

# Πολυδιάστατη στατιστική ανάλυση δεδομένων προερχόμενων από ακαδημαϊκές βιβλιοθήκες

*Αφροδίτη Μάλλιαρη*

## Εισαγωγή

Αντικειμενικός σκοπός των βιβλιοθηκών ανεξαρτήτως είδους ήταν ανέκαθεν η εξυπηρέτηση των χρηστών, η κάλυψη των πληροφοριακών τους αναγκών. Προκειμένου να επιτευχθεί ο στόχος αυτός, η σύγχρονη βιβλιοθηκονομία χρησιμοποιεί μια σειρά ερευνητικών μεθόδων που αποσκοπούν στον εντοπισμό των αναγκών των χρηστών, οι λεγόμενες έρευνες χρηστών, ώστε να βελτιώνονται οι παρεχόμενες υπηρεσίες.

Στο πλαίσιο των ερευνών χρηστών, και ακολουθώντας την προαναφερθείσα αντικειμενική ανάγκη των βιβλιοθηκών, η παρούσα έρευνα επιχειρεί τη συλλογή και επεξεργασία δεδομένων που αφορούν στη συμπεριφορά χρηστών. Συγκεκριμένα, συγκεντρώνει δεδομένα που προέρχονται από τις συνδέσεις των χρηστών με τον αυτοματοποιημένο κατάλογο μιας ακαδημαϊκής βιβλιοθήκης. Οι συνδέσεις αυτές είναι σε πραγματικό χρόνο και αναδεικνύουν στοιχεία συμπεριφοράς καθώς μελετάται ο τρόπος αναζήτησης και η αποδοτικότητα ανάκτησης πληροφοριών.

Τα δεδομένα της συγκεκριμένης έρευνας χρηστών μέσω μεταβλητών ποσοτικών αλλά κυρίως ποιοτικών, αποτέλεσαν αντικείμενα στατιστικής επεξεργασίας και ανάλυσης με μεθόδους πολυδιάστατης στατιστικής. Το αποδιδόμενο μοντέλο είναι ένα επιστημονικό αλλά απλό εργαλείο με αποτελέσματα εύκολα ερμηνεύσιμα. Προτείνεται, λοιπόν, ένας νέος τρόπος στατιστικής επεξεργασίας και ανάλυσης βιβλιοθηκονομικών δεδομένων, ο οποίος λόγω της μορφής του είναι άμεσα αξιοποιήσιμος ακόμη και από επιστήμονες που δεν είναι μαθηματικοί ή στατιστικοί, δηλαδή δεν απαιτεί την εξειδικευμένη γνώση, αλλά ακολουθώντας κάποιες βασικές και απλές οδηγίες δίνει τη δυνατότητα επεξεργασίας των δεδομένων.

## Στόχος

Η παρούσα έρευνα στοχεύει: α) στην εξέταση του τρόπου ανάκτησης δεδομένων στις βιβλιοθήκες και κυρίως β) στη δημιουργία ενός νέου μοντέλου στατιστικής ανάλυσης ποιοτικών δεδομένων προερχόμενων από βιβλιοθήκες με τη χρήση μεθόδων της πολυδιάστατης στατιστικής, οι οποίες επιτρέπουν την ανάλυση ενός πολύπλευρου φαινομένου με την ταυτόχρονη ανάλυση πολυάριθμων μεταβλητών. Η διαδικασία γίνεται με σκοπό την πλήρη αξιοποίηση των στοιχείων αυτών για τη βελτίωση των παρεχόμενων από τις βιβλιοθήκες υπηρεσιών.

Η χρηστικότητα του συγκεκριμένου μοντέλου θα αναδειχθεί μέσα από μια εφαρμογή σε ελληνική ακαδημαϊκή βιβλιοθήκη και συγκεκριμένα μέσα από τη μελέτη του τρόπου που αναζητούν και ανακτούν πληροφορίες οι χρήστες από έναν κατάλογο απευθείας σύνδεσης (online) προκειμένου να διαπιστώσει:

1. αν ο ίδιος ο κατάλογος μπορεί να παρέχει αξιοποιήσιμες πληροφορίες που μπορούν να επηρεάσουν την πολιτική της βιβλιοθήκης,
2. το βαθμό γνώσης αναζήτησης και ανάκτησης πληροφοριών από τον κατάλογο,
3. τον τρόπο με τον οποίο αναζητούν κατά κύριο λόγο και αν οι προτιμήσεις τους φανερώνουν κάποια στοιχεία συμπεριφοράς,
4. αν αξιοποιούν την απομακρυσμένη σύνδεση, κ.ο.κ.

### **Στατιστική και βιβλιοθηκονομία**

Σε αυτό το σημείο κρίθηκε απαραίτητο να γίνει μια σύντομη περιγραφή της σχέσης που διέπει τους δυο επιστημονικούς χώρους της στατιστικής και της βιβλιοθηκονομίας. Η στατιστική ως αντικείμενο διδασκαλίας στις σχολές βιβλιοθηκονομίας αλλά και ως αντικείμενο ενασχόλησης στο περιβάλλον εργασίας των βιβλιοθηκών, έχει παρατηρηθεί ότι αντιμετωπίζεται με αρνητική προδιάθεση. Οι λόγοι είναι ότι πολλές φορές οι σπουδαστές δυσκολεύονται αισθητά να κατανοήσουν το αντικείμενό της και το πεδίο εφαρμογής της. Οι δυσκολίες που αντιμετωπίζουν τόσο οι σπουδαστές όσο και οι εργαζόμενοι υπόκεινται στη μαθηματική υπόσταση της στατιστικής, περιγραφικής και επαγωγικής, γεγονός που αποτελεί ανασταλτικό παράγοντα για τους μη κατέχοντες το θεωρητικό υπόβαθρο. Επειδή όμως οι βιβλιοθήκες είναι ζωντανό οργανισμοί με στόχους, αποτελέσματα αλλά και αξιολογήσεις, η στατιστική είναι ένα απαραίτητο εργαλείο μέτρησης της αποδοτικότητας και αποτελεσματικότητας τους. Έχοντας, λοιπόν, ως δεδομένα τη δυσκολία της στατιστικής για τους βιβλιοθηκονόμους και την ανάγκη εφαρμογής της στις βιβλιοθήκες, προτείνονται με την παρούσα ανακοίνωση νέες για τα βιβλιοθηκονομικά δεδομένα μεθόδους πολυδιάστατης στατιστικής ανάλυσης άμεσα εφαρμόσιμες από ανθρώπους που δεν απαιτείται να έχουν υψηλό στατιστικό θεωρητικό υπόβαθρο.

Προβαίνοντας σε μια ανασκόπηση τόσο στη διεθνή όσο και στην ελληνική αρθρογραφία και βιβλιογραφία όσον αφορά στη χρήση της στατιστικής σε βιβλιοθηκονομικές εφαρμογές, καταλήγουμε στο συμπέρασμα ότι τα δημοσιεύματα που ασχολούνται με τη χρήση στατιστικών μεθόδων στους χώρους των βιβλιοθηκών κινούνται κατά κύριο λόγο γύρω από τις βασικές έννοιες της περιγραφικής στατιστικής δίνοντας ορισμούς αυτών και παραδείγματα. Γενικότερα συμπεραίνεται ότι η χρήση της στατιστικής στην επιστήμη της βιβλιοθηκονομίας εστιάζεται κατά κύριο λόγο στη συλλογή και επεξεργασία ποσοτικών δεδομένων με μεθόδους που ανήκουν στη μονοδιάστατη ανάλυση ή και σε λίγες περιπτώσεις στη δισδιάστατη ανάλυση εξαρτημένων κατά κύριο λόγο όμως μεταβλητών. Αυτή η συνθήκη παρατηρείται ακόμη και στο σημείο των εθνικών στατιστικών, καθώς οι μέθοδοι που ακολουθούνται κατά την επεξεργασία και ανάλυση των δεδομένων ανήκουν στη μονοδιάστατη ανάλυση ποσοτικών μεγεθών.

### **Μέθοδοι ανάλυσης δεδομένων**

Η επεξεργασία και ανάλυση των δεδομένων, τα οποία όπως έχει ήδη αναφερθεί προέρχονται από τις αναζητήσεις των χρηστών στον ορακό ακαδημαϊκής βιβλιοθήκης, πραγματοποιήθηκε με Μεθόδους Ανάλυσης Δεδομένων και συγκεκριμένα με την Παραγοντική Ανάλυση Αντιστοιχιών και την Ταξινόμηση Κατ' Αύξουσα Ιεραρχία. Οι Μέθοδοι Ανάλυσης Δεδομένων έχουν ευρεία εφαρμογή στο χώρο των κοινωνικών επιστημών καθώς αφενός μεν οι κοινωνικές επιστήμες κατακλύζονται από ποιοτικά δεδομένα και αφετέρου η χρήση και εφαρμογή τους από μη στατιστικούς δεν είναι απαγορευτική. Η βιβλιοθηκονομία ανήκοντας στον τομέα των κοινωνικών επιστημών αποτελείται πέρα από ποσοτικά δεδομένα και από πληθώρα ποιοτικών δεδομένων, τα οποία θα μπορούσαν εν δυνάμει να αποτελέσουν αντικείμενο της εδώ προτεινόμενης πολυδιάστατης στατιστικής ανάλυσης, όπως αποτελεί στη συγκεκριμένη έρευνα η συμπεριφορά των χρηστών απέναντι στον αυτοματοποιημένο κατάλογο μιας ακαδημαϊκής βιβλιοθήκης.

Εξάλλου ένα μεγάλο πλεονέκτημα τόσο για τον ερευνητή όσο και για τον αναγνώστη και αποδέκτη του τελικού προϊόντος της έρευνας είναι η απεικόνιση του υπό μελέτη φαινομένου με το σύνολο των μεταβλητών στο επίπεδο είτε με τη μορφή παραγοντικού επιπέδου στην περίπτωση της Παραγοντικής Ανάλυσης Αντιστοιχιών είτε με τη μορφή δενδρογράμματος στην περίπτωση της Ταξινόμησης Κατ' Αύξουσα Ιεραρχία και ως γνωστόν η εικόνα και η σχηματοποίηση ενός προβλήματος ή ενός θέματος συμβάλλει θετικά στην κατανόηση και ερμηνεία του.

Η επεξεργασία των δεδομένων με τις μεθόδους της Ανάλυσης Δεδομένων, Παραγοντική Ανάλυση Αντιστοιχιών (A.F.C.) και Ταξινόμηση Κατ' Αύξουσα Ιεραρχία (C.A.H.) επιτρέπει να

αντιμετωπισθεί το φαινόμενο συνθετικά. Στόχος της σύνθεσης αυτής είναι η παρουσίαση των χαρακτηριστικών κάθε ομάδας χρηστών καθώς και των χαρακτηριστικών των διαδικασιών που εφαρμόζουν οι χρήστες σε σχέση με το σύστημα. Η σύνθεση αυτή αναμένεται να δώσει αποτελέσματα που μπορούν να αξιοποιηθούν σε μια σειρά εφαρμογών για τη βιβλιοθήκη και ιδιαίτερα για θέματα που άπτονται της εκπαίδευσης των χρηστών, του βαθμού εξοικείωσης με το σύστημα, τις ιδιαιτερότητες της κάθε ομάδας, το βαθμό χρήσης συγκεκριμένων στοιχείων του συστήματος και κατ' επέκταση τη δημιουργία ενός εργαλείου υποδομής για την αξιολόγηση των παραπάνω παραμέτρων. Παρατίθεται η μεθοδολογία βήμα προς βήμα με παρουσίαση των σημαντικότερων σχεδιαγραμμάτων για κατανόηση της πορείας της μεθοδολογίας και τέλος, σημειώνονται τα συμπεράσματα στα οποία οδηγεί η στατιστική επεξεργασία χωρίς την παρέμβαση του ανθρώπινου-υποκειμενικού παράγοντα.

Στο παρελθόν έχουν γίνει έρευνες στις οποίες σκιαγραφείται το προφίλ των χρηστών μιας ακαδημαϊκής βιβλιοθήκης μέσω στατιστικών στοιχείων τα οποία χαρακτηρίζουν μονοδιάστατα το χρήστη. Η στατιστική μεθοδολογία που προτείνεται με τη συγκεκριμένη έρευνα παρέχει τη δυνατότητα να αναδειχθούν περισσότερα του ενός χαρακτηριστικά ταυτοχρόνως και κατά συνέπεια το αντικείμενο της μελέτης να περιγραφεί πολυδιάστατα, να ανακαλυφθούν ομοιότητες και να σχηματιστούν ομάδες με κοινά γνωρίσματα. Κατά συνέπεια, η στατιστική επεξεργασία δεδομένων προερχόμενων από βιβλιοθήκες έχει τη δυνατότητα να εξελιχθεί και να μην περιορίζεται μόνο σε αναλύσεις ποσοτικών δεδομένων.

Επομένως, η συγκεκριμένη έρευνα έχει ως σκοπό να συνεισφέρει στην επιστήμη της πολυδιάστατης στατιστικής ανάλυσης προσφέροντας ένα νέο πεδίο εφαρμογών από το χώρο της βιβλιοθηκονομίας και της επιστήμης της πληροφόρησης, ένα χώρο που ανήκει στις κοινωνικές επιστήμες αλλά που προς το παρόν δεν είχε αποτελέσει αντικείμενο παροχής δεδομένων προς πολυμεταβλητή ανάλυση.

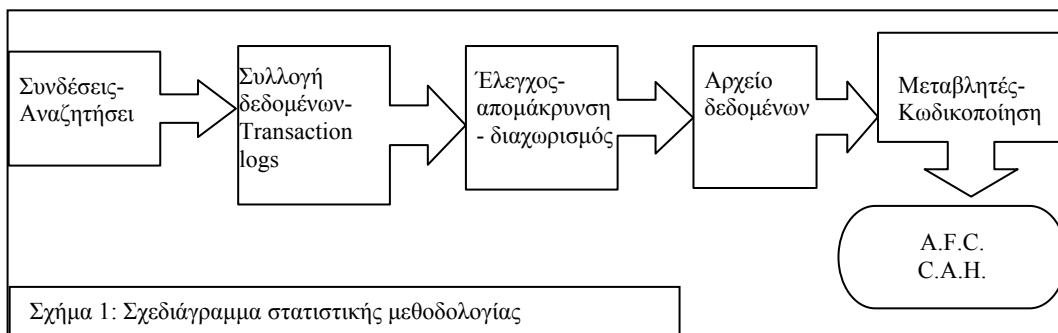
### **Στατιστική μεθοδολογία**

Σε αυτό το σημείο θα γίνει μια προσπάθεια να αποδοθεί με συνοπτικό τρόπο η όλη διαδικασία που ακολουθήθηκε κατά τη διάρκεια της στατιστικής μεθοδολογίας ώστε τα πρωτογενή δεδομένα να οδηγήσουν σε συμπεράσματα σχετικά με τους χρήστες της ακαδημαϊκής βιβλιοθήκης μέσα από μια νέα για τα βιβλιοθηκονομικά δρώμενα στατιστική ανάλυση.

Προκειμένου να ξεκινήσει η έρευνα χρειάζεται να ορισθούν τα αντικείμενα αυτής, ο τρόπος συγκέντρωσής τους, οι μεταβλητές που θα συμβάλλουν στην ανάλυσή τους και οι στατιστικές μέθοδοι μέσω των οποίων θα πραγματοποιηθεί η επεξεργασία και η ανάλυση. Αντικείμενα της έρευνας και συνεπώς ο στατιστικός πληθυσμός υπήρξαν οι αναζητήσεις των χρηστών στον ορακ της Βιβλιοθήκης του Πανεπιστημίου Μακεδονίας. Με τον προσδιορισμό του πληθυσμού και κατόπιν του δείγματος και αφού εισήχθη στην εισαγωγική οθόνη του καταλόγου ένα online ερωτηματολόγιο των πέντε ερωτήσεων κλειστού τύπου, άρχισε η συγκέντρωση των πληροφοριών για τις αναζητήσεις των χρηστών μέσα από τα αρχεία των transaction logs του συστήματος. Δημιουργήθηκε ένα πρόγραμμα αποκωδικοποίησης των πληροφοριών που «κρατούσε» το σύστημα στα αρχεία των transaction logs προκειμένου να υπάρχει η δυνατότητα επεξεργασίας. Ακολούθως πραγματοποιήθηκε χειρονακτικός έλεγχος σε όλες τις συνδέσεις, που σημειώθηκαν στο υπό εξέταση διάστημα, απομακρύνθηκαν όσες δεν προέβαιναν σε αναζητήσεις και διαχωρίστηκαν οι λοιπές σε τρία αρχεία ανάλογα με το αν παρείχαν τις επιρόσθετες πληροφορίες για την ιδιότητα του χρήστη ή όχι. Σε αυτό το σημείο πρέπει να σημειώσουμε ότι αντικείμενο έρευνας για τη συγκεκριμένη ανακοίνωση αποτέλεσε το αρχείο με τις αναζητήσεις των εξωτερικών χρηστών της βιβλιοθήκης, δηλαδή εκείνων των χρηστών που συμπλήρωσαν το προαναφερθέν ερωτηματολόγιο με την επιλογή «άλλο». Στη συνέχεια ορίζονται οι μεταβλητές και τα χαρακτηριστικά τους και πραγματοποιείται σταδιακά η κωδικοποίηση των πρωτογενών

δεδομένων με χειρονακτικό τρόπο και όχι με αυτοματοποιημένο προκειμένου να πραγματοποιηθεί παράλληλα και ο έλεγχος ορθότητας των απαντήσεων/ δεδομένων. Ο αρχικός πίνακας με τα κωδικοποιημένα πλέον δεδομένα εισάγεται στο στατιστικό πακέτο M.A.D. και μέσω αυτού κατασκευάζονται οι γενικευμένοι πίνακες συμπτώσεων σχετικών και απολύτων συχνοτήτων, οι λογικοί πίνακες 0-1 και ακολούθως πραγματοποιείται η πολυδιάστατη στατιστική ανάλυση με τις μεθόδους της Παραγοντικής Ανάλυσης Αντιστοιχιών (A.F.C.) και της Ταξινόμησης Κατ' Αύξουσα Ιεραρχία (C.A.H.). Το σύστημα με την ολοκλήρωση της επεξεργασίας εξάγει μια σειρά από πίνακες, οι οποίοι αποτελούν αντικείμενο ανάλυσης και μέσο για την εξαγωγή συμπερασμάτων.

Στο σχήμα που ακολουθεί γίνεται προσπάθεια μέσα από ένα απλό και όσο το δυνατόν πιο συνοπτικό σχεδιάγραμμα να αποδοθεί σχηματικά η προαναφερθείσα στατιστική μεθοδολογία.



Η στατιστική μεθοδολογία που ακολουθήθηκε για την επεξεργασία και ανάλυση των δεδομένων χαρακτηρίζεται κατά κύριο λόγο από τη δυνατότητα που παρέχει στον ερευνητή να μελετήσει κάθε υπό ανάλυση μεταβλητή σε σχέση με όλες τις υπόλοιπες και να μελετηθεί το φαινόμενο με τη χρήση πολλών μεταβλητών ταυτοχρόνως έχοντας όλες εξίσου την ίδια βαρύτητα.

### Αρχείο δεδομένων

Το αρχείο δεδομένων που υπήρξε αντικείμενο μελέτης, είναι μεγέθους πεντακοσίων δύο αντικειμένων (502) και αποτελείται από τις συνδέσεις που πραγματοποίησαν με τον κατάλογο οι εξωτερικοί κατά κύριο λόγο χρήστες της Βιβλιοθήκης. Οδηγηθήκαμε στο συμπέρασμα ότι πρόκειται για εξωτερικούς χρήστες της Βιβλιοθήκης και όχι για μελέτη της ακαδημαϊκής κοινότητας καθώς το κύριο δεδομένο στο συγκεκριμένο αρχείο είναι η συμπλήρωση της φόρμας σε όλα τα πεδία της με την επιλογή «άλλο».

### Μεταβλητές

Οι μεταβλητές οι οποίες συνέβαλλαν στην επεξεργασία και ανάλυση των δεδομένων είναι οι ακόλουθες τριάντα (30): 1) Ώρα αναζήτησης, 2) Τερματικό, 3) Γλώσσα αναζήτησης, 4) Είδος αρχικής αναζήτησης, 5) Αριθμός απλών αναζητήσεων, 6) Αριθμός σύνθετων αναζητήσεων, 7) Συνολικό αριθμός "search", 8) Συνολικός αριθμός επιλογών, 9) Χρόνος δράσης, 10) Τρόπος αρχικής αναζήτησης, 11) Αριθμός αναζητήσεων με *τίτλο*, 12) Αριθμός αναζητήσεων με *συγγραφέα*, 13) Αριθμός αναζητήσεων με *θέμα*, 14) Αριθμός αναζητήσεων με *σειρά*, 15) Αριθμός αναζητήσεων με *βασικό ευρετήριο*, 16) Αριθμός αναζητήσεων με *ISSN/ISBN*, 17) Χρήση συνδυασμών, 18) Αριθμός χρήσης του *και(and)*, 19) Αριθμός χρήσης του *όχι(not)*, 20) Αριθμός χρήσης του *ή(or)*, 21) Χρήση περιορισμών, 22) Αριθμός χρήσης του περιορισμού *βιβλιοθήκης*.

23) Αριθμός χρήσης του περιορισμού *συλλογής*, 24) Αριθμός χρήσης του περιορισμού *χρόνου έκδοσης*, 25) Επαναλαμβανόμενες λέξεις στις συνεχείς αναζητήσεις, 26) Αριθμός λανθασμένων ερωτήσεων, 27) Αριθμός ορθογραφικών λαθών, 28) Αριθμός λαθών στον τρόπο αναζήτησης, 29) Αριθμός λαθών από παραλήψεις, 30) Αριθμός λαθών του συστήματος.

### Κωδικοποίηση

Με τη συγκέντρωση των πρωτογενών πληροφοριών μέσα από τα αρχεία των transaction logs και το σχηματισμό του αρχείου δεδομένων, ξεκινά η διαδικασία κωδικοποίησης των πληροφοριών ώστε να καταταχθούν σε μορφή πινάκων και να ακολουθήσει η στατιστική επεξεργασία τους. Επιπρόσθετα, αξίζει να σημειωθεί ότι η κωδικοποίηση προσφέρεται και ως μέσο ελέγχου της ορθότητας των πληροφοριών/ απαντήσεων, καθώς επιτρέπει στον ερευνητή να διαπιστώσει αν υπάρχουν αλληλοσυγκρουόμενες πληροφορίες και να αποκαταστήσει τα αληθινά δεδομένα, αν είναι δυνατόν ή να απομακρύνει τελείως την υπό αμφισβήτηση αναζήτηση. Στη συνέχεια περιγράφεται η πορεία της κωδικοποίησης ενδεικτικά σε κάποιες ποιοτικές μεταβλητές αλλά και σε ποσοτικές με παράθεση παραδειγμάτων προς κατανόηση.

Η κωδικοποίηση των ποιοτικών μεταβλητών συνδέεται άμεσα με το σχηματισμό των κλάσεων τους και κατά την πρώτη φάση δίνεται ένας ακέραιος αριθμός για κάθε πιθανή επιλογή-κίνηση του χρήστη, δηλαδή για κάθε κλάση της μεταβλητής.

Με την ολοκλήρωση της πρώτης κωδικοποίησης στις ποιοτικές μεταβλητές επακολούθησε και δεύτερη, πέρα από την απομάκρυνση ορισμένων εξ' αυτών, κατά την οποία οι κωδικοί ορισμένων μεταβλητών συμπτύχθηκαν. Πιο συγκεκριμένα, οι κωδικοί της μεταβλητής «τρόπος αρχικής αναζήτησης» μειώθηκαν από δέκα σε τέσσερις, καθώς συμπτύχθηκαν οι κωδικοί του τίτλος αλφαβητικά με τον τίτλο λέξη-κλειδί, οι κωδικοί της σειράς αλφαβητικά με τη σειρά λέξη-κλειδί, το ISSN/ISBN και το βασικό ευρετήριο, οι κωδικοί του συγγραφέα αλφαβητικά με το συγγραφέα λέξη-κλειδί, και τέλος, οι κωδικοί που έφερε το θέμα αλφαβητικά με το θέμα λέξη-κλειδί δημιουργώντας αντίστοιχα τέσσερις νέους κωδικούς. Οι λόγοι που οδήγησαν στη σύμπτυξη των μεταβλητών είναι είτε οι μικρές τιμές σε ορισμένες εξ' αυτών με αποτέλεσμα αναλύοντας μεμονωμένα να μην παρουσιάζουν στατιστικό ενδιαφέρον, όπως είναι η περίπτωση των τμημάτων και του ISBN είτε το ίδιο το περιεχόμενο των μεταβλητών, όπως είναι για παράδειγμα ο τίτλος αλφαβητικά και ο τίτλος λέξη-κλειδί, αφού πρόκειται ουσιαστικά για το ίδιο αντικείμενο.

Οι ασυνεχείς ποσοτικές μεταβλητές ακολουθούν μια διαφορετική προσέγγιση ως προς την κωδικοποίηση. Πιο συγκεκριμένα, στην πρώτη φάση δόθηκαν σε κάθε μεταβλητή οι πραγματικοί αριθμοί που σημείωσε κατά την έρευνά του ο χρήστης ενώ στη δεύτερη φάση κάποιες εξ' αυτών συμπτύχθηκαν και δημιούργησαν νέες τάξεις.

### Εφαρμογή των μεθόδων της A.F.C. & C.A.H.

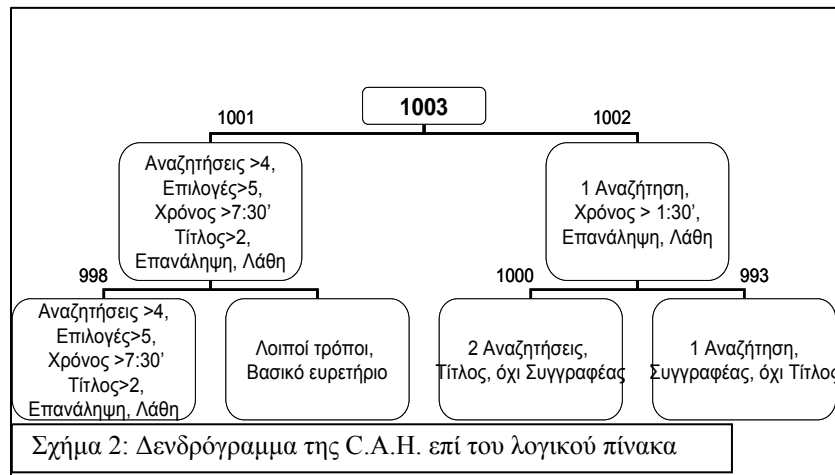
Για την εφαρμογή της μεθόδου της Ταξινόμησης κατ' Αύξουσα Ιεραρχία (C.A.H.) ο αρχικός πίνακας δεδομένων διαστάσεων 502 X 12, στον οποίο 502 είναι τα αντικείμενα και γραμμές του πίνακα και 12 είναι οι μεταβλητές και στήλες συγχρόνως του πίνακα, μετατράπηκε σε λογικό πίνακα διαστάσεων 502 X 38, όπου 38 πλέον είναι το σύνολο των κλάσεων των 12 μεταβλητών και αποτέλεσε τον πίνακα δεδομένων προκειμένου να εφαρμοσθεί η μέθοδος.

Όπως μπορεί να παρατηρήσει κανείς και στο δενδρόγραμμα (βλ. σχήμα 2) με την πρώτη μεγάλη διάσπαση δημιουργούνται δύο ομάδες. Η ομάδα A, τα άτομα της οποίας καλύπτουν το 38% του δείγματος, χαρακτηρίζεται κατά κύριο λόγο από τους χρήστες που παραμένουν αρκετή ώρα συνδεδεμένοι με τον ηλεκτρονικό κατάλογο της Βιβλιοθήκης προβαίνοντας σε αναζητήσεις και ίσως θα μπορούσαν να χαρακτηριστούν ως πιο «επίμονοι» σε αντιδιαστολή με την ομάδα B, τα άτομα της οποίας καλύπτουν το 62% του δείγματος, που χαρακτηρίζεται κυρίως από εκείνους τους χρήστες που παραμένουν συνδεδεμένοι ελάχιστη ώρα και εγκαταλείπουν έχοντας

πραγματοποιήσει μια μόνο αναζήτηση και ίσως θα μπορούσαν να χαρακτηριστούν ως πιο «επιπόλαιοι».

Στη συνέχεια η ομάδα Α διαχωρίζεται στην ομάδα Α1 και Α2. Τα άτομα της ομάδας Α1 αντιπροσωπεύουν το 33% του δείγματος και έχουν κατά κύριο λόγο το χαρακτηριστικό των «επίμονων» χρηστών με τις πολλές αναζητήσεις και επιπλέον, σε μικρότερο βαθμό και την ιδιότητα του θέματος ως τρόπο για την αρχική αναζήτηση. Σε αντιδιαστολή έρχονται τα άτομα της ομάδας Α2, τα οποία αντιστοιχούν στο 5% του δείγματος και διακρίνονται βασικά από τις αναζητήσεις τους μέσω των λοιπών τρόπων, δηλαδή του βασικού ευρετηρίου, της σειράς ή του ISBN και ίσως θα μπορούσαν να χαρακτηρισθούν ως γνώστες των πιο εξειδικευμένων τρόπων αναζήτησης, εκείνων δηλαδή που το ευρύτερο κοινό δεν τους επιλέγει καθώς δεν τους γνωρίζει. Αποτελούν όμως τη μειοψηφία των χρηστών στο συγκεκριμένο αρχείο δεδομένων. Σε δευτερεύοντα βαθμό χαρακτηρίζονται από τρεις αναζητήσεις και καμία επιλογή κατά τη διάρκεια μιας σύνδεσής τους με το σύστημα, η οποία πραγματοποιήθηκε μεταξύ των ωρών 15μ.μ. και 20μ.μ. και διήρκεσε από 2:30 έως 3:30 λεπτά και η οποία παρουσίασε το φαινόμενο των επαναλαμβανόμενων όρων.

Ακολουθώντας και η ομάδα Β διασπάται στις υποομάδες Β1 και Β2. Τα άτομα της ομάδας Β1 αντιπροσωπεύουν το 43% του δείγματος και έχουν κατά κύριο λόγο το χαρακτηριστικό της ολοκλήρωσης της έρευνας με δυο αναζητήσεις μέσω του πεδίου του τίτλου. Δευτερευόντως χαρακτηρίζονται από την απουσία κάποιας επιλογής στα ενδεχόμενα αποτελέσματα ή από την παρουσία μόνο μιας, από τη διάρκεια της αναζήτησης που άγγιζε μόλις το 1:30 λεπτό, από την αποφυγή τόσο επανάληψης των χρησιμοποιούμενων όρων όσο και εμφάνισης λαθών και από τη μεταβλητή που φανερώνει ότι η σύνδεση με τον αυτοματοποιημένο κατάλογο πραγματοποιήθηκε μέσω των παρεχόμενων από το Πανεπιστήμιο γραμμών για δωρεάν σύνδεση. Σε αντιδιαστολή έρχονται τα άτομα της ομάδας Β2, τα οποία αντιστοιχούν στο 19% του δείγματος και διακρίνονται βασικά από τη χρήση αντίστοιχα του συγγραφέα. Κατά δεύτερο λόγο χαρακτηρίζονται από τις δύο επιλογές που σημειώνουν στα παραγόμενα αποτελέσματα, από το χρόνο δράσης τους που ανέρχεται μέχρι τα 2:30 λεπτά και από την απουσία τόσο της χρήσης του θέματος όσο και από την επανάληψη, όπως είναι φυσικό αφού προβαίνουν σε μια μόνο αναζήτηση, και από την ύπαρξη λαθών.



Συνεχίζοντας τη στατιστική επεξεργασία των δεδομένων του συγκεκριμένου αρχείου και προκειμένου να εφαρμοσθεί η μέθοδος της Παραγοντικής Ανάλυσης (A.F.C.) μετατρέπεται ο λογικός πίνακας διαστάσεων 502 X 38, στον οποίο 502 είναι τα αντικείμενα και γραμμές του

πίνακα και 38 είναι το σύνολο των κλάσεων των 12 μεταβλητών και στήλες παράλληλα του πίνακα, στον πίνακα συμπτώσεων διαστάσεων 38 X 38, ο οποίος αποτέλεσε και τον πίνακα δεδομένων.

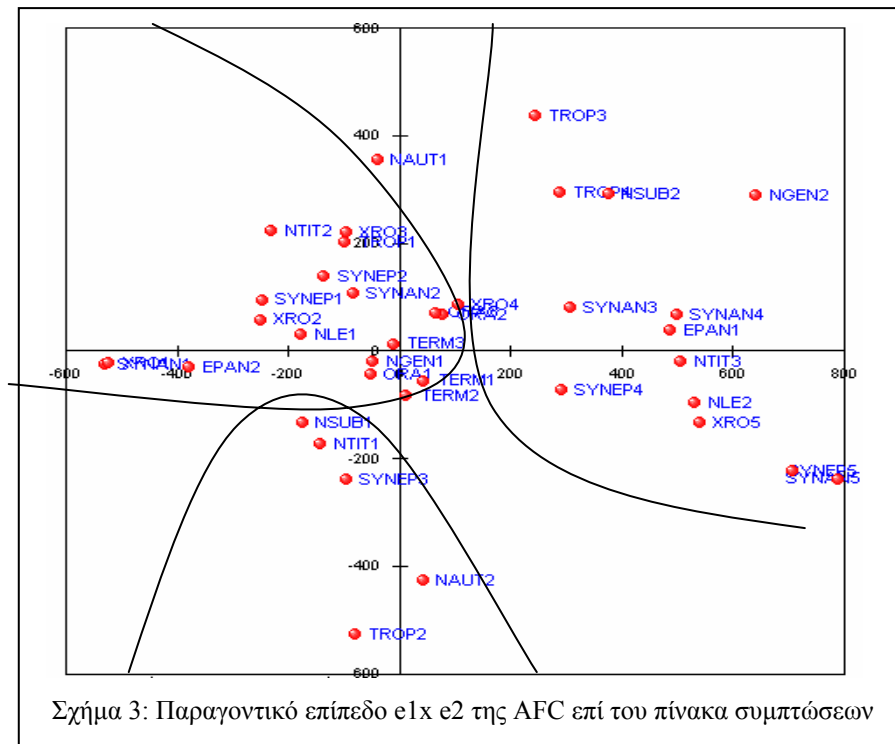
Συνοπτικά αναφέρουμε ότι ο πρώτος παραγοντικός άξονας επιτρέπει κυρίως να διαπιστωθούν οι διαφορές που υπάρχουν μεταξύ της μιας και των πολλών αναζητήσεων κατά τη διάρκεια μιας σύνδεσης με τον ορακ, ενώ ο δεύτερος άξονας επιτρέπει κυρίως να διαπιστωθούν οι διαφορές που υπάρχουν μεταξύ των τρόπων αρχικής αναζήτησης.

Στο επίπεδο δυο διαστάσεων που δημιουργείται από τον πρώτο (οριζόντιος) και το δεύτερο (κάθετος) παραγοντικό άξονα, όπως παρουσιάζεται και στο σχήμα 3, σχηματίζονται τρεις ομάδες.

Πιο συγκεκριμένα, στο αριστερό τμήμα του επιπέδου δημιουργείται η πρώτη ομάδα τα άτομα της οποίας αναζητούν αρχικά εισάγοντας όρο/ όρους στο πεδίο του τίτλου μια φορά, επιλέγουν μια φορά ή δεν επιλέγουν καθόλου κάποιο από τα παραγόμενα αποτελέσματα. Επιπλέον, η έρευνά τους έχει διάρκεια που κυμαίνεται έως και 7:30' χωρίς να παρουσιάζει ούτε το φαινόμενο της επανάληψης ούτε το αντίστοιχο των λαθών.

Στο δεξί τμήμα του επιπέδου σχηματίζεται η δεύτερη ομάδα, η οποία αποτελείται από εκείνους τους χρήστες που αναζητούν από τρεις φορές και άνω κατά τη διάρκεια μιας σύνδεσης με τον ηλεκτρονικό κατάλογο, επιλέγουν το λιγότερο τρεις φορές κάποιο από τα παραγόμενα αποτελέσματα, αναζητούν αρχικά μέσω του πεδίου του θέματος αλλά και των λοιπών τρόπων (βασικό ευρετήριο, σειρά, ISBN) και στη συνέχεια και μέσω του τίτλου και του βασικού ευρετηρίου επαναλαμβάνοντας τους όρους που εισάγουν και κάνοντας λάθη. Η έρευνά τους διαρκεί από 7:30' και άνω.

Στο κάτω μέρος του επιπέδου δημιουργείται η τρίτη ομάδα τα άτομα της οποίας χαρακτηρίζονται από το γεγονός ότι για τις αναζητήσεις που πραγματοποιούν, αξιοποιούν τη δυνατότητα του συγγραφέα και επιλέγουν δυο από τα παραγόμενα αποτελέσματα κατά τη διάρκεια μιας σύνδεσης με τον αυτοματοποιημένο κατάλογο.



Σχήμα 3: Παραγοντικό επίπεδο  $e_1 \times e_2$  της AFC επί του πίνακα συμπτώσεων

### Συμπεράσματα

Συνοψίζοντας και έχοντας ως δεδομένο ότι οι παραπάνω μέθοδοι είναι συμπληρωματικές, αξίζει να σημειωθούν από κοινού μερικά από τα πιο αντιπροσωπευτικά αποτελέσματα:

- Οι χρήστες κατά κύριο λόγο αποτυγχάνουν όταν αναζητούν στον κατάλογο χρησιμοποιώντας το θέμα ή το ISBN, τη σειρά και το βασικό ευρετήριο. Οι λόγοι που οδηγούν στη δυσλειτουργία καταρχάς του θέματος έγκεινται στην αδυναμία πλήρης ταύτισης των θεμάτων που εισαγάγει ο χρήστης στη φυσική γλώσσα και του ελεγχόμενου λεξιλογίου του συστήματος, δηλαδή ανάμεσα σε αυτό που σκέφτεται ο χρήστης και στο λεξιλόγιο καθιερωμένων όρων, όπως είναι οι Θεματικές Επικεφαλίδες της Βιβλιοθήκης του Κογκρέσου (LCSH) που είναι ενσωματωμένες στον κατάλογο. Η μη ταύτιση των όρων της φυσικής γλώσσας και των καθιερωμένων είναι ένα πρόβλημα που εμφανίζεται σε συστήματα ανάκτησης πληροφοριών, όταν η παραμετροποίηση του συστήματος δεν υποστηρίζει τη συγκεκριμένη λειτουργία. Όσον αφορά στις αποτυχημένες αναζητήσεις με βάση τους πρότυπους αριθμούς και τη σειρά ευθύνεται σε μεγάλο βαθμό η άγνοια των χρηστών για τους συγκεκριμένους τρόπους αναζήτησης. Δεν γνωρίζουν ποια είναι ακριβώς η χρησιμότητα και η λειτουργικότητα ούτε των πρότυπων αριθμών ούτε της σειράς και του βασικού ευρετηρίου. Έχει παρατηρηθεί να εισάγεται στο πεδίο του πρότυπου αριθμού ο ταξινομικός αριθμός ή ακόμη και λέξη-κλειδί ή πρόταση προς αναζήτηση. Η έννοια της σειράς και του βασικού ευρετηρίου είναι πιθανότατα άγνωστες στους χρήστες καθώς αφενός μεν η σειρά χρησιμοποιείται ελάχιστα και αφετέρου το βασικό ευρετήριο λιγότερο από ότι θα έπρεπε αφού είναι ο μοναδικός τρόπος μέσω του οποίου ο όρος που εισάγεται αναζητείται ως λέξη-κλειδί σε όλα τα πεδία. Εξάλλου σε αυτό το σημείο πρέπει να σημειωθεί πόσο σχετική είναι η έννοια της επιτυχίας και του βαθμού ικανοποίησης του χρήστη, αφού ο χρήστης θα είναι σε θέση να αξιολογήσει πλήρως το τεκμήριο που ανέκτησε μόνο όταν το εντοπίσει στο βιβλιοστάσιο και το μελετήσει.
- Αναλύοντας τους πίνακες συμπτώσεων των αρχείων δεδομένων, γίνεται φανερό ότι οι χρήστες της συγκεκριμένης βιβλιοθήκης δείχνουν μια σαφή προτίμηση στις «γνωστού-αντικειμένου» αναζητήσεις, δηλαδή σε εκείνες που πραγματοποιούνται με βάση τον τίτλο ενός τεκμηρίου ή τον συγγραφέα, υπεύθυνο πνευματικής παραγωγής του τεκμηρίου. Οι χρήστες πιθανότατα θεωρούν αυτούς τους τρόπους αναζήτησης πιο εύκολους και πιο «ασφαλείς» καθώς αποφεύγονται τα λάθη και συνήθως σημειώνονται περισσότερες επιλογές στα παραγόμενα αποτελέσματα.
- Επιπλέον από το πρώτο κιάλας στάδιο της επεξεργασίας των δεδομένων, παρατηρήθηκε ότι οι χρήστες δεν αξιοποιούν όλες τις δυνατότητες που τους παρέχει το σύστημα. Πιο συγκεκριμένα, χρησιμοποιούν σε ποσοστό μόλις 0,4% τόσο τους περιορισμούς, δηλαδή τους λογικούς τελεστές Boolean, την ημερομηνία έκδοσης, τους περιορισμούς της βιβλιοθήκης και της συλλογής, όσο και τους συνδυασμούς των όρων. Επιπλέον, μόλις το 2% των χρηστών συνδυάζει τους όρους που εισαγάγει, αφού δεν επιλέγει κατά κύριο λόγο τη σύνθετη αναζήτηση. Η προαναφερθείσα περιορισμένη χρήση οδήγησε και στην απομάκρυνση των μεταβλητών που σχετίζονται με τις συγκεκριμένες πτυχές του καταλόγου με την ολοκλήρωση της αρχικής επεξεργασίας. Επίσης, πρέπει να σημειωθεί, όσον αφορά στους τρόπους αναζήτησης, η περιορισμένη χρήση του ISBN και της σειράς σε ποσοστό 5% και 2% αντίστοιχα.
- Ολοκληρώνοντας την επεξεργασία των δεδομένων και σκιαγραφώντας τη συμπεριφορά των χρηστών στον αυτοματοποιημένο κατάλογο μιας ακαδημαϊκής βιβλιοθήκης, αναδεικνύονται στοιχεία που μπορούν να συμβάλλουν αποφασιστικά στην αναδιαμόρφωση της πολιτικής της βιβλιοθήκης έχοντας ως στόχο τη βελτίωση των



παρεχόμενων υπηρεσιών και την εξυπηρέτηση των χρηστών. Αναδείχθηκαν στοιχεία που φανερώνουν τις αδυναμίες αφενός μεν του χρήστη να αξιοποιήσει πλήρως τον κατάλογο και αφετέρου του συστήματος σε ορισμένες περιπτώσεις να υποστηρίζει και να διευκολύνει τον χρήστη. Έχοντας τα συγκεκριμένα δεδομένα η πολιτική της βιβλιοθήκης δύναται να επηρεαστεί τόσο στον ανθρωποκεντρικό χαρακτήρα του λειτουργικού της συστήματος όσο και στην αποδοτικότητα και αποτελεσματικότητά του.

Η έλλειψη κατανόησης του τρόπου με τον οποίο πρέπει οι χρήστες να αναζητήσουν στον κατάλογο το αντικείμενο που τους ενδιαφέρει ίσως σκιαγραφεί ακόμη μια ανάγκη για βιβλιογραφική εκπαίδευση των χρηστών και για κάποια αλλαγή στη δομή του συστήματος.

### Προτάσεις

Σε συνέχεια των συμπερασμάτων που παρατέθηκαν, ακολουθούν προτάσεις που αφορούν τόσο στις βιβλιοθηκονομικές εφαρμογές όσο και στη χρήση των μεθόδων της έρευνας για την ανάπτυξη νέων τρόπων ανάλυσης ποιοτικών δεδομένων στο χώρο των βιβλιοθηκών και των πληροφοριακών τους συστημάτων.

Η παρούσα έρευνα μας οδηγεί στο συμπέρασμα ότι τα transaction logs που χρησιμοποιήθηκαν για τη συλλογή των δεδομένων, αναδεικνύουν πρότυπα συμπεριφοράς. Είναι επίσης εμφανές ότι η χρήση των transaction logs στη διαπίστωση αυτών των προτύπων αποφέρει ικανοποιητικά αποτελέσματα. Τα πρότυπα συμπεριφοράς φανερώνουν και αποδεικνύουν με τεκμηριωμένα πλέον στοιχεία: α) τις αδυναμίες των χρηστών, β) τα προβλήματα που αντιμετωπίζουν, γ) τις ελλείψεις στη βιβλιογραφική τους εκπαίδευση, δ) τις συνήθειες που αναπτύσσουν καθώς και ε) τις αδυναμίες και ελλείψεις που έχει το ίδιο το σύστημα ανάκτησης. Μέρος των προβλημάτων πολλές φορές είναι γνωστά στους βιβλιοθηκονόμους, καθώς έρχονται σε καθημερινή επαφή με το χρήστη. Η έκταση των προβλημάτων δείχνει να μην είναι κατανοητή καθώς η έλλειψη τεκμηριωμένων δεδομένων δεν το επέτρεπε ενώ ταυτόχρονα δεν παρείχε μηχανισμό για τη διεκδίκηση αλλαγών από τους υπευθύνους του οργανισμού. Επομένως, πραγματοποιώντας μια εμπειρισταωμένη έρευνα, παρέχεται στους βιβλιοθηκονόμους ένα εργαλείο πίεσης με στόχο πάντα τη βελτίωση των παρεχόμενων υπηρεσιών και την ικανοποίηση των χρηστών.

Στηριζόμενη σε αυτά τα δεδομένα η Βιβλιοθήκη μπορεί να επανεξετάζει συγκεκριμένα σημεία των λειτουργιών της και να προβαίνει στην αναβάθμιση των συστημάτων της. Στη συνέχεια παρατίθενται προτάσεις βελτίωσης των ισχυουσών πρακτικών και ιδέες για την αντιμετώπιση των διαπιστωθέντων προβλημάτων. Πιο συγκεκριμένα προτείνεται:

- Συνεχής εκπαίδευση των χρηστών στη χρήση των συστημάτων της βιβλιοθήκης. Ειδικότερη έμφαση πρέπει να δοθεί σε θέματα βιβλιογραφικής εκπαίδευσης με στόχο την αναγνώριση των ανάλογων βιβλιογραφικών σημείων πρόσβασης στους καταλόγους και τις πληροφοριακές πηγές της βιβλιοθήκης. Αυτό θα έχει ως αποτέλεσμα την πληρέστερη αξιοποίηση του καταλόγου και των δυνατοτήτων ανάκτησης.
- Επανεξέταση της ορολογίας του ελεγχόμενου λεξιλογίου ανάκτησης και αναζήτησης τρόπων ενσωμάτωσης της φυσικής γλώσσας.
- Αναπροσαρμογή των παραμέτρων του συστήματος για συσχετισμό του λεξιλογίου της φυσικής γλώσσας και του ελεγχόμενου λεξιλογίου. Το γεγονός αυτό θα επιτρέψει την ανάκτηση πληροφοριών σε φυσική γλώσσα και θα απαλείψει σημαντικό μέρος των λαθών στη διαδικασία αναζήτησης.
- Online βοήθεια με σύντομες και σαφείς απαντήσεις.
- Αλλαγή στη δομή και εμφάνιση του καταλόγου, δηλαδή
- εισαγωγή online παραδειγμάτων για τη ορθή χρήση του κάθε τρόπου αναζήτησης και
- αυτόματη διόρθωση των ορθογραφικών λαθών με επιλογή.

- Τέλος, καλύτερη υποστήριξη του server ώστε να μην «κρεμάει» το σύστημα.
- Συνοψίζοντας, διαπιστώνουμε ότι η αποτύπωση της συμπεριφοράς των χρηστών στον αυτοματοποιημένο κατάλογο μιας ακαδημαϊκής βιβλιοθήκης, αναδεικνύει στοιχεία που μπορούν να συμβάλλουν αποφασιστικά στην αναδιαμόρφωση της πολιτικής της βιβλιοθήκης έχοντας ως στόχο τη βελτίωση των παρεχόμενων υπηρεσιών και την εξυπηρέτηση των χρηστών.
- Έχοντας τα συγκεκριμένα δεδομένα η πολιτική της βιβλιοθήκης δύναται να επηρεαστεί τόσο στον ανθρωποκεντρικό χαρακτήρα του λειτουργικού της συστήματος όσο και στην αποδοτικότητα και αποτελεσματικότητά του.

### Βιβλιογραφία

1. Anne Ciliberti [et.al] “Empty Handed?: a material availability study and transaction log analysis verification”. *The Journal of Academic Librarianship* 24 no 4 (1998) 282-289.
2. Deborah D. Blecic [et.al], ”Using Transaction Log Analysis to Improve OPAC Retrieval Results”. *College & Research Libraries* 59 no 1 (1998) 39-50
3. Bernard J. Jansen, Amanda Spink and Tefko Saracevic, “Real life, Real Users, and Real Needs: a study and analysis of user queries on the web”. *Information Processing & Management* 36 no 2 (2000) 207-227.
4. Bonnie Macewan, “Understanding Users’ Needs and Making Collections Choices”. *Library Collections, Acquisitions and Technical Services* 23 no 3 (1999) 315- 320.
5. Malika Mahoui and Sally Jo Cunningham, “A Comparative Transaction Log Analysis of Two Computing Collections”. *Lecture Notes in Computer Science* 1293 (2000) 418-423.
6. M.S. Sridhar, “OPAC vs Card Catalogue: a comparative study of user behaviour”. *The Electronic Library* 22 no 2 (2004) 175-183.
7. Thomas A Peters, “When Smart People Fail: an analysis of the transaction log of an online public access catalog”. *The Journal of Academic Librarianship* 15 no 5 (1989) 267-273.
8. \_\_\_\_\_. “Using Transaction Logs Analysis for Library Management Information”. *Library administration and management* 10 (1996) 20-25.
9. Steve Jones [et.al.], “A Transaction Log Analysis of a Digital Library”. *Human –Computer Interaction for DigitalLibraries* 3 (2000) 152-169.
10. Δημήτριος Ν. Καραπιστόλης, *Δημιουργία Λογισμικού για την Κατάρτιση Φερέγγου Χαρτοφυλακίου με Μεθόδους Ανάλυσης Δεδομένων*. Διδακτορική διατριβή, Θεσσαλονίκη: Πανεπιστήμιο Μακεδονίας Τμήμα Εφαρμοσμένης Πληροφορικής, 1996.
11. \_\_\_\_\_. *Ανάλυση Δεδομένων και Έρευνα Αγοράς*. Θεσσαλονίκη: Ανίκουλα, 2001?.
12. Α. Μάλλιαρη, «Η Εικόνα των Χρηστών των Ακαδημαϊκών Βιβλιοθηκών όπως Προκύπτει από τις Ηλεκτρονικές Υπηρεσίες τους», *Ακαδημαϊκές βιβλιοθήκες: επεκτείνοντας τα όρια: πρακτικά, 12<sup>ο</sup> Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών, Σέρρες, 12-14 Νοεμβρίου 2003*. Επιμ. Σοφία Π. Γρηγοριάδου. Σέρρες: Τμήμα Εκδόσεων & Βιβλιοθήκης Τ.Ε.Ι. Σερρών, 2004, 409-419. Αναρτημένη εργασία.
13. \_\_\_\_\_. «Ανάκτηση Πληροφοριών από τον Αυτοματοποιημένο Κατάλογο Ακαδημαϊκής Βιβλιοθήκης». *Τετράδια Ανάλυσης Δεδομένων* 4(2004): 123-135.
14. \_\_\_\_\_. *Ανάκτηση πληροφοριών από αυτοματοποιημένα συστήματα ακαδημαϊκών βιβλιοθηκών και στατιστική ανάλυση των δεδομένων αυτών*, Διδακτορική διατριβή, Θεσσαλονίκη: Πανεπιστήμιο Μακεδονίας Τμήμα Εφαρμοσμένης Πληροφορικής, 2005.
15. Α. Μάλλιαρη και Γ. Παπαδημητρίου, «Ταξινόμηση των Χρηστών της Βιβλιοθήκης του Πανεπιστημίου Μακεδονίας», *Πρακτικά 15<sup>ο</sup> Πανελλήνιου Συνεδρίου Στατιστικής: συμβολή της στατιστικής στην ανάπτυξη της περιφέρειας, 8-11 Μαΐου 2002, Ιωάννινα*. Επιμ. Ελληνικό Στατιστικό Ινστιτούτο. Ιωάννινα: Πανεπιστημιακό τυπογραφείο, 2002, 435-442.

16. Γ. Παπαδημητρίου και Γ. Αθανασίου, «Η Συμπληρωματικότητα της AFC με την CAH». *Τετράδια ανάλυσης δεδομένων* 3 (2003) 41-52.
17. Γιάννης Παπαδημητρίου, *Μέθοδοι Ανάλυσης Δεδομένων*. Θεσσαλονίκη: Πανεπιστήμιο Μακεδονίας, 1994.
18. \_\_\_\_\_. *Περιγραφική Στατιστική*. Θεσσαλονίκη: Παρατηρητής, 2001.
19. \_\_\_\_\_. «Η Ανάλυση Δεδομένων στην Ελλάδα». *Τετράδια Ανάλυσης Δεδομένων* 1 (2002) 5-11.