Ενισχύοντας σημασιολογικά τις διαδικασίες αναζήτησης σε ένα περιβάλλον μετα-αναζήτησης

Enabling search semantics in the library meta-search environments

Μιχάλης Σφακάκης και Σαράντος Καπιδάκης

Τμήμα Αρχειονομίας Βιβλιοθηκονομίας / Ιόνιο Πανεπιστήμιο
Πλατεία Ελευθερίας, Παλαιό Ανάκτορο, 491 00 Κέρκυρα
{sfakakis, sarantos}@ionio.gr

Περίληψη

Παρέχουμε ένα δυναμικό μηχανισμό αντικατάστασης μη υποστηριζόμενων Σημείων Πρόσβασης από τις Z39.50 πηγές. Η αντικατάσταση αυτή βασίζεται στη σημασιολογική συσχέτιση των Σημείων Πρόσβασης που εξάγεται έμμεσα από τα μεταδεδομένα που ορίζουν το κάθε ένα από αυτά. Με το μηχανισμό αυτό αποφεύγονται οι αποτυχημένες ή οι ασυνεπείς επερωτήσεις και βελτιώνεται η απόδοση στα περιβάλλοντα μετα-αναζήτησης των Z39.50 πηγών. Η λειτουργικότητα αυτή παρέχεται ως ελεύθερη υπηρεσία, για συγκεκριμένες Ελληνικές και επιλεγμένες Z39.50 πηγές, από τη διεύθυνση http://dlib.ionio.gr/zSAPN.

Εισαγωγή

Για την αντιμετώπιση της πληθώρας και ποικιλομορφίας των διαθέσιμων πηγών πληροφόρησης μια κοινή πρακτική αποτελεί η δημιουργία και χρήση περιβαλλόντων μετα-αναζήτησης. Η υλοποίηση ενός τέτοιου περιβάλλοντος προϋποθέτει την ύπαρξη των πρωταρχικών πηγών και ταυτόχρονα τη συνεχή και απρόσκοπτη διάθεσή τους μέσα από σαφώς καθορισμένες και σταθερές διαδικασίες και μεθόδους διάθεσης.

Ο βαθμός απόδοσης ενός τέτοιου περιβάλλοντος καθορίζεται από τη δυνατότητα που έχει να αναζητήσει ταυτόχρονα πολλές πηγές και επιπλέον την επιτυχία του σχετικά με την ανάκληση/ακρίβεια (recall/precision) κατά την αναζήτηση, χρησιμοποιώντας τις διαθέσιμες μεθόδους και διαδικασίες αναζήτησης των υφισταμένων πηγών. Αξίζει να σημειωθεί ότι τα προβλήματα αυτά έχουν ήδη τεθεί και αντιμετωπισθεί, με περισσότερη ή λιγότερη επιτυχία, κατά την υλοποίηση συστημάτων ενιαίας αναζήτησης και ολοκλήρωσης πληροφοριών (information integration), συλλογικών καταλόγων (εικονικών ή φυσικών), συγκομιδής πληροφοριών (information harvesting) κλπ.

Στην προσέγγιση των παραπάνω προβλημάτων είναι πλέον δεδομένη η αναγκαιότητα και η εφαρμογή αφενός κοινών κανόνων, προτύπων και πολιτικών υλοποίησης στις διαδικασίες που αφορούν τη δημιουργία των πηγών, και αφετέρου η διάθεσή τους με κοινά πρωτοκόλλα και μεθόδους υλοποίησης. Ειδικότερα, η δεδομένη καθολική χρήση του πρωτοκόλλου Z39.50 στο χώρο των βιβλιοθηκών το καθιστά ένα βασικό ρυθμιστή απόδοσης στα περιβάλλοντα μετα-αναζήτησης που υλοποιούνται.

Οι αποτυχημένες ή οι ασυνεπείς απαντήσεις είναι ένα κυρίαρχο πρόβλημα που δημιουργούν οι διαφορετικές υλοποιήσεις του Z39.50 πρωτοκόλλου. Οι αποτυχημένες ή οι ασυνεπείς

απαντήσεις προκύπτουν όταν μια Z39.50 πηγή δεν υποστηρίζει ένα Σημείο Πρόσβασης στις επερωτήσεις που δέχεται. Για την αντιμετώπιση του προβλήματος αυτού, στο εργαστήριο ψηφιακών βιβλιοθηκών και ηλεκτρονικής δημοσίευσης του Τμήματος Αρχειονομίας και Βιβλιοθηκονομίας του Ιονίου Πανεπιστημίου, έχουμε υλοποιήσει ένα σύστημα που αντικαθιστά τα μη υποστηριζόμενα Σημεία Πρόσβασης με άλλα που υποστηρίζει η Z39.50 πηγή. Η αντικατάσταση αυτή βασίζεται στη σημασιολογία που μπορεί να εξαχθεί έμμεσα από τη σχετική γνώση των μετα-δεδομένων δημιουργίας και των χαρακτηριστικών αναζήτησης της κάθε πηγής.

Σημείο Πρόσβασης

Σε έναν κατάλογο ή σε μία βάση δεδομένων ως σημείο πρόσβασης μπορεί να θεωρηθεί οποιοδήποτε μέρος της κάθε εγγραφής με το οποίο μπορούμε να αναζητήσουμε, είτε ακόμα και να ταυτίσουμε, τις οντότητες που περιγράφονται. Ένα όνομα, μία θεματική επικεφαλίδα, ένας ταξινομικός αριθμός είναι μερικά παραδείγματα σημείων πρόσβασης. Σε ένα πληροφοριακό σύστημα όμως, αυτά τα σημεία πρόσβασης ομαδοποιούνται με βάση κάποια κριτήρια, δημιουργώντας έτσι τα πεδία (ή ευρετήρια) τα οποία μπορούμε να χρησιμοποιήσουμε στις επερωτήσεις μας κατά τις διαδικασίες αναζήτησης. Είναι αρκετά συνηθισμένο στα περιβάλλοντα των πληροφοριακών συστημάτων τα πεδία αναζήτησης να αναφέρονται και αυτά ως Σημεία Πρόσβασης.

Κάθε πληροφοριακό σύστημα μπορεί να χρησιμοποιεί διαφορετικά κριτήρια επιλογής και ομαδοποίησης των σημείων πρόσβασης καθορίζοντας με αυτό το τρόπο τη σημασιολογία του πεδίου αναζήτησης. Κατά συνέπεια, ένα πληροφοριακό σύστημα μπορεί να χρησιμοποιεί τα σημεία πρόσβασης που δημιουργήθηκαν από τον κύριο τίτλο και τον παράλληλο τίτλο κάθε εγγραφής και να δημιουργεί το πεδίο αναζήτησης του τίτλου. Κάποιο άλλο σύστημα, για το πεδίο αναζήτησης του τίτλου, μπορεί να χρησιμοποιεί μόνο το σημείο πρόσβασης του κύριου τίτλου από κάθε εγγραφή. Συχνά, μολονότι χρησιμοποιούνται τα ίδια σημεία πρόσβασης κατά την επιλογή και ομαδοποίηση τους από διαφορετικά πληροφοριακά συστήματα οι ονομασίες των πεδίων αναζήτησης μπορεί να διαφέρουν.

Στην συνέχεια της παρούσας εργασίας, εκτός και αν αναφέρεται διαφορετικά, ο όρος *Σημείο Πρόσβασης* θα είναι αυτός όπως ορίζεται στο περιβάλλον των πληροφοριακών συστημάτων. Θα εννοούνται, δηλαδή, τα πεδία τα οποία μπορούν να χρησιμοποιηθούν σε επερωτήσεις κατά τις διαδικασίες αναζήτησης και τα οποία προκύπτουν από την ομαδοποίηση των σημείων πρόσβασης και όχι τα ίδια τα σημεία πρόσβασης.

Το πρωτόκολλο Z39.50 και τα Σημεία Πρόσβασης του Bib-1

Το Z39.50 [1, 3] είναι ένα πρωτόκολλο επικοινωνίας, το οποίο ορίζει τον τρόπο με τον οποίο θα επικοινωνήσουν δύο διαφορετικά υπολογιστικά συστήματα, με βάση την αρχιτεκτονική Client/Server, με απώτερο σκοπό την ανάκτηση πληροφοριών. Στο πλαίσιο μιας τέτοιας επικοινωνίας, διαφορετικές πληροφοριακές πηγές (targets), που διατίθενται από αντίστοιχους Z-servers, μπορούν να αναζητηθούν ταυτόχρονα από έναν Z-client (origin). Ουσιαστικά, το πρωτόκολλο θέτει τις βάσεις για τη λύση στο πρόβλημα διαλειτουργικότητας που υπάρχει ανάμεσα στις διαφορετικές πηγές λόγω ασυμβατότητας ή/και ανομοιογένειας των δεδομένων τους.

Με βάση το πρωτόκολλο Z39.50, ο client στέλνει την ερώτηση και οι servers αναλαμβάνουν την αποκωδικοποίηση και την εκτέλεσή της. Κάθε ερώτημα μπορεί να συνοδεύεται από σειρά χαρακτηριστικών, τα οποία το προσδιορίζουν και το επεξηγούν. Το πιο βασικό, και ουσιαστικά υποχρεωτικό, χαρακτηριστικό κάθε ερωτήματος είναι το πεδίο ή τα πεδία στα οποία θα πρέπει να διεξαχθεί η αναζήτηση (use attribute ή Access Point attribute Type). Το χαρακτηριστικό αυτό μας προσδιορίζει τα Σημεία Πρόσβασης με τα οποία θα γίνει ο έλεγχος ταύτισης του όρου αναζήτησης.

Τα υπόλοιπα χαρακτηριστικά που στέλνει ο client (σύμφωνα με το Bib-1) ανήκουν σε 5 κατηγορίες, κάθε μία από τις οποίες μπορεί να πάρει διάφορες τιμές. Οι κατηγορίες είναι οι εξής: relation = σχέση, truncation = αποκοπή, structure = δομή, position = θέση και completeness = πληρότητα. Κάθε μία από τις κατηγορίες περιέχει μία σειρά από τιμές οι οποίες ορίζονται στο Bib-1 attribute set. Είναι προφανές ότι η πληθώρα αυτή των χαρακτηριστικών αυξάνει την δυνατότητα αλλά και την πολυπλοκότητα της επικοινωνίας και οξύνει τα προβλήματα διαλειτουργικότητας μεταξύ των συστημάτων. Η παρούσα εργασία αναφέρεται μόνο στα προβλήματα που προκύπτουν από τη μη υποστήριξη των Σημείων Πρόσβασης. Μια εμπεριστατωμένη περιγραφή για τα προβλήματα που δημιουργούνται από τις διαφορετικές υλοποιήσεις των υπολοίπων χαρακτηριστικών γίνεται στο [7].

Σύμφωνα με το πρωτόκολλο, υπάρχουν δύο εναλλακτικοί τρόποι απάντησης όταν η πηγή δεν υποστηρίζει κάποιο από τα χαρακτηριστικά της επερώτησης. Ο πρώτος τρόπος απάντησης δίνει τη δυνατότητα στην πηγή να απορρίψει την επερώτηση και να στείλει το κατάλληλο διαγνωστικό μήνυμα λάθους (αποτυχημένη επερώτηση). Με τον τρόπο αυτό ο χρήστης, παρότι δε λαμβάνει απάντηση, ενημερώνεται για το ότι το σύστημα δεν υποστηρίζει το χαρακτηριστικό της επερώτησής του. Βέβαια, σε ένα περιβάλλον όπου η επερώτηση υποβάλλεται ταυτόχρονα σε πολλές πηγές είναι αρκετά πιθανό ο χρήστης να λάβει μόνο διαγνωστικά μηνύματα λάθους, ή να μη δει ή ξεχωρίσει τα μηνύματα αυτά ανάμεσα σε αποτελέσματα άλλων πηγών. Ο δεύτερος τρόπος απάντησης δίνει τη δυνατότητα στην πηγή να αντικαταστήσει αυθαίρετα το μη υποστηριζόμενο χαρακτηριστικό με κάποιο άλλο που υποστηρίζει (ασυνεπής απάντηση). Στην περίπτωση αυτή ο χρήστης θα λάβει απάντηση χωρίς να γνωρίζει πώς προήλθε και το σπουδαιότερο, δε θα ενημερωθεί για το ότι έγινε η αντικατάσταση. Το πλεονέκτημα στην περίπτωση αυτή, και ιδιαίτερα σε περιπτώσεις μη έμπειρων χρηστών που αναζητούν ταυτόχρονα πολλές πηγές, είναι ότι ο χρήστης θα λάβει τουλάχιστον κάποια σχετική απάντηση και δε θα αποτραπεί από την έρευνά του.

Σε σχέση με την υποστήριξη του χαρακτηριστικού του *Σημείου Πρόσβασης*, τα περισσότερα συστήματα διεθνώς υλοποιούν την πρώτη προσέγγιση. Ενδεικτικά αναφέρονται τα MELVYL, COPAC, Library and Archives Canada, κλπ. Το σύστημα της Library of Congress, αντίθετα, είναι ένα από τα συστήματα που έχουν υλοποιήσει τη δεύτερη προσέγγιση. Συνεπώς, σε περίπτωση που κάνουμε π.χ. την επερώτηση με το σημείο πρόσβασης *Author-Title-Subject = 'Shakespeare'*, δε θα μας ενημερώσει ότι δεν το υποστηρίζει[1] αλλά θα μας απαντήσει ότι βρήκε περισσότερες από 10.000 περιπτώσεις. Τα στοιχεία δείχνουν ότι πιθανότατα γίνεται αντικατάσταση από το σύστημα του σημείου πρόσβασης που ορίστηκε κατά την επερώτηση με το *Any*. Αυτό προκύπτει και από την παρατήρηση ότι η αναζήτηση με αυτό το σημείο πρόσβασης το οποίο υποστηρίζεται (*Any*) δίνει τα ίδια αποτελέσματα. Στην Ελλάδα, όλες οι πηγές που ελέγχθηκαν υλοποιούν την πρώτη προσέγγιση απορρίπτοντας την επερώτηση με αποστολή διαγνωστικού μηνύματος λάθους.

[1] http://www.loc.gov/z3950/lcserver.html

Στατιστικές μελέτες έχουν δείξει ότι υπάρχουν πολλές διαφορετικές υλοποιήσεις του πρωτοκόλλου Z39.50 και ότι όσο μεγαλύτερος είναι ο αριθμός των Z39.50 πηγών τόσο μειώνεται ο αριθμός των κοινών χαρακτηριστικών αναζήτησης. Ειδικότερα, αποτελέσματα που δημοσιεύονται στο δικτυακό τόπο της IndexData[2] και αφορούν στα δέκα περισσότερο υποστηριζόμενα Σημεία Πρόσβασης από τις περιγραφές 3.018 πηγών, δείχνουν ότι δεν υπάρχει ένα Σημείο Πρόσβασης που να υποστηρίζεται από το σύνολο των 1.697 πηγών που υλοποιούν την υπηρεσία αναζήτησης (search service). Το Σημείο Πρόσβασης που υποστηρίζεται από τις περισσότερες πηγές είναι το Title (use attribute 4) και το υποστηρίζουν 1.591 ή 93,75%, επόμενο είναι το Subject Heading (use attribute 21) που το υποστηρίζουν 1.555 ή 91,63%, και ακολουθείται από το Author (use attribute 1003) που το υποστηρίζουν 1.549 ή 91,28%.

Σε αντίστοιχη έρευνα, ελέγχου των Σημείων Πρόσβασης, που κάναμε σε δείγμα 24 Z39.50 πηγών στην Ελλάδα, διαπιστώθηκε ότι υπάρχει ένα μόνο Σημείο Πρόσβασης που το υποστηρίζουν όλες οι πηγές, το Author (use attribute 1003). Στη συνέχεια ακολουθούν τα Subject Heading (use attribute 21) και Title (use attribute 4) κάθε ένα από τα οποία υποστηρίζεται από 23 διαφορετικές πηγές. Σύμφωνα με τα παραπάνω γίνεται αντιληπτό ότι η κατάσταση στην Ελλάδα εμφανίζεται καλύτερη από τη διεθνή που δείχνουν τα στατιστικά της IndexData, παρά το ότι η σειρά των Σημείων Πρόσβασης είναι διαφορετική. Για τα υπόλοιπα χαρακτηριστικά αναζήτησης (π.χ. δομή, αποκοπή, κλπ), εκτός από το Σημείο Πρόσβασης, δεν υπάρχει αντίστοιχα ο ίδιος βαθμός υποστήριξης και ισχύουν αυτά που περιγράφονται στο [7]. Τέλος, πληροφοριακά αναφέρουμε ότι στο δείγμα της Ελλάδας ο μέγιστος αριθμός υποστηριζόμενων Σημείων Πρόσβασης είναι 37 και ο ελάχιστος είναι 7.

Από τα παραπάνω είναι εμφανές ότι όσο αυξάνεται ο αριθμός των πηγών τόσο αυξάνεται και η πιθανότητα για αποτυχημένη επερώτηση ή ασυνεπή απάντηση. Το πρόβλημα είναι ακόμα εντονότερο σε περιπτώσεις όπου είναι αναγκαίο να γίνουν πιο συγκεκριμένες επερωτήσεις, π.χ. Author-name conference ή use attribute 1006, και να χρησιμοποιηθούν πιο συγκεκριμένα Σημεία Πρόσβασης εκτός από τα τρία που αναφέρθηκαν στις προηγούμενες παραγράφους. Όπως ήδη αναφέραμε, με δεδομένο ότι όλες οι Ελληνικές Z39.50 πηγές απορρίπτουν την επερώτηση στέλνοντας διαγνωστικό μήνυμα όταν δεν υποστηρίζουν ένα Σημείο Πρόσβασης, ασυνεπείς απαντήσεις θα προκύψουν μόνο σε περιπτώσεις συνδυασμού τους με άλλες διεθνείς πηγές.

Σημασιολογική συσχέτιση Σημείων Πρόσβασης

Η σημασιολογία του κάθε Σημείου Πρόσβασης έχει ορισθεί στο [2], το οποίο εκφράζει την αμοιβαία συναίνεση μεταξύ των μελών της ομάδας Z39.50 Implementors Group (ZIG). Για τον προσδιορισμό της σημασιολογίας, δηλαδή ποιο θα είναι το σύνολο της πληροφορίας που θα επιλεχθεί, έχουν χρησιμοποιηθεί οι ετικέτες των πεδίων του MARC Bibliographic προτύπου. Για παράδειγμα, ο ορισμός της σημασιολογίας του Σημείου Πρόσβασης *Author-name-Personal* (ή *use* = 1004) καθορίζει ότι το Σημείο Πρόσβασης θα δημιουργηθεί από το σύνολο των δεδομένων με τις MARC ετικέτες πεδίων {100, 400, 700, 800}.

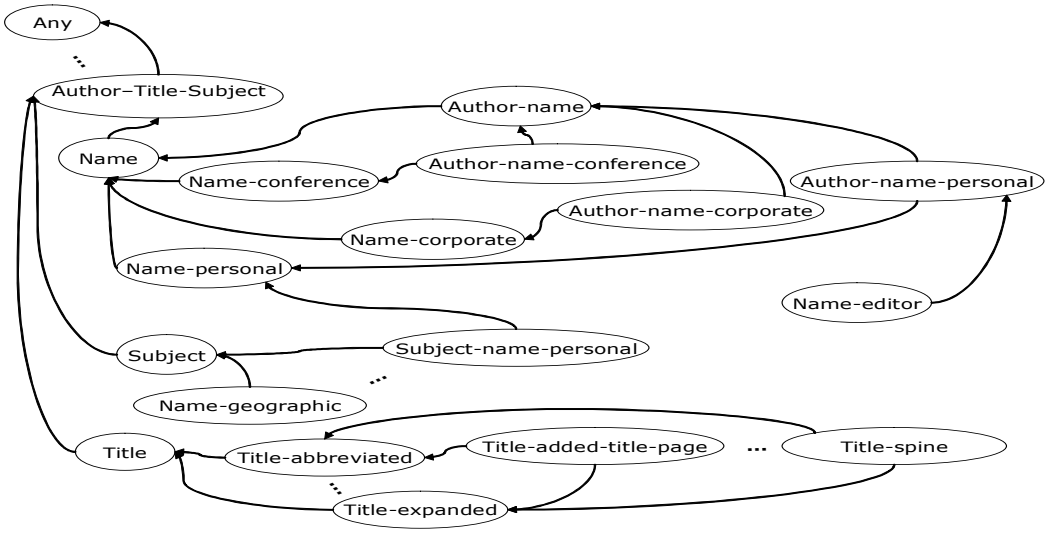
Βασισμένοι στις σχέσεις υποσυνόλων των Σημείων Πρόσβασης, όπως αυτές προκύπτουν από την προκαθορισμένη σημασιολογία των Bib-1 σημείων πρόσβασης του πρωτοκόλλου Z39.50,

[2] http://targettest.indexdata.com/stats.html. Τελευταία ημερομηνία πρόσβασης 23 Αυγούστου 2007

αποδώσαμε τις μεταξύ τους σημασιολογικές συσχετίσεις και τις περιγράφουμε με την γλώσσα RDF Schema [6] (εικόνα 1).

Θεωρούμε ένα Σημείο Πρόσβασης υποσύνολο ενός άλλου εάν το σύνολο των πεδίων που ορίζονται για να το δημιουργήσουν είναι υποσύνολο των πεδίων που ορίζονται για να δημιουργήσουν το άλλο. Για παράδειγμα, ας θεωρήσουμε το Σημείο Πρόσβασης *Author-name* το οποίο, σύμφωνα με τον ορισμό του στο [2], δημιουργείται από τα δεδομένα των πεδίων που ορίζουν το σύνολο {100, 110, 111, 400, 410, 411, 700, 710, 711, 800, 810, 811}, και ακόμα, το Σημείο Πρόσβασης *Author-name-personal* το οποίο δημιουργείται από τα δεδομένα των πεδίων που ορίζουν το σύνολο {100, 400, 700, 800}. Κατά συνέπεια, το Σημείο Πρόσβασης *Author-name- personal* θεωρείται ως υποσύνολο του *Author-name*.

Η εικόνα 1 δείχνει ένα αντιπροσωπευτικό δείγμα από τον RDFS γράφο της σημασιολογικής συσχέτισης των Bib-1 Σημείων Πρόσβασης του Z39.50 πρωτοκόλλου. Η πλήρης περιγραφή της μεθόδου για τη δημιουργία του γράφου είναι στο [4, 5], ενώ ο πλήρης γράφος σε διάταξη dot και εικόνα του βρίσκεται στη διεύθυνση http://dlib.ionio.gr/lab/graph.htm.



Εικόνα 1. Απλοποιημένο αντιπροσωπευτικό δείγμα από τον RDF Schema γράφο της σημασιολογικής συσχέτισης των Bib-1 Σημείων Πρόσβασης του Z39.50 πρωτοκόλλου.

Αντικατάσταση μη υποστηριζόμενων Σημείων Πρόσβασης

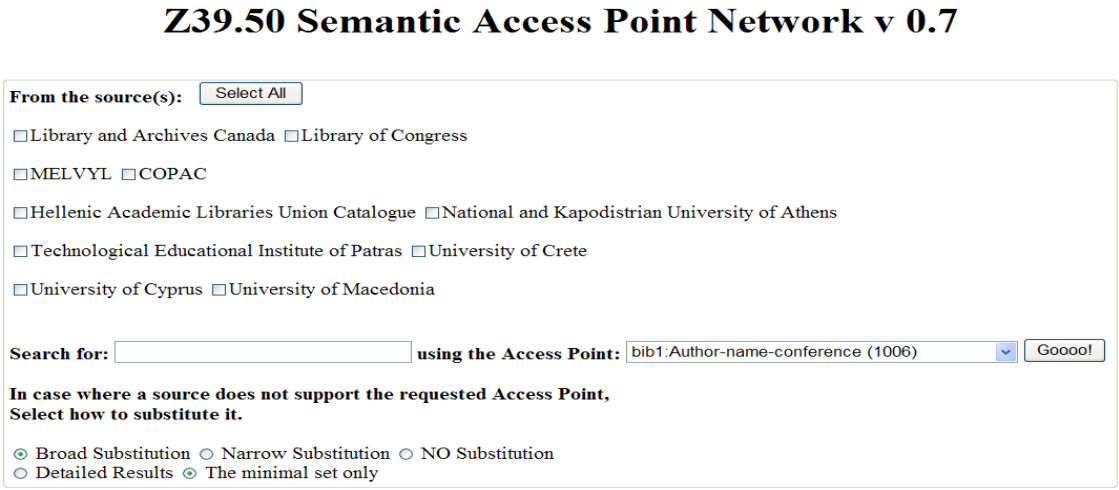
Με δεδομένα τα προβλήματα που τίθενται στην περίπτωση που μια πηγή δεν υποστηρίζει ένα Σημείο Πρόσβασης, καθώς επίσης το ότι είναι σχεδόν αδύνατον όλες οι πηγές να υποστηρίζουν ακριβώς τα ίδια χαρακτηριστικά, το ερώτημα που τίθεται είναι: Πώς θα μπορούσε μία μηχανή μετα-αναζήτησης να αντικαταστήσει το μη υποστηριζόμενο Σημείο Πρόσβασης με κάποια άλλα, έτσι ώστε η σημασιολογία τους να είναι αρκετά κοντά/συγγενική, αν όχι ίδια, με αυτή που έχει το μη υποστηριζόμενο Σημείο Πρόσβασης;

Η προσέγγισή μας στο πρόβλημα αυτό γίνεται χρησιμοποιώντας τη θέση κάθε Σημείου Πρόσβασης στο σημασιολογικό γράφο συσχετίσεων των Σημείων Πρόσβασης. Για παράδειγμα, ένα Σημείο Πρόσβασης που μπορεί να χρησιμοποιηθεί είναι το *Author-Title-Subject*, υποδηλώνοντας ότι ο όρος αναζήτησης θα πρέπει να ταυτιστεί μόνο με πληροφορίες από τα πεδία Author ή Title ή Subject. Από την εικόνα 1, το Σημείο Πρόσβασης *Author-Title-Subject* μπορεί να θεωρηθεί υπερσύνολο των Σημείων Πρόσβασης *Name, Title, Subject,*

Music-key, and *Identifier-thematic*. Επίσης, το *Author-Title-Subject* μπορεί να θεωρηθεί υποσύνολο του Σημείου Πρόσβασης *Any*. Σε περίπτωση επερώτησης με το *Author-Title-Subject* σε μια πηγή που δεν το υποστηρίζει, αυτό το Σημείο Πρόσβασης θα μπορούσε να αντικατασταθεί από τη λογική ένωση των υποστηριζόμενων Σημείων Πρόσβασης *Author*, *Title* και *Subject* (εφόσον η πηγή υποστηρίζει αυτά τα τρία Σημεία Πρόσβασης). Βέβαια, με τον τρόπο αυτό περιορίζουμε τα αποτελέσματα της ερώτησης (περιορίζουμε τη σημασιολογία της). Εναλλακτικά, σε περίπτωση που επιθυμούσαμε να επεκτείνουμε το σύνολο των αποτελεσμάτων θα μπορούσαμε να το αντικαταστήσουμε με το *Any* (επεκτείνουμε τη σημασιολογία της). Η πλήρης περιγραφή της μεθόδου και των αλγορίθμων που υλοποιούν την αντικατάσταση γίνεται στο [5].

Παρουσίαση συστήματος

Το σύστημα που υλοποιεί την αντικατάσταση των μη υποστηριζόμενων Σημείων Πρόσβασης είναι διαθέσιμο στη διεύθυνση http://dlib.ionio.gr/zSAPN. Η εικόνα 2 δείχνει την αρχική σελίδα του συστήματος, από όπου ο χρήστης μπορεί να επιλέξει τις πηγές στις οποίες θα αναζητήσει, τον όρο αναζήτησης, το Σημείο πρόσβασης και τέλος, σε περίπτωση που είναι επιθυμητό, τον τρόπο με τον οποίο θα γίνει η αντικατάσταση του Σημείου Πρόσβασης σε περίπτωση που δεν υποστηρίζεται από μία πηγή.



Εικόνα 2. Αρχική οθόνη συστήματος

Αρχικά, το σύστημα πριν στείλει την επερώτηση σε κάθε επιλεγμένη πηγή ελέγχει αν αυτή υποστηρίζει το Σημείο Πρόσβασης. Στην περίπτωση που το υποστηρίζει το στέλνει. Διαφορετικά, εφόσον έχουμε επιλέξει να γίνει αντικατάσταση και λαμβάνοντας υπόψη τον τρόπο που έχουμε επιλέξει (περιορισμό ή επέκταση) μετασχηματίζει την επερώτηση χρησιμοποιώντας Σημεία Πρόσβασης που υποστηρίζει η πηγή. Είναι πιθανό να μην είναι δυνατή η αντικατάσταση σύμφωνα με τις παραμέτρους που έχουν επιλεγεί, ενώ για κάθε πηγή να προκύψει διαφορετική αντικατάσταση σύμφωνα με τα Σημεία Πρόσβασης που υποστηρίζει.

Ο πίνακας 1 μας δείχνει μέρος από τη σύνοψη των αποτελεσμάτων της αντικατάστασης που εφαρμόζει το σύστημα επιλέγοντας αντικατάσταση με επέκταση της σημασιολογίας του Σημείου Πρόσβασης. Η επερώτηση αναζητά από τις πηγές τα συνέδρια της IEEE (*Author-name-conference_1006 = IEEE*). Όπως φαίνεται από τον πίνακα 1, η πηγή *Hellenic Academic Libraries Union Catalogue* είναι η μόνη που υποστηρίζει το Σημείο Πρόσβασης. Για τις

πηγές *National and Kapodistrian University of Athens* και *University of Crete* οι οποίες δεν το υποστηρίζουν, γίνεται αντικατάσταση με το Σημείο Πρόσβασης *Author-name_1003*. Αντίστοιχα, για την πηγή *Library of Congress*, η οποία επίσης δεν το υποστηρίζει, το Σημείο Πρόσβασης *Author-name-conference* θα αντικατασταθεί από την τομή των Σημείων Πρόσβασης *Author-name_1003* και *Name-conference_3*. Συγκρίνοντας σημασιολογικά τις νέες επερωτήσεις που προέκυψαν από την αντικατάσταση του μη υποστηριζόμενου Σημείου Πρόσβασης, παρατηρούμε ότι η νέα επερώτηση για την πηγή *Library of Congress* σημασιολογικά είναι 'πιο κοντά' στην αρχική επερώτηση από αυτήν που προέκυψε για τις άλλες δύο πηγές.

Πίνακας 1. Μέρος σύνοψης αποτελεσμάτων αντικατάστασης

Source	Hits	Query
Library of Congress	1661	Author-name-conference_1006 = IEEE From the Access Point Substitution... --The Minimal Set is: Author-name_1003 Name-conference_3
Hellenic Academic Libraries Union Catalogue	1036	Author-name-conference_1006 = IEEE The Access Point is supported by the source.
National and Kapodistrian University of Athens	95	Author-name-conference_1006 = IEEE From the Access Point Substitution... --The Minimal Set is: Author-name_1003
University of Crete	339	Author-name-conference_1006 = IEEE From the Access Point Substitution... --The Minimal Set is: Author-name_1003

Αναφορές

[1] ANSI/NISO: Z39.50 Information Retrieval: application service definition and protocol specification: approved May 10, 1995.

[2] Attribute Set BIB-1 (Z39.50-1995): Semantics. ftp://ftp.loc.gov/pub/z3950/defs/bib1.txt.

[3] Finnigan, S., Ward, N. Z39.50 Made Simple. Available from: http://archive.dstc.edu.au/DDU/projects/Z3950/zsimple.html.

[4] Sfakakis, M., Kapidakis, S. A Semantics-Based Graph for the Bib-1 Access Points of the Z39.50 Protocol. LNCS 4172, Proceedings of the 10th European Conference on Research and Advanced Technology for Digital Libraries, Vienna 2006, pp. 445 – 448.

[5] Sfakakis, M., Kapidakis, S. A Semantics-Based Substitution for Unsupported Z39.50 Bib-1 Access Points, Proceedings of the 2nd International Conference on Digital Information Management (ICDIM 2007), October 28-31, 2007, Lyon, France 2007, [υπό έκδοση].

[6] W3C. Resource Description Framework (RDF) Schema Specification 1.0. W3C Candidate Recommendation 27 March 2000. Available from: http://www.w3.org/TR/2000/CR-rdf-schema-20000327/.

[7] Μανόλης Πεπονάκης, Ντίνα Τρούτπεγλη, Μιχάλης Σφακάκης, «Προβλήματα διαλειτουργικότητας κατά την ταυτόχρονη πρόσβαση σε πηγές μέσω του πρωτοκόλλου Z39.50 και το περιβάλλον πρόσβασης *Η Αργώ*», 13ο Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών, Οκτώβριος 10-13, 2004, [τεκμήριο www, URL: http://www.ionio.gr/libconf/pdfs/PeponakisZ3950_Interoperability.pdf]