

Προβλήματα διαλειτουργικότητας κατά την ταυτόχρονη πρόσβαση σε πηγές μέσω του πρωτοκόλλου Z39.50 και το περιβάλλον πρόσβασης "Η ΑΡΓΩ"

Interoperability problems regarding the concurrent access in resources via the Z39.50 protocol & "The ARGO" Web Gate

Μανόλης Πεπονάκης, Ντίνα Τρούτπεγλη, Μιχάλης Σφακάκης
Εθνικό Κέντρο Τεκμηρίωσης (ΕΚΤ) / ΕΙΕ
Βασ. Κωνσταντίνου 48
11635, Αθήνα
{epepo, troutpegli, [msfaka](mailto:msfaka@ekt.gr)}@ekt.gr

Manolis Peponakis, Ntina Troutpegli, Michalis Sfakakis
National Documentation Centre (NDC) / NHRF
48 Vassileos Constantinou Av
GR-11635, Athens
{epepo, troutpegli, [msfaka](mailto:msfaka@ekt.gr)}@ekt.gr

Περίληψη

Στην εργασία αυτή εξετάζονται προβλήματα διαλειτουργικότητας στο πλαίσιο μιας ενιαίας πρόσβασης σε ένα κατακευματισμένο περιβάλλον εργασίας, όπου οι διαφορετικοί βιβλιογραφικοί κατάλογοι διατίθενται μέσω του Z39.50 πρωτοκόλλου. Το σημείο αναφοράς είναι η δυνατότητα ταυτόχρονης αναζήτησης σε διαφορετικούς καταλόγους βιβλιοθηκών (πηγές) με βάση το πρωτόκολλο, καθώς και η ένταξη των αποτελεσμάτων (βιβλιογραφικών εγγραφών) που προκύπτουν από κάθε πηγή σε ένα ενιαίο σύνολο αποτελεσμάτων. Κάποια από τα προβλήματα αυτά πηγάζουν από τις διαφορετικές υλοποιήσεις του πρωτοκόλλου ενώ άλλα προέρχονται από την ανομοιογένεια που παρουσιάζουν οι ίδιες οι βάσεις δεδομένων των καταλόγων αυτών, τόσο σε επίπεδο δομής (π.χ. διαφορετικά πρότυπα κωδικοποίησης) όσο και σε επίπεδο περιεχομένου.

Τα σημεία που χρήζουν ιδιαίτερης εξέτασης εντοπίζονται τόσο στη διαδικασία της αναζήτησης όσο και στη διαδικασία της ανάκτησης των εγγραφών. Οι δύο διαδικασίες εξετάζονται μεμονωμένα, τόσο γιατί αποτελούν δύο διακριτά στάδια, σύμφωνα με το πρωτόκολλο, αλλά κυρίως λόγω των διαφορετικών προβλημάτων που προκύπτουν σε κάθε στάδιο.

Κατά τη διαδικασία της αναζήτησης, τα προβλήματα εστιάζονται κυρίως γύρω από τον εντοπισμό εγγραφών και πολύ περισσότερο από το αν τελικά το σύνολο αποτελεσμάτων ανταποκρίνεται στα πραγματικά κριτήρια που θέτει ο χρήστης. Αυτά προκύπτουν, καταρχήν, λόγω της υλοποίησης διαφορετικών χαρακτηριστικών αναζήτησης (attributes) του Z39.50 πρωτοκόλλου από την εκάστοτε βιβλιοθήκη. Στα διαφορετικά χαρακτηριστικά αναζήτησης προστίθεται η έλλειψη χρήσης κοινών εργαλείων όπως καθιερωμένες αποδόσεις ή θεματικά και ταξινομικά συστήματα, με αποτέλεσμα ο εντοπισμός τεκμηρίων που εμπίπτουν σε συγκεκριμένες κατηγορίες να γίνεται ακόμη δυσχερέστερος.

Κατά τη διαδικασία της ανάκτησης, τα προβλήματα σχετίζονται με δυσκολίες ένταξης των εγγραφών που προέρχονται από διαφορετικές πηγές σε ένα ομοιόμορφο σύνολο

αποτελεσμάτων. Καθοριστικής σημασίας είναι η έλλειψη μιας κοινής διάταξης δεδομένων, καθώς μία μερίδα βιβλιοθηκών χρησιμοποιεί το MARC21 ενώ μία άλλη το UNIMARC. Επίσης, σημαντικό ρόλο κατέχει η ποικιλία των συνόλων χαρακτήρων (*character sets*) που υλοποιούνται από τα επιμέρους συστήματα. Ενδεικτική παρενέργεια των παραπάνω αποτελεί η αδυναμία αξιοποίησης του συνόλου αποτελεσμάτων στο πλαίσιο μιας διαδικασίας κοινής χρήσης πόρων.

Τέλος, παρουσιάζεται ο τρόπος αντιμετώπισης των παραπάνω προβλημάτων και οι λύσεις που δόθηκαν μέσα από την υλοποίηση του Z39.50 client που ανέπτυξε το Εθνικό Κέντρο Τεκμηρίωσης, καθώς επίσης και η εφαρμογή του στο περιβάλλον πρόσβασης ΑΡΓΩ.

Λέξεις κλειδιά: Αναζήτηση-Ανάκτηση πληροφοριών, Διαλειτουργικότητα, Πρωτόκολλο Z39.50, Κοινή χρήση πόρων.

Abstract

The adoption of the Z39.50 information retrieval protocol for use in the integrated library systems environment has put the possible ways of communication among them under a new perspective. It's been sometime since, in the context of delivering new services, almost every Hellenic academic library has incorporated the Z39.50 protocol in their automation systems. The issue rising is: in what extent the protocol's features have been exploited and whether the final goal is achieved according to its current implementations.

This study examines interoperability problems towards a one-stop access point within a distributed interface, where the various bibliographic catalogues are accessible using the Z39.50 protocol. The highlight is the possibility of conducting simultaneous searches in a number of library catalogues (resources) based on the protocol, as well as the incorporation of the results (bibliographic records) which occur from each, single resource in a unified result set. Some of these problems occur due to the heterogeneity of the databases themselves both in structure (different codification standards) and in respect of their content.

Issues in need for a thorough examination are identified both within the searching procedure, as well as within the records' retrieval procedure. These two stages are studied separately, because they are defined as separate functions according to the protocol, but, mainly, because of the distinct nature of the problems occurring during each of the above mentioned stages.

During the search procedure, benchmark problems are both the tracking of the records and, furthermore, the confident reassurance that the final result set comprises of responses relevant to the criteria, set by the user initially, as necessary in order to successfully meet the query needs. These problems occur, primarily, because of the various implementations of the Z39.50 protocol's search attributes adopted by each library. Additionally, there is a lack of commonly used tools, namely authorities or subject and classification systems. The latter makes locating the documents of certain categories even harder.

During the retrieval procedure, the problems occurring relate to difficulties in incorporating records deriving from various resources into a unified result set. Of major importance is the lack of a common data format, as a number of libraries use MARC21, whilst another uses

UNIMARC. The importance of the variety of character sets implemented by the various systems has, also, to be mentioned here. Consequently, there is an apparent weakness in exploiting the result set in the context of resource sharing.

Finally, the way for dealing with the above mentioned problems is presented in this study, along with the specific solutions provided by the Z39.50 client developed by the National Documentation Centre. Its implementation in the access environment of ARGO is, also, presented.

Keywords: *Information Retrieval, Interoperability, Z39.50 protocol, Resource sharing*

Εισαγωγή

Το πρόβλημα της ταυτόχρονης πρόσβασης σε πολλές πληροφοριακές πηγές απασχολεί τις βιβλιοθήκες και τα κέντρα πληροφόρησης εδώ και αρκετά χρόνια. Οι πρώτες προσεγγίσεις για τη λύση του προβλήματος αφορούσαν στη δημιουργία κεντρικών συλλογικών καταλόγων. Τα τελευταία χρόνια, και σε συνδυασμό με την αλματώδη ανάπτυξη της τεχνολογίας και των δικτύων, το κλίμα έχει μεταστραφεί και δίνεται βαρύτητα στη δημιουργία κατανεμημένων συλλογικών καταλόγων. Στην ουσία, πρόκειται για εικονικούς συλλογικούς καταλόγους. Σε μία τέτοια προσέγγιση, δυναμικής σύνδεσης σε πραγματικό χρόνο, κυρίαρχο ρόλο στην επικοινωνία κατέχει το πρωτόκολλο Z39.50.

Το Z39.50 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 (NISO 2002). 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.

Στο πλαίσιο αυτό, και με δεδομένο ότι οι περισσότερες ελληνικές ακαδημαϊκές βιβλιοθήκες υλοποιούν το Z39.50 πρωτόκολλο, εξετάζεται κατά πόσο είναι δυνατό να αξιοποιηθούν οι δυνατότητες του πρωτοκόλλου, ώστε να υπάρξει ένας εικονικός συλλογικός κατάλογος για τις ελληνικές ακαδημαϊκές βιβλιοθήκες, ικανός να ανταποκριθεί με συνέπεια στις απαιτήσεις των χρηστών του.

Για την έρευνα επιλέχθηκαν ενδεικτικά δέκα ακαδημαϊκές βιβλιοθήκες, από όλη την Ελλάδα, με κριτήρια την ποικιλία των συστημάτων αυτοματοποίησης και την διαφορετική διάταξη των βιβλιογραφικών εγγραφών. Επτά από τις βιβλιοθήκες είχαν ως εσωτερική διάταξη εγγραφής το UNIMARC και τρεις το MARC21 ενώ υπήρχαν πέντε διαφορετικά συστήματα αυτοματισμού. Τα ονόματα των βιβλιοθηκών και των συστημάτων αυτοματοποίησής τους προτιμήθηκε να μην ανακοινωθούν.

Έλεγχοι σχετικά με τη συμπεριφορά των servers γίνονται από την Άνοιξη του 2003. Τα στοιχεία που παρουσιάζονται στους πίνακες που ακολουθούν αντικατοπτρίζουν την κατάσταση που ίσχυε το Σεπτέμβριο του 2004. Οι Z39.50 clients που χρησιμοποιήθηκαν για την πρόσβαση στους servers των βιβλιοθηκών ήταν: το BookWhere, ο Z39.50 client του ABEKT, και ο Ζέφυρος. Σε εξαιρετικές περιπτώσεις, αναζητήσεις έγιναν από τους

ιστότοπους των βιβλιοθηκών. Αξίζει να σημειωθεί ότι δεν εντοπίστηκαν σε κανέναν ιστότοπο βιβλιοθήκης τα χαρακτηριστικά υλοποίησης του Z39.50 server της και όλα τα παρακάτω συμπεράσματα προέκυψαν από την υποβολή ερωτήσεων και την εξέταση των αποτελεσμάτων κάθε φορά.

Διαδικασία αναζήτησης

Γενικά

Το σημείο αναφοράς για τη διαδικασία της αναζήτησης είναι το πρωτόκολλο Z39.50 και τα χαρακτηριστικά (attributes) κάθε ερωτήματος, τα προβλεπόμενα από το BIB-1 Attribute set. Ο στόχος είναι να ελεγχθεί η δυνατότητα ταυτόχρονης αναζήτησης στους καταλόγους των ελληνικών ακαδημαϊκών βιβλιοθηκών και παράλληλα να εξετασθεί αν τα αποτελέσματα ανταποκρίνονται στην ερώτηση που υποβλήθηκε.

Με βάση το πρωτόκολλο Z39.50, ο client στέλνει την ερώτηση και οι servers αναλαμβάνουν την αποκωδικοποίηση της ερώτησης και την εκτέλεσή της. Το κάθε ερώτημα μπορεί να συνοδεύεται από μια 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). 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, completeness = πληρότητα. Κάθε μία από τις κατηγορίες περιέχει μία σειρά από τιμές οι οποίες ορίζονται στο Bib-1 attribute set.

Επειδή υπάρχει μεγάλος αριθμός χαρακτηριστικών σε κάθε κατηγορία, με αποτέλεσμα οι πιθανοί συνδυασμοί να μπορούν να ανέλθουν σε αρκετές εκατοντάδες, ορίστηκε ως σημείο εκκίνησης η απόκριση των servers στα πέντε πιο διαδεδομένα ευρετήρια σύμφωνα με την Index Data (Σεπτέμβριος 2004). Στον πίνακα 1 παρουσιάζεται η απόκριση ή όχι των servers όταν τους υποβάλλεται ερώτημα με μόνο χαρακτηριστικό αυτό του ευρετηρίου (Use Attribute).

Πίνακας 1

Βιβλιοθήκη	Οποιοδήποτε Any (1016)	Τίτλος Title (4)	Συγγραφέας Author (1003)	Θέμα Subject Heading (21)	ISBN (7)
Βιβλιοθήκη Α	Όχι	Ναι	Ναι	Ναι	Ναι
Βιβλιοθήκη Β	Όχι	Ναι	Ναι	Ναι	Ναι
Βιβλιοθήκη Γ	Ναι	Ναι	Ναι	Ναι	Όχι
Βιβλιοθήκη Δ	Όχι	Ναι	Ναι	Ναι	Ναι
Βιβλιοθήκη Ε	Ναι	Ναι	Ναι	Ναι	Ναι
Βιβλιοθήκη ΣΤ	Ναι	Ναι	Ναι	Ναι	Ναι
Βιβλιοθήκη Ζ	Ναι	Ναι	Ναι	Ναι	Ναι
Βιβλιοθήκη Η	Ναι	Ναι	Ναι	Ναι	Ναι
Βιβλιοθήκη Θ	Όχι	Ναι	Ναι	Ναι	Ναι
Βιβλιοθήκη Ι	Ναι	Ναι	Ναι	Ναι	Ναι

Από τον πίνακα 1 είναι προφανές ότι μόνο για τα τρία από τα πέντε ευρετήρια (Τίτλος, Συγγραφέας, Θέμα) μπορεί να διεξαχθεί αναζήτηση σε όλες τις βιβλιοθήκες. Το ευρετήριο "Οποιοδήποτε" χρησιμοποιείται από το 80% των Z39.50 servers σε παγκόσμιο επίπεδο (Index Data 2004). Στις ελληνικές ακαδημαϊκές βιβλιοθήκες του δείγματος το ποσοστό πέφτει στο 60%.

Στη συνέχεια παρατίθενται τα αποτελέσματα από έναν εκτενή και διεξοδικό έλεγχο συμβατότητας σε πεδία τα οποία παραδοσιακά δομούνται ως σημεία πρόσβασης.

Αναζήτηση με βάση το συγγραφέα

Το πρώτο σημείο που εξετάστηκε είναι οι προεπιλεγμένες τιμές που ισχύουν στους servers του δείγματος όταν διεξάγεται μια αναζήτηση στα πεδία του συγγραφέα χωρίς να αποστέλλει ο client κανένα άλλο χαρακτηριστικό. Στον πίνακα 2 εμφανίζονται οι προεπιλεγμένες τιμές που υιοθετούν οι servers των βιβλιοθηκών για την αναζήτηση με βάση το συγγραφέα.

Πίνακας 2

Βιβλιοθήκη	Ευρετήριο Συγγραφέα		
	Δομή	Αποκοπή	Πληρότητα
Βιβλιοθήκη Α	Λέξεις	Όχι	Incomplete Subfield
Βιβλιοθήκη Β	Λέξεις	Όχι	Incomplete Subfield
Βιβλιοθήκη Γ	Λέξεις	Όχι	Incomplete Subfield
Βιβλιοθήκη Δ	Λέξεις	Όχι	Incomplete Subfield
Βιβλιοθήκη Ε	Λέξεις	Όχι	Incomplete Subfield
Βιβλιοθήκη ΣΤ	Λέξεις	Όχι	Incomplete Subfield
Βιβλιοθήκη Ζ	Λέξεις	Όχι	Incomplete Subfield
Βιβλιοθήκη Η	Φράση	Δεξιά Αποκοπή	Incomplete Subfield
Βιβλιοθήκη Θ	Λέξεις	Όχι	Incomplete Subfield
Βιβλιοθήκη Ι	Φράση	Δεξιά Αποκοπή	Incomplete Subfield

Στον πίνακα 2 εμφανίζεται μια σχετική ομοιομορφία στον τρόπο που αντιμετωπίζεται ο συγγραφέας, αφού οι οκτώ στους δέκα servers δείχνουν να συμπεριφέρονται με τον ίδιο τρόπο. Ως παράδειγμα τέθηκε η ερώτηση : *Συγγραφέας = Smith, John* χωρίς κανένα άλλο χαρακτηριστικό. Διαπιστώθηκε ότι σχεδόν όλοι απάντησαν με ομοιόμορφο τρόπο. Ωστόσο, τα αποτελέσματα δεν ήταν μόνο εγγραφές όπου συγγραφέας ήταν ο Smith, John αλλά και έργα που είχαν γραφτεί από δύο συγγραφείς με ονόματα όπως Smith, Douglas και Nelsen, John κ.τ.λ.. Αναμενόμενο, μια και στους περισσότερους servers η ερώτηση εκτελέστηκε ως λέξεις. Ενδεικτικά αναφέρεται ότι εντοπίστηκαν 235 εγγραφές.

Για να περιοριστούν τα αποτελέσματα σε έργα του "*Smith, John*" υποβλήθηκε ξανά η ίδια ερώτηση, αλλά προστέθηκε και το χαρακτηριστικό της δομής και ζητήθηκε ως φράση. Το αποτέλεσμα ήταν να μειωθούν οι εγγραφές σε 48. Ο θόρυβος που είχε η αρχική απάντηση δεν υπάρχει πλέον. Όμως η σιωπή από τα συστήματα είναι μεγάλη. Κύρια αιτία είναι η αλλαγή των χαρακτηριστικών σε σχέση με τον Πίνακα 2. Συγκεκριμένα, για την ερώτηση *Συγγραφέας = "Smith, John"*, δομή = Φράση διαπιστώθηκε:

- Σε 5 βιβλιοθήκες η απάντηση περιελάμβανε μόνο τις εγγραφές όπου υπήρχε η ακριβής φράση *Smith, John* και μόνον αυτή (complete field). Αυτό σημαίνει ότι αν, για παράδειγμα, μετά το ονοματεπώνυμο ακολουθούσε και η ημερομηνία γέννησης οι εγγραφές δεν συμπεριλαμβάνονταν στο αποτέλεσμα.
- Στις άλλες 5 βιβλιοθήκες οι εγγραφές συμπεριλαμβάνονταν στα αποτελέσματα, αρκεί να υπήρχε η φράση *Smith, John* (incomplete subfield)
 - Σε 2 βιβλιοθήκες αυτής της ομάδας έπρεπε να ξεκινάει έτσι το πεδίο (first in field) ενώ στις άλλες 3 δεν ήταν απαραίτητο να βρίσκονται οι δύο λέξεις στην αρχή του πεδίου (any position in field).
- Στις 2 από τις 10 ακόμη και αν προσθέτονταν στο χαρακτηριστικό πληρότητας η τιμή Incomplete Subfield, η συμπεριφορά του server παρέμενε η ίδια και δεν θεωρούσε ότι οι εγγραφές εμπίπτουν στα κριτήρια της αναζήτησης.

Γίνεται, λοιπόν, κατανοητό ότι η αρχική ομοιομορφία πάυει να υφίσταται όταν ο χρήστης επιχειρήσει να γίνει πιο συγκεκριμένος και να υποβάλει την ερώτηση με διαφορετικά χαρακτηριστικά. Στάθηκε αδύνατο να βρεθεί ένας ενιαίος τρόπος αναζήτησης ώστε να εντοπιστούν τα έργα ενός συγγραφέα του οποίου δεν ήταν γνωστή η καθιερωμένη ονομασία.

Επιχειρήθηκε, λοιπόν, να εξετασθεί κατά πόσο ήταν εφικτή μια αναζήτηση με το πλήρες – καθιερωμένο – όνομα ενός συγγραφέα. Η διαδικασία που ακολουθήθηκε ήταν η εξής: Αναζητήθηκε ένα συγκεκριμένο βιβλίο. Το βιβλίο είχε τίτλο "*Two studies in ancient Macedonian topography*" και είναι γραμμένο από δύο Έλληνες συγγραφείς. Το έργο εντοπίστηκε σε τέσσερις από τις δέκα βιβλιοθήκες του δείγματος. Το όνομα του πρώτου συγγραφέα ως σημείο πρόσβασης βρέθηκε καθιερωμένο με τους παρακάτω τρόπους:

Χατζόπουλος, Μ.Β. (Μιλτιάδης Β.) (Βιβλιοθήκη Δ)

Χατζόπουλος, Μιλτιάδης Β. (Βιβλιοθήκη Ε & Βιβλιοθήκη ΣΤ)

Χατζόπουλος, Μιλτιάδης Β., 1944- (Βιβλιοθήκη Α)

Από τα όσα προηγήθηκαν είναι προφανές ότι η πολυμορφία των καθιερωμένων ονομάτων των συγγραφέων σε συνδυασμό με την υλοποίηση διαφορετικών χαρακτηριστικών από κάθε server καθιστά ιδιαίτερα δύσκολη την ταυτόχρονη αναζήτηση. Στις περιπτώσεις δε όπου και το όνομα και το επώνυμο του συγγραφέα είναι ιδιαίτερα κοινά (π.χ. Παπαδόπουλος, Γιώργος), αν ο χρήστης είναι ιδιαίτερα γενικός και δεν δηλώσει συγκεκριμένα χαρακτηριστικά, θα πάρει μεγάλο αριθμό μη σχετικών εγγραφών. Αν, από την άλλη, γίνει συγκεκριμένος και προσθέσει το χαρακτηριστικό της φράσης, θα χάσει τις εγγραφές από μερίδα βιβλιοθηκών που έχουν προσθέσει και άλλες πληροφορίες στο πεδίο του συγγραφέα - όπως την ημερομηνία γέννησης - και ταυτόχρονα το σύστημά τους θα θεωρήσει ότι το ονοματεπώνυμο πρέπει να αναζητηθεί ως complete field.

Θεματική αναζήτηση (με βάση λεκτική απόδοση ή ταξινομικό αριθμό)

Η προσέγγιση του υλικού με θεματικά κριτήρια παρουσίασε περισσότερες δυσκολίες. Οι καθιερωμένες αποδόσεις των θεμάτων παρουσίαζαν τόσο μεγάλη ποικιλία που στάθηκε αδύνατο να βγουν ασφαλή συμπεράσματα. Σε πολλές εγγραφές τα θέματα ήταν μόνο στην αγγλική γλώσσα ενώ ο κατάλογος της βιβλιοθήκης περιείχε και ελληνικά θέματα. Σε κάποιες περιπτώσεις υπήρχαν μη ελεγχόμενοι θεματικοί όροι (πεδίο 610 του UNIMARC) παράλληλα με θεματικές επικεφαλίδες (πεδίο 606 του UNIMARC). Σε άλλες περιπτώσεις υπήρχαν μόνο

μη ελεγχόμενοι θεματικοί όροι και σε άλλες μόνο επικεφαλίδες. Όσες φορές εντοπιζόταν το ίδιο τεκμήριο σε δύο ή περισσότερες βιβλιοθήκες οι θεματικές επικεφαλίδες συνέπιπταν σπάνια. Παρατηρήθηκε αρκετές φορές ότι οι εγγραφές, κυρίως για ξενόγλωσσα έργα, προέρχονταν από άλλα καταλογογραφικά κέντρα, κατά βάση του εξωτερικού, και επιπλέον σε αυτές τις περιπτώσεις υπήρχε μια τάση να διατηρούνται οι αρχικές – ξενόγλωσσες - θεματικές επικεφαλίδες.

Η εξέταση των χαρακτηριστικών έδειξε μια σχετική ομοιογένεια όσων αφορά την υλοποίησή τους από τις βιβλιοθήκες του δείγματος. Οι οχτώ στις δέκα θεωρούσαν τους προς αναζήτηση όρους ως λέξεις χωρίς αποκοπή και οι δύο φράση με δεξιά αποκοπή ενώ και για τις δέκα ήταν incomplete subfield. Όμως, η ταυτόχρονη αναζήτηση είχε ήδη καταστεί ουσιαστικά αδύνατη λόγω της ποικιλίας των λεκτικών αποδόσεων.

Αφού η θεματική προσέγγιση με βάση τις λεκτικές αποδόσεις στάθηκε ιδιαίτερα δυσχερής έγινε προσπάθεια θεματικής προσέγγισης του υλικού με βάση τα ταξινομικά συστήματα. Ωστόσο, και η προσέγγιση με βάση το κριτήριο αυτό παρουσίασε πολλά προβλήματα. Καταρχήν, εντοπίστηκαν τρία διαφορετικά ταξινομικά συστήματα. Στη συνέχεια, διαπιστώθηκε ότι οι μισοί από τους servers του δείγματος δεν υποστήριζαν αναζήτηση με βάση το ταξινομικό αριθμό. Στον πίνακα 3 εμφανίζονται τα ταξινομικά σχήματα που χρησιμοποιούνται καθώς και αν γίνεται ή όχι αναζήτηση σε αυτά.

Πίνακας 3

Βιβλιοθήκη	Ταξινομικό σύστημα	Αναζήτηση
Βιβλιοθήκη Α	DDC, UDC	Δεν Υποστηρίζεται για κανένα από τα δύο
Βιβλιοθήκη Β	DDC	Υποστηρίζεται
Βιβλιοθήκη Γ	DDC	Δεν Υποστηρίζεται
Βιβλιοθήκη Δ	LCC	Δεν Υποστηρίζεται
Βιβλιοθήκη Ε	LCC	Ο server δείχνει ότι εκτελεί την ερώτηση αλλά δεν επιστρέφει ποτέ αποτελέσματα, ακόμη και όταν υπάρχει ο ζητούμενος αριθμός.
Βιβλιοθήκη ΣΤ	LCC	Υποστηρίζεται
Βιβλιοθήκη Ζ	DDC	Υποστηρίζεται
Βιβλιοθήκη Η	DDC	Υποστηρίζεται
Βιβλιοθήκη Θ	LCC	Δεν Υποστηρίζεται
Βιβλιοθήκη Ι	DDC	Υποστηρίζεται

Αξίζει να σημειωθεί ότι σε βιβλιοθήκες που χρησιμοποιούν μόνο ένα ταξινομικό σύστημα εντοπίστηκαν εγγραφές και άλλου. Μια πρόχειρη ανάλυση έδειξε ότι επρόκειτο, κατά κανόνα, για εγγραφές που προέρχονταν από άλλη βιβλιοθήκη (πρωτότυπη καταλογογράφηση).

Με όσα αναφέρθηκαν για τη διαδικασία της αναζήτησης είναι προφανές ότι οι ερωτήσεις που υποβάλλονται, αν μπορούν να εκτελεστούν, αποκτούν αρκετά διαφορετικό νόημα σε κάθε server. Επίσης, διαφαίνεται έλλειψη βιβλιοθηκονομικής σύμπνοιας στο τρόπο αντιμετώπισης των κλασικών βιβλιοθηκονομικών θεμάτων, όπως είναι η διαμόρφωση κοινών καθιερωμένων αποδόσεων. Όλα αυτά φανερώνουν ότι η ταυτόχρονη αναζήτηση καθίσταται εξαιρετικά δυσχερής και αναποτελεσματική.

Διαδικασία ανάκτησης

Ως διαδικασία ανάκτησης ορίζεται το στάδιο κατά το οποίο τα επιμέρους σύνολα αποτελεσμάτων καταλήγουν στον client. Το κεντρικό ζητούμενο σε αυτή τη φάση είναι εάν τα επιμέρους αποτελέσματα μπορούν να αποτελέσουν ένα ενιαίο σύνολο.

Όπως αναφέρθηκε και στην αρχή του άρθρου, από τις βιβλιοθήκες που εξετάστηκαν οι τρεις χρησιμοποιούν ως εσωτερική διάταξη των βιβλιογραφικών εγγραφών τους το MARC21 και οι επτά το UNIMARC. Ως εκ τούτου, τα επιμέρους αποτελέσματα δεν αποτελούν ένα ομογενοποιημένο ενιαίο σύνολο από την πλευρά της διάταξης των εγγραφών.

Ένα ακόμη σημαντικό πρόβλημα για την ένταξη των επιμέρους αποτελεσμάτων σε ένα κοινό σύνολο ήταν η πολλαπλότητα των συνόλων χαρακτήρων. Στις δέκα βιβλιοθήκες του δείγματος απαντήθηκαν πέντε διαφορετικά σύνολα χαρακτήρων.

Πρόσθετο σημείο ασυμβατότητας είναι οι πληροφορίες για τα αντίτυπα. Οι τρεις στους δέκα servers δεν στέλνουν καθόλου πληροφορίες αντιτύπων. Οι υπόλοιποι επτά, αν και στέλνουν πληροφορίες, αυτές δεν είναι δομημένες με τον ίδιο τρόπο. Στο σημείο των αντιτύπων υπάρχουν βέβαια αντικειμενικές δυσκολίες αφού το UNIMARC bibliographic format δεν έχει ακόμη συμπεριλάβει πεδία αντιτύπων από το πολύ πρόσφατο UNIMARC holdings. Το πεδίο 850, που υπάρχει στην παρούσα έκδοση του UNIMARC bibliographic format, προφανώς δεν μπορεί να δώσει λύση στο πρόβλημα. Αντίστοιχα και στο MARC21 οι υλοποιήσεις που έχουν γίνει, στις ελληνικές ακαδημαϊκές βιβλιοθήκες, δεν είναι απόλυτα συμβατές με τα Holdings του MARC21.

Δυσκολίες δημιουργούνται κατά την ανάκτηση και από την ποιότητα των εγγραφών. Σε κάποιες περιπτώσεις οι εγγραφές που επιστρέφει ο server δεν είναι έγκυρες. Οι εγγραφές μπορεί να μην είναι έγκυρες κατά το ISO 2709, δηλαδή να έχουν λάθος μήκος, λάθος ετικέτα εγγραφής κ.τ.λ., ή μπορεί να μην είναι έγκυρες σύμφωνα με το MARC το οποίο υλοποιούν (MARC21 ή UNIMARC), να τους λείπουν, για παράδειγμα, υποχρεωτικά πεδία όπως ο τίτλος. Τέτοιου είδους προβλήματα παρουσιάζονται σε σχετικά μικρό αριθμό δεδομένων. Πρέπει όμως να αναφερθεί ότι η μία από τις δέκα βιβλιοθήκες του δείγματος από την πρώτη απόπειρα σύνδεσης, τον Ιούνιο του 2003, έως και την τελευταία, το Σεπτέμβριο του 2004, έστειλε MARC εγγραφές οι οποίες δεν ήταν έγκυρες σύμφωνα με το ISO 2709.

Το περιβάλλον πρόσβασης "Η ΑΡΓΩ"

Γενικά

Λαμβάνοντας υπόψη τα παραπάνω, το Εθνικό Κέντρο Τεκμηρίωσης προχώρησε στη δημιουργία της ΑΡΓΩ, ενός web περιβάλλοντος πρόσβασης πηγών πληροφόρησης που διατίθενται μέσω του πρωτοκόλλου Z39.50.

Ο στόχος ήταν να δημιουργηθεί μια web εφαρμογή που να μπορεί να αξιοποιηθεί από δύο βασικές κατηγορίες χρηστών. Πρώτον, να μπορεί να χρησιμοποιηθεί από όλους εκείνους που αναζητούν πληροφορίες στο διαδίκτυο, χωρίς κατ' ανάγκη να είναι ειδικοί σε αυτό. Δεύτερον, να μπορεί να χρησιμοποιηθεί από ειδικούς της πληροφόρησης, είτε για να εντοπιστούν τεκμήρια στο διαδίκτυο είτε για να εντοπιστούν βιβλιογραφικές εγγραφές οι οποίες θα

χρησιμοποιηθούν για την ανάπτυξη του καταλόγου της βιβλιοθήκης τους. Στο σχεδιασμό δόθηκε ιδιαίτερη βαρύτητα ώστε το σύστημα να μπορεί να εξυπηρετεί και τις δύο αυτές κατηγορίες χρηστών.

Το περιβάλλον πρόσβασης

Η ανάλυση των log files του Z39.50 server του Εθνικού Κέντρου Τεκμηρίωσης έδειξε ότι είχε μεγάλη σημασία η ομαδοποίηση σχετικών μεταξύ τους πηγών. Παρατηρήθηκε ότι όταν οι ερωτήσεις προέρχονταν από Z39.50 clients οι οποίοι διέθεταν μια εκτενή, μη κατηγοριοποιημένη λίστα με πηγές, οι ερωτήσεις που υποβάλλονταν ήταν στις περισσότερες περιπτώσεις χωρίς νόημα. Για παράδειγμα, γίνονται αναζητήσεις για τις βροχοπτώσεις στην Αττική σε μία βάση δεδομένων η οποία περιέχει μόνον τίτλους περιοδικών δημοσιευμάτων, ή γίνονται ερωτήσεις με κυριλλικούς χαρακτήρες σε βάσεις δεδομένων που περιέχουν μόνο ελληνική αρθρογραφία. Το μόνο που επιτυγχάνεται με τέτοιες ερωτήσεις είναι η αύξηση της κίνησης στο δίκτυο και ο φόρτος του server. Προτιμήθηκε να δοθεί στο χρήστη η δυνατότητα να επιλέξει καταρχήν ομάδα βάσεων δεδομένων και μέσα από την ομάδα να επιλέξει τις βάσεις που επιθυμεί για αναζήτηση.

Η ομαδοποίηση των βάσεων προκύπτει περισσότερο από τη συνάφεια περιεχομένου παρά από τη θεματική συνάφεια. Για παράδειγμα, μία ομάδα βάσεων δεδομένων περιέχει μόνο βάσεις που καταλογραφούν τίτλους περιοδικών. Ο στόχος είναι ο χρήστης να υποβάλει κάθε φορά ερωτήματα τα οποία έχουν πιθανότητες να απαντηθούν και όχι ερωτήματα που εξ ορισμού δεν θα μπορούσαν να απαντηθούν.

Ο Z39.50 client

Ο Z39.50 client, ο οποίος αναπτύχθηκε από το Εθνικό Κέντρο Τεκμηρίωσης, και λειτουργεί στο περιβάλλον της ΑΡΙΓΩ είχε να αντιμετωπίσει τα προβλήματα που αναφέρθηκαν παραπάνω, τόσο στο επίπεδο της αναζήτησης όσο και της ανάκτησης.

Για να αντιμετωπιστούν τα ζητήματα που σχετίζονται με την ποικιλομορφία των χαρακτηριστικών που υλοποιεί ο κάθε server και παράλληλα για να είναι επιτυχής η αναζήτηση, παρέχεται η δυνατότητα να υποβάλλεται το ερώτημα με μόνο χαρακτηριστικό αυτό του ευρετηρίου (use attribute). Παράλληλα, προσφέρεται η δυνατότητα στο χρήστη που το επιθυμεί να υποβάλλει πιο συγκεκριμένες ερωτήσεις συνδυάζοντας περισσότερα χαρακτηριστικά και όρους αναζήτησης. Πάντα, σε αυτή την περίπτωση, υπάρχει η πιθανότητα οι servers που αναλαμβάνουν να εκτελέσουν την ερώτηση να μην μπορούν να απαντήσουν σε κάποιους από τους συνδυασμούς χαρακτηριστικών. Σε αυτή την περίπτωση, οι servers είτε θα επιστρέψουν διαγνωστικό μήνυμα, είτε θα αγνοήσουν τα χαρακτηριστικά που δεν υλοποιούν, οπότε και η απάντηση δεν θα είναι σύμφωνη με την ερώτηση.

Μία σημαντική εργασία που επιτελεί ο Z39.50 client κατά την ανάκτηση των δεδομένων είναι η ομογενοποίηση των επιμέρους αποτελεσμάτων που παραλαμβάνει. Η ομογενοποίηση αυτή γίνεται σε δύο επίπεδα.

- Πρώτον, ομογενοποιείται η διάταξη των εγγραφών, όπου αυτό είναι εφικτό. Η διάταξη στην οποία εμφανίζονται οι εγγραφές είναι UNIMARC, ανεξάρτητα από το αν ο server τις στέλνει σε UNIMARC ή MARC21.

- Δεν είναι εφικτή η μετατροπή εγγραφών οι οποίες είναι σε διάταξη SUTRS (Simple Unstructured Text Record Syntax).
- Δεύτερον, ομογενοποίηση γίνεται στο σύνολο χαρακτήρων. Ανεξάρτητα από το σύνολο χαρακτήρων στο οποίο η πηγή στέλνει τα δεδομένα, η ΑΡΙΓΩ τα μετατρέπει στο πρότυπο UNICODE (UTF-8).

Ο Z39.50 client διαχειρίζεται τις εγγραφές σε σύνολο χαρακτήρων UTF-8, όμως κατά την αποθήκευσή τους, ο χρήστης μπορεί να επιλέξει το σύνολο χαρακτήρων που εξυπηρετεί το σύστημά του.

Έτσι, παρέχεται η δυνατότητα στο χρήστη να αποθηκεύσει στο δίσκο του ένα αρχείο ISO 2709 – UNIMARC, το οποίο μπορεί να εισαγάγει στο σύστημά του χωρίς να επηρεάζεται από την αρχική διάταξη και το αρχικό σύνολο χαρακτήρων των εγγραφών. Με αυτό τον τρόπο, οι χρήστες του UNIMARC αποκτούν πρόσβαση σε μία πληθώρα βιβλιογραφικών εγγραφών οι οποίες προέρχονται από MARC21 και τις οποίες μπορούν να αξιοποιήσουν με άμεση εισαγωγή στο σύστημά τους, δίχως να απαιτείται η χρήση άλλων εργαλείων μετατροπής.

Συμπεράσματα

Όπως προκύπτει από τα παραπάνω, η ταυτόχρονη πρόσβαση στους καταλόγους των ελληνικών ακαδημαϊκών βιβλιοθηκών με βάση το πρωτόκολλο Z39.50 είναι πλέον εφικτή. Ωστόσο, εξακολουθούν να υπάρχουν θέματα που χρήζουν επιπλέον εξέτασης, μια και η ανομοιογένεια και η ασυμβατότητα των διαφορετικών βάσεων δεδομένων δεν παρακάμπτονται σε ικανοποιητικό βαθμό. Το πρωτόκολλο μπορεί να καλύψει σε πολύ μεγάλο βαθμό θέματα διαλειτουργικότητας τέτοιου είδους. Ο τρόπος, όμως, με τον οποίο έχει υλοποιηθεί στον ελληνικό ακαδημαϊκό χώρο μάλλον καταδεικνύει τα προβλήματα που ήδη υπάρχουν παρά τα λύνει.

Τα προβλήματα θα μπορούσαν να συνοψιστούν σε τέσσερις βασικές κατηγορίες:

- Δε χρησιμοποιούνται από όλους και με συνέπεια παραδοσιακά βιβλιοθηκονομικά εργαλεία, όπως είναι τα αρχεία καθιερωμένων αποδόσεων για ονόματα φυσικών προσώπων.
- Δεν υιοθετούνται κοινά εργαλεία προς χρήση, όπως κοινά ταξινομικά συστήματα, κοινές διατάξεις δεδομένων, κοινά σύνολα χαρακτήρων.
- Δεν υπάρχει απόλυτη συνέπεια και συμμόρφωση σε ό,τι ορίζουν τα πρότυπα και οι κανόνες που έχουν επιλεγεί για να χρησιμοποιηθούν (π.χ. οι AACR δεν επιτρέπουν τη χρήση της τιμής "χ.χ." για να δηλωθεί ότι ένα βιβλίο δεν έχει χρόνο έκδοσης, αλλά προτείνουν να υπολογιστεί ο χρόνος έκδοσης και να γραφτεί μέσα σε αγκύλες, παρόλα αυτά η ένδειξη "χ.χ" υπάρχει σε αρκετές βιβλιοθήκες οι οποίες χρησιμοποιούν τους AACR).
- Δεν υιοθετείται σε ικανοποιητικό βαθμό κοινή πολιτική, εκεί όπου τα εργαλεία επιδέχονται διαφορετικές θεωρήσεις, όπως στον τρόπο με τον οποίο υλοποιείται το Z39.50 πρωτόκολλο.

Αξίζει να σημειωθεί ότι τα παραπάνω προβλήματα είναι κυρίαρχα και στη δημιουργία κεντρικών συλλογικών καταλόγων, όπου και σε αυτές τις περιπτώσεις η προσέγγισή τους είναι απαραίτητη.

Για τον υποσκελισμό των προβλημάτων αυτών είναι ανάγκη να υπάρξει συνεργασία των εμπλεκόμενων φορέων. Προτείνεται, όπου είναι δυνατό, η υιοθέτηση κοινών εργαλείων και προτύπων καθώς και κοινής πολιτικής, για θέματα που δεν προβλέπονται από πρότυπα και κανόνες. Στο πλαίσιο υλοποίησης του Z39.50 πρωτοκόλλου, και με γνώμονα την αντιμετώπιση του προβλήματος της διαλειτουργικότητας, η θέσπιση και υιοθέτηση από όλους συγκεκριμένων κανόνων για τον ακριβή τρόπο υλοποίησής του κρίνεται απαραίτητη.

Αξίζει να σημειωθεί ότι έχουν γίνει πολλές προσπάθειες δημιουργίας των λεγόμενων Προφίλ υλοποιήσεων (Profile Implementations). Τα Profile, σχεδιασμένα να καλύψουν τις ανάγκες των διαφορετικών υπηρεσιών πληροφόρησης, καθορίζουν λεπτομερώς τις λειτουργίες και τα χαρακτηριστικά του πρωτοκόλλου που πρέπει να υποστηρίζονται. Ενδεικτικά, αναφέρεται το διεθνώς καταχωρημένο Bath Profile, το οποίο έχει υιοθετηθεί από ένα μεγάλο αριθμό υλοποιήσεων του πρωτοκόλλου διεθνώς και που μια πιθανή προσαρμογή του στον ελληνικό χώρο να έθετε τις βάσεις για την ουσιαστική ομογενοποίηση των υλοποιήσεων του Z39.50 πρωτοκόλλου.

Βιβλιογραφία

ANSI/NISO: Z39.50 Information Retrieval: application service definition and protocol specification (1995)

Husby, Ole (1999) "Real and virtual union catalogues" [Διαθέσιμο στο <http://www.caslin.cz:7777/caslin99/a2.htm>]

Index Data (2004) "The Z39.50 Target Directory" [Διαθέσιμο στο <http://www.indexdata.dk/targettest/>]

Index Data (2004) "Z39.50 Target Statistics", [Διαθέσιμο στο <http://www.indexdata.dk/targettest/stat.php>]

Lynch, Clifford (1997) "Building the Infrastructure of Resource Sharing: Union Catalogs, Distributed Search and Cross-Database Linkage", Library Trends , Vol. 45:3, pp. 448-461

NISO (2002) "Z39.50: a primer on the protocol" [Διαθέσιμο στο http://www.niso.org/standards/resources/Z3950_primer.pdf]

Καπιδάκης, Σαράντος και Σφακάκης, Μιχάλης (2002) "Σχεδιασμός αποτελεσματικών portal στο Διαδίκτυο", Στο *Πρακτικά 11^{ου} Πανελληνίου Συνεδρίου Ακαδημαϊκών Βιβλιοθηκών*. σσ. 189-200.