

ΠΟΙΟΤΙΚΟΙ ΔΕΙΚΤΕΣ ΥΠΗΡΕΣΙΩΝ ΒΙΒΛΙΟΘΗΚΩΝ & ΔΙΑΧΕΙΡΗΣΗ ΠΟΡΩΝ: ΜΕΘΟΔΟΛΟΓΙΕΣ ΑΝΑΛΥΣΗΣ & ΣΤΡΑΤΗΓΙΚΟΣ ΣΧΕΔΙΑΣΜΟΣ

ΑΡΙΣΤΕΙΔΗΣ ΜΕΛΕΤΙΟΥ - ΑΝΘΗ ΚΑΤΣΙΡΙΚΟΥ

Περίληψη

Είναι γεγονός ότι τα τελευταία χρόνια ο όγκος των πληροφοριών που έχουν σχέση με δεδομένα διαχείρισης βιβλιοθηκών αυξάνει με πολύ γρήγορους ρυθμούς, κάτι που καθιστά αρκετά δύσκολη τη διαδικασία ανάλυσής τους με απώτερο σκοπό τη λήψη αποφάσεων που σχετίζονται με την βελτίωση της ποιότητας των υπηρεσιών.

Η παρούσα εργασία στοχεύει στον ορισμό μιας νέας μεθοδολογίας ανάλυσης των αποτελεσμάτων αποτίμησης ποιότητας των ελληνικών βιβλιοθηκών, για τη διάγνωση της υφιστάμενης οργάνωσης και διοίκησης των ανθρώπινων και οικονομικών πόρων μιας βιβλιοθήκης, αλλά και για την συμβολή στην ευκολότερη και αποδοτικότερη λήψη αποφάσεων.

Η χρήση και ανάλυση με σωστό και μεθοδευμένο τρόπο, κάθε είδους δεδομένων που σχετίζονται με την λειτουργία μια βιβλιοθήκης, η χρήση δεικτών αποτίμησης της ποιότητας των βιβλιοθηκών και ο συνδυασμός των αποτελεσμάτων τους δίνει σημαντικά συμπεράσματα για τη δομή των Βιβλιοθηκών, γεγονός που οδηγεί στη διαμόρφωση του κατάλληλου στρατηγικού σχεδιασμού, για κάθε περίπτωση.

Ο όρος δομή αναφέρεται στη λειτουργία της βιβλιοθήκης και περιλαμβάνει τόσο την οργάνωση των προσφερόμενων υπηρεσιών προς τους χρήστες, όσο και την οργάνωση των διαδικασιών εντός της Βιβλιοθήκης. Περιλαμβάνει επίσης τη διαχείριση του ανθρώπινου δυναμικού, με σκοπό την κατά το δυνατόν άριστη αξιοποίησή του, καθώς και τη διαχείριση των οικονομικών πόρων με τον αποτελεσματικότερο και αποδοτικότερο τρόπο.

Στη συγκεκριμένη ανάλυση θα αναφερθούν μέθοδοι εξόρυξης πληροφορίας από ένα μεγάλο σύνολο λειτουργικών δεδομένων που χρησιμοποιούνται κατά την προσφορά υπηρεσιών και οργάνωσης των σύγχρονων βιβλιοθηκών. Απώτερος σκοπός είναι η εξαγωγή χρήσιμων συμπερασμάτων από την διαδικασία εξόρυξης που θα βοηθήσουν στη βελτίωση των παρεχόμενων υπηρεσιών και του τρόπου οργάνωσης και λειτουργίας τους (π.χ. συλλογές / αιτούμενες ανάγκες ομάδων χρηστών / πολιτικές κατανομής κονδυλίων για απόκτηση υλικού).

Εφαρμογές αυτής της μεθοδολογίας αφορούν την αποτίμηση της συλλογής σε συνδυασμό με τις θεματικές υποδιαιρέσεις, τις σχολές, τα ακαδημαϊκά τμήματα που εξυπηρετεί μια Βιβλιοθήκη, αφορούν επίσης το είδος των πηγών, σε συνδυασμό με τις κατηγορίες των χρηστών, αλλά και την επιστημονική κατεύθυνση αυτών.

Μια άλλη εφαρμογή της μεθόδου αφορά τις επιπτώσεις που μπορεί να εμφανιστούν σε περίπτωση δυσανάλογης αυξομείωσης ενός υποσυνόλου της συλλογής, π.χ. των ηλεκτρονικών έναντι των έντυπων πηγών ή μιας θεματικής κατηγορίας εις βάρος άλλων ή του συνδυασμού των δύο και που οι επιπτώσεις αυτές επηρεάζουν το σύνολο της παρουσίας της βιβλιοθήκης, σε επίπεδο πόρων, οργάνωσης, υποκατάστασης ή συμπλήρωσης υπηρεσιών, ακόμα και εξάσκησης του προσωπικού σε νέες δραστηριότητες.

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

1. Εισαγωγή

Είναι γεγονός ότι συνεχώς το μέγεθος των πληροφοριών που αποθηκεύονται συνολικά σε καθημερινή βάση, στις βάσεις δεδομένων που εφαρμόζονται σε μια βιβλιοθήκη αυξάνει με γρήγορους ρυθμούς. Είναι λοιπόν αυτονόητο ότι ο όγκος των λειτουργικών δεδομένων (δεδομένα που χρησιμοποιούνται ή δύναται να χρησιμοποιηθούν) που υπάρχει σε μια βιβλιοθήκη είναι μεγάλος και συνεχώς αυξάνεται. Παράλληλα δεν πρέπει να ξεχνάμε το γεγονός ότι τα δεδομένα (τα οποία υπάρχει περίπτωση να αναφέρονται στο ίδιο αντικείμενο) συνήθως είναι αποθηκευμένα σε παραπάνω από μία βάσεις δεδομένων, κάτι που μας οδηγεί να συνειδητοποιήσουμε αμέσως το βαθμό ανομοιογένεια της εν λόγω πληροφορίας (πολλές βάσεις δεδομένων με διαφορετικά στοιχεία η καθεμιά).

Το πρόβλημα λοιπόν έγκειται στη συγκέντρωση της πληροφορίας αυτής με τέτοιο τρόπο που θα επιτρέψει την επεξεργασία της με σκοπό την εξαγωγή χρήσιμων συμπερασμάτων που οδηγούν στην λήψη αποφάσεων (π.χ. διοικητικές, οικονομικές).

Η τεχνολογία που επιτρέπει το παραπάνω είναι η τε-

χνολογία εξόρυξης γνώσης από βάσεις δεδομένων (data mining). Θέλοντας λοιπόν να δώσουμε ένα πιο εμπειριστατικό ορισμό θα μπορούσαμε να πούμε ότι η εξόρυξη γνώσης (data mining technology) (Kao και Chang και Lin 2003) είναι μια διαδικασία κατά την οποία ανακαλύπτεται υπονοούμενη γνώση μέσα από μεγάλες βάσεις δεδομένων. Η τεχνολογία αυτή έχει την ικανότητα να αποκαλύπτει κρυμμένες σχέσεις, κρυμμένους συσχετισμούς, κρυμμένα πρότυπα και τάσεις σε δεδομένα που είναι αποθηκευμένα με παραδοσιακούς τρόπους σε μεγάλες βάσεις δεδομένων. Με άλλα λόγια η τεχνολογία αυτή αποκαλύπτει σχέσεις των δεδομένων που υπάρχουν μεν αλλά δεν είναι εμφανή (Han και Fu 1999; Hirota και Pedrycz 1999).

Το παρόν άρθρο περιγράφει τη διαδικασία αυτή και προτείνει μεθοδολογίες για την αξιοποίηση και χρήση των τεχνικών που εφαρμόζονται σε χώρους διαφορετικούς των Βιβλιοθηκών (π.χ. οικονομικοί, εμπορικοί, ιατρικοί, τεχνολογικοί) στα δεδομένα που αφορούν βιβλιοθήκες.

2. Γενική περιγραφή της διαδικασίας εξερεύνησης γνώσης

Στην ενότητα αυτή θα περιγράψουμε γενικά, τη διαδικασία εξερεύνησης γνώσεων από ένα σύνολο δεδομένων, αναφέροντας μια μικρή περιγραφή των διαφορετικών βημάτων που τη συνθέτουν και πρέπει να ακολουθηθούν.

Όπως φαίνεται από το παραπάνω σχήμα (Dunham 2004) η διαδικασία εξερεύνησης γνώσεων χωρίζεται σε έξι στάδια. Παρόλο που μπορεί κανείς να υποθέσει ότι τα στάδια αυτά είναι ανεξάρτητα μεταξύ τους, στην πράξη η διαδικασία μπορεί να γυρίσει ένα ή περισσότερα βήματα πίσω. Για παράδειγμα, στη φάση της κωδικοποίησης μπορεί να διαπιστωθεί ότι η διαδικασία του καθαρισμού δεν έγινε ικανοποιητικά και το σύστημα γυρνά δυο επίπεδα πίσω, ή κατά την εξερεύνηση προτύπων μπορεί να ανακαλύφθηκαν νέα δεδομένα και αυτά να χρησιμοποιηθούν για να εμπλουτίσουν ήδη υπάρχοντα δεδομένα.

Στην επιλογή των δεδομένων συλλέγουμε τα δεδομένα που απαιτούνται για τη διαδικασία της εξόρυξης και ανακάλυψης γνώσης (λειτουργικά). Τα δεδομένα αυτά μπορούν να προέλθουν από πολλές διαφορετικές και ετερογενείς πηγές δεδομένων (βάσεις δεδομένων, μη ηλεκτρονικές πηγές κτλ.).

Στο επόμενο στάδιο γίνεται ένα φιλτράρισμα των δεδομένων από πιθανή περιττή πληροφορία, λάθος δεδομένα και μόλυνση, φαινόμενα που επηρεάζουν σημαντικά τη διαδικασία εξερεύνησης νέας πληροφορίας οδηγώντας πολλές φορές σε ακραία αποτε-

λέσματα. Τα λανθασμένα δεδομένα μπορούν να διορθωθούν ή αφαιρεθούν, ενώ τα ελλιπή να εκτιμηθούν και να συμπληρωθούν. Για το φιλτράρισμα της μόλυνσης από τα δεδομένα τα συστήματα εξερεύνησης γνώσεων χρησιμοποιούνται συνήθως data mining και pattern recognition τεχνικές (τεχνικές εξόρυξης γνώσης και αναγνώρισης προτύπων) ή πιο απλές λύσεις όπως η αφαίρεση των εγγραφών (records) που έχουν λίγα δεδομένα, ο έλεγχος του πεδίου τιμών των ιδιοτήτων (attributes) των εγγραφών (records), η αντικατάσταση μη λογικών (invalid) τιμών με μηδενικές (null) κ.α.

Θα πρέπει να σημειωθεί ότι στη διαδικασία εξερεύνησης γνώσεων δε μας πειράζει να αφαιρέσουμε ένα πολύ μικρό ποσοστό μολυσμένων records αφού δεν πρόκειται να επηρεαστούν σημαντικά τα τελικά συμπεράσματα. Αντίθετα αν συμπεριληφθούν στην εξαγωγή συμπερασμάτων μπορεί να προκαλέσουν λάθος συμπεράσματα. Πάντως σε πολλές περιπτώσεις εξερεύνησης γνώσεων μας ενδιαφέρουν τα στοιχεία της μόλυνσης όπως για π.χ. το ποσοστό των χρηστών που αρνήθηκαν να δώσουν τα πραγματικά τους στοιχεία κ.α.

Στο στάδιο «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 λαμβάνουν χώρα κάποιοι μετασχηματισμοί που κάνουν πιο αποδοτική την διαδικασία της εξερεύνησης γνώσης. Τα δεδομένα που προέρχονται από διαφορετικές πηγές χρειάζεται να μετατραπούν σε ένα κοινό σχήμα για την περαιτέρω επεξεργασία τους. Μερικά δεδομένα ίσως απαιτείται να κωδικοποιηθούν ή να μετασχηματιστούν σε χρήσιμα σχήματα (π.χ. μετατροπή της ημερομηνία γέννησης σε ηλικία, διαίρεση των χρηματικών ποσών με 1000, μετατροπή των τιμών yes/no σε 1/0). Μπορεί επίσης να μειωθούν τα δεδομένα για να ελαττωθεί ο αριθμός των πιθανών τιμών των δεδομένων που θα ληφθούν υπόψη.

Στο στάδιο αυτό, γίνεται η εξερεύνηση των δεδομένων εφαρμόζοντας (ανάλογα με το είδος της εξόρυξης) μοντέλα και τεχνικές εξόρυξης γνώσης μέσω του οποίων πλέον λαμβάνονται οι πληροφορίες που έχουν σχέση με το συσχετισμό των δεδομένων, μέσω των οποίων θα παραχθούν οι χρήσιμες πληροφορίες της όλης διαδικασίας. Ενδεικτικά οι γνώσεις που

προκύπτουν από την εφαρμογή των μοντέλων και των τεχνικών αυτών μπορεί να είναι:

- **Κανόνες συσχέτισης (association rules):** Πρόκειται για εξαγωγή κανόνων που συσχετίζουν διάφορα αντικείμενα που ενδιαφέρουν το χρήστη (data miner), όπως για π.χ. «το 60% των πελατών που αγοράζουν περιοδικά αυτοκινήτων αγοράζουν και αθλητικά περιοδικά».
- **Κατηγοριοποίηση (classification):** Επιτυγχάνει την ταξινόμηση των δεδομένων της βάσης σύμφωνα με τις τιμές που έχουν σε κάποιες παραμέτρους, για π.χ. ταξινόμηση των αυτοκινήτων βάσει της κατανάλωσης τους ή του αριθμού χιλιομέτρων που έχουν κάνει.
- **Συσταδοποίηση (clustering):** Κατηγοριοποίηση των δεδομένων σε μικρό αριθμό κλάσεων με κάποια ιδιαίτερα κοινά χαρακτηριστικά, για π.χ. κατηγοριοποίηση των τρόπων προώθησης ενός προϊόντος ανάλογα με τις συνήθειες και τα εισοδήματα των πληθυσμών διαφορετικών περιοχών.
- **Ανάλυση τάσεων και αποκλίσεων (trend and deviation analysis):** Ανακάλυψη των κυρίων τάσεων και των αποκλίσεων των δεδομένων της βάσης, για π.χ. ανακάλυψη των μετοχών που αποδίδουν καλύτερα ή χειρότερα από το μέσο όρο σε ένα χρηματιστήριο.
- **Ανάλυση προτύπων (pattern analysis):** Ανακάλυψη συγκεκριμένων προτύπων στα δεδομένα της βάσης που καθορίζονται από το παρατηρητή (data miner), για π.χ. ανακάλυψη των πιο δημοφιλών μονοπατιών που ακολουθούν οι επισκέπτες ενός web site.

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 σημαντικό το πώς θα παρουσιαστούν στους χρήστες τα αποτελέσματα της εξόρυξης γνώσης, επειδή η χρησιμότητα ή μη των αποτελεσμάτων μπορεί να εξαρτάται ακριβώς από αυτήν την παρουσίαση. Για το σκοπό αυτό χρησιμοποιούνται διάφορες στρατηγικές οπτικοποίησης (οπτικής παρουσίασης των δεδομένων) (π.χ. γραφικές παραστάσεις όπως ραβδογράμματα, πίτες, ιστογράμματα, γραμμογράμματα, τρισδιάστατες απεικονίσεις).

3. Η διαδικασία εξόρυξης γνώσης σε δεδομένα που σχετίζονται με βιβλιοθήκες

Στην περίπτωση εφαρμογής της συγκεκριμένης διαδικασίας σε δεδομένα που σχετίζονται με βιβλιοθήκες,

περιγράφεται ακολούθως η διαδικασία προσαρμοσμένη στο τρόπο που στοχεύει να παρουσιάσει η εν λόγω εργασία όσον αφορά τη λήψη και επεξεργασία των δεδομένων αυτών με μεθόδους εξόρυξης γνώσης.

3.1. Επιλογή δεδομένων προς επεξεργασία

Αρχικά θα πρέπει να αποφασίσουμε ποια δεδομένα από αυτά που έχουν αποθηκευτεί στις βάσεις που εφαρμόζονται στη βιβλιοθήκη, θα συμπεριλάβουμε στη διαδικασία της εξερεύνησης γνώσης.

Όσον αφορά τα λειτουργικά δεδομένα (operational data) δεδομένα που μπορούν να χρησιμοποιηθούν κατά τη διαδικασία εξόρυξης γνώσης, και συγκεκριμένα στο στάδιο της συλλογής δεδομένων (data selection), αναφέρουμε ενδεικτικά τα σπουδαιότερα από αυτά:

3.1.1 Δεδομένα από τον OPAC

Τα δεδομένα που συλλέγονται εδώ έχουν σχέση με το είδος των βιβλίων (τίτλος, κατηγορία-θεματική συλλογή) που αναζητεί ο χρήστης. Δεν έχει σχέση προφανώς με το τι δανείζεται τελικά ο χρήστης μιας και μπορεί να αναζητήσει κάποιο βιβλίο και τελικά να μην το δανειστεί. Τα δεδομένα αυτά βρίσκονται μέσω των αρχείων ημερολογίου και ερωτήσεων (log files με τα queries που γίνονται στη βάση του OPAC). Ενδεικτικές τιμές που θα μπορούσαν να επιλεγούν για επεξεργασία είναι ο κωδικός του βιβλίου, ο τίτλος του και η κατηγορία στην οποία ανήκει και η ημερομηνία αναζήτησης.

3.1.2 Δανειστική κατάσταση χρηστών και δεδομένα χρήσης υλικού (έντυπου ή ηλεκτρονικού)

Τα δεδομένα που συλλέγονται εδώ έχουν σχέση με την ταυτότητα, κατηγορία και τμήμα του χρήστη, ποια βιβλία έχει δανειστεί ο καθένας, πόσο καιρό έχει κρατήσει το υλικό, πόσες φορές υπέστη δανεισμό ένα βιβλίο σε μια περίοδο και πόσες φορές το βιβλίο αυτό δεν ήταν διαθέσιμο και υπέστη κράτηση, πόσα και ποια βιβλία χαρακτηρίστηκαν εκπρόθεσμα κατά την επιστροφή τους, πόσοι δανεισμοί έχει κάνει κάθε χρήστης, πόσες εμπρόθεσμες και πόσες εκπρόθεσμες επιστροφές έχει κάνει κάθε χρήστης. Ενδεικτικές τιμές που θα μπορούσαν να επιλεγούν για επεξεργασία είναι ο κωδικός του βιβλίου, ο κωδικός και η κατηγορία του χρήστη (π.χ. προπτυχιακός, μεταπτυχιακός, ΔΕΠ), το τμήμα του χρήστη, αριθμός εκπρόθεσμων και εμπρόθεσμων επιστροφών χρήστη, πόσα και ποια βιβλία έχει δανειστεί, ο κωδικός και η κατηγορία-θεματική συλλογή του βιβλίου, αριθμός δανεισμών κάθε βιβλίου, αριθμός κρατήσεων που έχουν γίνει στο βιβλίο, πόσες φορές κάθε βιβλίο επι-

στράφηκε εμπρόθεσμα ή εκπρόθεσμα.

Όταν πρόκειται για ηλεκτρονικό υλικό ενδεικτικές τιμές που θα μπορούσαν να επιλεγούν για επεξεργασία είναι ο κωδικός/τίτλος/θεματική κατηγορία του βιβλίου, ο κωδικός και η κατηγορία του χρήστη (π.χ. προπτυχιακός, μεταπτυχιακός, ΔΕΠ) που το διάβασε, το τμήμα του χρήστη. Τα δεδομένα αυτά λαμβάνονται από το σύστημα που εξυπηρετεί τα ηλεκτρονικά βιβλία (π.χ. log files από Web Servers).

3.1.3 Αρχεία ημερολογίου Web εξυπηρετητών (Web logs)

Τα δεδομένα που συλλέγονται εδώ έχουν σχέση με τις διαδρομές που ακολουθεί ο χρήστης μέσα στον ιστοχώρο (site) της Βιβλιοθήκης σε μια επίσκεψή του. Αν ονοματίσουμε τις σελίδες που υλοποιούν τον ιστοχώρο (π.χ. A,B,C,D,...) λαμβάνονται οι διαδρομές που ακολουθεί ο κάθε χρήστης (π.χ. Από τη σελίδα A στη C, μετά στη D, μετά στην B, δηλαδή η εν λόγω διαδρομή μπορεί να απεικονισθεί ως ACDB). Από τα δεδομένα αυτά μπορούν να προκύψουν οι δημοφιλέστεροι προορισμοί και οι δημοφιλέστερες διαδρομές. Τα δεδομένα αυτά ανακτώνται από τα αρχεία ημερολογίου (log files) των Web Servers της Βιβλιοθήκης και ουσιαστικά προκύπτει ένας πίνακας μίας στήλης στον οποίο κάθε γραμμή αναφέρει τις διαδρομές που ακολούθησε κάθε εικονική επίσκεψη χρήστη στον ιστοχώρο της Βιβλιοθήκης.

3.1.4 Περιοδικά συλλογής

Τα δεδομένα που συλλέγονται εδώ έχουν σχέση με τις παραγγελίες περιοδικών (έντυπων ή ηλεκτρονικών) που παραγγέλλονται από την κοινότητα και συνθέτουν τη συλλογή των περιοδικών της Βιβλιοθήκης και τις χρήσεις τους. Ενδεικτικές τιμές που θα μπορούσαν να επιλεγούν για επεξεργασία είναι ο τίτλος του περιοδικού, ποιος το παρήγγειλε (πρόσωπο, τμήμα), το κόστος του, το είδος του (έντυπο, ηλεκτρονικό), η θεματική του ενότητα, ο προμηθευτής, το διάστημα για το οποίο ανήκει στη συλλογή (π.χ. για 2 χρόνια και μετά διακόπηκε). Επιπλέον τιμές θα μπορούσαν να είναι ένας αριθμός που δείχνει τη χρήση περιοδικών (δηλαδή πόσες φορές αναγνώστηκε από κάποιον χρήστη). Η τιμή αυτή που θα μπορούσε να ληφθεί, για έντυπο υλικό είτε από τον OPAC (αν επιτρέπεται ο δανεισμός περιοδικών, είτε από τη συμπλήρωση εντύπου που ενδεχομένως να βρίσκεται μέσα στο ίδιο το περιοδικό). Αν το περιοδικό είναι ηλεκτρονικό θα μπορούσε να ληφθεί από τα στατιστικά που κρατούνται από τους Web Servers όπου φιλοξενείται η ηλεκτρονική μορφή των περιοδικών αυτών.

3.1.5 Διαδανεισμός

Τα δεδομένα που συλλέγονται εδώ έχουν σχέση με τα στοιχεία του υλικού προς διαδανεισμό: από ποια ομάδα χρηστών παραγγέλλεται (π.χ. φοιτητές, καθηγητές), από πού (προμηθευτής), πόσο χρόνο απαιτήθηκε για τη λήψη, το κόστος απόκτησης. Ενδεικτικές τιμές που θα μπορούσαν να επιλεγούν για επεξεργασία είναι ο κωδικός και η κατηγορία του χρήστη (π.χ. προπτυχιακός, μεταπτυχιακός, ΔΕΠ), το τμήμα του χρήστη, ο προμηθευτής από τον οποίο ζητήθηκε το υλικό και ο χρόνος (π.χ. σε ημέρες) που πέρασε για την λήψη, το κόστος.

Από το πόσος χρόνος απαιτήθηκε για να ληφθεί το υλικό από τον προμηθευτή που ζητήθηκε, μπορούμε να εξάγουμε ένα δείκτη που έχει σχέση με την καταλληλότητά του. Επιπλέον ο λόγος (κόστος / χρόνος λήψης) μας δείχνει συνολικά μια καλή εικόνα όσον αφορά την ποιότητα ενός προμηθευτή από τον οποίο ζητάμε υλικό. Ο δείκτης αυτός μπορεί κάλλιστα να εφαρμοστεί όχι μόνο στο διαδανεισμό (για τον χαρακτηρισμό του προμηθευτή από τον οποίο ζητάμε υλικό) αλλά και όσον αφορά βιβλία, περιοδικά ή οποιοδήποτε άλλο υλικό. Σε αυτόν τον τομέα ο χρόνος απόκρισης του προμηθευτή από τον οποίο ζητάμε υλικό (δηλαδή μετά από πόσο χρόνο αφού παραγγέλθηκε θα φτάσει στη Βιβλιοθήκη) σε σχέση με το συνολικό κόστος του υλικού δίνει μια καλή εικόνα για να χαρακτηριστεί ποσοτικά η ποιότητα (και η καταλληλότητα) ενός φορέα, κάτι που μπορεί να οδηγήσει σε αποφάσεις του τύπου «δεν επιλέγουμε τον προμηθευτή Α για βιβλία στον τομέα της Μηχανικής αλλά επιλέγουμε τον προμηθευτή Β γιατί είναι καλύτερος (λαμβάνοντας υπόψιν τον παραπάνω λόγο για τα βιβλία Μηχανικής που έχουμε προμηθευτεί).

3.1.6 Κόστος υλικού

Τα δεδομένα εδώ έχουν σχέση με το κόστος του υλικού είτε είναι έντυπο είτε ηλεκτρονικό: κόστος βιβλίων, περιοδικών (έντυπων, ηλεκτρονικών), οπτικοακουστικού υλικού, ηλεκτρονικών συνδρομών, ηλεκτρονικών βιβλίων (e-books). Ενδεικτικές τιμές που θα μπορούσαν να επιλεγούν για επεξεργασία είναι ο τίτλος του υλικού (π.χ. τίτλος βιβλίου), το είδος του (π.χ. ηλεκτρονικό βιβλίο), η ημερομηνία κτήσης του, ο προμηθευτής του.

3.1.7 Παράμετροι από Οργανισμό / Ίδρυμα:

Αναφερόμαστε σε παραμέτρους που δίνονται από τον οργανισμό (π.χ. Πανεπιστημιακές αρχές) και αφορούν συνολικά το Ίδρυμα ή και την Βιβλιοθήκη. Ενδεικτικές τιμές αυτών των παραμέτρων (οι οποίοι

αρκετές φορές λειτουργούν ως κριτήρια για τη λήψη σημαντικών αποφάσεων, όπως θα δούμε στη συνέχεια), είναι ο συνολικός αριθμός των μελών της κοινότητας ανά κατηγορία (π.χ. ΔΕΠ, προπτυχιακοί, μεταπτυχιακοί) και ανά τμήμα, ο αριθμός των τμημάτων του Ιδρύματος, ο αριθμός και το είδος των μαθημάτων του κάθε τμήματος, τα αντικείμενα που διδάσκονται σε κάθε τμήμα (π.χ. Φυσική, Μαθηματικά, Αρχαία, Λατινικά).

3.1.8 Ερωματολογία χρηστών

Εδώ αναφερόμαστε σε δεδομένα που προκύπτουν από ερωματολογία που έχουν συμπληρωθεί από τους ίδιους τους χρήστες και τα αποτελέσματά τους δύναται να έχουν καταχωρηθεί σε μια βάση δεδομένων (π.χ. Excel ή Access). Τα δεδομένα αυτά έχουν σχέση συνήθως με την ποσοτικοποίηση της ικανοποίηση χρηστών για μια υπηρεσία. Ενδεικτικές τιμές που θα μπορούσαν να επιλεγούν για επεξεργασία είναι το όνομα της υπηρεσίας, η διαβάθμιση του χαρακτηρισμού ικανοποίησης της υπηρεσίας (π.χ. ικανοποιημένος, πολύ ικανοποιημένος), το ποσοστό ή ο αριθμός των χρηστών που ανήκει σε κάθε διαβάθμιση (π.χ. 30% (ή 120 άτομα) απάντησαν ότι είναι πολύ ικανοποιημένα), η κατηγορία στην οποία ανήκει ο χρήστης (π.χ. προπτυχιακός), το τμήμα του, η ημερομηνία γέννησής του, το έτος σπουδών του (αν είναι φοιτητής).

3.1.9 Δείκτες που υποδεικνύουν ποιότητα υπηρεσιών αλλά και ποσοτικά μεγέθη

Αναφερόμαστε σε δείκτες γενικά αποδεκτούς από την Βιβλιοθηκονομική Κοινότητα (π.χ. δείκτες ΜΟΠΑΒ)

3.2. Καθάρισμα δεδομένων

Στη συνέχεια θα πρέπει να «καθαρίσουμε» (φιλτράρουμε) τις πληροφορίες που προκύπτουν από περιττά ή μολυσμένα στοιχεία που μπορούν να αλλάξουν και να επηρεάσουν τα αποτελέσματά μας.

Οι πιο συχνές πηγές μόλυνσης δεδομένων είναι:

- τα duplications των records της βάσης δεδομένων, που συνήθως οφείλονται σε λόγους όπως τα τυπογραφικά λάθη των χρηστών, η αλλαγή της τιμής μιας μεταβλητής σε μία μόνο βάση δεδομένων και όχι παντού ενώ χρειάζεται (π.χ. η διεύθυνση ενός χρήστη δεν άλλαξε σε όλες τις βάσεις δεδομένων που τον αφορούν).
- η εσκεμμένη άρνηση από την πλευρά του χρήστη να εισάγει όλα τα στοιχεία του για λόγους ασφαλείας, ή ακόμα να εισάγει και λάθος στοιχεία δίνοντας για παράδειγμα λάθος διεύθυνση.
- η εισαγωγή μολυσμένων στοιχείων στα πεδία της

βάσης (για π.χ. τιμές της μορφής 111111)

Γίνεται λοιπόν διόρθωση ή αφαίρεση των δεδομένων αυτών και έτσι στο τέλος τα δεδομένα που προκύπτουν δεν περιέχουν λανθασμένη πληροφορία που δεν έχει ιδιαίτερο νόημα και μπορεί να επηρεάσει τα αποτελέσματα που θα προκύψουν.

3.3. Εμπλουτισμός δεδομένων

Στο στάδιο εμπλουτισμού των δεδομένων μπορούν να μπουν περαιτέρω στοιχεία για τα δεδομένα εξερεύνησης (π.χ. από ερωματολογία χρηστών σχετικά με βαθμό χρησιμοποίησης μιας υπηρεσίας ή ενός τύπου υλικού, δείκτες ΜΟΠΑΒ).

3.4. Κωδικοποίηση δεδομένων

Στο στάδιο κωδικοποίησης μπορούμε να μετασχηματίσουμε τα δεδομένα σε μία μορφή (π.χ. ημερομηνίες σε μία μορφή, άνδρας/γυναίκα σε μία μορφή, τμήματα σε μία μορφή (ΜΗΧΟΠ αντί Μηχανικών Ορυκτών Πόρων)). Δηλαδή αναπαριστούμε τις τιμές των δεδομένων με μια μορφή εκεί που υπάρχουν δύο ή περισσότερες. Έτσι στο τέλος η αναπαράσταση ίδιων τύπων δεδομένων (ίδιων χαρακτηριστικών) γίνεται με μία και μόνο μορφή, γεγονός που διευκολύνει κατά πολύ την επεξεργασία.

3.5. Εξόρυξη γνώσης – Εφαρμογή μοντέλων / τεχνικών

Στο στάδιο αυτό εφαρμόζουμε μεθόδους, μοντέλα και τεχνικές εξόρυξης γνώσης για να πάρουμε την πληροφορία που θέλουμε. Οι τεχνικές αυτές όχι μόνο απαιτούν μόνο ειδικούς τύπους δομής των δεδομένων, αλλά συνεπάγονται επίσης και ειδικούς τύπους αλγοριθμικών προσεγγίσεων (Dunham 2004). Η λεπτομερή τους ανάλυση δεν ανήκει στα πλαίσια της παρούσας εργασίας απλά αναφέρονται εν συντομία μερικές από τις συνηθισμένες τεχνικές εξόρυξης γνώσης. Εφαρμόζονται λοιπόν τεχνικές που έχουν ως στόχο να δημιουργήσουν ένα μοντέλο που βασίζεται στην είσοδο δεδομένων (που ουσιαστικά είναι τα δεδομένα προς επεξεργασία). Η διαδικασία της μοντελοποίησης δημιουργεί το μοντέλο ψάχνοντας μέσα στα δεδομένα. Οι τεχνικές που χρησιμοποιούνται ευρέως είναι:

- Η κατηγοριοποίηση των δεδομένων σε ομάδες ανάλογα με κάποιο χαρακτηριστικό τους. Εδώ εφαρμόζονται αλγόριθμοι βασισμένοι σε στατιστική (παλινδρόμηση, Bayesian κατηγοριοποίηση), βασισμένοι στην απόσταση (κάθε στοιχείο που απεικονίζεται στην ίδια κατηγορία μπορεί να θεωρηθεί ότι είναι πιο κοντά σε στοιχεία της ίδιας κατηγορίας από όσο σε στοιχεία τα οποία ανήκουν σε άλλες κατηγορίες), βασισμένοι σε δένδρα αποφάσεων (κατασκευή ενός δένδρου για να μοντελοποιηθεί η διαδικασία της

κατηγοριοποίησης), βασισμένοι σε νευρωνικά δίκτυα, βασισμένοι σε κανόνες (π.χ. if-then κανόνες), βασισμένοι σε συνδυαστικές τεχνικές.

- Η συσταδοποίηση (clustering), η οποία είναι παρόμοια με την κατηγοριοποίηση (καθώς και στις δύο περιπτώσεις τα δεδομένα οργανώνονται σε ομάδες) με βασική διαφορά ότι οι ομάδες στις οποίες θα χωριστούν / κατηγοριοποιηθούν τα δεδομένα δεν είναι προκαθορισμένες αλλά προκύπτουν στην πορεία της επεξεργασίας. Ουσιαστικά η συσταδοποίηση επιτυγχάνεται βρίσκοντας ομοιότητες μεταξύ των δεδομένων βάσει των χαρακτηριστικών που υπάρχουν σε αυτά. Εδώ εφαρμόζονται ιεραρχικοί αλγόριθμοι (συσσωρευτικοί, διαιρετικοί), διαμεριστικοί αλγόριθμοι, ειδικοί αλγόριθμοι που εφαρμόζονται σε δυναμικές (συνεχώς μεταβαλλόμενες) βάσεις δεδομένων, αλγόριθμοι με κατηγορικά (μη αριθμητικά) γνωρίσματα.

- Η δημιουργία κανόνων συσχέτισης, όπου τα δεδομένα ανιχνεύονται με τέτοιο τρόπο ώστε να προκύψουν συσχετίσεις ανάμεσα σε αυτά εξετάζοντας συγκεκριμένα χαρακτηριστικά τους (π.χ. συσχετισμός που μπορεί να υπάρξει ανάμεσα στο χρόνο δανεισμού ενός βιβλίου, στην κατηγορία που ανήκει και στο είδος του αναγνώστη του). Εδώ εφαρμόζονται αλγόριθμοι δειγματοληψίας, διαμέρισης, παραλληλισμού δεδομένων.

Οι παραπάνω τεχνικές εφαρμόζονται από το λογισμικό που έχουμε προμηθευτεί για να διεξάγουμε την όλη διαδικασία της επεξεργασίας και εξόρυξης γνώσης των δεδομένων που μας ενδιαφέρουν και δεν είναι στα πλαίσια της εν λόγω εργασίας να επεκταθούμε περισσότερο στον τρόπο λειτουργίας τους. Παραδείγματα τέτοιων λογισμικών είναι το Clementine (εταιρεία SPSS) και το Darwin (εταιρεία Oracle).

Από τις αναλύσεις λοιπόν αυτές μπορούν να προκύψουν πληροφορίες που δεν είναι εμφανείς αρχικά αλλά μόνο μετά το συσχετισμό των δεδομένων προς επεξεργασία. Δηλαδή προκύπτει ουσιαστικά πληροφορία που ήταν «κρυμμένη» και χρειάζοταν εφαρμογή τεχνικών και μεθόδων εξόρυξης γνώσης για να αποκαλυφθεί.

Ενδεικτικά λοιπόν αναφέρουμε ένα πλήθος πληροφοριών που μπορούν να προκύψουν από τα δεδομένα που περιγράψαμε στο 3.1 και εμφανίζουν πολύ χρήσιμες πληροφορίες στον παρατηρητή :

- Από την ανάλυση των «δεδομένων δανειστικών καταστάσεων χρηστών και δεδομένων χρήσης υλικού (έντυπου ή ηλεκτρονικού)» μπορούμε να βρούμε συσχετισμούς μεταξύ χαρακτηριστικών που έχουν σχέση με τον αριθμό κρατήσεων και δανεισμού ενός τίτλου. Μπορούμε έτσι να αναφερθούμε στον «συντελεστή

κρατήσεων σε σχέση με το πλήθος των δανεισμών ενός τίτλου για μια περίοδο», ο οποίος δείχνει το ποσοστό άμεσης διαθεσιμότητας του τίτλου αυτού.

Ορίζουμε λοιπόν το δείκτη «άμεσης διαθεσιμότητας» ενός βιβλίου ως εξής:

Δείκτης Άμεσης Διαθεσιμότητας (ΔΑΔ) = (αριθμός αιτήσεων για κράτηση) / (αριθμός δανεισμών)

Όσο μεγαλύτερος είναι ο δείκτης αυτός τόσο η ζήτηση του τίτλου αυξάνει. Π.χ. αν σε περίοδο ενός χρόνου το βιβλίο Α δανείστηκε 10 φορές και στον ίδιο χρόνο του έγιναν 5 κρατήσεις, σημαίνει ότι το εν λόγω βιβλίο δεν ήταν διαθέσιμο, όταν ζητήθηκε για δανεισμό, 5 φορές και γι' αυτό του έγινε κράτηση. Ο δείκτης ΔΑΔ λοιπόν είναι $5/10 = 0.5$. Οπότε ίσως θα ήταν καλύτερο να αυξηθούν τα αντίτυπα σε αυτό το βιβλίο. Με το δείκτη «άμεσης διαθεσιμότητας» που ορίσαμε, βάζοντας ένα κατώφλι, μπορούμε να πούμε ότι αν ο δείκτης υπερβαίνει το κατώφλι αυτό τότε είναι καλή πρακτική να αυξήσουμε τα αντίτυπα του τίτλου στον οποίο αναφέρεται ο δείκτης αυτός για να αυξήσουμε τη διαθεσιμότητά του προς τους χρήστες.

Παράλληλα μπορούμε να δούμε στοιχεία σχετικά με το ποιες ομάδες χρηστών δανειζονται συγκεκριμένες κατηγορίες υλικού συχνότερα (π.χ. οι προπτυχιακοί φοιτητές ενός τμήματος δανειζονται συχνότερα βιβλία σχετικά με ένα συγκεκριμένο αντικείμενο σε σύγκριση με φοιτητές άλλων τμημάτων).

- Από την ανάλυση των «δεδομένων που προέκυψαν από τα αρχεία ημερολογίου (Web Logs) του Web Server» μπορούμε να εξάγουμε άμεσα συμπεράσματα τόσο για τις πιο δημοφιλείς σελίδες του ιστοχώρου όσο και για τις πιο δημοφιλείς διαδρομές που ακολουθούνται σε αυτόν με σκοπό τον καλύτερο ανασχεδιασμό του. Π.χ. αν διαπιστώσουμε ότι πολύ συχνά ακολουθούνται οι διαδρομές A,B,C,E και A,D,C και A,E,C, τότε κάλλιστα μπορεί να φτιαχτεί ένας σύνδεσμος (link) από τη σελίδα A στη C για να κάνει ευκολότερη και γρηγορότερη την εν λόγω μετάβαση μιας και διαπιστώνουμε ότι οι δύο σελίδες A και C συναντώνται πολύ συχνά στις διαδρομές που ακολουθούνται στον ιστοχώρο της Βιβλιοθήκης.

- Τέλος, ένας ακόμη τομέας που μπορεί να εφαρμοστεί η τεχνολογία εξόρυξης γνώσης, μετά την επεξεργασία των δεδομένων που αναφέρθηκαν, αφορά στην απόφαση για την κατανομή των κονδυλίων κάθε βιβλιοθήκης στα επιμέρους πανεπιστημιακά τμήματα το οποίο είναι ένα θέμα πολύ σημαντικό και ταυτόχρονα αρκετά πολύπλοκο. Οι παράμετροι που παίζουν το σημαντικότερο ρόλο στο συγκεκριμένο θέμα είναι (Graves 1974, Kao-Chang-Lin 2003) (1) το μέγεθος του τμήματος, (2) ο αριθμός των φοιτητών, (3)

το κόστος του βιβλιοθηκονομικού υλικού, (4) η καταλληλότητα της συλλογής σε σχέση με τα αντικείμενα που διδάσκει το τμήμα, (5) ο αριθμός και το είδος των μαθημάτων του τμήματος, (6) το σύνολο των ερευνητικών δραστηριοτήτων του τμήματος, (7) το ποσοστό απορρόφησης των κονδυλίων που δόθηκαν σε παρελθόντα έτη στο τμήμα και (8) οι δανειστικές στατιστικές (circulation statistics). Εφαρμόζοντας λοιπόν τεχνικές εξόρυξης γνώσης, στα δεδομένα αυτά, μπορούμε να εξάγουμε αποτελέσματα που έχουν σχέση με τον τρόπο κατανομής των κονδυλίων στα διάφορα τμήματα (αλγόριθμος ABAMDM, acquisition budget allocation modem via data mining) (Graves 1974, Kao-Chang-Lin 2003).

3.6. Παρουσίαση αποτελεσμάτων – λήψη αποφάσεων

Τα εργαλεία (λογισμικό) εξόρυξης γνώσης διαθέτουν δυνατότητες που απεικονίζουν τα αποτελέσματα που προκύπτουν από τη διαδικασία εξόρυξης γνώσης με τέτοιο τρόπο που δίνει την παρουσίαση μιας όσο το δυνατόν πιο εποπτικής εικόνας στον παρατηρητή για να βρεθεί σε θέση να εκτιμήσει αποδοτικά τα αποτελέσματα και να πάρει αποφάσεις με βάση αυτά. Έτσι, στον τελικό χρήστη αυτών των συστημάτων παρουσιάζονται οθόνες που έχουν γραφικές παραστάσεις, πίνακες, γραφήματα, ιστογράμματα και πολλού είδους οπτικοποιημένες πληροφορίες..

4. Συμπεράσματα

Η συγκεκριμένη εργασία αναφέρθηκε στην επιτακτική ανάγκη που συναντάται πλέον στην επεξεργασία δεδομένων με σκοπό την λήψη χρήσιμων πληροφοριών στον τομέα των Βιβλιοθηκών. Περιγράφηκε εποπτικά η διαδικασία της εξερεύνησης γνώσης και τα βήματα που ακολουθούνται σε αυτή με σκοπό τη λήψη χρήσιμης πληροφορίας (που αρχικά δεν είναι εμφανή) από αυτά. Παρουσιάστηκαν οι κατηγορίες των δεδομένων στον τομέα των Βιβλιοθηκών, οι οποίες μπορούν να υποστούν επεξεργασία με βάση μοντέλα εξόρυξης γνώσης.

Στη συνέχεια περιγράφηκε πώς μπορεί η γενική διαδικασία και τα βήματα που ακολουθούνται στην εξερεύνηση και εξόρυξη γνώσης να εφαρμοστεί στις περιπτώσεις δεδομένων από Βιβλιοθήκες.

Μέσα από τη διαδικασία τονίστηκε με παραδείγματα η χρησιμότητα των πληροφοριών που ανέκλυψαν, κάτι που αυτόματα σημαίνει ότι ο τομέας της εξερεύνησης γνώσης σε δεδομένα Βιβλιοθηκών είναι πολύ σημαντικός τόσο για την λήψη χρήσιμης και πολλές φορές μη εμφανούς πληροφορίας, όσο και για τη χάραξη στρατηγικών και λήψη αποφάσεων στα

πλαίσια σωστής λειτουργίας, οργάνωσης και βελτίωσης των προσφερόμενων υπηρεσιών τους.

Βιβλιογραφία

Λίστα αναφορών. Αλφαβητική σειρά των αναφορών.

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

Greaves, F. L., Jr. 1974. The allocation formula as a form of book fund management in selected state-supported academic libraries, Florida State University, unpublished doctoral dissertation.

Han, J. και Fu, Y. Mining multiple-level association rules in large databases. IEEE Transactions on Knowledge and Data Engineering 11(5) (1999): 798-805

Hirota, K. και Pedrycz, W. Fuzzy computing for data mining. Proceedings of the IEEE 87(9) (1999): 1575-1600

Kao, S.-C. και Chang, H.-C. και Lin, C.-H. Decision support for the academic library acquisition budget allocation via circulation database mining. Information Processing and Management 39 (2003): 133-147.

Kim, Hyunki. 2005. Developing Semantic Digital Libraries Using Data Mining Techniques, UMI Dissertation Services.

Αριστείδης Μελετίου

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

Ανθή Κατσιρίκου

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

