδο ειδικότερων όρων. Μόνον το «κλαδί» όπου ανήκει ο επιλεγμένος όρος παρουσιάζεται σε ανάπτυξη.

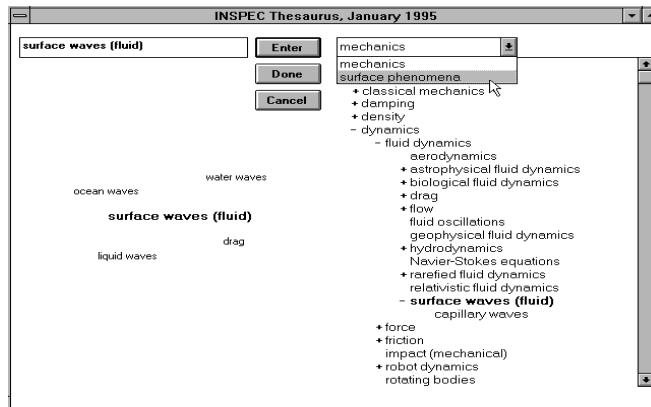


Εικόνα 1

Κατόπιν, ο χρήστης μπορεί να εκδιπλώσει ή να συρρικνώσει όποιο «κλαδί» του «δέντρου» επιθυμεί και σε όποιον βαθμό επιθυμεί, επιλέγοντας αντίστοιχα τα σύμβολα «+» [συν] ή «-» [πλην] που προηγούνται των όρων (όταν αυτά απουσιάζουν σημαίνει ότι απλώς ο όρος δεν έχει ειδικότερους όρους). Παράλληλα με την ιεραρχική παρουσίαση, εμφανίζεται και το «δελτίο» του εκάστοτε επιλεγμένου όρου (βλ. δεξιό μέρος της οθόνης στην Εικόνα 1), όπου περιλαμβάνεται το σύνολο της σχετικής πληροφορίας (π.χ. Συνώνυμοι Όροι, Διευκρινιστικές Σημειώσεις κ.λπ.). Έτσι, με την επιλογή ενός όρου ο χρήστης έχει μπροστά του όλες τις διασυνδέσεις του και κάθε άλλη χρήσιμη πληροφορία. Η εμφάνιση αυτή μπορεί να δώσει μια καλή εικόνα του περιβάλλοντος του όρου.

Σε περίπτωση πολυιεραρχίας, όταν δηλαδή ένας όρος έχει περισσότερους από έναν Γενικότερους όρους (BT) ή όρους κορυφής (TT), υπάρχει επίσης η παραδοχή ότι πρέπει να εμφανίζονται όλα τα «δέντρα» στα οποία ανήκει ο όρος. Μια καλή λύση προτείνεται από την

προσπάθεια που έγινε στα πλαίσια του University of Illinois Digital Library Initiative (Εικόνα 2). Όταν ένας όρος έχει πάνω από έναν ΒΤ, οι οποίοι ανήκουν στην ίδια ιεραρχία, τότε αυτοί εμφανίζονται στην ίδια οθόνη. Όταν ένας όρος έχει πάνω από έναν ΤΤ, όταν δηλαδή εμφανίζεται κάτω από διαφορετικές ιεραρχίες, το λογισμικό διαπαφής τοποθετεί τους ΤΤ σε ένα «pull-down box» στην κορυφή της ιεραρχικής εμφάνισης, επιτρέποντας στον χρήστη να επιλέξει τον ΤΤ για τον οποίον θα ήθελε να δει την ιεραρχία.



Εικόνα 2

Η παρουσίαση του πλέγματος των σχέσεων μεταξύ των όρων υπηρετείται κατά τον καλύτερο τρόπο με την τεχνική του υπερκειμένου. Το υπερκείμενο είναι κατάλληλο για καταστάσεις όπου υπάρχει μεγάλη ποσότητα πληροφορίας οργανωμένη σε τμήματα που σχετίζονται μεταξύ τους και ο χρήστης χρειάζεται μόνον ένα μικρό τμήμα της πληροφορίας κάθε φορά. Επίσης έχει την ικανότητα να παρέχει μη γραμμική πρόσβαση στην πληροφορία, πράγμα που αναδεικνύει και καθιστά περισσότερο κατανοητή στους χρήστες την επίσης μη γραμμική δομή του θησαυρού [15].

Με το υπερκείμενο βελτιώνεται θεαματικά η πολυπλοκότητα ευελιξία: ο χρήστης μπορεί να κινηθεί όπου και όπως θέλει. Έχει τη δυνατότητα να επιλέγει διαδοχικά όρους, **να πλοηγείται** δηλαδή ανάμεσα στους όρους ακολουθώντας τις διασυνδέσεις τους, είτε για εμφάνιση της αντίστοιχης πληροφορίας, είτε για χρήση τους περαιτέρω σε μια βιβλιογραφική βάση (χωρίς να απαιτείται επαναπληκτρολόγησή τους).

Η διεπαφή του Θησαυρού Ελληνικών Όρων (Εικόνα 1) είναι υπερκειμενική. Η πλοήγηση στον Θησαυρό γίνεται με απλό κλικ σε οποιονδήποτε περιγραφέα εμφανίζεται στην οθόνη. Το λογισμικό της διεπαφής αλλάζει αμέσως την εμφάνιση του Θησαυρού για να προβάλει την πληροφορία που αφορά τον εκάστοτε επιλεγμένο όρο.

Είναι φανερό ότι όσο πιο εξελιγμένος είναι ο τρόπος παρουσίασης του λεξιλογίου και οι δυνατότητες πρόσβασης του χρήστη σε αυτό, τόσο ανεβαίνει ο πήχυς και για τη διεργασία συγκρότησης του θησαυρού: ιδιαίτερη βαρύτητα αποκτούν η σωστή δόμηση των ιεραρχιών, ο πλούτος στις οριζόντιες διασυνδέσεις («σχέσεις συσχέτισης») μεταξύ των όρων και ο περιορισμός του αριθμού των «ορφανών» (χωρίς γενικότερο) όρων.

2. Πρόσβαση

Ένα σωστά σχεδιασμένο σύστημα φροντίζει για την ομαλή πρόσβαση του χρήστη στην πληροφορία που του παρέχει, τον προσανατολίζει και τον εισάγει στη «λογική» του. Έρευνες έχουν δείξει ότι όταν ο χρήστης γνωρίζει επακριβώς τι θέλει, οι αναζητήσεις με λέξεις-κλειδιά

προτιμώνται από την πλοήγηση σε ιεραρχίες όρων. Εάν, όμως, δεν έχει διαμορφωμένο εκ των προτέρων το ερώτημά του, τότε η μέθοδος της πλοήγησης προτιμάται έναντι της αναζήτησης με λέξεις-κλειδιά. Επιπλέον, η πλοήγηση είναι ένας τρόπος πιο χαλαρής γνωριμίας του χρήστη με τον εννοιολογικό χώρο του θησαυρού.

Είναι φανερό ότι οι δύο αυτοί βασικοί τρόποι θεματικής πρόσβασης στην πληροφορία, δηλαδή η πλοήγηση και η αναζήτηση με λέξεις-κλειδιά, συμπληρώνουν ο ένας τον άλλο, και ως εκ τούτου κρίνεται σημαντικό να παρέχονται παράλληλα στον χρήστη εξ *αρχής*, ώστε να μπορεί να επιλέξει ανάμεσα στους δύο με βάση τις ανάγκες του, ή να μεταπηδήσει ανά πάσα στιγμή από τον έναν στον άλλον.

2.1. Πλοήγηση

Είδαμε ήδη με ποιον τρόπο η δενδρική παρουσίαση, σε συνδυασμό με την τεχνική του υπερκειμένου, ξεδιπλώνει μπροστά στα μάτια του χρήστη τη δομή ενός θησαυρού, προσφέροντάς του τη δυνατότητα της πλοήγησης και καθοδηγώντας τον στην επιλογή του κατάλληλου όρου για την εκτέλεση μιας αναζήτησης σε ένα σύστημα ανάκτησης πληροφοριών. Αυτό που υποστηρίζεται στην ενότητα αυτήν είναι ότι το σύστημα ανάκτησης πρέπει να έχει με τέτοιο τρόπο διαμορφωμένη την πρώτη οθόνη εμφάνισης του θησαυρού, που να παρέχει τη δυνατότητα πλοήγησης από την αρχή. Τα πρότυπα δόμησης θησαυρών βοηθούν πολύ σε μια εύκολα κατανοητή παρουσίαση του τμήματος αυτού της Πλοήγησης. Μάλιστα, ο χρήστης διευκολύνεται περισσότερο στην κατανόηση του εννοιολογικού περιβάλλοντος εάν η δομή του θησαυρού ακολουθεί τη φασετική ανάλυση.

Η φασετική ανάλυση, που παρουσιάστηκε για πρώτη φορά τη δεκαετία του 1930 από τον Ινδό μαθηματικό και τεκμηριωτή S. Ranganathan, περιλαμβάνει *ανάλυση* των σύνθετων θεμάτων σε ομάδες απλούστερων εννοιών. Οι ομάδες αυτές των όρων είναι γνωστές ως φασέτες ή απόψεις. Κατά τη θεματική περιγραφή γίνεται, αν χρειάζεται, *σύνθεση* με συνδυασμούς των απλών εννοιών. Για τον λόγο αυτόν τα φασετικά ταξινομικά σχήματα αναφέρονται επίσης και ως αναλυτικο-συνθετικά, σε αντίθεση με τα κλασικά, απαριθμητικά (enumerative) συστήματα.

Κατά τη δόμηση των θησαυρών, η ανάλυση αυτή χρησιμοποιείται για την σωστότερη δήλωση των ιεραρχικών σχέσεων, αλλά και των σχέσεων συσχέτισης (RT), καθώς ένα βασικό κριτήριο είναι αν οι υπό επεξεργασία έννοιες ανήκουν σε βασικές κατηγορίες εννοιών, που έχουν χαρακτηριστικά φασέτας (π.χ. «αντικείμενα», «ενέργειες»).

Όταν η οργάνωση του θησαυρού ακολουθεί τη φασετική ανάλυση, υπάρχουν δύο περιπτώσεις. Η πρώτη είναι ο διαχωρισμός των όρων εξ αρχής σε φασέτες (σε κάθε φασέτα παρουσιάζεται μία και μοναδική ιεραρχία όρων), οπότε η πρώτη οθόνη παρουσίασης θα πρέπει να εμφανίζει τις φασέτες (συγκεκριμένα, τις ετικέτες δεσμού που περιέχουν τα ονόματα των φασετών) ή και τις ιεραρχίες.¹ Η δεύτερη, είναι η ομαδοποίηση των όρων σε θεματικά πεδία πριν εφαρμοστεί η φασετική ανάλυση, οπότε η πρώτη οθόνη παρουσίασης θα πρέπει να εμφανίζει τα θεματικά πεδία [7].

Με την βοήθεια της πλοήγησης, γενικά, και σε συνδυασμό με τη χρήση τελεστών Μπουλ, ο χρήστης έχει τη δυνατότητα να διαμορφώσει συνδυαστικά ερωτήματα. Μπορεί να επιλέγει διάφορους όρους από διαφορετικά δένδρα (διαφορετικές φασέτες). Σε αυτή την περίπτωση, για τον συνδυασμό των όρων χρησιμοποιείται κυρίως ο τελεστής Μπουλ «AND». Θα μπορούσε βέβαια να επιλέγει για το ερώτημά του και όρους από τους Ειδικότερους των ήδη επιλεγμένων όρων. Σε αυτή την περίπτωση η αναζήτηση γίνεται κυρίως με τη χρήση του τελεστή «OR».

¹ Για παράδειγμα τέτοιου είδους εμφάνισης, βλ. στην ιστοσελίδα του AAT Thesaurus την επιλογή “Browse the AAT Hierarchies” (http://www.getty.edu/research/conducting_research/vocabularies/aat/)

Οι πληροφορίες σχετικά με τη χρήση των όρων σε μια βάση δεδομένων είναι πολύ σημαντικές. Κατά την πλοήγηση στο λεξιλόγιο, ο χρήστης θέλει να βλέπει εάν και πόσο συχνά έχει χρησιμοποιηθεί ένας όρος κατά την ευρετηρίαση της συλλογής. Για τον σκοπό αυτό, καλό είναι δίπλα σε κάθε όρο να αναγράφεται ο αριθμός των εγγραφών τεκμηρίων στις οποίες ο συγκεκριμένος όρος έχει χρησιμοποιηθεί, έστω και αν ο αριθμός αυτός είναι το 0 (μηδέν). Έτσι ο χρήστης μπορεί να συμπεράνει, λόγου χάρη, ότι ο όρος είναι πολύ εξειδικευμένος (όταν δεν χρησιμοποιείται συχνά), ή ότι η συλλογή δεν έχει πλούσιο υλικό ή και κανένα τεκμήριο πάνω στο θέμα που τον ενδιαφέρει.

Πρόσθετο πλεονέκτημα αποτελεί η δυνατότητα υπολογισμού εγγραφών (επιτυχιών) που περιέχουν συνδυασμούς επιλεγμένων όρων, κατά τη διαμόρφωση του ερωτήματος. Όταν δηλαδή ο χρήστης επιλέγει έναν δεύτερο όρο για συνδυασμό με τον πρώτο, το σύστημα «ενημερώνει» τον αριθμό των επιτυχιών με βάση τα τεκμήρια όπου έχουν χρησιμοποιηθεί και οι δύο όροι. Έτσι ο χρήστης αποφεύγει την εκτέλεση μιας αναζήτησης με μηδενικό αποτέλεσμα, και μπορεί να συνεχίσει κάνοντας μια νέα επιλογή σε υψηλότερο επίπεδο των φασετών ή όρων, ή ακόμη και αφήνοντας τελείως τη συγκεκριμένη φασέτα έξω από το ερώτημα.

2.2. Αναζήτηση

Εναλλακτικά, ο χρήστης μπορεί να εισέλθει στο λεξιλόγιο με μια αναζήτηση, διαμορφώνοντας δηλαδή ένα ερώτημα με μια λέξη ή φράση. Ιδιαίτερη προσοχή χρειάζεται το γεγονός ότι εάν η αρχική αυτή αναζήτηση αποβεί άκαρπη, είναι πιθανόν ο χρήστης να απογοητευθεί και να εγκαταλείψει την προσπάθεια. Για να μη συμβεί κάτι τέτοιο, το λογισμικό δεν θα πρέπει να προσπαθεί να ταιριάζει με ακρίβεια τη λέξη ή φράση που εισάγει ο χρήστης μόνο με έναν όρο (περιγραφέα) στη βάση, αλλά, είτε απευθείας είτε μέσω των παραπομπών, με κάθε όρο που μπορεί να περιγράψει σχετικό υλικό. Γενικά, κάθε λειτουργία που επιτρέπει στον χρήστη να εκφράσει το ερώτημά του με δικές του λέξεις, και μάλιστα που επιτρέπει τη διαμόρφωση ερωτήματος σε φυσική γλώσσα, είναι ένα πλεονέκτημα που όμως συνεπάγεται περισσότερο πολύπλοκη επεξεργασία ερωτήματος.

Ας δούμε πώς ο μηχανισμός αναζήτησης που έχει χρησιμοποιηθεί στον Θησαυρό Ελληνικών Όρων αξιοποιεί τη γλωσσική τεχνολογία. Ο συγκεκριμένος μηχανισμός βασίζεται σε τεχνικές συγχώνευσης λέξεων (conflation). Με τον όρο συγχώνευση εννοούμε τη διαδικασία αναγνώρισης και συγκέντρωσης λέξεων οι οποίες είναι πιθανό να αποτελούν μορφολογικές παραλλαγές του ίδιου λεκτικού θέματος. Ειδικά στον Θησαυρό Ελληνικών Όρων έχει υιοθετηθεί η τεχνική της αφαίρεσης της κατάληξης (stemming). Η μέθοδος αυτή ενδείκνυται ιδιαίτερα για ένα σύστημα που επεξεργάζεται δεδομένα στην ελληνική γλώσσα, η οποία είναι εξαιρετικά πλούσια μορφολογικά. Το μεγάλο της πλεονέκτημα έναντι άλλων τεχνικών που εξαρτώνται άμεσα από τη χρήση μορφολογικών λεξικών, είναι η πληρέστερη κάλυψη των περισσότερο εξειδικευμένων όρων που δεν περιλαμβάνονται στα λεξικά αυτά.

Πιο συγκεκριμένα, αρχικά εντοπίζεται και αφαιρείται η κατάληξη των λέξεων που απαρτίζουν έναν όρο και στη συνέχεια «κανονικοποιείται» η μορφή των λέξεων (απαλοιφή τόνου, μετατροπή σε κεφαλαία). Αυτό σημαίνει ότι οι μορφολογικά συγγενικές λέξεις αναπαριστώνται και αποθηκεύονται με την ίδια μορφή, βάσει της οποίας και ανακτώνται μετέπειτα. Δηλαδή, ένας όρος ανακτάται όταν έστω μία από τις λέξεις που τον απαρτίζουν χρησιμοποιηθεί σε οποιαδήποτε μορφή της στο ερώτημα. Σημειώνεται ότι επιπλέον ο μηχανισμός εντοπίζει και όρους που «περιέχουν» τη λέξη του ερωτήματος (π.χ. ερώτημα: μαγνητισμός > αποτέλεσμα: ηλεκτρομαγνητισμός). Έτσι μειώνονται αισθητά οι πιθανότητες να μην εντοπιστεί ένας «συναφής» με το ερώτημα όρος, χωρίς μάλιστα ο χρήστης να χρειάζεται να καταφεύγει σε ειδική επεξεργασία της διατύπωσης του ερωτήματος (π.χ. με «αποκοπή» [truncation]). Η σειρά εμφάνισης των αποτελεσμάτων είναι από το πιο σχετικό (όρος με μεγαλύτερη ομοιότητα ή με περισσότερες κοινές

λέξεις με το ερώτημα) στο λιγότερο σχετικό (αντικαθιστώντας ουσιαστικά και την επιλογή για «ακριβή ανάκληση» [exact match]).

Είναι φανερό ότι ένας τέτοιος μηχανισμός αναζήτησης καθιστά περιττά τα ειδικά ευρετήρια των έντυπων θησαυρών (KWIC, KWOC κ.λπ.), συμπεριλαμβανομένης και της αλφαβητικής παρουσίασης, η οποία ωστόσο μπορεί να αποτελεί μια συμπληρωματική επιλογή για την παρουσίαση των αποτελεσμάτων. Γενικά, υπάρχουν και άλλες λύσεις που προσφέρει η γλωσσική τεχνολογία, κάθε μία από τις οποίες είναι περισσότερο ή λιγότερο κατάλληλη ανάλογα με την περίπτωση.

Πέρα από τα παραπάνω, την αναζήτηση μπορεί να απλοποιήσει και ταυτόχρονα να βελτιώσει σημαντικά και η εφαρμογή τεχνικών διαμόρφωσης και επεξεργασίας ερωτήματος όπως η χρήση τελεστών Μπουλ, η οποία διευκολύνει επίσης την εφαρμογή της συντακτικής διαίρεσης κατά τη δόμηση ενός θησαυρού. Έτσι, ένας σύνθετος όρος όπως π.χ. «λογισμικό διαχείρισης θησαυρού» μπορεί, αντί να ενσωματωθεί αυτούσιος στον θησαυρό, να εκφραστεί κατά την ευρετηρίαση και την αναζήτηση με τον συνδυασμό των όρων που περιλαμβάνει: «λογισμικό» AND «διαχείριση» AND «θησαυρού», ή ίσως με τον συνδυασμό των όρων «διαχείριση θησαυρού» AND «λογισμικό». Η συντακτική διαίρεση υποδεικνύεται με σχέση ισοδυναμίας:

λογισμικό διαχείρισης θησαυρού
USE (ή USE+)

λογισμικό + διαχείριση + θησαυροί

Καθώς οι σχετικές οδηγίες στα πρότυπα δεν είναι απόλυτες, η απόφαση εάν θα γίνει αποδεκτός ή όχι ένας σύνθετος όρος, αλλά και η απόφαση για τον τρόπο επανέκφρασης του σύνθετου όρου, ίσως να είναι από τις πιο δύσκολες και υποκειμενικές αποφάσεις που έχουν να αντιμετωπίσουν οι συντάκτες ενός θησαυρού. Το επίπεδο προσυνδυασμού βέβαια των όρων ενός θησαυρού σχετίζεται στενά με την εξειδίκευση του λεξιλογίου που θέλουμε να επιτύχουμε. Όσο πιο περίπλοκοι είναι οι σύνθετοι όροι, τόσο πιο εξειδικευμένο είναι το λεξιλόγιο αλλά και τόσο μεγαλύτερος είναι ο αριθμός των όρων που θα χρησιμοποιηθούν. Όταν χρησιμοποιείται η συντακτική διαίρεση περιορίζεται ο αριθμός των όρων του θησαυρού, αλλά και το ποσοστό προσυνδυασμού.

Από όλα όσα αναφέραμε γίνεται φανερή η συμβολή των διαφόρων στρατηγικών αναζήτησης που είναι σήμερα διαθέσιμες, στην αποτελεσματικότητα ενός συστήματος ανάκτησης. Οι θησαυροί από την πλευρά τους μπορούν να συμβάλλουν αποτελεσματικά τόσο στην αύξηση της ακρίβειας όσο και της ανάκλησης. Η ακρίβεια (το ποσοστό δηλαδή των συναφών εγγραφών από το σύνολο των ανακτηθέντων) αυξάνεται διότι ο κάθε όρος του θησαυρού αποδίδει αποκλειστικά μία και μόνο έννοια, π.χ. με τη χρήση προσδιοριστών, ή και περισσότερων σύνθετων όρων. Η ανάκληση (το ποσοστό των συναφών εγγραφών που ανακτήθηκαν από τον συνολικό αριθμό των συναφών εγγραφών της βάσης) αυξάνεται όσο περισσότερο ο θησαυρός καλύπτει τους διαφορετικούς τρόπους με τους οποίους ο κάθε χρήστης θα μπορούσε να αναζητήσει κάποιο θέμα, με τον εντοπισμό π.χ. όσο το δυνατόν περισσότερων συνωνύμων και τη δημιουργία παραπομπών προς τους προτιμώμενους όρους.

Δεν πρέπει να υποβαθμισθεί τέλος, η συμβολή των ευρετηριαστών στο συγκεκριμένο θέμα. Ο βαθμός εξειδίκευσης και εξαντλητικότητας της ευρετηρίασης επηρεάζουν την ανάκληση και την ακρίβεια. Όσον αφορά στην εξειδίκευση, ισχύει ότι όσο πιο εξειδικευμένοι είναι οι όροι που χρησιμοποιούνται κατά την ευρετηρίαση, τόσο μεν λιγότερες θα είναι οι εγγραφές που θα ανακτηθούν αλλά οπωσδήποτε θα είναι πιο συναφείς. Σχετικά με την εξαντλητικότητα, όσο περισσότεροι όροι περιλαμβάνονται σε μια εγγραφή, (όσο δηλαδή περισσότερες από τις πληροφορίες που περιλαμβάνονται σε ένα τεκμήριο ευρετηριάζονται), τόσο αυξάνεται και ο αριθμός των εγγραφών που θα ανακτηθούν.

2.3. Επέκταση ερωτήματος

Επιπλέον των προαναφερόμενων τεχνικών που θεωρούνται απαραίτητες για την επιτυχία μιας αναζήτησης, αξίζει να γίνει ιδιαίτερη μνεία και στην επέκταση ερωτήματος ή έκρηξη (query expansion, explosion). Πρόκειται για μια λειτουργία που χρησιμοποιείται στα συστήματα ανάκτησης πληροφοριών, η οποία μπορεί να υλοποιηθεί και χωρίς τη χρήση θησαυρού, αλλά με διαφορετική τεχνική από αυτή που θα μας απασχολήσει εδώ.

Λέγοντας επέκταση ερωτήματος ή έκρηξη σε κάποιον θησαυρό, εννοούμε την επέκταση ενός όρου ή όρων που τίθενται ως ερώτημα σε **ομάδες** όρων κοντινών στην ιεραρχία ή συνδεδεμένων με κάποιο σύνδεσμο π.χ. "Σχετικός Όρος" ("Related Term"). Με αυτόν τον τρόπο ανακαλούνται όλα τα τεκμήρια που είναι ευρητηριασμένα όχι μόνο με τον συγκεκριμένο όρο του ερωτήματος αλλά και με όλες τις ομάδες των προαναφερόμενων όρων. Η έκρηξη δηλαδή, επιτρέπει αποτελέσματα μερικού ταιριάσματος με όρους του θησαυρού, εννοιολογικά πολύ κοντά αλλά **όχι ταυτόσημους** με τους όρους του ερωτήματος. Βασίζεται σε ένα "μέτρο" εννοιολογικής συγγένειας μεταξύ των όρων του θησαυρού και αντιπροσωπεύει, όπως είναι φανερό, τη δύναμη της σχέσης τους. Αρκετά είναι τα συστήματα που συμπεριλαμβάνουν στη διαδικασία της έκρηξης και σχετικούς όρους πέραν των γενικότερων και ειδικότερων, ακολουθώντας τις προτιμήσεις των χρηστών που προέκυψαν από αντίστοιχες έρευνες. Πάντως, κατά γενική συμφωνία, η χρήση των σχέσεων πρέπει να γίνεται με την εξής σειρά: πρώτα να γίνεται επέκταση στους ειδικότερους όρους, κατόπιν στους σχετικούς και στο τέλος στους γενικότερους.

Για να γίνει πιο κατανοητή η διαδικασία, αξίζει να αναφέρουμε ότι όταν γίνεται «επέκταση», είναι σαν να χρησιμοποιείται ο τελεστής Μπουλ for π.χ. για όλους τους ειδικότερους όρους, του όρου που έχουμε επιλέξει. Έτσι, για παράδειγμα, επιλέγοντας τον όρο «καταδύσεις» από μια φασέτα «δραστηριοτήτων» και τον όρο «φωτογραφικός εξοπλισμός» από μια φασέτα «αντικειμένων», ο χρήστης μπορεί να διατυπώσει ένα ερώτημα και να ανακτήσει όλα τα τεκμήρια που αναφέρονται σε φωτογραφικό εξοπλισμό που χρησιμοποιείται στις καταδύσεις.

Η λειτουργία της επέκτασης αποδεικνύεται πολύ χρήσιμη για περίπλοκα ερωτήματα κυρίως σε μεγάλους θησαυρούς όπου η χειροκίνητη πλοήγηση για επιλογή όρων μπορεί να είναι κουραστική. Για να κατανοεί ο χρήστης όσο γίνεται περισσότερο τη διαδικασία και για να μην μπερδεύεται, πρέπει να υπάρχει ένας συνδυασμός αυτόματης και αλληλεπιδραστικής επέκτασης. Ο χρήστης πρέπει να ενημερώνεται για την ύπαρξη της συγκεκριμένης λειτουργίας, η οποία να ενεργοποιείται αφού το ζητήσει με κάποιο τρόπο ο ίδιος σε περιπτώσεις που κρίνει απαραίτητο όπως:

- μετά από εμφάνιση μηδενικού αποτελέσματος, όταν δηλαδή το σύστημα αδυνατεί να εντοπίσει όρους από το ερώτημά του που να ταυτίζονται ακριβώς με όρους του θησαυρού
- για τη βελτίωση του ερωτήματος που έχει θέσει, όταν π.χ. το σύστημα του προσφέρει περιορισμένο αριθμό αποτελεσμάτων συγκριτικά με τις προσδοκίες του.

Οι φασετικοί δείκτες που προκύπτουν κατά τη διαδικασία της φασετικής ανάλυσης θεωρούνται πολύ βοηθητικοί σε ένα σύστημα που χρησιμοποιεί επέκταση αφού πρόκειται για σημείο σήμανσης όχι ενός μόνο όρου αλλά μιας ομάδας όρων με κοινά χαρακτηριστικά (παρόμοιων ή κοντινών στην ιεραρχία). Η εξαντλητικότητα και η εξειδίκευση του λεξιλογίου, που σχετίζονται με το πλήθος των όρων και το βάθος στο οποίο μπορεί να φθάνουν οι ιεραρχίες αντίστοιχα, επηρεάζουν άμεσα όπως είναι φανερό τη διαδικασία της επέκτασης.

Η επιτυχία όμως της «έκρηξης» πέραν από τη σωστή δόμηση του λεξιλογίου εξαρτάται κυρίως από τον τρόπο ευρητηρίασης. Ιδιαίτερος στις περιπτώσεις που το σύστημα περιλαμβάνει επιλογή επέκτασης ερωτήματος, είναι φανερό πως οι ευρητηριαστές πρέπει να είναι προσεκτικοί ώστε να ακολουθούν την αρχή της θεματικής ευρητηρίασης που αναφέρει ότι στα τεκμήρια πρέπει να αποδίδονται ακριβώς τα θέματα που τα περιγράφουν· ούτε ευρύτερα ούτε ειδικότερα. Επιπλέον,

θα πρέπει να επιλέγουν την σε βάθος ευρετηρίαση και να αποδίδουν στα τεκμήρια όλα τα υποθέματα που τυχόν αυτό περιέχει. Αυτό ίσως αρχικά να σημαίνει έναν βαθμό δυσκολίας, λόγω του εθισμού από την ευρετηρίαση βιβλίων, όπου σύμφωνα με τους κανόνες της βιβλιοθήκης του Κογκρέσου οι ευρετηριαστές πρέπει να αποδίδουν στα βιβλία έναν όρο (θεματική επικεφαλίδα) ο οποίος να αντιπροσωπεύει το σύνολο του περιεχομένου ενός βιβλίου (επιλεκτική ευρετηρίαση). Ο τρόπος ευρετηρίασης βέβαια είναι συνυφασμένος και με τη δομή του θησαυρού, αφού η εξειδίκευση και η εξαντλητικότητα του θησαυρού βοηθά τον ευρετηριαστή να προχωρήσει σε εις βάθος ευρετηρίαση.

3. Διαλειτουργικότητα

Η πληθώρα των πληροφοριών που είναι διαθέσιμες σε έναν χρήστη ηλεκτρονικών πληροφοριακών πηγών είναι μια πραγματικότητα που χαρακτηρίζει την εποχή μας. Παραδείγματα τέτοιων πληροφοριακών πηγών είναι οι βάσεις δεδομένων, οι ψηφιακές βιβλιοθήκες, η βιβλιογραφική βάση μιας βιβλιοθήκης, ακόμη και ο Ιστός. Ο μεγάλος αριθμός όμως των πηγών αντί να διευκολύνει τους χρήστες, αποδεικνύεται ότι τους οδηγεί σε ένα χαοτικό περιβάλλον που αδυνατεί να τους προσφέρει σαφείς απαντήσεις στα πληροφοριακά ερωτήματά τους. Το πρόβλημα ξεκινά από το γεγονός ότι οι χρήστες είναι αναγκασμένοι να συμβουλευονται κάθε μια από αυτές τις πηγές χωριστά, και διογκώνεται από τη διαφορετικότητα των συστημάτων. Τα συστήματα μπορεί να διαφέρουν τόσο στα τεχνικά χαρακτηριστικά τους (πρότυπα που υιοθετούν, τρόποι αναζήτησης που χρησιμοποιούν) όσο και στα λεξιλόγια που χρησιμοποιούν για την θεματική οργάνωσή τους. Δεν είναι πλέον αρκετό για τους χρήστες να αποφασίζουν μόνο εάν θα χρησιμοποιήσουν λέξεις κλειδιά ή τελεστές Μπουλ, ή να δοκιμάσουν την τύχη τους με αποκοπή, αλλά επιπλέον θα πρέπει να προσέχουν και την ορολογία που χρησιμοποιούν στο κάθε σύστημα που εισέρχονται.

Ιδανικό θα μπορούσε να χαρακτηριστεί ένα περιβάλλον που θα προσέφερε αναζήτηση σε πολλαπλές πηγές ταυτοχρόνως. Αντί ο χρήστης να είναι αναγκασμένος να κάνει ξεχωριστές αναζητήσεις στις διάφορες βάσεις δεδομένων ή συλλογές, θα του δινόταν η δυνατότητα, θέτοντας μια ερώτηση, να ανακτήσει αποτελέσματα από διάφορες, ακόμη και ετερογενείς πηγές. Αυτό σημαίνει εξοικονόμηση χρόνου, κόστους και κόπου.

Κάτι τέτοιο όμως μπορεί να επιτευχθεί μόνο όταν υφίσταται διαλειτουργικότητα μεταξύ των συστημάτων, όταν δηλαδή αυτά μπορούν να επικοινωνούν μεταξύ τους. Ο τρόπος επικοινωνίας των συστημάτων σε τεχνικό επίπεδο είναι μεγάλο πρόβλημα που δεν θα μας απασχολήσει εδώ. Το πρόβλημα όμως που θα μας απασχολήσει είναι η διαλειτουργικότητα που σχετίζεται με τα λεξιλόγια που χρησιμοποιούν τα συστήματα.

Αν όλα τα συστήματα (ή οι υπηρεσίες πληροφόρησης) χρησιμοποιούσαν το ίδιο λεξιλόγιο, τα πράγματα θα ήταν απλά. Ένας θησαυρός, για παράδειγμα, θα μπορούσε να χρησιμοποιηθεί για κατανεμημένη αναζήτηση σε πολλές πηγές ταυτοχρόνως. Από τη στιγμή που ένα ερώτημα θα έχει διαμορφωθεί σωστά στο ένα σύστημα χρησιμοποιώντας έναν συγκεκριμένο θησαυρό, τότε θα μπορούσε να υποβληθεί σε όλες τις βάσεις δεδομένων ή άλλες πηγές ευρετηριασμένες με τον ίδιο θησαυρό, για παράλληλες αναζητήσεις.

Η πραγματικότητα όμως δυστυχώς, είναι τελείως διαφορετική. Οι υπηρεσίες πληροφόρησης κατ' αρχήν χρησιμοποιούν διαφορετικούς τύπους Συστημάτων Οργάνωσης Γνώσης (π.χ. θεματικές επικεφαλίδες, θησαυρούς, οντολογίες). Αλλά και τον ίδιο τύπο ΣΟΓ να χρησιμοποιούν, π.χ. θησαυρούς, προβλήματα εμφανίζονται λόγω της ύπαρξης θησαυρών όμοιων μεν από άποψη θεματικού περιεχομένου, διαφορετικών δε, λόγω της διαφορετικής τους προέλευσης (π.χ. θησαυροί διαφορετικών γλωσσών), έκδοσης ή στόχου κατασκευής. Ειδικότερα, προβλήματα δημιουργούνται όταν λαμβάνονται διαφορετικές αποφάσεις για τη διαδικασία κατασκευής του κάθε θησαυρού (μεγάλη ή μικρή εξαντλητικότητα και εξειδίκευση, χρήση ή μη σύνθετων όρων,

διαφορετικά ιεραρχικά επίπεδα) και όταν η ίδια έννοια αντιπροσωπεύεται από διαφορετικούς όρους ή ο ίδιος όρος έχει διαφορετική σημασία σε διάφορους θησαυρούς (συνώνυμα).

Συμπερασματικά, οι διαφορές στα λεξιλόγια που χρησιμοποιούνται προκαλούν μεγάλες δυσκολίες στη διαλειτουργικότητα των συστημάτων και κατ' επέκταση στην ανάκληση της πληροφορίας. Διαρκείς είναι οι έρευνες και οι μελέτες για την αναζήτηση τρόπων που θα διευκολύνουν την δια-θεματική αναζήτηση και την συνδυασμένη χρήση όλων αυτών των εργαλείων στοχεύοντας στην ενοποίηση των πληροφοριακών πηγών. Η ενοποίηση αυτή μπορεί να πραγματοποιηθεί μέσω συνδετικών κρίκων που είναι εφικτό να δημιουργηθούν από τους όρους και τις σχέσεις ενός ή πολλών ειδών θησαυρών, (μονόγλωσσων, πολύγλωσσων και διαφορετικού περιεχομένου) και από άλλων ειδών Συστημάτων Οργάνωσης Γνώσης. Η επίλυση του προβλήματος σχετίζεται κατά κύριο λόγο με την εύρεση τρόπων δημιουργίας τέτοιων συνδετικών κρίκων.

Ενδεικτικά θα αναφέρουμε δύο από αυτούς τους τρόπους ή μεθόδους συνδυαστικής χρήσης των λεξιλογίων:

- αντιστοιχισμός με σχέσεις ισοδυναμίας (mapping) από όρους ενός λεξιλογίου σε όρους ενός ή περισσότερων άλλων λεξιλογίων. Τα στοιχεία που κρατούνται στις αντιστοιχίσεις παρέχουν πληροφορίες οι οποίες κυμαίνονται από το εάν οι όροι είναι ταυτόσημοι ή συνώνυμοι έως το αν συνδέονται με κάποιο άλλο είδος σχέσης (π.χ. ενικός – πληθυντικός αριθμός, ευρύτερη ή στενότερη έννοια κ.λπ.).
- σχήματα μεταπήδησης (switching systems) που χρησιμοποιούν μια ενδιάμεση γλώσσα στην οποία αντιστοιχίζονται τα συμμετέχοντα λεξιλόγια.

Μέχρι στιγμής πάντως η διαδικασία αυτή ερμηνεύεται κυρίως με τρόπους αντιστοιχισμών μεταξύ των όρων. Συνεπώς, για να υποβληθεί ένα ερώτημα σε πηγές ευρετηριασμένες με διαφορετικά λεξιλόγια, θα πρέπει οι περιγραφείς των λεξιλογίων να έχουν διασυνδεθεί μεταξύ τους με αντιστοιχίσεις (mappings) που βεβαίως θα πρέπει να έχουν καθιερωθεί από πριν. Πρόκειται για μια δύσκολη εργασία που καθώς φαίνεται αυτή τη στιγμή ο ανθρώπινος παράγοντας είναι απαραίτητος για την επίτευξή της.

Η δημιουργία αντιστοιχίσεων με αυτόν τον τρόπο μπορεί επίσης να προσφέρει καλύτερα αποτελέσματα όσον αφορά στη δαπανηρή λύση της δημιουργίας πολύγλωσσων θησαυρών. Συζητείται και μελετάται η εναλλακτική λύση της διασύνδεσης θησαυρών διαφόρων γλωσσών, μέσω δήλωσης εκφράσεων ισοδυναμίας. Έτσι, μπορεί να χρησιμοποιούνται μεν πολύγλωσσοι θησαυροί αλλά και συνδυασμοί αντίστοιχου θέματος μονόγλωσσων θησαυρών, για να μετατρέπουν την αναζήτηση των ίδιων εννοιών από τη μια γλώσσα στην άλλη.

Όλες οι παραπάνω αναφερόμενες δυνατότητες είναι οι ζητούμενες και για τη σωστή λειτουργία του πολλά υποσχόμενου Σημασιολογικού Ιστού. Είναι ανάγκη όλα τα ΣΟΓ που θα χρησιμοποιηθούν εκεί να μπορούν να μιλούν μεταξύ τους.

Οι θησαυροί πρέπει και μπορούν να χρησιμοποιηθούν αρκετά στο μελλοντικό περιβάλλον πληροφόρησης όπως αυτό προδιαγράφεται, με διάφορους τρόπους. Οι τεχνικές σύμφωνα με τις οποίες αναπτύχθηκαν και υλοποιήθηκαν με επιτυχία τόσα χρόνια διάφοροι θησαυροί αξιοποιούνται και εμπλουτίζονται πλέον για τη δημιουργία οντολογιών. Οι οντολογίες αποτελούν ένα από τα θεμελιώδη συστατικά στοιχεία του Σημασιολογικού Ιστού και πρέπει να είναι οι φυσικοί συνεχιστές των θησαυρών. Λέγοντας φυσικοί συνεχιστές εννοούμε ότι οι οντολογίες θα πρέπει με κάποιο τρόπο να απορροφήσουν όλα τα δεδομένα που μπορούν να προσφέρουν οι υπάρχοντες θησαυροί και όχι να δημιουργηθούν από το μηδέν.

Οι κύριες διαφορές θησαυρών και οντολογιών εντοπίζονται στο επίπεδο αφαίρεσης και σχέσεων μεταξύ των εννοιών. Οι όροι των θησαυρών είναι λέξεις ή φράσεις της φυσικής γλώσσας και όχι κωδικοί που εκφράζουν πιο αφηρημένες έννοιες όπως στις οντολογίες. Οι οντολογίες επίσης εκφράζονται σε πρότυπες γλώσσες όπως η RDF, DAML/OIL και OWL και έχουν τη

δυνατότητα χρήσης ενός προσδιοριστή μοναδικής έννοιας για την προτυποποίηση μιας έννοιας παγκοσμίως. Επιπλέον, ορίζουν πολύ αυστηρά και με μεγαλύτερη ακρίβεια τις σχέσεις μεταξύ των εννοιών ώστε να θέτονται πιο εύκολα κανόνες και να διευκολύνεται η εξαγωγή συμπερασμάτων από τις μηχανές.

Όσο πιο καλά και αυστηρά δομημένος είναι ένας θησαυρός τόσο πιο εύκολα μπορεί να ενσωματωθεί ή να επικοινωνήσει με μια οντολογία. Γιατί λοιπόν ένας θησαυρός να μην επεκταθεί ώστε να περιλαμβάνει ορισμούς, λεπτομερείς σημειώσεις για τη χρήση των όρων και πιο ξεκάθαρα καθορισμένες σχέσεις; Ακολουθούν μερικές προτάσεις προς αυτή την κατεύθυνση:

- Ο κάθε όρος θα πρέπει να αντιπροσωπεύει μια και μοναδική έννοια με σαφή ορισμό και περιγραφή (π.χ. χρήση προσδιοριστών).
- Οι σχέσεις μεταξύ των εννοιών μπορούν να επεκταθούν, κατ' αρχήν διαχωρίζοντας τις σχέσεις ιεραρχίας γένους – είδους από τις σχέσεις ιεραρχίας όλου – μέρους και από αυτές της περιστασης (BTI = "Γενικότεροι Όροι περίστασης", BTG = "Γενικότεροι Όροι γένους", BTP = "Γενικότεροι Όροι μέρους" NPI = "Ειδικότεροι Όροι περίστασης", NTG = "Ειδικότεροι Όροι γένους", NTP = "Ειδικότεροι Όροι μέρους").
- Οι σχέσεις συσχέτισης μπορούν επίσης να ορίζονται πιο αυστηρά π.χ. ως σχετικοί όροι (RT) με σχέση αιτίας/αποτελέσματος, υλικού/ιδιότητας κ.λπ.
- Οι σχέσεις ισοδυναμίας επίσης χρειάζονται ιδιαίτερη προσοχή. Πρέπει να γίνεται ομαδοποίηση των μη προτιμώμενων όρων σύμφωνα με το είδος της σχέσης που τους ενώνει με τους προτιμώμενους, αφού μπορεί να περιλαμβάνονται πέραν των συνωνύμων και οiwνεί συνώνυμα (αντίθετα, παραπομπή προς το γένος κ.λπ) [2].

Επίσης:

- Ο διαχωρισμός των όρων με βάση τη φασετική ανάλυση θεωρείται πλέον απαραίτητος.
- Η κωδικοποίηση των θησαυρών σε κάποια γλώσσα, π.χ. RDF, είναι πολύ χρήσιμη. Ήδη έχουν αρχίσει προσπάθειες προς αυτή την κατεύθυνση όπως αυτή του SKOS Core schema από το W3C.

Επίλογος

Η επίδραση που έχουν στη δομή του θησαυρού οι εξελίξεις στην τεχνολογία των συστημάτων ανάκτησης πληροφοριών και η ανάγκη εξέλιξης των θησαυρών για την καλύτερη συμμετοχή τους στο σύγχρονο περιβάλλον πληροφόρησης αντανακλώνται στα πρότυπα συγκρότησης και διαχείρισής τους. Ήδη το British Standard Institution και ο NISO των ΗΠΑ έχουν προχωρήσει σε αναθεώρηση του βρετανικού και του αμερικανικού προτύπου αντίστοιχα για μονόγλωσσους θησαυρούς (και τα δύο βρίσκονται στη φάση της δημόσιας κρίσης). Οι κυριότερες διαφορές με τις παλαιότερες εκδόσεις εντοπίζονται στα ακόλουθα:

- Δίνεται μεγαλύτερη προσοχή στην λειτουργία των θησαυρών με τη χρήση ηλεκτρονικών μέσων
- Περιλαμβάνονται συστάσεις που αφορούν τη λειτουργικότητα ενός λογισμικού για την καλύτερη διαχείριση των δεδομένων ενός θησαυρού
- Γίνεται αναφορά στα ελεγχόμενα λεξιλόγια με πιο ευρεία έννοια και όχι αποκλειστικά στους θησαυρούς
- Γίνεται εκτενής παρουσίαση του θέματος της διαλειτουργικότητας
- Συστήνεται στους συντάκτες τέτοια δόμηση των εργαλείων οργάνωσης γνώσης έτσι ώστε αυτά να γίνουν αυστηρά κωδικοποιημένα
- Δίνεται ιδιαίτερη βαρύτητα στην οργάνωση ενός θησαυρού με βάση τη φασετική ανάλυση.

Τέλος, αξίζει να αναφερθεί ότι από την ενότητα της διαλειτουργικότητας γίνεται φανερή η ανάγκη άμεσης εμπλοκής των βιβλιοθηκών στη δημιουργία ενιαίου περιβάλλοντος πληροφόρησης. Η αρχή μπορεί να γίνει από τα Ολοκληρωμένα Συστήματα Αυτοματισμού Βιβλιοθηκών (ΣΑΒ), τα οποία αποτελούν τα κατ' εξοχήν συστήματα οργάνωσης βιβλιοθηκονομικού υλικού. Στην πλειονότητα των ΣΑΒ δεν έχει προβλεφθεί η ενσωμάτωση υποστηρικτικών υποσυστημάτων θησαυρού, αλλά και σε αυτά που έχει προβλεφθεί, απ' ότι φαίνεται, δεν λειτουργούν ικανοποιητικά, κυρίως λόγω της εσωτερικής διάταξής τους σε *MARC formats*, πράγμα που τα καθιστά δύσκαμπτα. Για παράδειγμα, «το MARC format επιτρέπει τον ορισμό ιεραρχικών συνδέσμων μεταξύ δύο ή περισσότερων αντικειμένων (βιβλιογραφικών εγγραφών). Αυτή η διασύνδεση όμως, αποτελεί ένα μόνο ιεραρχικό επίπεδο (είτε προς τα πάνω είτε προς τα κάτω). Έτσι, δεν μπορεί να ικανοποιηθεί η ανάγκη για αναπαράσταση των συνεχών και διαδοχικών ιεραρχικών επιπέδων που απαιτούνται σε έναν θησαυρό» [16].

Αυτό που πρέπει να γίνει είναι να δημιουργηθούν γέφυρες για ολοκληρωμένη επικοινωνία και κοινή χρήση των ΣΑΒ με λογισμικά ή εξυπηρετητές (servers) διαχείρισης Θησαυρών. Αν το ίδιο συμβεί και με τις βάσεις δεδομένων, τις ψηφιακές βιβλιοθήκες και γενικότερα τα πληροφοριακά συστήματα, ένα μεγάλο βήμα προς το ενιαίο πληροφοριακό περιβάλλον θα έχει επιτευχθεί [3].

Βιβλιογραφία

1. Εθνικό Κέντρο Τεκμηρίωσης, *Θησαυρός ελληνικών όρων: μεθοδολογία ανάπτυξης*, Αθήνα: Εθνικό Κέντρο Τεκμηρίωσης / Αθήνα: ΕΙΕ, 2005
2. ΕΛΟΤ 1321, *Τεκμηρίωση – Κατευθυντήριες οδηγίες για τη συγκρότηση και ανάπτυξη μονόγλωσσων Θησαυρών*, Αθήνα: ΕΛΟΤ, 1993
3. Χαρίκλεια Μπρίντζη, *Φιλική χρήση οντολογιών / θησαυρών από πληροφοριακά συστήματα / ψηφιακές βιβλιοθήκες: μεταπτυχιακή εργασία*, Κέρκυρα: Ιόνιο Πανεπιστήμιο, Τμήμα Αρχαιολογίας – Βιβλιοθηκονομίας, 2005.
4. Jean Aitchison, Alan Gilchrist, and David Bawden, *Thesaurus Construction and Use: a Practical Manual*, London: ASLIB, ³1997.
5. Jean Aitchison, Stella Dextre Clarke, "The thesaurus: a historical viewpoint, with a look to the future", *Cataloging & Classification Quarterly* 37 no 3,4 (2004) 5-21.
6. Marcia J Bates, "Subject Access in Online Catalogs: A design Model", *Journal of the American Society for Information Science*, 37 no 6 (1986) 357-376.
7. BSI committee IDT/2/2, *Structured vocabularies for information retrieval — Guide. Part 2: Thesauri*, BRITISH STANDARD BS 8723 Part 2 (DRAFT), (2004)
8. Emma Culloch, "Multiple terminologies: an obstacle to information retrieval", *Library review*, 53 no 6 (2004) 297-300.
9. W. B. Frakes "Term Conflation for Information Retrieval". Research and Development in Information Retrieval. New York: Cambridge University Press, 1984
10. Alan Gilchrist, "Thesauri, taxonomies and ontologies – an etymological note", *Journal of Documentation*, 59 no 1 (2003) 7-18.
11. Susan Jones, "A thesaurus data model for an intelligent retrieval system". *Journal of Information Science* 19 (1993) 167-178
12. Eric H Johnson, Pauline Cochrane "A hypertextual interface for a searcher's thesaurus". *Second Annual Conference on the Theory and Practice of Digital Libraries*, Austin, TX, 1995 [τεκμήριο www, URL <http://csdl.tamu.edu/DL95/papers/johncoch/johncoch.html>, ημερομηνία πρόσβασης: 29.9.2004].
13. National Information Standards Organization (U.S.) *Guidelines for the construction, format, and management of monolingual controlled vocabularies* (ANSI/NISO Z39.19-200x). Bethesda, Maryland, U.S.A: NISO Press, 2005 (draft standard)

14. Toni Petersen, Patricia Barnett (ed). *Guide to indexing and cataloging with the Art & Architecture thesaurus*. New York: Oxford University Press, 1994.
15. Richard Pollard "Hypertext presentation of thesauri used in online searching", *Electronic Publishing*, 3 no 3 (1990) 155–172.
16. Georgia Prokopiadou, Christos Papatheodorou, Dionysis Moschopoulos, "Integrating knowledge management tools for government information". *Government Information quarterly*, 2004
17. Louise F Spiteri, "The essential elements of faceted thesauri" *Cataloging & Classification Quarterly*. 28 no 4 (1999) 31-51
18. B. J. Wielinga et al., "From thesaurus to ontology" International Conference On Knowledge Capture, 2001. New York: ACM Press, 2001.
19. Marcia Lei Zeng, Lois Mai Chan "Trends and issues in establishing interoperability among knowledge organization systems". *Journal of the American Society for information science and technology*, 55 no 5 (2004) 377-395.