

**Μεθοδολογία Ανάλυσης και Συσχετισμού Δεδομένων Ποιοτικών και Ποσοτικών
Δεικτών Υπηρεσιών Βιβλιοθηκών χρησιμοποιώντας μεθόδους Εξόρυξης Γνώσης.**

Αριστείδης Μελετίου¹

Ανθή Κατσιρίκου²

Περίληψη

Στην εργασία αυτή περιγράφεται μια μεθοδολογία ανάλυσης δεδομένων χρησιμοποιώντας τεχνικές Εξόρυξης Δεδομένων και Γνώσης (Data Mining) με σκοπό την εξόρυξη γνώσης από αυτό, δηλαδή την απόρροια χρήσιμων συμπερασμάτων που θα βοηθήσουν τους ιθύνοντες μιας Βιβλιοθήκης στις αποφάσεις τους και στον στρατηγικό τους σχεδιασμό.

Αποτελεί ουσιαστικά συνέχεια της εργασίας που παρουσιάστηκε στο συνέδριο Ακαδημαϊκών Βιβλιοθηκών 2006 στην Πάτρα με τίτλο: «Ποιοτικοί Δείκτες Υπηρεσιών Βιβλιοθηκών και Διαχείριση Πόρων: Μεθοδολογίες Ανάλυσης και στρατηγικός σχεδιασμός», στο οποίο δινόταν μια γενική περιγραφή μεθοδολογιών ανάλυσης δεδομένων και εξόρυξης γνώσης και πώς αυτές μπορούν να εφαρμοστούν στον τομέα διαχείρισης Βιβλιοθηκών.

Η παρούσα εργασία ασχολείται αποκλειστικά πλέον με πραγματικά δεδομένα που προέρχονται από το χώρο των σύγχρονων Βιβλιοθηκών. Δίνεται αρχικά εμπειριστατωμένη περιγραφή και αιτιολόγηση των δεδομένων που θα χρησιμοποιηθούν και στη συνέχεια εφαρμόζεται συγκεκριμένη μεθοδολογία σε αυτά, από την οποία τελικά προκύπτουν αποτελέσματα, συσχετίσεις και συμπεράσματα που αρχικά ήταν πολύ δύσκολο ή αδύνατο να διαπιστωθούν. Γίνεται λεπτομερής περιγραφή της μεθοδολογίας και περιγράφονται σε κάθε βήμα της οι διαδικασίες που εφαρμόζονται ενώ σε συνδυασμό με ειδικό λογισμικό ανάλυσης δεδομένων το οποίο χρησιμοποιεί μεθόδους εξόρυξης γνώσης, προκύπτουν και παρουσιάζονται αποτελέσματα συσχετίσεων δεδομένων που μπορούν να χρησιμοποιηθούν σε μεγάλο βαθμό στο στρατηγικό σχεδιασμό μελλοντικών κινήσεων σε μια Βιβλιοθήκη.

Με άλλα λόγια, αρχικά ορίζεται ποια δεδομένα θα χρησιμοποιηθούν, γίνεται περιγραφή της διαδικασίας που εφαρμόζεται σε αυτά και τέλος παρουσιάζονται συσχετίσεις μέσω των

¹ Πολυτεχνείο Κρήτης, Πολυτεχνειούπολη, 73100 Χανιά, amlet@library.tuc.gr.

² Πανεπιστήμιο Πειραιά, Καραολή και Δημητρίου 80, 18534, Πειραιάς, anthi@unipi.gr.

οποίων εξάγονται συμπεράσματα που μπορούν να βοηθήσουν στη λήψη αποφάσεων διοικητικού χαρακτήρα.

Τα δεδομένα που χρησιμοποιούνται προέρχονται από βάσεις δεδομένων της Βιβλιοθήκης του Πολυτεχνείου Κρήτης που χρησιμοποιούνται σε καθημερινή βάση, κάτι όμως που δεν σημαίνει ότι στη μεθοδολογία που περιγράφεται δεν μπορούν να χρησιμοποιηθούν παρόμοια δεδομένα από οποιαδήποτε Βιβλιοθήκη.

Keywords: Διαχείριση Ανθρωπίνων Πόρων, Εξόρυξη Πληροφορίας, Δείκτες Αποτίμησης, Αξιολόγηση, Στρατηγικός Σχεδιασμός

1. Εισαγωγή

Οι παραδοσιακές βιβλιοθήκες είναι δύσκολο να ικανοποιήσουν τους χρήστες τους σε τέτοιο βαθμό και σε ταχύτητα σε σχέση με αυτές που διαθέτουν ένα πλήρως μηχανογραφημένο και αυτοματοποιημένο σύστημα διαχείρισης υλικού (Yu and Chen 2001). Επομένως, είναι ουσιώδες για μια Βιβλιοθήκη να διαθέτει ένα τρόπο που να βοηθάει στο μέγιστο και τάχιστο βαθμό τους χρήστες της να βρουν το υλικό που ψάχνουν αλλά και να είναι σε θέση να βελτιώνει συνεχώς την ποιότητα των υπηρεσιών που τους προσφέρει. Για να επιτευχθεί αυτό είναι επιτακτική ανάγκη η ύπαρξη ενός ευέλικτου και σύγχρονου συστήματος διαχείρισης υλικού και ανάλυσης δεδομένων. Η τεχνολογία της «εξόρυξης γνώσης από δεδομένα βάσεων δεδομένων» (data mining) χρησιμοποιείται για αυτή τη επεξεργασία με σκοπό τη λήψη πληροφοριών και γνώσης από τα δεδομένα που διαχειρίζεται ένα σύστημα μηχανοργάνωσης, οι οποίες δεν είναι εμφανείς. Παράλληλα, μια τέτοια διαδικασία προκαλεί την εξαγωγή πληροφοριών που εκφράζουν τάσεις και συσχετισμούς μεταξύ των δεδομένων που είναι αδύνατο να παρατηρηθούν χωρίς αυτή (Chang and Chen 2006). Ο σκοπός της «εξόρυξης γνώσης από δεδομένα» είναι ουσιαστικά να βελτιώσει την ποιότητα της αλληλεπίδρασης μεταξύ μιας Βιβλιοθήκης και των χρηστών της. Τα δεδομένα που υπάρχουν αποθηκευμένα σε μια Βιβλιοθήκη περιέχουν πολύ χρήσιμη πληροφορία που με την κατάλληλη επεξεργασία παρέχουν χρήσιμη γνώση που μπορεί να χρησιμοποιηθεί από τους υπευθύνους για τη χάραξη σωστής και εποικοδομητικής στρατηγικής (Chang and Chen 2006). Για την επεξεργασία αυτή απαιτούνται εργαλεία ανάλυσης δεδομένων που χρησιμοποιούν μοντέλα και τεχνικές βασισμένες στην τεχνολογία της «εξόρυξης δεδομένων» (data mining).

Θέλοντας, να δώσουμε έναν πιο περιγραφικό ορισμό θα μπορούσαμε να πούμε ότι ο όρος «ανακάλυψη γνώσης» (knowledge discovery) σε βάσεις δεδομένων δηλώνει την ολοκληρωμένη διαδικασία μετατροπής χαμηλού επιπέδου και σημασίας αρχικής

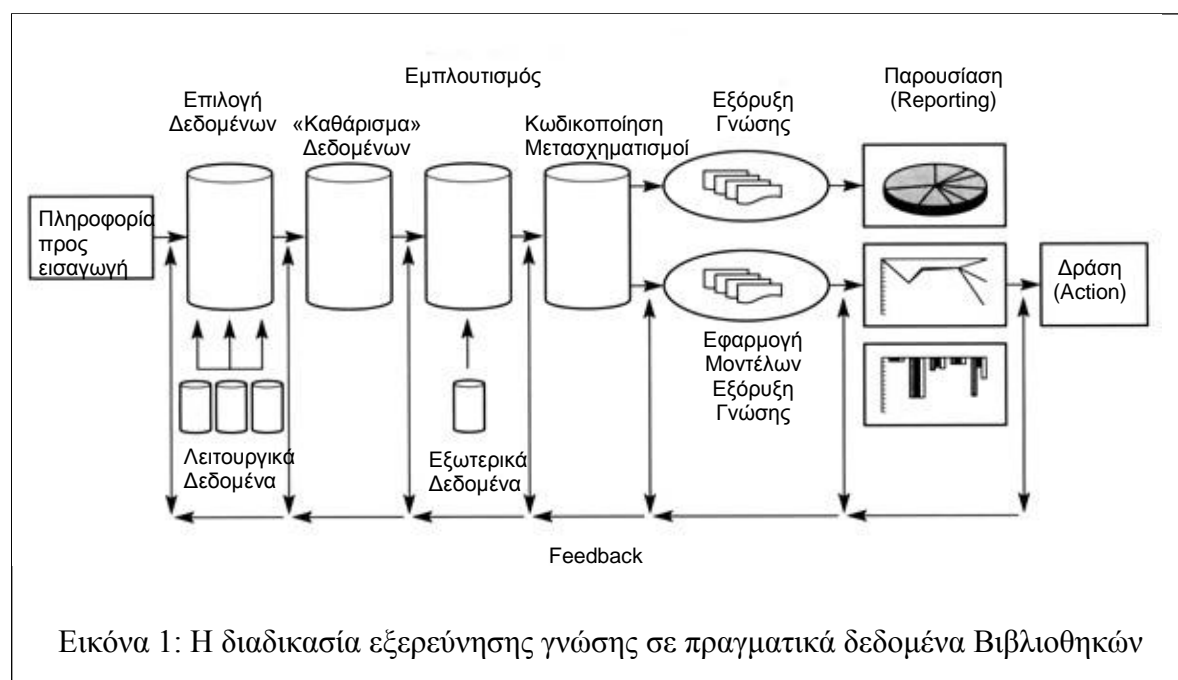
πληροφορίας σε χρήσιμη πληροφορία και γνώση με σκοπό τη βελτιστοποίηση της διαδικασίας λήψης αποφάσεων.

Στην εργασία αυτή χρησιμοποιούνται δεδομένα από βάσεις δεδομένων Βιβλιοθηκών, γίνεται επεξεργασία τους και εξάγονται πληροφορίες που εκφράζουν σχέσεις και συσχετίσεις μεταξύ τους που ήταν αδύνατο να παρατηρηθούν αρχικά.

Συγκεκριμένα, περιγράφεται η διαδικασία που θα ακολουθηθεί (data mining process), ο τρόπος επιλογής των δεδομένων που θα συμμετάσχουν σε αυτή σε συνδυασμό με τους επιθυμητούς στόχους, η ανάλυση των δεδομένων και τέλος τα συμπεράσματα που προκύπτουν. Ουσιαστικά γίνεται μια μελέτη περίπτωσης, η οποία στοχεύει να καταδείξει πρακτικά τη διαδικασία «εξερεύνησης δεδομένων και γνώσης» (data and knowledge mining) σε δεδομένα που προέρχονται από σύγχρονες Βιβλιοθήκες.

2. Η διαδικασία εξερεύνησης γνώσης και η εφαρμογή της σε δεδομένα Βιβλιοθηκών

Η διαδικασία εξερεύνησης γνώσεων από ένα σύνολο δεδομένων, απεικονίζεται συνοπτικά στο παρακάτω σχήμα (Dunham 2004):



Εικόνα 1: Η διαδικασία εξερεύνησης γνώσης σε πραγματικά δεδομένα Βιβλιοθηκών

Μπορεί να θεωρηθεί ότι χωρίζεται σε έξι στάδια, τα οποία δεν είναι ανεξάρτητα μεταξύ τους αλλά ουσιαστικά αποτελούν ένα συνεχόμενο λογικό μοντέλο.

Αρχικά λοιπόν, γίνεται επιλογή των δεδομένων προς επεξεργασία. Το στάδιο αυτό είναι από τα σημαντικότερα της όλης διαδικασίας και περιγράφεται εκτενέστερα σε επόμενη ενότητα. Πρέπει να επιλεγούν τα δεδομένα εκείνα που μας δώσουν τελικά πληροφορίες για τα μεγέθη που μας ενδιαφέρουν (π.χ. αν θέλουμε να πάρουμε

πληροφορίες για τη χρήση ενός βιβλίου θα πρέπει να επιλέξουμε δεδομένα που σχετίζονται με αυτό, όπως ο αριθμός των δανεισμών που αυτό υπέστη. Τα δεδομένα αυτά προέρχονται κατά κύριο λόγο από λειτουργικές βάσεις δεδομένων.

Κατόπιν γίνεται ένα ξεκαθάρισμα των αρχικών δεδομένων με σκοπό την απομόνωση και αφαίρεση ή συμπλήρωση τιμών που δεν έχουν νόημα και μπορούν να προκαλέσουν πρόβλημα στην επεξεργασία (π.χ. λάθος πληκτρολογημένες τιμές από τους χειριστές της βάσης δεδομένων, ελλιπή στοιχεία για κάποιο αντικείμενο της βάσης).

Στο επόμενο στάδιο του «εμπλουτισμού» γίνεται εμπλουτισμός των δεδομένων εξερεύνησης από «εξωτερικά δεδομένα», δεδομένα δηλαδή που μπορούν να συμπληρώσουν τα ήδη υπάρχοντα με σκοπό την αποδοτικότερη επεξεργασία και την λήψη καλύτερων αποτελεσμάτων (π.χ. λεπτομερή πληροφορία για κάποια κατηγορία αντικειμένων εξερεύνησης, στατιστικές μετρήσεις σε μικρά δείγματα της βάσης δεδομένων).

Στο στάδιο κωδικοποίησης και μετασχηματισμού των δεδομένων γίνονται κάποιοι μετασχηματισμοί δεδομένων που προέρχονται από διαφορετικές πηγές και απαιτείται να μετατραπούν σε ένα κοινό σχήμα για την περαιτέρω επεξεργασία τους. (π.χ. μετατροπή της ημερομηνία γέννησης σε ηλικία, διαίρεση των χρηματικών ποσών με 1000, μετατροπή των τιμών yes/no σε 1/0).

Στο επόμενο στάδιο, εφαρμόζονται μοντέλα και τεχνικές εξόρυξης γνώσης μέσω του οποίων πλέον λαμβάνονται οι πληροφορίες που έχουν σχέση με το συσχετισμό των δεδομένων και οι οποίες θα παράγουν την τελική χρήσιμη πληροφορία. Οι πληροφορίες που λαμβάνονται στο τέλος του εν λόγω σταδίου μπορεί να είναι : **πληροφορίες συσχέτισης (association rules)** αντικειμένων (π.χ. «το 30% των προπτυχιακών φοιτητών δεν επιστρέφει το δανεισμένο υλικό εμπρόθεσμα»), **πληροφορίες κατηγοριοποίησης (classification)** (π.χ. ταξινόμηση των αναγνωστών βάσει του αριθμού των βιβλίων που έχουν δανειστεί), **πληροφορίες σχετικές με ανάλυση τάσεων και αποκλίσεων (trend and deviation analysis)** (π.χ. ανακάλυψη του υλικού που θεωρείται περισσότερο ή λιγότερο ενδιαφέρον από τους χρήστες της Βιβλιοθήκης), **πληροφορίες ανάλυσης προτύπων (pattern analysis)** (π.χ. ανακάλυψη των πιο δημοφιλών μονοπατιών που ακολουθούν οι επισκέπτες του δικτυακού τόπου (web site) μιας Βιβλιοθήκης).

Στο τελευταίο στάδιο της διαδικασίας εξόρυξης γνώσης, γίνεται παρουσίαση των αποτελεσμάτων της εξερεύνησης (που παρήχθησαν στο προηγούμενο στάδιο) στον παρατηρητή, με τέτοιο τρόπο ώστε να τον βοηθήσουν στο μέγιστο για την κατανόηση της φύσης τους και τον τρόπο ερμηνείας και χρήσης/αξιοποίησής τους για την λήψη των αναγκαίων αποφάσεων. Είναι πολύ σημαντικός ο τρόπος παρουσίασης των αποτελεσμάτων της εξόρυξης γνώσης (γραφικές απεικονίσεις, σχεδιαγράμματα), επειδή η χρησιμότητά τους μπορεί να εξαρτάται ακριβώς από αυτόν. Μέσω της παρουσίασης αυτής απεικονίζονται καθαρά οι συσχετίσεις των δεδομένων και τα συμπεράσματα που προκύπτουν με βάση αυτές τις συσχετίσεις.

3. Επιλογή δεδομένων, επεξεργασία αυτών και λήψη χρήσιμης πληροφορίας

Κατά την επιλογή των δεδομένων προς επεξεργασία, η μεθοδολογία επικεντρώνεται στην επιλογή αυτών που θα χρησιμεύσουν στη λήψη των συμπερασμάτων που τελικά αναζητούμε. Πρέπει να γίνει ένας συσχετισμός οντοτήτων και ομάδων αντικειμένων με σκοπό την εξαγωγή όσο το δυνατόν περισσότερων και χρησιμότερων συμπερασμάτων. Είναι λοιπόν αναγκαίο να αποφασιστεί τι ακριβώς θέλουμε να δούμε στο τέλος και σχετικά με ποιες συσχετίσεις οντοτήτων και μεγεθών ενδιαφερόμαστε. Ουσιαστικά γίνεται επικέντρωση στις οντότητες που θέλουμε να παρατηρήσουμε. Π.χ. αν θέλουμε να δούμε τις συσχετίσεις ομάδων χρηστών ως προς την κατηγορία του υλικού που δανείζονται, θα πρέπει οπωσδήποτε να συμπεριλάβουμε στη διαδικασία δεδομένα που έχουν άμεση σχέση με την ομάδα που ανήκει ο κάθε χρήστης καθώς και την κατηγορία που ανήκει το δανεισμένο υλικό. Στην παρούσα εργασία χρησιμοποιήσαμε δεδομένα που προέρχονται από τη βάση δεδομένων διαχείρισης υλικού και χρηστών της Βιβλιοθήκης του Πολυτεχνείου Κρήτης. Στη μελέτη περίπτωσης που εξετάζουμε θέλουμε να δούμε:

- τη σχέση που έχουν οι διάφορες κατηγορίες χρηστών με το είδος και την κατηγορία του υλικού που χρησιμοποιούν και δανείζονται
- τη σχέση που έχουν οι διάφορες κατηγορίες χρηστών με το χρόνο δανεισμού του υλικού και το γεγονός εμπρόθεσμης ή εκπρόθεσμης επιστροφής του από αυτούς.

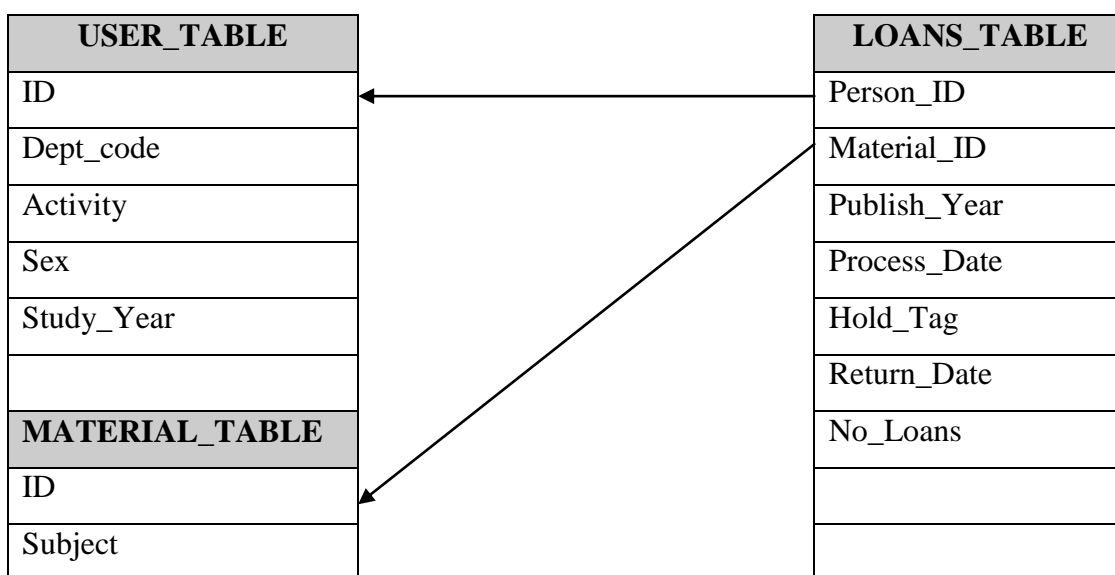
Με βάση λοιπόν το τι θέλουμε ουσιαστικά να παρατηρήσουμε, επιλέγουμε τα δεδομένα εκείνα που σχετίζονται με αυτό. Στη συγκεκριμένη περίπτωση λοιπόν μας ενδιαφέρουν τα δεδομένα που περιγράφουν τους χρήστες της Βιβλιοθήκης (user's data), το υλικό της (material's data) και τα δανειστικά δεδομένα (user's loan registration). Οι παρακάτω πίνακες περιγράφουν τα δεδομένα αυτά.

Πίνακας I. Στοιχεία χρηστών (USER_TABLE)	
Πεδίο	Περιγραφή πεδίου
ID	Μοναδική ταυτότητα χρήστη
Dept_code	Τμήμα στο οποίο ο χρήστης ανήκει
Activity	Ιδιότητα που έχει ο χρήστης (Προπτυχιακός, Μεταπτυχιακός, Διδακτορικός Φοιτητής, Καθηγητής, Υπάλληλος, Εξωτερικός χρήστης)
Sex	Το φύλο του χρήστη (άνδρας, γυναίκα)
Study_Year	Το έτος φοίτησης αν πρόκειται για φοιτητή. Σε άλλη περίπτωση αφήνεται κενό.

Πίνακας II. Στοιχεία Υλικού (MATERIAL_TABLE)	
Πεδίο	Περιγραφή πεδίου
ID	Μοναδική ταυτότητα του υλικού
Subject	Πεδίο στο οποίο ανήκει το υλικό

Πίνακας III. Στοιχεία δανεισμών υλικού (LOANS_TABLE)	
Πεδίο	Περιγραφή πεδίου
Material_ID	Η μοναδική ταυτότητα του υλικού (προέρχεται από τον πίνακα II)
Person_ID	Η μοναδική ταυτότητα του χρήστη που δανείστηκε το υλικό (προέρχεται από τον πίνακα I)
Publish_Year	Η ημερομηνία που εισήχθη στη συλλογή υλικού της Βιβλιοθήκης
Process_Date	Η ημερομηνία του δανεισμού
Hold_Tag	Δείχνει αν το υλικό είναι προς κράτηση
Return_Date	Η ημερομηνία που πρέπει να επιστραφεί το υλικό
No_Loans	Ο συνολικός αριθμός δανεισμών του

Η συσχέτιση των παραπάνω πινάκων απεικονίζεται στο παρακάτω σχήμα:



Με βάση τους παραπάνω πίνακες αντλήθηκαν δεδομένα από τη βάση της Βιβλιοθήκης και αυτά εισήχθησαν στους παραπάνω πίνακες. Έτσι, πραγματοποιήθηκε το 1^ο στάδιο της επιλογής και εισαγωγής δεδομένων.

Στη συνέχεια (2^ο στάδιο) ελέγχθηκαν οι τιμές των δεδομένων στους πίνακες με σκοπό τη συμπλήρωση / διόρθωση αυτών που δεν είχαν κάποιο νόημα ή ήταν ελλιπής. Συγκεκριμένα έγινε συμπλήρωση τιμών σε πεδία από όπου αυτές έλειπαν όπως για παράδειγμα σε κάποιους χρήστες το τμήμα που ανήκουν ή το φύλο τους.

Στο επόμενο στάδιο (στάδιο «εμπλουτισμού») ελέγχθηκε αν θα πρέπει να γίνει εμπλουτισμός των δεδομένων από κάποια άλλα εξωτερικά δεδομένα. Για την συγκεκριμένη ανάλυση δεν απαιτήθηκε κάτι τέτοιο. Σε άλλες όμως αναλύσεις δεδομένων μπορεί να κριθεί ότι στα αρχικά δεδομένα θα πρέπει να προστεθούν και κάποια άλλα από άλλες πηγές (π.χ. ερωτηματολόγια χρηστών, στοιχεία για το κόστος απόκτησης του υλικού) με σκοπό τη συμπλήρωση των ήδη υπαρχόντων για αποδοτικότερη επεξεργασία και λήψη καλύτερων αποτελεσμάτων.

Στο 4^ο στάδιο πραγματοποιήθηκαν κάποιοι μετασχηματισμοί σε ορισμένες τιμές δεδομένων με σκοπό τη μετατροπή τους σε ένα κοινό σχήμα για καλύτερη επεξεργασία. Έτσι, κάποιες τιμές που περιέγραφαν το γένος και ήταν εκφρασμένες είτε με «Άνδρας / Γυναίκα» είτε με «Α/Γ» μετασχηματίστηκαν σε «0/1» (0: Άνδρας, 1: Γυναίκα). Επίσης, έγινε μετασχηματισμός των τμημάτων που ανήκει ο κάθε χρήστης χρησιμοποιώντας αριθμούς για κάθε ένα από αυτά (HMMY: 10, ΜΠΔ: 11, ΜΗΧΟΠ: 12, ΜΗΠΕΡ: 13, ΑΡΧ: 14,

ΥΠΟΛΟΙΠΑ: 16) και μετασχηματισμός της ιδιότητας που έχει ο χρήστης με χρήση μοναδικών χαρακτήρων (Προπτυχιακός: Π, Μεταπτυχιακός: Μ, Διδακτορικός Φοιτητής: Δ, Καθηγητής: Κ, Υπάλληλος: Υ, Εξωτερικός χρήστης: ΕΞ). Με αυτό τον τρόπο γίνεται ταχύτερη η επεξεργασία καθώς μειώνεται κατά πολύ το μέγεθος των τιμών. Όσον αφορά το μετασχηματισμό του χαρακτηρισμού του υλικού επιλέχθηκε να δοθεί ως TECHNICAL (υλικό τεχνικού ενδιαφέροντος, όπως μαθηματικά, υπολογιστές, μηχανική κτλ.) και LITERATURE (υλικό λογοτεχνικού ενδιαφέροντος)

Στο 5^ο στάδιο εφαρμόστηκαν μοντέλα εξόρυξης γνώσης στα δεδομένα των πινάκων. Στη διαδικασία που περιγράφεται χρησιμοποιήθηκε το πρόγραμμα Clementine (SPSS). Η λεπτομερής διαδικασία που έγινε για να παραμετροποιηθεί το εν λόγω πρόγραμμα και να κάνει την ανάλυση των δεδομένων που περιγράφηκαν δεν είναι σκοπός της συγκεκριμένης εργασίας και δεν περιγράφεται σε αυτή. Απλά, δίνονται οι γενικές αρχές που δηλώθηκαν στο πρόγραμμα για να γίνει η περαιτέρω επεξεργασία των δεδομένων.

Αρχικά λοιπόν έγινε ορισμός των πινάκων στο πρόγραμμα και κατόπιν εισαγωγή των αρχικών δεδομένων σε αυτό. Κατόπιν έγινε «καθαρισμός» των δεδομένων, μέσω ανάλογου μενού επιλογών στο Clementine καθώς επίσης και οι μετασχηματισμοί που αναφέρθηκαν. Στη συνέχεια δηλώθηκαν ποια δεδομένα θα θέλαμε να συσχετίσουμε για τη λήψη των ανάλογων συμπερασμάτων. Έτσι, δηλώθηκε ότι η επιθυμητή συσχέτιση είναι μεταξύ των χρηστών και του είδους υλικού που δανείζονται (ποσοστό ανά ομάδα χρηστών σε σχέση με το υλικό) καθώς επίσης και μεταξύ των χρηστών και του χρόνου επιστροφής του υλικού (ποσοστό ανά ομάδα χρηστών σε σχέση με το χρόνο επιστροφής).

Κατόπιν το πρόγραμμα εφάρμοσε τεχνικές ανάλυσης και εξόρυξης γνώσεων, σύμφωνα με το τι συσχετισμοί δεδομένων ζητήθηκαν στο προηγούμενο βήμα. Συγκεκριμένα το πρόγραμμα χρησιμοποίησε μεθόδους κατηγοριοποίησης και ομαδοποίησης δεδομένων. Οι αλγόριθμοι και οι λεπτομερής τεχνικές δεν είναι σκοπός της εν λόγω εργασίας να περιγραφούν.

Τελικά, προέκυψαν πίνακες συσχετίσεων δεδομένων από την όλη διαδικασία και σύμφωνα με τα εισαγόμενα και επεξεργασμένα δεδομένα οι ομάδες των χρηστών που προέκυψαν καθώς επίσης και οι συσχετισμοί τους με το υλικό παριστάνονται στους παρακάτω πίνακες:

ΟΜΑΔΑ	ΣΥΣΧΕΤΙΣΜΟΣ	ΠΟΣΟΣΤΟ
Π (προπτυχιακοί)	TECHNICAL	77%
	LITERATURE	33%
Μ (μεταπτυχιακοί)	TECHNICAL	64%
	LITERATURE	36%
Δ (διδασκτορικοί)	TECHNICAL	82%
	LITERATURE	18%
Κ (καθηγητές)	TECHNICAL	88%
	LITERATURE	12%
Υ (υπάλληλοι)	TECHNICAL	44%
	LITERATURE	56%
ΕΞ (εξωτ. χρήστες)	TECHNICAL	38%
	LITERATURE	62%

Από τον παραπάνω πίνακα μπορεί να διαπιστωθεί ότι π.χ. το 64% των μεταπτυχιακών φοιτητών δανείζονται βιβλία τεχνικού ενδιαφέροντος ενώ το 36% βιβλία λογοτεχνικού ενδιαφέροντος.

ΟΜΑΔΑ	ΣΥΣΧΕΤΙΣΜΟΣ	ΠΟΣΟΣΤΟ
Π (προπτυχιακοί)	Εκπρόθεσμοι	37%
	Εμπρόθεσμοι	63%
Μ (μεταπτυχιακοί)	Εκπρόθεσμοι	39%
	Εμπρόθεσμοι	61%
Δ (διδασκτορικοί)	Εκπρόθεσμοι	35%
	Εμπρόθεσμοι	65%
Κ (καθηγητές)	Εκπρόθεσμοι	83%
	Εμπρόθεσμοι	17%
Υ (υπάλληλοι)	Εκπρόθεσμοι	37%
	Εμπρόθεσμοι	63%
ΕΞ (εξωτ. χρήστες)	Εκπρόθεσμοι	21%
	Εμπρόθεσμοι	79%

Από τον παραπάνω πίνακα μπορεί να διαπιστωθεί ότι π.χ. το 83% των καθηγητών αργούν να επιστρέψουν το υλικό όταν πρέπει. Ο χαρακτηρισμός Εκπρόθεσμοι ή Εμπρόθεσμοι, ουσιαστικά προέκυψε κατόπιν σύγκρισης (που δηλώθηκε στο πρόγραμμα Clementine) μεταξύ του Return_Date (τυπικής ημερομηνίας επιστροφής του υλικού) και του Current_Date (σημερινή ημερομηνία).

Από τα δεδομένα που αναφέρθηκαν στην αρχική επιλογή κάποια δεν χρησιμοποιήθηκαν στη μελέτη περίπτωσης που αναφερθήκαμε (π.χ. Publish_Year, Study_Year, Sex). Μπορούν όμως κάλλιστα να χρησιμοποιηθούν σε διαφορετικές ζητούμενες συσχετίσεις (π.χ. υλικό που δανείζεται σε σχέση με το φύλλο ή το έτος του αναγνώστη). Από τη στιγμή που υπάρχουν τα αναγκαία δεδομένα, η αλλαγή των ζητούμενων συσχετίσεων είναι πολύ εύκολη.

Παρόμοιοι πίνακες συσχετίσεων είναι πολύ απλό να παραχθούν από το πρόγραμμα επεξεργασίας δεδομένων που χρησιμοποιήθηκε αρκεί να δηλωθούν οι ανάλογες παράμετροι σε αυτό. Το σημαντικό βήμα ουσιαστικά είναι η επιλογή των αρχικών δεδομένων, ο ορισμός των επιθυμητών συσχετίσεων και ο ορισμός του είδους των τελικών πληροφοριών που θέλουμε να λάβουμε.

Τέτοιες συσχετίσεις είναι πολύ σημαντικές λαμβάνοντας υπόψη ότι τα δεδομένα που χρησιμοποιήθηκαν προέρχονται από Ακαδημαϊκή Βιβλιοθήκη Πολυτεχνείου. Από τις συσχετίσεις λοιπόν αυτές (οι οποίες ουσιαστικά αποτελούν τη τελική γνώση) μπορούν να προκύψουν χρήσιμα συμπεράσματα ικανά να βοηθήσουν τους ιθύνοντες να πάρουν σημαντικές αποφάσεις με σκοπό την βελτίωση των προσφερόμενων υπηρεσιών (π.χ. αύξηση του υλικού λογοτεχνικού ενδιαφέροντος, αύξηση του χρόνου δανεισμού).

Τέλος, θα πρέπει να επισημάνουμε ότι είναι δυνατή και συνίσταται η χρησιμοποίηση δεδομένων με σκοπό τον ορισμό δεικτών ικανών να δώσουν χρήσιμες πληροφορίες για τις διαδικασίες που εφαρμόζονται σε μια Βιβλιοθήκη.

Ένας σημαντικός τέτοιος δείκτης είναι ο (Μελετίου-Κατσιρίκου 2004) *Δείκτης Άμεσης Διαθεσιμότητας (ΔΑΔ)* = (αριθμός αιτήσεων για κράτηση) / (αριθμός δανεισμών), ο οποίος εκφράζει το «συντελεστή κρατήσεων σε σχέση με το πλήθος των δανεισμών ενός τίτλου για μια περίοδο» και δείχνει το ποσοστό άμεσης διαθεσιμότητας του τίτλου αυτού. Όσο μεγαλώνει με το χρόνο ο δείκτης αυτός τόσο η ζήτηση του τίτλου αυξάνει.

Άλλος δείκτης που έχει σχέση με το χαρακτηρισμό προμηθευτών είναι (Μελετίου-Κατσιρίκου 2004) ο *Δείκτης καταλληλότητας προμηθευτή*, ο οποίος εκφράζει το χρόνο απόκρισης του προμηθευτή σε σχέση με το συνολικό κόστος του υλικού («λόγος κόστους/χρόνου λήψης»).

Οι δείκτες αυτοί, με κατάλληλες συσχετίσεις είναι σε θέση να δώσουν σημαντικές πληροφορίες τόσο για την οργάνωση του υλικού όσο και για την ποιότητα των προσφερομένων υπηρεσιών σε μια Βιβλιοθήκη.

4. Συμπεράσματα

Η συγκεκριμένη εργασία επικεντρώθηκε στην ανάλυση δεδομένων που προέρχονται από το χώρο των Βιβλιοθηκών και τα οποία παράγονται σε καθημερινή βάση.

Περιέγραψε μια γενική μεθοδολογία στηριζόμενη σε διαδικασίες «εξερεύνησης και εξόρυξης γνώσης από δεδομένα» (data mining). Χρησιμοποίησε πραγματικά δεδομένα και αναλύοντάς τα με βάση τις τεχνικές και τις μεθόδους «Εξόρυξης γνώσης» (data mining methods) εξήγαγε χρήσιμα συμπεράσματα συσχετίσεων αρχικών δεδομένων, οι οποίες ήταν αδύνατο να εξαχθούν δίχως την ανάλυση και επεξεργασία αυτή.

Τέλος, έδωσε τις γενικές αρχές με βάση τις οποίες θα πρέπει να γίνεται η επιλογή των δεδομένων προς επεξεργασία και τόνισε τον τρόπο με τον οποίο θα πρέπει να συνδέονται αυτά τα δεδομένα. Απώτερος σκοπός της διαδικασίας αυτής είναι η πληροφορία που θα προκύψει από την ανάλυση και επεξεργασία τους να είναι σε θέση να κατανοηθεί, να αξιολογηθεί και να χρησιμοποιηθεί με τέτοιο τρόπο που θα βοηθήσει στη χάραξη στρατηγικών και στη λήψη αποφάσεων με σκοπό τη σωστή λειτουργία, οργάνωση και βελτίωση των προσφερόμενων υπηρεσιών μιας Βιβλιοθήκης.

Βιβλιογραφία

Λίστα αναφορών. Αλφαβητική σειρά των αναφορών.

Μελετίου, Α. και Κατσιρίκου Α.: *Ποιοτικοί Δείκτες Υπηρεσιών Βιβλιοθηκών και Διαχείριση Πόρων: Μεθοδολογίες Ανάλυσης και στρατηγικός σχεδιασμός*, 15^ο συνέδριο Ακαδημαϊκών Βιβλιοθηκών, 2006

Anand, S.S. and Buchner, A.G. (1997), *Decision Support Using Data Mining*, Prentice-Hall, Englewood Cliffs, Nj.

Chang, Chan-Chine και Chen, Ruey-Shun (2006): *Using data mining technology to solve classification problems (A case study of campus digital library)*, The Electronic Library, Vol.24 No 3, pp. 129-139

Dunham, Margaret H. 2004. *Data Mining: Εισαγωγικά και Προηγμένα Θέματα Εξόρυξης Γνώσης από Δεδομένα*, Εκδόσεις Νέων Τεχνολογιών.

Yu, S-C. and Chen, R.-S. (2001), *Development an XML framework for an electronic document delivery system*, The Electronic Library, Vol. 19 No 2, pp. 102-105.