

# Εννοιολογική Διεύρυνση Ερωτημάτων με τη Χρήση Θησαυρού: μια εμπειρική μελέτη

**Άννα Μάστορα<sup>1</sup>, Μαρία Μονόπωλη<sup>2</sup>, Σαράντος Καπιδάκης<sup>1</sup>**

<sup>1</sup>Εργαστήριο Ψηφιακών Βιβλιοθηκών & Ηλεκτρονικής Δημοσίευσης, Τμήμα Αρχαιονομίας – Βιβλιοθηκονομίας, Ιόνιο Πανεπιστήμιο, Ιωάννου Θεοτόκη 72, 49100, Κέρκυρα

<sup>2</sup>Βιβλιοθήκη, Τράπεζα της Ελλάδος, Ελ. Βενιζέλου 21, 102 50, Αθήνα  
{mastora, sarantos}@ionio.gr, mariamonopoli@hotmail.com

## Περίληψη

Η παρούσα μελέτη αποτελεί μέρος της έρευνάς μας με αντικείμενο την *εννοιολογική διεύρυνση ερωτημάτων* (conceptual query expansion). Στόχος είναι να δούμε κατά πόσο οι υπάρχουσες δομές των Συστημάτων Οργάνωσης της Γνώσης (ΣΟΓ) ανταποκρίνονται στην αντίληψη για την αναπαράσταση της γνώσης που έχουν οι χρήστες, ώστε να εξετάσουμε σε ποιες περιπτώσεις μπορούν αυτά να χρησιμεύσουν στη διεύρυνση των ερωτημάτων. Κατά την έρευνα επικεντρωθήκαμε σε μη-ειδικούς χρήστες και σε ένα σχεσιακό ΣΟΓ, το Θησαυρό EUROVOC. Για τη διερεύνηση των υποθέσεων της έρευνάς μας εφαρμόσαμε εμπειρική μελέτη στην οποία οι συμμετέχοντες καλούνταν να εκτελέσουν αναζητήσεις στη βάση δεδομένων της *Ευωνύμου Οικολογικής Βιβλιοθήκης*. Η έρευνα έδειξε ότι μόλις το 15,0% των όρων που χρησιμοποιήθηκαν ήταν μοναδικοί καθώς και ότι, στην πλειοψηφία τους, οι χρήστες δομούν όρους αναζήτησης που αποτελούνται από ελάχιστες λέξεις. Επιπλέον, από τους όρους που χρησιμοποιήθηκαν από τους χρήστες, το 60,5% εντοπίστηκε στο Θησαυρό. Καταλήγουμε ότι για τη διεύρυνση των ερωτημάτων με τη χρήση Θησαυρού αποτελεί σημαντικό εμπόδιο η πολυσημία των λέξεων, η οποία δεν είναι εύκολο να αντιμετωπιστεί καθώς οι περιορισμένες λέξεις που δομούν τους όρους αναζήτησης δεν προσφέρονται για ασφαλή εννοιολογική συσχέτιση, αρχικά μεταξύ των χρησιμοποιούμενων όρων και κατόπιν με όρους από την ιεραρχία του Θησαυρού. Μελλοντική έρευνα αποτελεί το αν αυτό αντιμετωπίζεται με την ταυτόχρονη χρήση περισσότερων ΣΟΓ κατά τη διεύρυνση των ερωτημάτων ή με την *εννοιολογική συσταδοποίηση ερωτημάτων*.

*Λέξεις κλειδιά: Σύνθεση & ανασύνθεση ερωτημάτων, Εννοιολογική διεύρυνση ερωτημάτων, Συστήματα Οργάνωσης της Γνώσης (ΣΟΓ), Γνωσιακή πολυ-αναπαράσταση*

## Abstract

This study is part of our broader research focusing on conceptual query expansion. Our goal is to examine whether current structures of Knowledge Organisation Systems (KOSs) reflect users' perception of knowledge representation in order to verify under which circumstances the KOSs can ameliorate the query expansion process. For our study we focused on non-expert users using a relational KOS, namely the EUROVOC thesaurus. To test our hypotheses we proceeded with an empirical study during which the participants had to conduct specific searches in the *Evonymos Ecological Library* database. Our metrics showed that only 15.0% of the submitted terms were unique and that the search terms mainly contained one or two words. Furthermore, 60.5% of the submitted terms were matched to Thesaurus' terms. We conclude that for conceptual query expansion using a thesaurus the polysemy of words is a serious obstacle, which is hard to deal with because of the limited search words used for each query. This is because the limited search words do not serve for safe conceptual matching both among the search terms and between the search terms and terms from the hierarchy of the Thesaurus. Future research will try to answer whether this problem is solved with the synchronous use of multiple KOSs during the query expansion or with conceptual query clustering techniques.

*Keywords: Query formulation & reformulation, Conceptual query expansion, Knowledge Organisation Systems (KOSs), Cognitive poly-representation*

## 1. Εισαγωγή

Η παρούσα μελέτη αποτελεί μέρος της έρευνάς μας με αντικείμενο την *εννοιολογική διεύρυνση ερωτημάτων* (conceptual query expansion). Τα *Συστήματα Οργάνωσης της Γνώσης* (ΣΟΓ) αποτελούν σημαντικό εργαλείο για το σκοπό αυτό. Θελήσαμε να διερευνήσουμε κατά πόσο οι χρήστες μπορούν να ενταχθούν σε πρότυπα συμπεριφοράς συνδυάζοντας την αποτύπωση της γνώσης στα ΣΟΓ και την αντίληψη για την αναπαράσταση της γνώσης που έχουν οι χρήστες. Επιπλέον, εξετάσαμε σε ποιες περιπτώσεις μπορούν να χρησιμεύσουν τα ΣΟΓ για *διεύρυνση των ερωτημάτων*. Ειδικότερα, επικεντρωθήκαμε σε μη-ειδικούς χρήστες και σε ένα *σχεσιακό ΣΟΓ*, το Θησαυρό EUROVOC.

Βασική διαφοροποίηση της έρευνάς μας από άλλες στον ίδιο χώρο αποτελεί το γεγονός ότι αξιοποιήσαμε δεδομένα (ερωτήσεις χρηστών και ΣΟΓ) στην ελληνική γλώσσα. Επιπλέον διαφοροποίηση αποτελεί το ότι χρησιμοποιήσαμε τον ηλεκτρονικό κατάλογο μιας βιβλιοθήκης και όχι το περιβάλλον του Διαδικτύου και κάποια μηχανή αναζήτησης, όπως οι περισσότερες ανάλογες μελέτες. Τόσο οι δεδομένοι παράγοντες, όσο και τα προβλήματα καθώς και η αντιμετώπισή τους διαφέρουν μεταξύ των δύο αυτών χώρων.

Για την επίτευξη του στόχου μας, αρχικά διερευνήσαμε τη συμπεριφορά των χρηστών κατά την επιλογή όρων τόσο για τη *σύνθεση* όσο και την *ανασύνθεση* των ερωτημάτων που θέτουν. Κατόπιν, μελετήσαμε τους *μοναδικούς όρους* που εντοπίστηκαν ώστε να καταδειχτεί αν οι χρήστες είναι αρκετά προβλέψιμοι ώστε να στηρίξουμε επάνω τους πρότυπα (patterns) για τη *διεύρυνση των ερωτημάτων*. Περαιτέρω, εξετάσαμε το πλήθος των λέξεων που χρησιμοποιούνται σε ένα ερώτημα και τη σχέση που έχει αυτό με τις υποθέσεις της έρευνάς μας. Στη συνέχεια, εντοπίσαμε πόσοι από τους όρους που χρησιμοποιούν οι χρήστες υπήρχαν στο Θησαυρό έχοντας εννοιολογική αντιστοιχία.

Η μελέτη ολοκληρώνεται με την παρουσίαση των συμπερασμάτων μας και προτάσεις για μελλοντική έρευνα πάνω στο συγκεκριμένο πεδίο.

## 2. Σχετική βιβλιογραφία

Στο πλαίσιο της ευρύτερης έρευνας που λαμβάνει χώρα για την αποτελεσματικότερη *Ανάκτηση Πληροφοριών* (Information Retrieval), το πεδίο της *διεύρυνσης των ερωτημάτων* που υποβάλλει ο χρήστης αποτελεί ιδιαίτερα σημαντικό κλάδο. Στόχος της *διεύρυνσης των ερωτημάτων* είναι ο εμπλουτισμός του αρχικού ερωτήματος με επιπλέον όρους οι οποίοι θα αυξήσουν την απόδοση της ανάκτησης [Shiri και Revie, 2006].

Το αποτέλεσμα της διεργασίας αυτής έρχεται αντιμέτωπο με το *σημασιολογικό χάσμα* που παρατηρείται μεταξύ των ερωτημάτων που υποβάλλουν οι χρήστες και της αποτύπωσης του περιεχομένου [Belkin, 1982]. Στην προσπάθεια να εξασφαλιστεί η *σχετικότητα* των αποτελεσμάτων, παρά τη δεδομένη αύξηση της *ανάκλησης*, η διαδικασία *διεύρυνσης των ερωτημάτων*, όποιους μηχανισμούς -αυτόματους, ημι-αυτόματους, διαδραστικούς ή χειρωνακτικούς- και αν υιοθετεί (Efthimiadis, 1996), έχει μελετηθεί αρκετά, και εξακολουθεί να μελετάται, από πολλές οπτικές. Κάποιες από αυτές συνδυάζουν την οπτική του χρήστη [Joho κ.α., 2004] ή της φιλοσοφίας δημιουργίας των ΣΟΓ [Doerr, 2001], την αντιστοιχία των ΣΟΓ μεταξύ τους και τη συνδυαστική χρήση τους για την επίτευξη της *γνωσιακής πολυ-αναπαράστασης* [Ingwersen, 1996], όπως και την εκ των προτέρων συσχέτιση του περιεχομένου με ΣΟΓ (Tudhope κ.α., 2006).

Οι χρήστες, είναι αποδεδειγμένο, πως είναι απρόβλεπτοι [Wilson, 1997] και επιρρεπείς σε συμπεριφορικές αλλαγές [Spink και Cole, 2006] και ίσως για το λόγο αυτό η έρευνα έχει στραφεί σε παράγοντες λιγότερο ευμετάβλητους. Κατά τη διαδικασία αναζήτησης, το στάδιο της *ανασύνθεσης* ή *βελτίωσης των ερωτημάτων* έχει μελετηθεί ιδιαίτερω, όπως προαναφέραμε. Αντίθετα, το πρώτο στάδιο, εκείνο της *σύνθεσης* του πρώτου ερωτήματος, δεν έχει τύχει ανάλογης μελέτης εξαιτίας, ίσως, της πολυπλοκότητάς του και των πολλών αστάθμητων παραγόντων που εμπλέκονται, με πρώτο και βασικότερο τον ίδιο το χρήστη. Η μόρφωση, η εξοικείωση τόσο με το περιβάλλον αναζήτησης αλλά και με το αντικείμενο που αναζητά, η αντιληπτική, αναλυτική αλλά και συνδυαστική του ικανότητα, καθώς και εξωγενείς, περιβαλλοντικοί παράγοντες είναι μερικά από τα στοιχεία που επηρεάζουν τη συμπεριφορά του χρήστη κατά τη διαδικασία αναζήτησης πληροφοριών.

### 3. Μεθοδολογία

Οι συμμετέχοντες στην εμπειρική μελέτη καλούνταν να εκτελέσουν αναζητήσεις στη βάση δεδομένων της *Ευωνύμου Οικολογικής Βιβλιοθήκης*<sup>1</sup> και να καταγράψουν σε συνοδευτικό έντυπο τις κινήσεις τους. Εκτός από το συνοδευτικό έντυπο, διατηρήσαμε και ιστορικό (log files) των κινήσεών τους ώστε να είναι δυνατή η επιβεβαίωση των αποτελεσμάτων μας. Οι συμμετέχοντες έκαναν *Είσοδο* στο περιβάλλον αναζήτησης με συγκεκριμένους λογαριασμούς ώστε να είναι δυνατή η αναγνώριση συγκεκριμένων κινήσεων κατά την επεξεργασία του ιστορικού.

#### 3.1. Συμμετέχοντες

Στην έρευνα συμμετείχαν 48 φοιτητές (27 προπτυχιακοί και 21 μεταπτυχιακοί) του Τμήματος Αρχαιονομίας και Βιβλιοθηκονομίας του Ιονίου Πανεπιστημίου. Η επιλογή του παραπάνω δείγματος βασίστηκε στο γεγονός ότι αποτελούσαν ενεργά μέλη του Πανεπιστημίου και συνεπώς ήταν ευκολότερο να εντοπιστούν για να συμμετέχουν και σε μελλοντική μας έρευνα προσφέροντας συγκριτικά αποτελέσματα.

#### 3.2. Σενάριο πειράματος

Οι συμμετέχοντες καλούνταν να υποβάλλουν ερωτήσεις, ώστε να εντοπίσουν σχετικά τεκμήρια, αναφορικά με τα εξής θέματα: *Αποδημητικά πουλιά* (E1), *Οπωροφόρα δέντρα* (E2), *Προστασία του περιβάλλοντος* (E3), *Φαινόμενο του θερμοκηπίου* (E4) και *Εναλλακτικές μορφές ενέργειας* (E5). Πριν ξεκινήσει το πείραμα, δόθηκε στους συμμετέχοντες έντυπο στο οποίο παρουσιάζόταν αναλυτικά το σενάριο των αναζητήσεων, καθώς και οδηγίες για τις δυνατότητες αναζήτησης. Αναφορικά με τις οδηγίες για το πείραμα οι συμμετέχοντες καλούνταν να ακολουθήσουν συγκεκριμένους κανόνες. Για παράδειγμα, οι αναζητήσεις έπρεπε να γίνουν στην Ελληνική γλώσσα και με αυστηρή σειρά εκτέλεσης. Το πρώτο μέρος της έρευνάς μας, το οποίο παρουσιάζεται εδώ, έλαβε χώρα το Μάιο του 2007.

#### 3.3. Περιβάλλον αναζήτησης

Για τις ανάγκες της έρευνας τροποποιήσαμε τόσο τις δυνατότητες του περιβάλλοντος αναζήτησης όσο και τις διαθέσιμες βιβλιογραφικές εγγραφές. Απλοποιήσαμε το περιβάλλον αναζήτησης με διπλό σκοπό. Αφενός θέλαμε να μην αντιμετωπίσουν οι χρήστες δυσκολίες με το σύστημα και να εστιάσουν, έτσι, στα ζητούμενα θέματα. Αφετέρου θέλαμε να περιορίσουμε την επίδραση που ενδεχομένως είχε στα αποτελέσματά μας το γεγονός ότι οι χρήστες μας ήταν, λόγω ειδικότητας, εξοικειωμένοι τόσο με περιβάλλοντα όσο και με τεχνικές αναζήτησης. Για το λόγο αυτό καταργήσαμε τη χρήση των τελεστών της Boolean λογικής (και, ή, και όχι) και παρείχαμε στους χρήστες μόνο ένα πεδίο αναζήτησης, εκείνο του θέματος, εξαλείφοντας τις δυνατότητες εφαρμογής εξειδικευμένων τεχνικών

---

<sup>1</sup> Η βάση της Ευωνύμου είναι διαθέσιμη στο <http://www.evonymos.org> (Τελευταία πρόσβαση Ιούνιος 2010)

αναζήτησης. Η επιλογή υποστηρίζεται και από τους Nicholas κ.α. (2008, σ.22), όπου σημειώνεται ότι το πεδίο του *θέματος* είναι το πιο σημαντικό. Αυτό, ωστόσο, μπορεί να διαφοροποιείται ανάλογα με το γνωστικό αντικείμενο. Οι χρήστες, τέλος, μπορούσαν να χρησιμοποιήσουν μόνο τη δεξιά αποκοπή.

Αναφορικά με τις βιβλιογραφικές εγγραφές, παρείχαμε περίπου 14.400 επιλεγμένες, από τη συνολική βάση της Ευωνύμου, βιβλιογραφικές εγγραφές, τις οποίες φιλοξενούσε ένας z39.50 server. Από αυτές διαγράφηκαν εγγραφές που περιείχαν θέματα στην Αγγλική γλώσσα καθώς και εγγραφές που θα επέστρεφαν αμφισβητήσιμα αποτελέσματα αναφορικά με την εννοιολογική συσχέτιση των όρων που χρησιμοποιούνταν, όπως για παράδειγμα τεκμήρια Λογοτεχνίας όπου οι λέξεις μπορεί να έχουν παραπλανητική σημασία. Σημειώνεται εδώ ότι η βάση δεν ακολουθεί κάποιο εργαλείο για τη λεκτική θεματική απόδοση του περιεχομένου.

#### **4. Αποτελέσματα**

Βασικό εργαλείο για το χαρακτηρισμό των όρων αναζήτησης και την εννοιολογική τους συσχέτιση αποτέλεσε ο διεπιστημονικός Θησαυρός EUROVOC<sup>2</sup>, ο οποίος είναι χωρισμένος σε θεματικές ενότητες, χωρίς να εμβαθύνει σε βαθμό εξειδίκευσης στην ιεραρχία των όρων. Στο πλαίσιο της παρούσας έρευνας η αναζήτηση των όρων έγινε στη σχετική ενότητα του Θησαυρού ανάλογα με την έννοια κάθε αναζήτησης. Για παράδειγμα, για το θέμα αναζήτησης *Εναλλακτικές μορφές ενέργειας* χρησιμοποιήθηκε η ενότητα του Θησαυρού με τίτλο *Ενέργεια*. Έτσι θεωρήσαμε ότι θα αποφύγουμε ένα μεγάλο σκόπελο της διαδικασίας της εννοιολογικής συσχέτισης των όρων που χρησιμοποιούν οι χρήστες με εκείνους που υπάρχουν στα ΣΟΓ και ειδικά αν αυτά δεν είναι περιορισμένα θεματικά. Το εμπόδιο έχει να κάνει με το φαινόμενο της *πολυσημίας* των λέξεων που αποτελεί σημαντικό λόγο αποτυχίας κάποιων συστημάτων που χρησιμοποιούνται για την εννοιολογική διεύρυνση ερωτημάτων. Για παράδειγμα, η λέξη *αλλαγή* συσχετίζεται με έξι έννοιες που υπάρχουν στο Θησαυρό.

##### *4.1. Ορισμοί και Διευκρινίσεις*

Ως *όρος* αναφέρεται η λέξη ή οι λέξεις που αποτελούν το υποβαλλόμενο, κάθε φορά, ερώτημα στο σύστημα. Σε σύγκριση, λοιπόν, με τον υποβαλλόμενο όρο σε ένα ερώτημα, ο όρος που χρησιμοποιείται μπορεί να είναι:

- *Ειδικότερος όρος*: οι χρήστες εξειδικεύουν το ερώτημα είτε προσθέτοντας περισσότερες λέξεις σε αυτό είτε αντικαθιστώντας τον αρχικό όρο με πιο ειδικούς, όπως αυτοί εντοπίζονται μέσα στην ιεραρχία του Θησαυρού.
- *Γενικότερος όρος*: οι χρήστες γενικεύουν το ερώτημα είτε απαλείφοντας κάποια από τις λέξεις του προηγούμενου όρου είτε αντικαθιστώντας τον αρχικό όρο με πιο γενικούς, όπως αυτοί εντοπίζονται μέσα στην ιεραρχία του Θησαυρού.

---

<sup>2</sup> Χρησιμοποιήθηκε η έκδοση 4.2 του Θησαυρού EUROVOC και, συγκεκριμένα, η ελληνική έκδοσή του. Διαθέσιμο στο <http://europa.eu/eurovoc/> (Τελευταία πρόσβαση: Ιούνιος 2010)

- *Συνώνυμος όρος*: ο χρήστης αντικαθιστά τον όρο με άλλον που έχει την ίδια ή περίπου την ίδια σημασία.
- *Παράλληλος όρος*: οι χρήστες ούτε γενικεύουν ούτε εξειδικεύουν τον προηγούμενο όρο ούτε χρησιμοποιούν κάποιο συνώνυμο. Υπάρχει, ωστόσο, μία σχέση μεταξύ των όρων, οι οποίοι βρίσκονται, κατά κανόνα, στο ίδιο επίπεδο της ιεραρχίας με τον αρχικά υποβαλλόμενο όρο, όπως *πορτοκαλιά – μανταρινιά ή χελιδόνη - πελαργός*.

Επιπλέον, ορίσαμε και τα παρακάτω:

- *Παρεχόμενος όρος*: όρος που παρέχεται στην περιγραφή του θέματος προς αναζήτηση.
- *Λανθασμένος όρος*: όρος που δεν υπάρχει στην ελληνική γλώσσα<sup>3</sup>.
- *Ακατάλληλος όρος*: όρος που κρίνεται ως ακατάλληλος να περιγράψει το θέμα προς αναζήτηση, δηλαδή δεν υπάρχει εννοιολογική συσχέτιση μεταξύ του όρου που εισάγει ο χρήστης και του θέματος προς αναζήτηση. Για παράδειγμα, ένας χρήστης χρησιμοποίησε τον όρο *έντομα* στο ερώτημα που αφορούσε τα *Αποδημητικά πουλιά*. Στους *ακατάλληλους* όρους περιλαμβάνονται και όροι στην Αγγλική γλώσσα καθώς είχαμε καταστήσει σαφές ότι οι αναζητήσεις έπρεπε να γίνουν μόνο στην Ελληνική, ώστε να μπορέσουμε να προχωρήσουμε στην εννοιολογική συσχέτιση με τη χρήση του επιλεγμένου Θησαυρού εξαλείφοντας εμπόδια που προκύπτουν από τη χρήση διαφορετικών γλωσσών στα διάφορα στάδια μιας εμπειρικής έρευνας. Τέτοια εμπόδια, τα οποία είναι αυτονόητα, αναφέρονται στο Barsky και Bar-Plan (2005).

Εφόσον η μελέτη μας στηρίζεται στην εννοιολογική συσχέτιση των όρων, δε λάβαμε υπόψη ορθογραφικά λάθη στην αποτύπωσή των εννοιών ούτε και διαφοροποιήσεις στον αριθμό, ενικό – πληθυντικό, των λέξεων. Αυτά, κατά κύριο λόγο, αντιμετωπίζονται με διάφορες προσεγγίσεις, όπως είναι η χρήση λεξιλογίων και διαδικασιών *stemming*. Η αναγωγή που κάναμε αφορούσε στην αντιστοιχία της έννοιας, για αυτό λέξεις που χρησιμοποιήθηκαν με αποκοπή εντάχθηκαν στον όρο που θα προέκυπτε αν η λέξη αποτυπωνόταν ολόκληρη, πάντα σε εννοιολογική συσχέτιση με το αντικείμενο του πειράματος.

Η χρήση του *παρεχόμενου όρου* κατά τη διαμόρφωση ερωτημάτων, ιδιαίτερα κατά το στάδιο της *σύνθεσης*, παίζει σημαντικό ρόλο στα αποτελέσματα της έρευνας, ιδιαίτερα σε περιπτώσεις εργαστηριακών πειραμάτων. Αναφέρεται ειδικά το στάδιο της *σύνθεσης* καθώς η εκκίνηση της διαδικασίας αναζήτησης και η αρχική αποτύπωση της πληροφοριακής ανάγκης είναι πολύ σημαντική για τη συνέχεια. Κάποια από τα θέματα που προκύπτουν με τη χρήση του παρεχόμενου όρου για τη διαμόρφωση των ερωτήσεων μπορούν να αντιμετωπιστούν με τη μελέτη των επόμενων σταδίων διαμόρφωσης ερωτημάτων, δηλαδή της *ανασύνθεσης*.

Αναφορικά με τη χρήση του *παρεχόμενου όρου* στο στάδιο της *σύνθεσης*, τα αποτελέσματα έδειξαν ότι το 41,6% των αναζητήσεων περιλάμβαναν τον *παρεχόμενο όρο*. Δηλαδή οι συμμετέχοντες ξεκίνησαν τη διαδικασία αναζήτησης κατά κύριο λόγο χρησιμοποιώντας τον όρο που τους παρείχαμε στην περιγραφή του θέματος αναζήτησης. Και κατά το στάδιο των

---

<sup>3</sup> Σύμφωνα με το Μπαμπινιώτης, Γ. *Λεξικό της Νέας Ελληνικής Γλώσσας: με σχόλια για τη σωστή χρήση των λέξεων*. Κέντρο Λεξικολογίας: Αθήνα, 1998.

ανασυνθέσεων έγινε χρήση του παρεχόμενου όρου γεγονός που το μετρούμε ξεχωριστά καθώς οι όροι αυτοί εντάχθηκαν ταυτόχρονα και κάτω από κάποια από τις υπόλοιπες κατηγορίες, δηλαδή μπορεί να ήταν ταυτόχρονα και παρεχόμενος όρος αλλά και ειδικότερος όρος σε σύγκριση με τον προηγούμενο. Τα αποτελέσματα έδειξαν ότι το 5,3% των ανασυνθέσεων αποτελούνταν από τον παρεχόμενο όρο. Αυτή η παρατήρηση, για τη χρήση του παρεχόμενου όρου, είναι συμβατή και με τη μελέτη που έκαναν οι Barsky και Bar-Ilan (2005) όπου αναφέρουν ότι οι συμμετέχοντες στην έρευνά τους προτίμησαν κατά κύριο λόγο όρους από την περιγραφή της διαδικασίας που είχαν να εκτελέσουν παρά όρους από τα ανακτημένα αποτελέσματα. Το ίδιο επιβεβαιώνουν και άλλες μελέτες, όπως των Colaric (2003), Aula (2003) και Shiri και Revie (2006).

#### 4.2. Πρότυπα Ερωτημάτων: στάδιο σύνθεσης ερωτήματος

Με τον όρο *σύνθεση ερωτήματος* ορίζουμε το πρώτο ερώτημα που υποβάλλουν οι ενδιαφερόμενοι στο σύστημα με σκοπό να εντοπίσουν πληροφορίες σχετικές με ένα θέμα. Οι όροι αναζήτησης που χρησιμοποιήθηκαν στην πρώτη αναζήτηση κατηγοριοποιήθηκαν σε επτά βασικές κατηγορίες: *γενικότερος όρος*, *ειδικότερος όρος*, *παράλληλος όρος*, *συνώνυμος όρος*, *λανθασμένος όρος*, *ακατάλληλος όρος* και *παρεχόμενος όρος*. Η σύγκριση έγινε με τον *παρεχόμενο όρο* από την περιγραφή του θέματος αναζήτησης. Ο *παρεχόμενος όρος* θεωρήθηκε το σημείο κατά το οποίο ο χρήστης ξεκινά να συλλαμβάνει την πληροφοριακή ανάγκη και να τη *μεταφράζει* σε όρους αναζήτησης [Wilson, 1997].

Τα αποτελέσματα για κάθε κατηγορία παρουσιάζονται παρακάτω (και συνοπτικά στον Πίνακα 1).

- **Γενικότερος όρος:** το 18,1% των συνθέσεων περιλάμβανε όρο γενικότερο του παρεχόμενου.
- **Ειδικότερος όρος:** το 25,2% των συνθέσεων περιλάμβανε όρο ειδικότερο του παρεχόμενου.
- **Παράλληλος όρος:** το 8,4% των συνθέσεων περιλάμβανε όρο παράλληλο του παρεχόμενου.
- **Συνώνυμος όρος:** το 5,9% των συνθέσεων περιλάμβανε όρο συνώνυμο του παρεχόμενου.
- **Λανθασμένος όρος:** δεν εντοπίστηκαν λανθασμένοι όροι στη σύνθεση των ερωτημάτων.
- **Ακατάλληλος όρος:** το 0,8% των συνθέσεων περιλάμβανε όρο ακατάλληλο για την περιγραφή των θεμάτων αναζήτησης, δηλαδή ο συγκεκριμένος όρος δεν επρόκειτο να λειτουργήσει κατά κανένα τρόπο ως συνδετικός κρίκος μεταξύ του αρχικού θέματος και σχετικών αποτελεσμάτων.
- **Παρεχόμενος όρος:** το 41,6% των συνθέσεων περιλάμβανε κάποια μορφή του παρεχόμενου όρου, είτε αυτούσιου είτε με χρήση αποκοπής.

Γενικότερος όρος	Ειδικότερος όρος	Παράλληλος όρος	Συνώνυμος όρος	Λανθασμένος όρος	Ακατάλληλος όρος	Παρεχόμενος όρος
18,1	25,2	8,4	5,9	0,0	0,8	41,6

**Πίνακας Νº1:** Κατηγοριοποίηση όρων Σύνθεσης Ερωτημάτων (%)

#### 4.3. Πρότυπα ερωτημάτων: στάδιο ανασύνθεσης ερωτήματος

Με τον όρο *ανασύνθεση ερωτήματος* ορίζουμε την αναθεώρηση (ή αναθεωρήσεις) του πρώτου ερωτήματος. Οι χαρακτηρισμοί έγιναν με βάση τον αμέσως προηγούμενο υποβαλλόμενο όρο. Στην περίπτωση που ένας όρος χαρακτηρίστηκε *λανθασμένος* ή *ακατάλληλος*, η σύγκριση έγινε με τον αμέσως προηγούμενο διαθέσιμο όρο.

Τα αποτελέσματα των ανασυνθέσεων παρουσιάζονται παρακάτω (και συνοπτικά στον Πίνακα 2).

- **Γενικότερος όρος:** το 20,0% των ανασυνθέσεων περιλάμβανε όρο γενικότερο του προηγούμενου.
- **Ειδικότερος όρος:** το 20,3% των ανασυνθέσεων περιλάμβανε όρο ειδικότερο του προηγούμενου.
- **Παράλληλος όρος:** το 47,6% των ανασυνθέσεων περιλάμβανε όρο παράλληλο του προηγούμενου.
- **Συνώνυμος όρος:** το 5,3% των ανασυνθέσεων περιλάμβανε όρο συνώνυμο του προηγούμενου.
- **Λανθασμένος όρος:** το 0,8% των ανασυνθέσεων περιλάμβανε όρο λανθασμένο, δηλαδή οι συμμετέχοντες χρησιμοποίησαν όρους που δεν υπάρχουν στην Ελληνική γλώσσα, όπως σημειώσαμε και παραπάνω στην ενότητα των *Ορισμών*.
- **Ακατάλληλος όρος:** το 6,0% των ανασυνθέσεων περιλάμβανε όρο ακατάλληλο για την περιγραφή των θεμάτων αναζήτησης, δηλαδή ο συγκεκριμένος όρος δεν επρόκειτο να λειτουργήσει κατά κανένα τρόπο ως συνδετικός κρίκος μεταξύ του αρχικού θέματος και σχετικών αποτελεσμάτων.

Γενικότερος όρος	Ειδικότερος όρος	Παράλληλος όρος	Συνώνυμος όρος	Λανθασμένος όρος	Ακατάλληλος όρος
20,0	20,3	47,6	5,3	0,8	6,0

Πίνακας Ν°2: Κατηγοριοποίηση όρων Ανασύνθεσης Ερωτημάτων (%)

#### 4.4. Μοναδικοί όροι και Αριθμός λέξεων ανά όρο

Ο συνολικός αριθμός των ερωτημάτων που υποβλήθηκαν από τους συμμετέχοντες και για τα πέντε θέματα ήταν 843. Τα ερωτήματα αποτελούνταν συνολικά από 1372 λέξεις. Από αυτές διαπιστώσαμε ότι μόλις 205 ήταν μοναδικές. Οι υπόλοιπες επαναλαμβάνονταν κατά την υποβολή των ερωτημάτων, δηλαδή μόνο ένα ποσοστό 15,0% του συνολικού αριθμού των λέξεων που χρησιμοποιήθηκαν ήταν μοναδικές. Αυτό είναι αρκετά αξιοσημείωτο αν ληφθεί υπόψη ότι η θεματολογία των ζητούμενων αναζητήσεων δεν ήταν αλληλο-επικαλυπτόμενη σε τέτοιο βαθμό ώστε να δικαιολογεί την υποβολή τόσο όμοιων ερωτημάτων σε κάθε περίπτωση.

Αναφορικά με τον αριθμό των λέξεων που χρησιμοποίησαν οι συμμετέχοντες ανά ερώτημα, τα αποτελέσματα έδειξαν ότι σε ποσοστό 57,7%, τα ερωτήματα αποτελούνταν από μία (1)



μόνο λέξη. Το 33,6% των ερωτημάτων περιλάμβανε δύο (2) λέξεις και το 8,8% τρεις (3) λέξεις. Τα αποτελέσματα αυτά είναι συμβατά με άλλες μελέτες που δείχνουν ότι, τα ερωτήματα των χρηστών, κατά κύριο λόγο, δε δομούνται από πολλές λέξεις. Για παράδειγμα, οι Jansen κ.α. (1998) σημειώνουν ότι κατά μέσο όρο κάθε ερώτημα περιλάμβανε 2,35 λέξεις, όπως και σε μεταγενέστερη μελέτη (Jansen, 2000) σημειώνεται ότι το 93,0% των ερωτημάτων στο Διαδίκτυο φτάνουν να περιέχουν έως τέσσερις λέξεις. Επίσης, σε έρευνα που περιλάμβανε τη μελέτη ενός δισεκατομμυρίου ερωτημάτων στη μηχανή αναζήτησης AltaVista επισημαίνεται ότι το 85,0% των ερωτημάτων περιείχαν λιγότερες από τρεις λέξεις ενώ ο μέσος όρος άγγιζε τις 2,35 λέξεις (Silverstein, 1999).

#### *4.5. Εννοιολογική συσχέτιση όρων με το Θησαυρό*

Το επόμενο βήμα αφορούσε στους όρους που εντοπίστηκαν στο Θησαυρό. Κρίναμε σημαντικό να μετρήσουμε και να διευκρινίσουμε πόσοι από τους όρους που χρησιμοποιήθηκαν από τους χρήστες για την υποβολή των ερωτημάτων στο σύστημα συσχετίζονταν εννοιολογικά με όρους του Θησαυρού EUROVOC. Η μέτρησή μας αφορά στους μοναδικούς όρους που εντοπίστηκαν και όχι στο σύνολο των όρων που υποβλήθηκαν. Σύμφωνα με τα αποτελέσματα, από τους 205 μοναδικούς όρους που αναγνωρίστηκαν, 124 συσχετίστηκαν εννοιολογικά με το Θησαυρό, δηλαδή ποσοστό 60,5%.

Ο εννοιολογικός χαρακτηρισμός, για το στάδιο της σύνθεσης και της ανασύνθεσης, των όρων αναζήτησης οι οποίοι δεν περιέχονταν στο θησαυρό έγινε από δύο κριτές, ειδικούς σε τεχνικές αναζήτησης. Για το λόγο αυτό έγινε έρευνα σε εργαλεία του γνωστικού πεδίου της μελέτης, δηλαδή του περιβάλλοντος, για τον εντοπισμό των συσχετίσεων. Δεχόμαστε ότι η συμβολή κριτών από άλλα γνωστικά αντικείμενα, όπως η γλωσσολογία και οι περιβαλλοντικές επιστήμες, θα προσέδιδαν επιπλέον προβληματισμούς και οπτικές στην ανάλυση των δεδομένων όχι όμως ανατροπή των αποτελεσμάτων. Αυτό το υποστηρίζουμε καθώς οι όροι που χρησιμοποιήθηκαν ήταν στο μεγαλύτερο ποσοστό τους ξεκάθαροι, εντοπίστηκαν εύκολα σε σχετικές πηγές και δεν ήταν εξειδικευμένοι ώστε να γεννούν αμφιβολίες για τη σημασία τους. Ωστόσο, σχετική επιβεβαίωση είναι στο μελλοντικό προγραμματισμό της έρευνάς μας.

### **5. Συμπεράσματα**

Τα αποτελέσματά μας βασίζονται στην επεξεργασία ελληνικών δεδομένων εφόσον και τα υποβαλλόμενα ερωτήματα και το εργαλείο εννοιολογικής συσχέτισης, δηλαδή ο Θησαυρός, ήταν στα ελληνικά ενώ αποδεικνύονται συμβατά με αντίστοιχες έρευνες που έχουν προκύψει από τις μελέτες που αξιοποίησαν ξενόγλωσσα δεδομένα. Κρίνεται, ωστόσο, χρήσιμη περαιτέρω εξέταση των δεδομένων καθώς μπορεί να επηρεάστηκαν από τα συγκεκριμένα θέματα που επιλέχτηκαν προς αναζήτηση. Αναγνωρίζοντας, επίσης, ότι σημαντικό ρόλο παίζει η επιλογή του ΣΟΓ για την εννοιολογική συσχέτιση, η διερεύνηση των υποθέσεών μας με τη χρήση επιπλέον ΣΟΓ αποτελεί μέρος του μελλοντικού μας σχεδιασμού.

Οι μετρήσεις αποκάλυψαν ότι, κατά τη *σύνθεση του ερωτήματος*, οι χρήστες, μετά την εμφανή συχνότητα χρήσης του *παρεχόμενου όρου* (41,6%), έχουν την τάση να στραφούν είτε σε ειδικότερο (25,2%) είτε σε κάποιο γενικότερο όρο (18,1%). Κατά την *ανασύνθεση* παρατηρείται ότι ο *παρεχόμενος* όρος της προηγούμενης φάσης δίνει τη σκυτάλη στον *παράλληλο* όρο με ποσοστό αντίστοιχα υψηλό (47,6%) και ακολουθούν αντίστοιχες παρατηρήσεις για τους *ειδικότερους* (20,3%) και *γενικότερους* όρους (20,0%). Διαπιστώνεται, επίσης, ότι η χρήση *συνώνυμων* όρων δεν είναι πολύ δημοφιλής μεταξύ των χρηστών, όπου μόλις το 5,9% των *συνθέσεων* και το 3,7% των *ανασυνθέσεων* περιλάμβανε χρήση *συνώνυμου* όρου. Κρίνεται, συνεπώς, χρήσιμη η πρόταση *συνώνυμων* όρων από ένα ΣΟΓ για τη διεύρυνση των ερωτημάτων αν αυτό γίνεται χωρίς την άμεση συμμετοχή των χρηστών καθώς δε διαπιστώνεται εξοικείωσή τους με *συνώνυμους* όρους εννοιών.

Τα υψηλά ποσοστά χρήσης *γενικότερου* όρου είναι ενθαρρυντικά για την αξιοποίηση της ιεραρχίας του Θησαυρού καθώς, όπως αναφέρεται στο Gray κ.α. (2010), κάθε πηγή που ανακτάται μέσω ενός συγκεκριμένου όρου, θα ανακτηθεί και χρησιμοποιώντας το *γενικότερο* όρο του. Η χρήση του *ειδικότερου* όρου, όπως σημειώνεται στην ίδια μελέτη, είναι λίγο πιο περίπλοκη καθώς ένας *ειδικότερος* όρος δεν είναι πάντα ειδίκευση του *γενικότερου*. Για να ενισχύσουν αυτή την άποψη δίνουν ως παράδειγμα τον όρο *τραπεζικός λογαριασμός*, ο οποίος μπορεί να βρεθεί κάτω από τον όρο *τράπεζα* αλλά δεν αποτελεί ειδίκευση του όρου καθώς δεν είναι κάποιο είδος τράπεζας, ώστε να προτεινόταν ως σχετικός αν κάποιος αναζητούσε κάποιο είδος τράπεζας ή κάποια τράπεζα με την επωνυμία της.

Η αποκάλυψη των περιορισμένων μοναδικών λέξεων που χρησιμοποιήθηκαν από τους χρήστες, μόλις 15,0% από το σύνολο των λέξεων, οδηγούν στο συμπέρασμα ότι οι χρήστες μπορεί να είναι γενικά απρόβλεπτοι και επιρρεπείς σε συμπεριφορικές αλλαγές, ωστόσο, κάτω από συγκεκριμένες συνθήκες, όπως η αναζήτηση σε συγκεκριμένη θεματική περιοχή, μπορεί να γίνουν προβλέψιμοι. Η παρατήρηση αυτή ενθαρρύνει την έρευνα για την ομαδοποίηση των χρηστών.

Εξετάζοντας το στοιχείο αυτό από μια άλλη οπτική μπορεί να υποστηριχτεί ότι ο περιορισμένος αριθμός μοναδικών όρων που χρησιμοποιήθηκαν θα μπορούσε να οφείλεται στις απαιτήσεις των θεμάτων που είχαν επιλεγεί για αναζήτηση. Για παράδειγμα, αν τα θέματα προς αναζήτηση ήταν αρκετά ειδικά, τότε δεν άφηναν περιθώρια στους χρήστες για υποβολή πρωτότυπων ερωτημάτων. Είναι αναμενόμενο ότι τα σημεία πρόσβασης περιορίζονται όταν η θεματολογία είναι εξειδικευμένη. Ακόμη κι έτσι όμως, πάλι στηρίζεται το συμπέρασμά μας καθώς μπορεί να αξιοποιηθεί σε συγκεκριμένο γνωστικό πεδίο και για τις ανάγκες συγκεκριμένου κοινού.

Οι μετρήσεις έδειξαν, επίσης, ότι οι χρήστες χρησιμοποιούν λίγες λέξεις για τη δόμηση ενός ερωτήματος. Το 57,7% των όρων περιλάμβανε μία (1) λέξη, το 33,6% περιλάμβανε δύο (2) λέξεις ενώ μόλις το 8,8% περιλάμβανε τρεις (3) λέξεις. Επιπλέον, κατά τη συσχέτιση των όρων που χρησιμοποιήθηκαν από τους χρήστες με όρους που υπάρχουν στο Θησαυρό, διαπιστώσαμε ότι, παρά το ότι χρησιμοποιήσαμε τη σχετική, κατά περίπτωση, ένότητα του Θησαυρού για την εννοιολογική συσχέτιση, το ποσοστό των όρων που ταυτίστηκε με όρους του Θησαυρού έφτασε μέχρι το 60,5%.

Ας θεωρήσουμε ότι ο χρήστης υποβάλλει μία λέξη στο σύστημα, π.χ. *περιβάλλον*. Αν θελήσουμε να διευρύνουμε το ερώτημά του χρησιμοποιώντας όρους του Θησαυρού, χωρίς να περιοριστούμε εκ των προτέρων στη σχετική θεματική ενότητα, εντοπίζουμε, μετά από λεξιλογική συσχέτιση του όρου, τη λέξη *περιβάλλον* σε 19 έννοιες του Θησαυρού. Μεταξύ αυτών υπάρχουν και οι όροι *πανεπιστημιακό περιβάλλον*, *σχολικό περιβάλλον*, *δίκαιο του περιβάλλοντος* και άλλοι. Αν ο όρος αναζήτησης αποτελείται από περισσότερες λέξεις, π.χ. *προστασία του περιβάλλοντος*, μετά από απλή συσχέτισή τους στο Θησαυρό, οδηγούμαστε πλέον σε δύο (2) μόνο έννοιες, δηλαδή στην επιλογή μεταξύ των όρων *οικονομικό μέσο για την προστασία του περιβάλλοντος* και *προστασία του περιβάλλοντος*. Συνεπώς, οι περισσότερες λέξεις για τη δόμηση ενός όρου ισοδυναμούν με παροχή επιπλέον στοιχείων για την ασφαλέστερη συσχέτιση της αναπαράστασης μιας έννοιας με σχετική της έννοια καθώς, με τον τρόπο αυτό, προσδιορίζεται ασφαλέστερα η έννοια την οποία αντιπροσωπεύουν. Αυτή η τεχνική είναι εξαιρετικά χρήσιμη και αποτελεσματική για την κάλυψη του ποσοστού των λέξεων που εισάγονται ως όροι αναζήτησης αλλά δε εντοπίζονται στο Θησαυρό ώστε να συσχετιστούν άμεσα κατά τη διαδικασία της διεύρυνσης.

Εφαρμόζοντας τεχνικές εγγύτητας των δεδομένων όρων με όρους του Θησαυρού δίνεται η δυνατότητα χρήσης ενός ικανού σημείου εκκίνησης για την εννοιολογική διεύρυνση του ερωτήματος του χρήστη. Με τον τρόπο αυτό αντιμετωπίζεται το εμπόδιο της *πολυσημίας* των λέξεων. Δίνουν, δηλαδή περισσότερες ευκαιρίες εντοπισμού των σχετικών εννοιών μέσα από τα συμφραζόμενα. Έτσι, εξασφαλίζονται ακόμη πιο σχετικά αποτελέσματα από ό,τι με την απλή λεξιλογική συσχέτιση των όρων. Αναλύοντας τα προηγούμενα συμπεράσματα καταλήγουμε στο ότι η χρήση ενός μόνο ΣΟΓ για την εννοιολογική *διεύρυνση των ερωτημάτων*, ιδιαίτερα αν δεν είναι θεματικά απόλυτα σχετικό με το θέμα αναζήτησης, δεν εγγυάται υψηλά ποσοστά επιτυχίας της μεθόδου.

## **6. Μελλοντική Έρευνα**

Οι μέθοδοι που ακολουθήσαμε και τα συμπεράσματα στα οποία καταλήξαμε αξιοποιώντας ελληνικά δεδομένα στο περιβάλλον ενός ηλεκτρονικού καταλόγου βιβλιοθήκης επιβεβαιώνονται από παρόμοιες μελέτες που έχουν χρησιμοποιήσει ξενόγλωσσα δεδομένα, κυρίως από το χώρο της Cross-Lingual Information Retrieval (Allan κ.α., 2003), όπου αναφέρεται πως ό,τι έχει εφαρμογή σε μονόγλωσσα δεδομένα έχει εφαρμογή και σε πολύγλωσσα. Για τη συνέχεια σημαντικό ρόλο για την περαιτέρω επιβεβαίωση των αρχικών μας αποτελεσμάτων θα αποτελέσει η εξέταση περισσότερων *Συστημάτων Οργάνωσης της Γνώσης* (ΣΟΓ) με συμμετοχή περισσότερων χρηστών και περισσότερων ερωτημάτων.

Αναφορικά με το μέλλον της γενικότερης έρευνάς μας θα μελετήσουμε τη δυνατότητα εφαρμογής συγκεκριμένων προτύπων αναζήτησης, τα οποία σχετίζονται με τη χρήση ΣΟΓ για *εννοιολογική διεύρυνση ερωτημάτων*, με ειδικότερη έμφαση στο αν αυτά μπορούν να αξιοποιηθούν για τα δεδομένα των ελληνικών βιβλιοθηκών και συγκεκριμένα για περιεχόμενο στην ελληνική γλώσσα. Στο πλαίσιο αυτό θα εξετάσουμε και την αξιοποίηση μοντέλων για *εννοιολογική συσταδοποίηση ερωτημάτων* (conceptual query clustering).

## Βιβλιογραφία

- ALLAN, J. [et al] (2003). Challenges in information retrieval and language modeling: report of a workshop held at the center for intelligent information retrieval, University of Massachusetts Amherst, September 2002. *SIGIR Forum* 37, 1 (Apr. 2003), p31-47.
- AULA, A. (2003). Query formulation in web information search. In: Isaias, P. and Karmakar, N. *Proceedings of IADIS International Conference WWW/Internet 2003*. Portugal: IADIS Press. p403-410.
- BARSKY, E. & BAR-ILAN, J. (2005). From the search problem through query formulation to results on the web. *Online Information Review*. 29 (1), p75-89.
- BELKIN, N. J., ODDY, R. & BROOKS, H. (1982). ASK for Information Retrieval. *Journal of Documentation*. 38 (1), p61-71. & 38 (2) p145-164
- COLARIC, S.M.. (2003). Instruction for web searching: an empirical study. *College & Research Libraries*. 64 (2), p111-122.
- DOERR, M. (2001). Semantic problems of thesaurus mapping. *Journal of Digital Information*. 1 (8).
- EFTHIMIADIS, E.N. (1996). Query expansion. *ARIST*. **31**, pp.121-187.
- GRAY, J.G., GRAY, N., HALL, C.W. & OUNIS, I. (2010). Finding the right term: retrieving and exploring semantic concepts in astronomical vocabularies. *Information Processing and Management*. 46 (4), p470-478.
- INGWERSEN, P. (1996). Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation*. 52 (1), 3-50.
- JANSEN, B.J., SPINK, A., BATEMAN, J. & SARACEVIC, T. (1998). Real life information retrieval: a study of user queries on the web. *SIGIR Forum*. 32 (1), p5-17.
- JANSEN, B. (2000). An investigation into the use of simple queries on Web IR systems. *Information Research: an Electronic Journal*. 6 (1).
- JOHO, H., SANDERSON, M. & BEAULIEU, M. (2004). A Study of User Interaction with a Concept-Based Interactive Query Expansion Support Tool. In: *Advances in Information Retrieval*. Berlin / Heidelberg: Springer. p42-56.
- NICHOLAS, D., HUNTINGTON, P. & JAMALI, H.R. (2008). User diversity: as demonstrated by deep log analysis. *The Electronic Library*. 26 (1), p21-38.
- RIEH, S.Y. & XIE, H. (2006). Analysis of multiple query reformulations on the web: the interactive information retrieval context. *Information Processing and Management*. **42**, p751-768.
- SHIRI, A. & REVIE, C. (2006). Query expansion behavior within a thesaurus-enhanced search environment: a user-centered evaluation. *JASIST*. 57 (4), p462-478.
- SILVERSTEIN, C., HENZINGER, M., MARAIS, H. & MORICZ, M. (1999). Analysis of a very large AltaVista query log. *SIGIR Forum*. 33 (1), p6-12.
- SPINK, A. & COLE, C. (2006). Human Information Behavior: integrating diverse approaches and information use. *JASIST*. 57 (1), p25-35.
- TUDHOPE, D., BINDING, C., BLOCKS, D. & CUNLIFFE, D. (2006). Query expansion via conceptual distance in thesaurus indexed collections. *Journal of Documentation*. 62 (4), p509-533.
- WILSON, T.D. (1997). Information behaviour: an interdisciplinary perspective. *Information Processing and Management*. 33 (4), p551-572.