

**16° Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών
(Αθήνα-Πανεπιστήμιο Πειραιώς: 1-3 Οκτωβρίου 2006)**

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

Αριστείδης Μελετίου

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

Ανθή Κατσιρικού

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

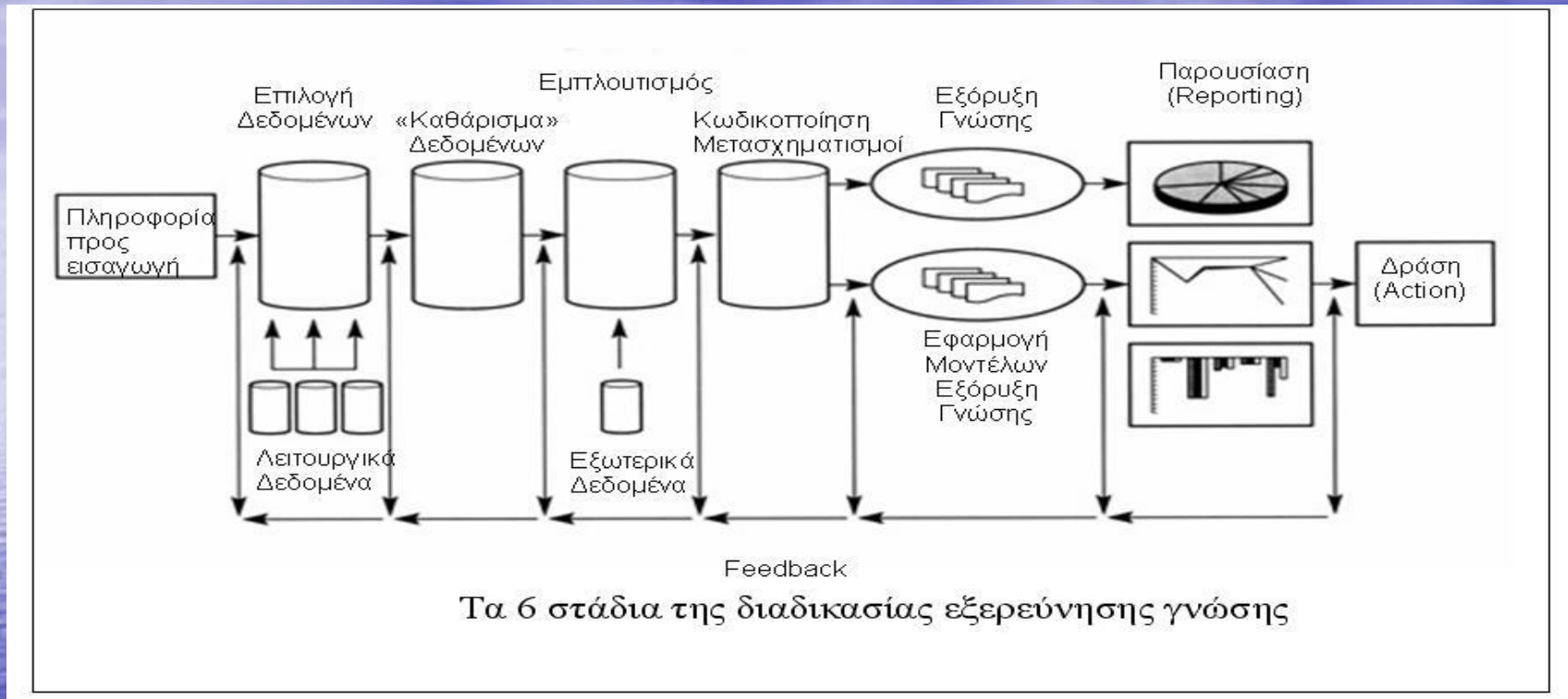
Εισαγωγή

- 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
- 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
- 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
- Η «ανακάλυψη γνώσης» (knowledge discovery) σε βάσεις δεδομένων δηλώνει την ολοκληρωμένη διαδικασία μετατροπής «μικρής σημασίας» αρχικής πληροφορίας σε χρήσιμη πληροφορία και γνώση με σκοπό τη βελτιστοποίηση της διαδικασίας λήψης αποφάσεων.

Δομή – Στόχοι

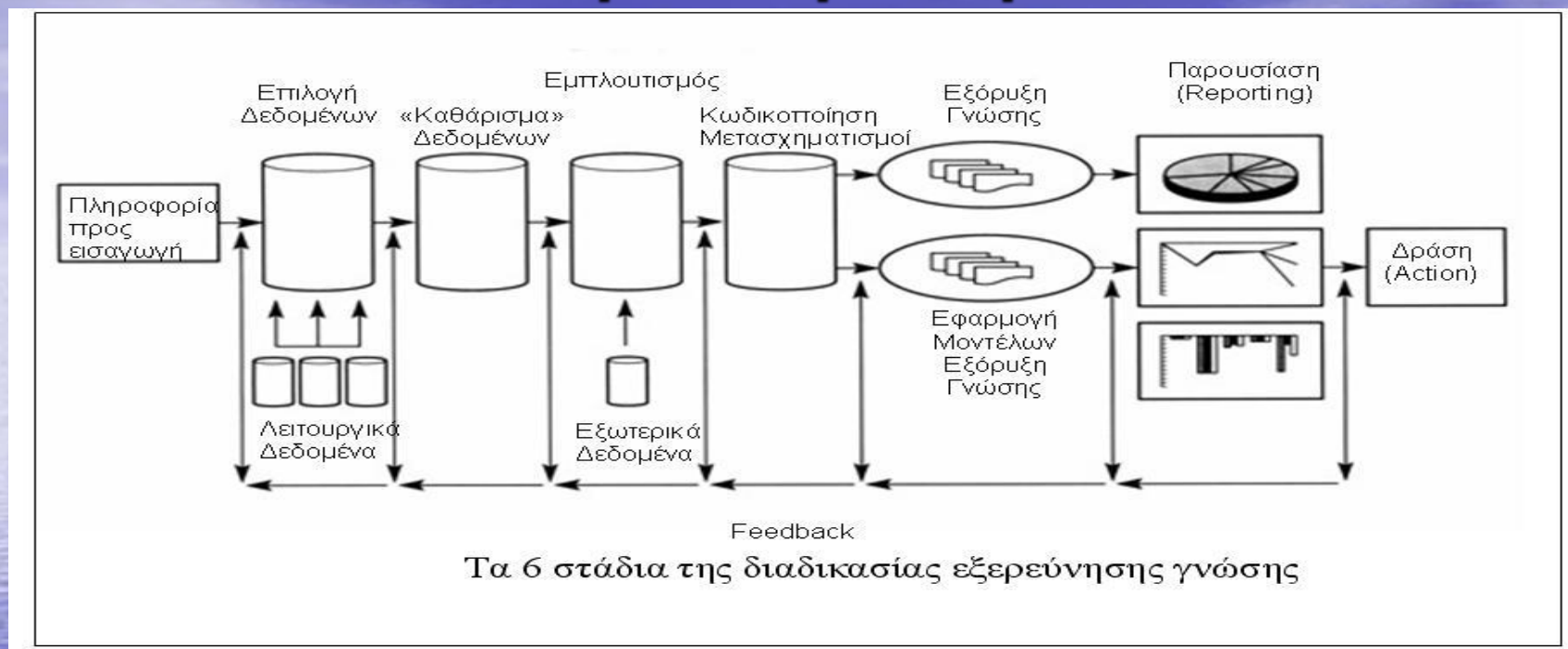
- Περιγράφεται η διαδικασία που θα ακολουθηθεί (data mining process)
 - Περιγράφεται ο τρόπος επιλογής των δεδομένων που θα συμμετάσχουν σε αυτή σε συνδυασμό με τους επιθυμητούς στόχους
 - Περιγράφεται ο τρόπος ανάλυσης των δεδομένων
 - Παρουσιάζονται τα συμπεράσματα που προκύπτουν από την ανάλυση (μελέτη περίπτωσης)
-
- ❖ Να δώσει τις γενικές κατευθύνσεις στην ανάλυση δεδομένων που προέρχονται από Βιβλιοθήκες
 - ❖ Να καταδείξει πώς μπορεί η επιστήμη της «εξόρυξης πληροφορίας και γνώσης» να εφαρμοστεί σε πραγματικά δεδομένα Βιβλιοθηκών
 - ❖ Να δείξει πώς μπορεί να κατανοηθεί και να αξιολογηθεί η παραγόμενη πληροφορία και γνώση για τη λήψη αποφάσεων στρατηγικού σχεδιασμού

Η διαδικασία εξερεύνησης γνώσης και η εφαρμογή της σε δεδομένα Βιβλιοθηκών



- **1ο στάδιο: Επιλογή δεδομένων** (ποια δεδομένα θα συμμετάσχουν στη διαδικασία)
- **2ο στάδιο: Φιλτράρισμα δεδομένων** (π.χ. συμπλήρωση ή διαγραφή ελλιπών στοιχείων)
- **3ο στάδιο. Εμπλουτισμός δεδομένων**
- **4ο στάδιο. Κωδικοποίηση και μετασχηματισμός δεδομένων** (π.χ. μετατροπή των τιμών yes/no σε 1/0)

Η διαδικασία εξερεύνησης γνώσης και η εφαρμογή της σε δεδομένα Βιβλιοθηκών



- **5ο στάδιο. Εξερεύνηση των δεδομένων :** Εφαρμογή μοντέλων και τεχνικών εξόρυξης γνώσης για τη λήψη **πληροφοριών** σχετικών με το συσχετισμό των δεδομένων, όπως:
 - ✓ **Κανόνες συσχέτισης (association rules):** Π.χ. «το 30% των προπτυχιακών φοιτητών δεν επιστρέφει το δανεισμένο υλικό εμπρόθεσμα».
 - ✓ **Κατηγοριοποίηση (classification):** Π.χ. «ταξινόμηση των αναγνωστών βάσει του αριθμού των βιβλίων που έχουν δανειστεί»
 - ✓ **Ανάλυση τάσεων και αποκλίσεων (trend and deviation analysis):** Π.χ. «Ανακάλυψη του υλικού που θεωρείται περισσότερο ή λιγότερο ενδιαφέρον από τους χρήστες της Βιβλιοθήκης»
 - ✓ **Ανάλυση προτύπων (pattern analysis):** Π.χ. ανακάλυψη των πιο δημοφιλών μονοπατιών που ακολουθούν οι επισκέπτες του δικτυακού τόπου (web site) μιας Βιβλιοθήκης
- **6ο στάδιο: Παρουσίαση των αποτελεσμάτων της εξερεύνησης**
(π.χ. πίνακας συσχετίσεων, διαγράμματα)

Επιλογή δεδομένων, επεξεργασία 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
- 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)*
- 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 το τι θέλουμε ουσιαστικά να παρατηρήσουμε, επιλέγουμε τα δεδομένα εκείνα που σχετίζονται με αυτό

Επιλογή δεδομένων, επεξεργασία και λήψη χρήσιμης πληροφορίας

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)

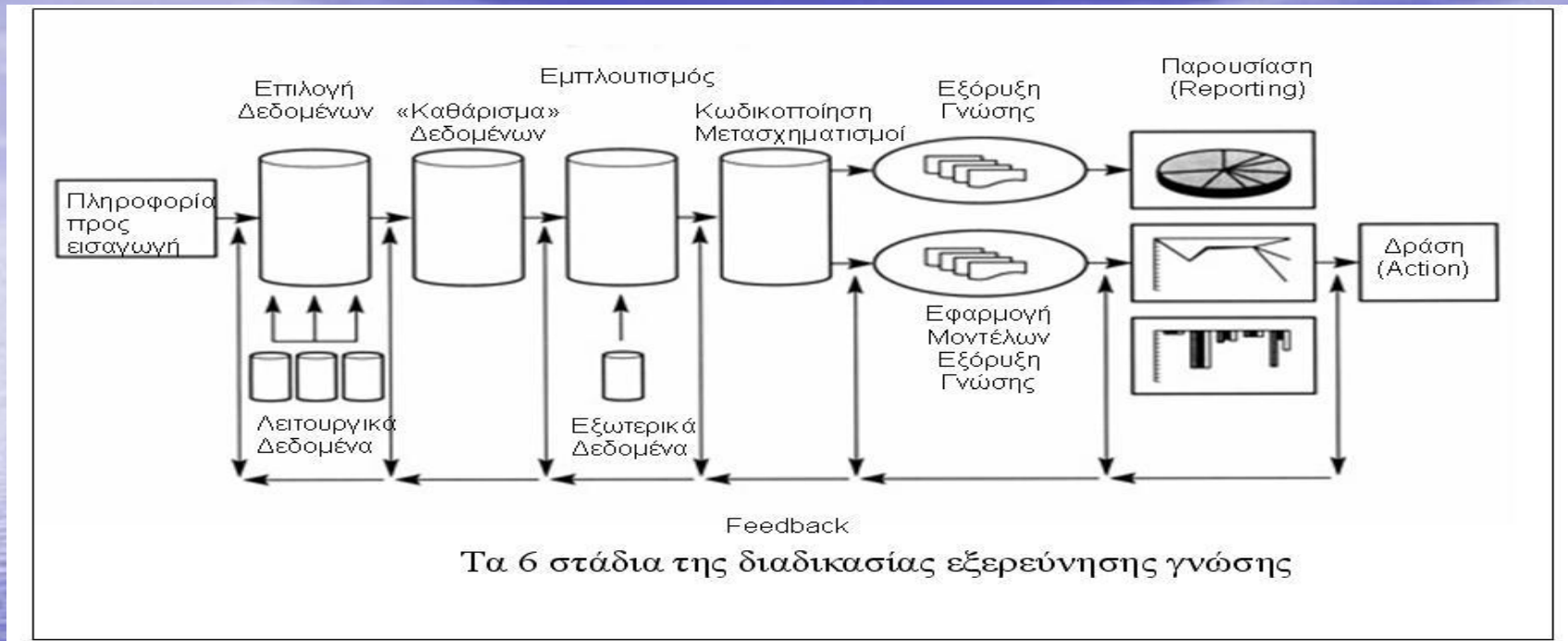
| Πεδίο | Περιγραφή πεδίου |
|--------------|---|
| Material_ID | Η μοναδική ταυτότητα του υλικού (προέρχεται από τον πίνακα II) |
| Person_ID | Η μοναδική ταυτότητα του χρήστη που δανείστηκε το υλικό (προέρχεται από τον πίνακα I) |
| Publish_Year | Η ημερομηνία που εισήχθη στη συλλογή υλικού της Βιβλιοθήκης |
| 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 |

Επιλογή δεδομένων, επεξεργασία και λήψη χρήσιμης πληροφορίας

Συσχετίσεις δεδομένων:

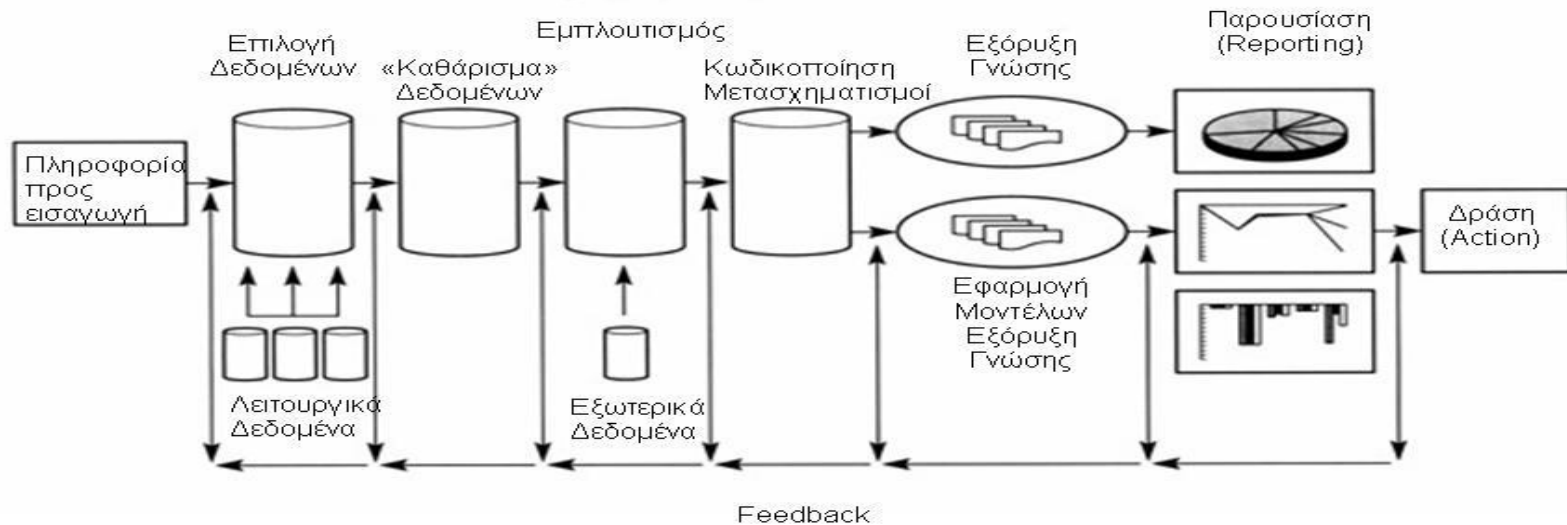


Περιγραφή των σταδίων επεξεργασίας



- **1ο στάδιο:** *Επιλογή των κατάλληλων δεδομένων ανάλογα με τους συσχετισμούς που θέλουμε να παρατηρήσουμε (κατηγορία χρηστών, είδος υλικού, κατηγορία αναγνωστών, χρόνος επιστροφής δανεισμένου υλικού)*
- **2ο στάδιο:** *Έλεγχος, διόρθωση ή συμπλήρωση τιμών (π.χ. συμπλήρωση τμήματος, ιδιότητας ή φύλου χρήστη)*
- **3ο στάδιο:** *Δεν απαιτήθηκε εμπλουτισμός από εξωτερικά δεδομένα (σε άλλες περιπτώσεις ίσως είναι αναγκαίο (π.χ. ερωτηματολόγια χρηστών, στοιχεία για το κόστος απόκτησης του υλικού))*

Περιγραφή των σταδίων επεξεργασίας



Τα 6 στάδια της διαδικασίας εξερεύνησης γνώσης

- **4ο στάδιο:** μετασχηματισμοί σε ορισμένες τιμές δεδομένων με σκοπό τη μετατροπή τους σε ένα κοινό σχήμα για καλύτερη επεξεργασία (π.χ. «Άνδρας / Γυναίκα» -> «0/1», ΗΜΜΥ: 10, ΜΠΔ: 11, ΜΗΧΟΠ: 12, ΜΗΠΕΡ: 13, ΑΡΧ: 14, ΥΠΟΛΟΙΠΑ: 16, Προπτυχιακός: Π, Μεταπτυχιακός: Μ, Διδακτορικός Φοιτητής: Δ, Καθηγητής: Κ, Υπάλληλος: Υ, Εξωτερικός χρήστης: ΕΞ). Ταχύτερη επεξεργασία
- **5ο στάδιο:** πρόγραμμα (Clementine (SPSS)), ορίστηκαν οι επιθυμητές συσχετίσεις εφαρμόστηκαν μοντέλα εξόρυξης γνώσης στα δεδομένα των πινάκων (μέθοδοι κατηγοριοποίησης και ομαδοποίησης δεδομένων)
- **6ο στάδιο:** πίνακες συσχετίσεων δεδομένων

Πίνακες συσχετίσεων δεδομένων

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

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

- Παρόμοιοι πίνακες συσχετίσεων είναι πολύ απλό να παραχθούν από το πρόγραμμα επεξεργασίας δεδομένων που χρησιμοποιήθηκε αρκεί να δηλωθούν οι ανάλογες παράμετροι σε αυτό (π.χ. έτος φοιτητή ή φύλλο χρήστη σε σχέση με το είδος του υλικού που δανείζονται)
- 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.

Δημιουργία και χρησιμοποίηση δεικτών

Συνίσταται η χρησιμοποίηση δεδομένων με σκοπό τον ορισμό δεικτών ικανών να δώσουν χρήσιμες πληροφορίες για τις διαδικασίες που εφαρμόζονται σε μια Βιβλιοθήκη.

Π.χ. από την ανάλυση «δεδομένων δανειστικών καταστάσεων χρηστών και δεδομένων χρήσης υλικού (έντυπου ή ηλεκτρονικού)» -> **Συσχετισμοί μεταξύ χαρακτηριστικών που έχουν σχέση με τον αριθμό κρατήσεων και δανεισμού ενός τίτλου :**

«Συντελεστής κρατήσεων σε σχέση με το πλήθος των δανεισμών ενός τίτλου για μια περίοδο» : δείχνει το ποσοστό άμεσης διαθεσιμότητας του τίτλου αυτού.

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

Όσο μεγαλύτερος είναι ο δείκτης αυτός σημαίνει ότι η ζήτηση του τίτλου είναι αυξημένη

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.

Συμπεράσματα

- 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.
- Έδωσε **τις γενικές αρχές** με βάση τις οποίες θα πρέπει να γίνεται η επιλογή των δεδομένων προς επεξεργασία
- 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.