

Ανάπτυξη Ψηφιακών Συλλογών στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Development of digital collections at Aristotle University of Thessaloniki

Ευστάθιος Αμανατίδης, Κωνσταντία Πελτέκη, Αντώνιος Σαραγιώτης,
Ανέστης Σίτας
Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Efstathios Amanatidis, Konstantia Pelteki, Antonios Saragiotis, Anestis Sitas
Aristotle University of Thessaloniki

Περίληψη

Αντικείμενο της εισήγησης αυτής είναι η παρουσίαση του έργου της ψηφιοποίησης του Α.Π.Θ. στο πλαίσιο του Β' ΕΠΕΑΕΚ. Αφορά δύο κατηγορίες τεκμηρίων: α) πρωτογενές ψηφιακό υλικό, και β) έντυπο υλικό που ψηφιοποιείται. Η πρώτη κατηγορία αφορά στις τρέχουσες διδακτορικές διατριβές που εκπονούνται στο σύνολο των Τμημάτων του Α.Π.Θ. και η δεύτερη στη ψηφιοποίηση της συλλογής άρθρων εφημερίδων και περιοδικών. Αναφέρονται, η πολιτική επιλογής και απόκτησης υλικού, οι διαδικασίες που αφορούν τα πνευματικά δικαιώματα, τα μορφότυπα και όλες οι τεχνικές λεπτομέρειες διαχείρισης των τεκμηρίων, καθώς και τα πρότυπα βιβλιογραφικής περιγραφής τους. Παρουσιάζονται τα κριτήρια και οι διαδικασίες επιλογής κατάλληλου λογισμικού και η παραμετροποίησή του. Γίνεται σύντομη περιγραφή του λογισμικού που επιλέχθηκε και δίνονται αντιπροσωπευτικά παραδείγματα. Γίνεται αναφορά στα έως τώρα αποτελέσματα, στις νέες υπηρεσίες που προσφέρονται στους χρήστες, στα προβλήματα που ενέκυψαν και στον τρόπο που επιλέχθηκε για την επίλυσή τους.

Λέξεις-Κλειδιά: Δημιουργία ψηφιακών συλλογών, Λογισμικά ψηφιοποίησης, Μεταδεδομένα, Γκρίζα λογοτεχνία

Abstract

The goal of this paper is to present the digitization project of Aristotle University. The material is consisted of the following categories: a) original digital material, and b) printed material that is being digitized. The first category includes the current phd dissertations from all Departments of Aristotle University, the second category is concerned with the digitization of articles, book reviews from the newspaper and magazine collection. Selection and acquisition policy, copyright clearance issues, formats and all the technical details for the treatment of the documents, as well as the bibliographical description standards, are presented. Criteria and procedures of software selection and adequate setting of parameters are discussed. Short description of the selected software and characteristic examples are presented. Special references to the results of the project, the new services, the upcoming problems and the selected method for resolving are included.

Keywords: *Digital collections, Digitization projects, Digitization softwares, Metadata, Grey literature*

Εισαγωγή

Η δημιουργία ψηφιακών βάσεων δεδομένων στο πλαίσιο των ακαδημαϊκών βιβλιοθηκών στοχεύει στην ανάδειξη και διατήρηση των ειδικών και σπανίων συλλογών που διαθέτουν, στην αξιοποίηση των αποτελεσμάτων του ερευνητικού έργου των εκπαιδευτικών ιδρυμάτων, και τέλος στη βελτίωση της εκπαιδευτικής διαδικασίας με την περαιτέρω αύξηση των πληροφοριών που διατίθενται μέσω ενός ψηφιακού περιβάλλοντος μάθησης (Virtual Learning Environment). Επιπλέον, αποτελεί κοινό τόπο, πως το Διαδίκτυο είναι μια πηγή υλικού που οριοθετείται από ελάχιστο αριθμό γλωσσών (κυρίως τα αγγλικά). Στο πλαίσιο λοιπόν των ελληνικών βιβλιοθηκών, ιδιαίτερα σημαντικός στόχος είναι η διάθεση και ο εμπλουτισμός του Διαδικτύου με ψηφιακό υλικό στην ελληνική γλώσσα.

Η ψηφιοποίηση υλικού για τη δημιουργία βάσεων δεδομένων είτε για εκπαιδευτικούς, ερευνητικούς ή πολιτιστικούς σκοπούς, είτε για τη διατήρηση σπάνιων συλλογών, είναι μια αρκετά δύσκολη και δαπανηρή διαδικασία. Αν και η εμπειρία που υπάρχει μέσα στο πλαίσιο των βιβλιοθηκών των ελληνικών ακαδημαϊκών ιδρυμάτων είναι σχετικά μικρή, τα τελευταία χρόνια παρουσιάζονται όλο και περισσότερες ελπιδοφόρες προσπάθειες. Παρ' όλες τις τυχόν δυσκολίες η απόκτηση εμπειρίας στην ψηφιοποίηση υλικού είναι απαραίτητη μιας και η ανάγκη ψηφιοποίησης συλλογών είναι μεγάλη.

Σχεδιασμός και διαχείριση

Για τη σχεδίαση ενός έργου ψηφιοποίησης πρέπει εξ αρχής να ορίζεται η σκοπιμότητα και η δυνατότητα υλοποίησής του, όσον αφορά την τεχνική πλευρά του. Μόνο με τον τρόπο αυτό είναι δυνατόν να επιτευχθεί μια ρεαλιστική και ορθολογική διαχείρισή του. Στα έργα που βασίζονται σε μεγάλο βαθμό σε τεχνολογία που αλλάζει γρήγορα είναι πολύ σημαντικό η σχεδίασή τους να ακολουθεί μια στρατηγική που εξασφαλίζει την επιβίωση του έργου και δεν περιορίζει τις μελλοντικές επιλογές του.

Η πιθανή ροή των εργασιών μπορεί να είναι η εξής:

- Επιλογή των συλλογών προς ψηφιοποίηση
- Συμφωνία με τους κατόχους των πνευματικών δικαιωμάτων, αν αυτό είναι απαραίτητο
- Σύνταξη οικονομοτεχνικής μελέτης (feasibility study)
- Ψηφιοποίηση
- Δημιουργία μεταδεδομένων (metadata)
- Ποιοτικός έλεγχος
- Παροχή πρόσβασης στους χρήστες

Κρίνεται αναγκαίο η οικονομοτεχνική μελέτη να περιλαμβάνει και μια μελέτη «χρόνου και κίνησης». Αυτή καθορίζει τον χρόνο που θα διατεθεί για τη ψηφιοποίηση και αρχειοθέτηση των εικόνων, την εφαρμογή οποιασδήποτε επεξεργασίας, την προσθήκη των μεταδεδομένων και την πραγματοποίηση ελέγχων ποιότητας για κάθε τεκμήριο. Όλες οι διαδικασίες που προαναφέρθηκαν πρέπει να αναπτυχθούν και να εφαρμοστούν σχολαστικά. Η εφαρμογή τους

επιτυγχάνει την ανάπτυξη μιας ψηφιακής συλλογής υψηλής ποιότητας, βραχυπρόθεσμα τη διασφάλιση της αξιοποίησης και προβολής της, και μακροπρόθεσμα την πιθανή επέκταση και αναβάθμισή της.

Ψηφιοποίηση στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Στο πλαίσιο του Β' ΕΠΕΑΕΚ, Γ' ΚΠΣ στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης (Α.Π.Θ.), οι συλλογές που επιλέχθηκαν για το έργο της ψηφιοποίησης ήταν:

- Συλλογή διδακτορικών διατριβών του Α.Π.Θ. Ο αρχικός στόχος ήταν 100 διατριβές, σύμφωνα όμως με την έως τώρα ροή των εργασιών υπολογίζεται πως θα ξεπεράσουν τις 500 έως το Δεκέμβριο του 2006.
- Συλλογή ψηφιοποιημένων άρθρων εφημερίδων και περιοδικών της περιόδου 1976-2000. Ο αρχικός στόχος ήταν τα 9000 άρθρα, ο οποίος έχει ήδη επιτευχθεί και με βάση την έως τώρα ροή των εργασιών υπολογίζεται πως θα ξεπεράσουν τα 15000 έως το Δεκέμβριο του 2006.

Οι δύο αυτές συλλογές καταχωρούνται στο ίδιο λογισμικό το οποίο έχει τη δυνατότητα να αναγνωρίζει ξεχωριστές συλλογές, να επιδέχεται εξειδικευμένη παραμετροποίηση για την κάθε συλλογή ανάλογα με τις ανάγκες της και όσον αφορά την αναζήτηση, παρέχει τη δυνατότητα ξεχωριστής ή ταυτόχρονης αναζήτησης σε όλες τις συλλογές. Τα χαρακτηριστικά που αναφέρονται εδώ αφορούν και τις δύο συλλογές, αλλά όπου αυτά ή άλλες διαδικασίες διαφέρουν, γίνεται η απαραίτητη διευκρίνηση.

Πολιτική επιλογής υλικού

Για τη δημιουργία μιας δυναμικής και αξιόλογης ψηφιακής συλλογής αποτελεί ανάγκη αυτή να αποτελείται από τεκμήρια που απαρτίζουν έναν ενιαίο και ομοειδές σύνολο.

Όσον αφορά στις **διδακτορικές διατριβές** δεν εφαρμόζεται καμιά πολιτική επιλογής, αφού καταχωρούνται στο σύνολό τους. Οι υποψήφιοι διδάκτορες μετά από πρυτανική απόφαση που λήφθηκε τον Σεπτέμβριο του 2001, είναι υποχρεωμένοι να καταθέτουν την διδακτορική τους διατριβή, εκτός από έντυπη και σε ψηφιακή μορφή. Η απόφαση αυτή τέθηκε ουσιαστικά σε εφαρμογή τον Μάρτιο του 2003. Να σημειωθεί πως καταχωρούμε μόνο τις τρέχουσες διατριβές οι οποίες είναι πρωτογενώς ψηφιακές και δεν σκανάρονται, σε αντίθεση με τα άρθρα τα οποία σκανάρονται.

Όσον αφορά όμως τη συλλογή **ψηφιοποιημένων άρθρων**, πρόκειται για άρθρα ημερησίων και εβδομαδιαίων ελληνικών εφημερίδων και περιοδικών ποικίλης ύλης. Τα άρθρα αυτά, αποκτώνται από το 1976 και μετά, από τη Βιβλιοθήκη του Τομέα ΜΝΕΣ του Τμήματος Φιλολογίας μέσω της αγοράς τους από εταιρίες αποδελτίωσης τύπου. Το αντικείμενό τους σχετίζεται με κριτικές για εκδόσεις βιβλίων ανθρωπιστικών επιστημών και συνεντεύξεις συγγραφέων, με ιδιαίτερη έμφαση στη λογοτεχνία. Δεν είχε εφαρμοστεί ποτέ μια διαδικασία εκκαθάρισής της, με αποτέλεσμα να περιέχει μεγάλο αριθμό τεκμηρίων που δεν κρίθηκε αναγκαίο να ψηφιοποιηθούν. Για τον λόγο αυτό δημιουργήθηκε γραπτή πολιτική επιλογής του υλικού προς ψηφιοποίηση, ώστε να διασφαλίζεται και στο παρόν αλλά και για το μέλλον, η εφαρμογή κοινής φιλοσοφίας επιλογής από όσους εργάζονται για το έργο αυτό. Η πολιτική αυτή δεν είναι κάτι στατικό αλλά προσαρμόζεται και βελτιώνεται συνεχώς.

Δικαιώματα χρήσης (copyright)

Για τα ψηφιακά τεκμήρια που δημιουργούνται και για τα οποία το ίδρυμα δεν έχει τα πνευματικά δικαιώματα, έπρεπε να διασφαλιστεί το δικαίωμα χρήσης τους.

Όσον αφορά τις **διδασκτορικές διατριβές**, ζητείται εγγράφως το δικαίωμα χρήσης από τους ίδιους τους συγγραφείς. Στην περίπτωση τους, διακρίνονται διαφορετικοί χρονικοί περιορισμοί που καθορίζονται από τους ίδιους τους συγγραφείς. Ο καθένας έχει το δικαίωμα να επιλέξει είτε την άμεση διάθεση της διατριβής του είτε τη διάθεσή της μετά από συγκεκριμένο χρονικό διάστημα το οποίο το ορίζει ο ίδιος. Έχει όμως και τη δυνατότητα να ζητήσει την πλήρη απαγόρευση της διάθεσής της διατριβής σε ψηφιακή μορφή. Μία ακόμη δυνατότητα που έχει ο συγγραφέας, είναι να παραχωρεί ή να απαγορεύει τη δυνατότητα εκτύπωσης της διατριβής του, η οποία επιτυγχάνεται με «κλειδώμα» του pdf, διαδικασία που γίνεται κατά την επισύναμη του στην εγγραφή.

Σύμφωνα με τα διαθέσιμα στοιχεία, από το σύνολο των 354 διατριβών που έχουν καταχωρηθεί ως τώρα:

Καταχωρημένες Διατριβές/ποσοστά πρόσβασης πλήρους κειμένου

Τμήματα	Αριθμός διατριβών	Άμεση %	Καθόλου %	3 μήνες %	6 μήνες %	1 χρόνο %	18 μήνες %	2 χρόνια %	3 χρόνια %	5-10 χρόνια%
Αγγλικής Γλώσσας	7	42,9	28,6	0,0	14,3	14,3	0,0	0,0	0,0	0,0
Αγρονόμων Τοπογράφων Μηχανικών	4	75,0	0,0	0,0	0,0	25,0	0,0	0,0	0,0	0,0
Αρχιτεκτονικής	2	50,0	50,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Βιολογίας	21	47,6	9,5	0,0	14,3	9,5	4,8	4,8	9,5	0,0
Γαλλικής Γλώσσας	3	33,3	0,0	0,0	33,3	0,0	0,0	33,3	0,0	0,0
Γενικό	6	33,3	16,7	0,0	16,7	33,3	0,0	0,0	0,0	0,0
Γεωλογίας	18	50,0	5,6	11,1	5,6	22,2	0,0	5,6	0,0	0,0
Γεωπονίας	19	36,8	5,3	0,0	10,5	15,8	0,0	10,5	21,1	0,0
Δασολογίας	12	33,3	33,3	0,0	0,0	0,0	0,0	8,3	25,0	0,0
Εικαστικών Τεχνών	3	0,0	33,3	0,0	0,0	33,3	0,0	0,0	33,3	0,0
Ηλεκτρολόγων Μηχανικών & Μηχανικών Η/Υ	8	62,5	25,0	0,0	0,0	0,0	0,0	0,0	12,5	0,0
Θεάτρου	1	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0
Θεολογίας	6	16,7	16,7	0,0	16,7	33,3	0,0	0,0	16,7	0,0
Ιστορίας Αρχαιολογίας	24	0,0	50,0	0,0	0,0	8,3	8,3	4,2	25,0	4,2
Ιταλικής Γλώσσας	7	42,9	14,3	14,3	0,0	28,6	0,0	0,0	0,0	0,0
Καλών Τεχνών	1	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Κτηνιατρικής	21	38,1	19,0	4,8	9,5	4,8	0,0	19,0	4,8	0,0
Μαθηματικών	6	16,7	50,0	0,0	0,0	0,0	0,0	16,7	16,7	0,0
Μηχανολόγων Μηχανικών	5	60,0	20,0	0,0	0,0	20,0	0,0	0,0	0,0	0,0
Μουσικών Σπουδών	2	0,0	50,0	0,0	0,0	0,0	0,0	0,0	0,0	50,0
Νομικής	14	7,1	71,4	0,0	14,3	0,0	0,0	0,0	7,1	0,0
Οδοντιατρική	11	27,3	36,4	0,0	18,2	18,2	0,0	0,0	0,0	0,0
Οικονομικών Επιστημών	1	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

Παιδαγωγικό	4	75,0	0,0	0,0	25,0	0,0	0,0	0,0	0,0	0,0
Πληροφορικής	5	60,0	0,0	20,0	0,0	0,0	0,0	0,0	20,0	0,0
Ποιμαντικής	18	16,7	33,3	0,0	16,7	0,0	0,0	11,1	22,2	0,0
Πολιτικών Μηχανικών	21	33,3	14,3	9,5	19,0	4,8	4,8	0,0	14,3	0,0
Προσχολικής Αγωγής	10	10,0	10,0	0,0	30,0	20,0	0,0	10,0	20,0	0,0
ΤΕΦΑΑ	23	34,8	8,7	0,0	13,0	17,4	4,3	0,0	21,7	0,0
Φαρμακευτικής	2	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Φιλολογίας	25	12,0	32,0	0,0	8,0	24,0	4,0	8,0	12,0	0,0
Φιλοσ.- Παιδαγωγικής	6	50,0	16,7	0,0	16,7	0,0	0,0	0,0	16,7	0,0
Φυσικής	15	60,0	20,0	0,0	6,7	6,7	0,0	0,0	6,7	0,0
Χημείας	12	58,3	0,0	8,3	0,0	0,0	0,0	0,0	33,3	0,0
Χημικών Μηχανικών	10	30,0	10,0	10,0	20,0	10,0	0,0	0,0	20,0	0,0
Ψυχολογίας	1	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Σύνολο	354	33,9	21,8	2,5	10,5	11,0	1,7	4,8	13,3	0,6

Από το σύνολο των 354 διατριβών που έχουν καταχωρηθεί ως τώρα:

- 77 συγγραφείς απαγόρευσαν εντελώς τη διάθεση της διατριβής τους (21,7%), και
- 277 συγγραφείς επιτρέψανε τη διάθεσή της διατριβής τους (78,3%)

Από αυτούς:

- 120 συγγραφείς επέλεξαν να επιτρέψουν την άμεση διάθεση της διατριβής τους (33,9%)
- 77 την απαγόρευσαν εντελώς (21,8%)
- 9 την επέτρεψαν έπειτα από 3 μήνες (2,5%)
- 37 την επέτρεψαν έπειτα από 6 μήνες (10,5%)
- 39 την επέτρεψαν έπειτα από 1 χρόνο (11%)
- 6 την επέτρεψαν έπειτα από 18 μήνες (1,7%)
- 17 την επέτρεψαν έπειτα από 2 χρόνια (4,8%)
- 47 την επέτρεψαν έπειτα από 3 χρόνια (13,3%)
- 2 την επέτρεψαν έπειτα από 5-10 χρόνια (0,6%)
- 116 από του 277 που επέτρεψαν την διάθεσή της, ζήτησαν την αφαίρεση της δυνατότητας εκτύπωσής της (32,7%)

Όσον αφορά τη συλλογή **άρθρων εφημερίδων και περιοδικών**, από τους ιδιοκτήτες των εφημερίδων ή περιοδικών στα οποία έχουν δημοσιευτεί, δεν ζητήθηκαν τα πνευματικά δικαιώματα του περιεχομένου των άρθρων, αλλά τα δικαιώματα χρήσης τους όσον αφορά την ψηφιοποίησή τους και την ελεύθερη πρόσβασή τους από την ακαδημαϊκή κοινότητα. Σύμφωνα με την κείμενη νομοθεσία, αυτό δεν είναι πάντα επαρκές, αφού σε πολλά έντυπα είναι πιθανό κάποιοι συγγραφείς να μην έχουν παραχωρήσει τα πνευματικά δικαιώματα. Στην περίπτωση όμως που οι συντάκτες είναι έμμισθοι, τότε το μέσο δημοσίευσης κατέχει τα πνευματικά δικαιώματα. Το γεγονός αυτό καθιστά αδύνατον να εξακριβωθεί σε ποια κατηγορία ανήκουν τα χιλιάδες άρθρα που ψηφιοποιούνται. Όπως είναι γνωστό όμως, από τα δυσκολότερα και τα πιο χρονοβόρα μέρη ενός έργου ψηφιοποίησης, είναι η απόκτηση των δικαιωμάτων χρήσης. Η προσπάθεια αναζήτησης των δικαιούχων για την απόκτηση των δικαιωμάτων χρήσης ξεκίνησε τον Σεπτέμβριο του 2002. Μετά από αλλεπάλληλες προσπάθειες με επιστολές, ηλεκτρονικό ταχυδρομείο, fax, αλλά και με τηλεφωνική επαφή, αποφασίστηκε να σταματήσει η προσπάθεια

αυτή τον Δεκέμβριο του 2003 και να ψηφιοποιηθούν μόνο τα άρθρα των εντύπων εκείνων που ανταποκρίθηκαν θετικά. Στο σύνολό της η προσπάθεια αυτή διήρκεσε 16 μήνες.

Εκτός από τα παραπάνω υπάρχει και η περίπτωση της ιδιοκτησίας πνευματικών δικαιωμάτων για παλαιό ή ασαφές υλικό που υπάρχει στο αρχείο των άρθρων. Η περίπτωση αυτή αφορά έντυπα που έχει διακοπεί η κυκλοφορία τους και είναι σχεδόν αδύνατον να εντοπιστούν οι ιδιοκτήτες τους ή αυτοί που κατέχουν τα δικαιώματα του συγκεκριμένου τίτλου.

Συγκεκριμένα στη συλλογή αυτή:

- Εντοπίστηκαν άρθρα από 101 τίτλους εφημερίδων και περιοδικών
- Παραχωρήθηκε το δικαίωμα χρήσης από 64 εκδότες (63,4%), στους οποίους ανήκει το μεγαλύτερο και πιο αξιόλογο μέρος του υλικού
- Αρνητικές απαντήσεις: καμία
- Δεν εντοπίστηκαν 30 εκδότες (Διακοπή έκδοσης. Δεν εντοπίστηκε το καθεστώς ιδιοκτησίας τους παρά τις αλλεπάλληλες προσπάθειες επικοινωνίας με τους αρμόδιους φορείς) (29,7%)
- Δεν απάντησαν 7 εκδότες (άρα μπορούμε να θεωρήσουμε ότι η απάντησή τους είναι αρνητική) (6,9%)

Αναλυτικότερα, από το σύνολο των 9000 άρθρων που έχουν καταχωρηθεί έως τώρα, η αναλογία ανά έντυπο είναι η εξής:

1. Το Βήμα: 1369 άρθρα (15%)
2. Ελευθεροτυπία: 1307 άρθρα (14,2%)
3. Η Καθημερινή: 1078 άρθρα (12%)
4. Τα Νέα: 827 άρθρα (9%)
5. Μακεδονία: 499 άρθρα (5,5 %)
6. Η Αυγή: 465 άρθρα (5%)
7. Ριζοσπάστης: 381 άρθρα (4,2 %)
8. Ελεύθερος Τύπος: 349 άρθρα (4%)
9. Αγγελιοφόρος: 326 άρθρα (3,5%)
10. Αδέσμευτος Τύπος: 246 άρθρα (2,6%)
11. Υπόλοιπα 54 έντυπα: 2153 άρθρα (25%)

Από τα παραπάνω στοιχεία γίνεται φανερό πως από το σύνολο των παραπάνω 64 τίτλων εφημερίδων και περιοδικών που τελικά ψηφιοποιούνται, 6847 άρθρα ανήκουν μόνο σε 10 συγκεκριμένους τίτλους. Δηλαδή στο 15% του συνόλου των τίτλων εφημερίδων και περιοδικών, ανήκει το 76% των άρθρων. Από τους υπόλοιπους 54 τίτλους, προέρχονται 2153 άρθρα. Δηλαδή στο 85% του συνόλου των τίτλων εφημερίδων και περιοδικών, ανήκει το 24% των άρθρων.

Επιλογή χαρακτηριστικών ποιότητας και αποθήκευση ψηφιοποιημένων τεκμηρίων

Οι **διδασκορικές διατριβές** παραδίδονται από τους συγγραφείς τους σε οποιαδήποτε ψηφιακή μορφή επιθυμούν οι ίδιοι (συνήθως word) και μετατρέπονται σε PDF από το προσωπικό της βιβλιοθήκης με το πρόγραμμα Adobe Acrobat.

Όσον αφορά την ψηφιοποίηση και την επεξεργασία της συλλογής **άρθρων εφημερίδων και περιοδικών** για την δημιουργία του ανάλογου αρχείου, η συσκευή που χρησιμοποιείται είναι ο σαρωτής (scanner). Παράγοντες που πρέπει να ληφθούν υπόψη στην επιλογή του σαρωτή είναι η οπτική ανάλυση, η οπτική πυκνότητα (bit depth ή optical density), η περιοχή σάρωσης και ο

χρόνος σάρωσης. Κάθε ένας από αυτούς τους παράγοντες επηρεάζει την ποιότητα της τελικής εικόνας. Όσον αφορά το συγκεκριμένο έργο για το σκανάρισμα των άρθρων χρησιμοποιούνται τα εξής:

- Επιτραπέζιος σαρωτής A3 Epson Expression 1640XL με μέγιστη οπτική ανάλυση 1600 dpi και δυνατότητα επιλογής έγχρωμης ή ασπρόμαυρης ανάλυσης.
- Επιτραπέζιος σαρωτής A4 HP Scanjet C5490A με μέγιστη οπτική ανάλυση 2400 dpi.
- Εκτυπωτές HP Scanjet 4200 dtn με ανάλυση 1200 X 1200 dpi. Μέσω των εκτυπώσεων γίνεται ο έλεγχος ποιότητας των σκαναρισμένων άρθρων, ενώ αυτές χρησιμοποιούνται και για την διαδικασία της καταλογογράφησης τους.
- H/Y Pentium(R) 4 CPU 200 HGHZ με RAM 512 MB.

Ο αριθμός των άρθρων που μπορούν να σκαναριστούν ημερησίως αποτελεί άμεση συνάρτηση του φυσικού μεγέθους τους, της έκτασης του περιεχομένου τους και των τεχνικών παραμέτρων στις οποίες αυτά σκανάρονται. Εάν δηλαδή το σκανάρισμα γίνεται στα 24-bit color ή σε 8-bit grey και σε 600 ή 300 dpi. Το μέγεθος των ψηφιακών τεκμηρίων έχει σημαντικό αντίκτυπο στο μέγεθος του χώρου αποθήκευσης που απαιτείται (και κατά συνέπεια και στο κόστος) αλλά και στον χρόνο που απαιτείται για να «κατέβουν» και να διαβαστούν αυτά από τους χρήστες. Το μεγαλύτερο μέρος του ανάλογου ψηφιοποιημένου υλικού που υπάρχει στο Διαδίκτυο είναι σκαναρισμένο με ανάλυση 300 dpi.

Στο έργο κατά τον αρχικό σχεδιασμό και στην πρώτη φάση υλοποίησης του τα άρθρα αποθηκεύονταν σε jpeg στα 600 dpi ώστε να είναι διαθέσιμα σε υψηλότερη ανάλυση για πιθανές μελλοντικές χρήσεις (π.χ. OCR), ενώ όπου υπήρχαν πρωτότυπα έγχρωμα τεκμήρια να ψηφιοποιούνται σε 24-bit color. Το αρχείο pdf αποθηκεύονταν με ανάλυση 300 dpi για να αποφευχθεί ο μεγάλος όγκος των αρχείων που αναπόφευκτα θα καθυστερούσε το «κατέβασμά τους» (downloading) από τους χρήστες. Τελικά αποφασίστηκε το συγκεκριμένο υλικό να μη επεξεργαστεί με OCR (Optical Character Recognition), και για να μειωθεί ο χρόνος εργασίας, αποφασίστηκε να σκανάρονται απευθείας στα 300 dpi.

Ως λογισμικό επεξεργασίας εικόνας χρησιμοποιείται το PaintShop Pro 8 και η εικόνα αποθηκεύεται σε μορφή jpeg. Η αποθήκευση γίνεται στον κεντρικό server. Ο τύπος αρχείων εικόνας jpeg επιλέχτηκε λόγω του πλήθους των πλεονεκτημάτων που αυτός προσφέρει (ικανοποιητική ποιότητα ψηφιακής απεικόνισης των τεκμηρίων, δυνατότητα συμπίεσης τους με αποτέλεσμα την δημιουργία μικρών σχετικά σε όγκο ηλεκτρονικών αρχείων κατάλληλων για χρήση μέσα σε περιβάλλον τοπικού ή παγκοσμίου δικτύου). Με τη χρήση του προγράμματος ABBYY Fine Reader 7.0 professional edition, τα αρχεία μετατρέπονται και αποθηκεύονται και σε μορφή pdf. Η μορφή αρχείου pdf είναι εκείνη που επισυνάπτεται στη βάση δεδομένων (ουσιαστικά δηλαδή αυτό διατίθεται στους χρήστες) στο πεδίο 856 του MARC21.

Πολιτική τεχνικής επεξεργασίας υλικού

Κατά την ψηφιοποίηση της συλλογή **άρθρων εφημερίδων και περιοδικών**, εφαρμόζονται οι εξής τεχνικές:

- Παραμένει πάντα η ημερομηνία (είτε πρωτότυπη είτε η «έντυπη» σφραγίδα που δίνει η εταιρεία αποδελτίωσης η οποία προμηθεύει τα άρθρα)
- Προστίθεται σε κάθε άρθρο το λογότυπο της εφημερίδας
- Αφαιρούνται οι διαφημίσεις που συνοδεύουν το άρθρο

- Συχνά (ιδιαίτερα σε άρθρα μεγέθους A3) για λόγους μεγέθους του αρχείου, κάποιες φωτογραφίες υπάρχει περίπτωση να αφαιρεθούν, να μειωθεί το μέγεθος των αρχικών τίτλων, να αφαιρεθεί οτιδήποτε περιττό, να μειωθούν οι στήλες (π.χ. από 3 σε 2), ή ένα μεγάλο μονόστηλο να μετατραπεί σε 2 στήλες

Επιλογή και προμήθεια λογισμικού

Από τα σημαντικότερα στάδια του σχεδιασμού και της υλοποίησης ψηφιακών βάσεων, είναι η επιλογή και η εφαρμογή του κατάλληλου λογισμικού. Βασική επιδίωξη του προγράμματος του Α.Π.Θ., ήταν να χρησιμοποιηθεί λογισμικό που παραχωρείται δωρεάν (ανοικτού κώδικα/open source) και το λογισμικό αυτό να διαθέτει και να ικανοποιεί συγκεκριμένα βασικά χαρακτηριστικά, όπως για παράδειγμα να χρησιμοποιεί το MARC21 για την δημιουργία των μεταδεδομένων και να είναι συμβατό με το πρωτόκολλο συγκομιδής μεταδεδομένων OAI (Open Archives Initiative). Στο τέλος του 2002, εντοπίστηκαν διάφορα λογισμικά, που το καθένα προσέφερε διαφορετικές δυνατότητες. Εδώ αναφέρεται η αξιολόγηση των επικρατέστερων:

1. LAURIN (<http://laurin.uibk.ac.at>)

- Χρηματοδοτήθηκε από την Ευρωπαϊκή Ένωση
- Συμμετέχουν πανεπιστήμια, βιβλιοθήκες και ιδιωτικές εταιρείες. Συμμετέχουν 10 ευρωπαϊκές βιβλιοθήκες. Έχουν εκφράσει ενδιαφέρον 30 βιβλιοθήκες από 18 χώρες
- Εφαρμόζεται για καταχώρηση αποκομμάτων εφημερίδων που βρίσκονται στο Τμήμα Γερμανικής Γλώσσας και Φιλολογίας του University of Innsbruck
- Πολύ καλή οργάνωση πολύγλωσσου θησαυρού
- Η βάση του διαθέτει το ψηφιοποιημένο κείμενο σε μορφή pdf και σε πλήρες κείμενο που προήλθε μετά από επεξεργασία με OCR. Το ενδιαφέρον είναι πως στους χρήστες διατίθεται μόνο η pdf μορφή. Το πλήρες κείμενο παραμένει μη εμφανίσιμο στους χρήστες και χρησιμοποιείται μόνο για την διαδικασία της αναζήτησης. Αυτό γίνεται και για λόγους copyright ώστε να μην μπορεί κανείς να κάνει εύκολα copy-paste το κείμενο.
- Ο μελλοντικός σχεδιασμός του προβλέπει και καταχώρηση και άλλων μορφών τεκμηρίων πέρα από τα αποκόμματα εφημερίδων (κάρτες, χάρτες, χειρόγραφα, κτλ)
- Η υποστήριξη παρέχεται μόνο με αμοιβή
- Τεχνικά χαρακτηριστικά (linux, oracle)
- Δεν υποστηρίζει πρότυπα όπως MARC, Dublin Core, XML
- Έκδοση 2.0.2. Έχει σταματήσει η ανάπτυξή του το 2002 και το ενδιαφέρον τους πλέον επικεντρώνεται κυρίως τον θησαυρό

2. ePrints2 (University of Southampton) (<http://eprints.org/>)

- Μεγάλη βάση χρηστών (140)
- Συμβατό με το OAI (Open Archives Metadata Set XML DTD)
- Υποστηρίζει αναζήτηση πλήρους κειμένου
- Πολύ καλή τεχνική υποστήριξη από τους δημιουργούς
- Καλό αναλυτικό manual (120 σελ.)
- Λιτό και καλό interface, εύκολα παραμετροποιήσιμο
- Σχετικά ώριμο (έκδοση 2.3.0)

3. ETD-db (Virginia Tech) (<http://shcolar.lib.vt.edu/ETD-db>)

- Μεγάλη βάση χρηστών
- Δεν υποστηρίζει MARC, Dublin Core, XML
- Δεν υποστηρίζει αναζήτηση πλήρους κειμένου
- Δεν ανανεώνεται τακτικά με νέες εκδόσεις
- Δεν παρέχει καλή τεχνική υποστήριξη
- Πολύ λιτό interface
- Δύσκολα παραμετροποιήσιμο

4. Dspace (MIT) (<http://libraries.mit.edu/dspace-mit/>)

- Μικρή βάση χρηστών
- Συμβατό με το OAI
- Υποστηρίζει μόνο το Dublin Core
- Υποστηρίζει αναζήτηση πλήρους κειμένου
- Δεν είναι εύκολα παραμετροποιήσιμο
- Έκδοση 1.2

5. CDSware (CERN) (<http://cdsware.cern.ch>)

- Υποστηρίζει MARC/DC/XML
 - Συμβατό με το OAI
 - Υποστηρίζει αναζήτηση πλήρους κειμένου
 - Προσφέρει μεγάλες δυνατότητες παραμετροποίησης (αν και η διαδικασία αυτή τελικά είναι αρκετά δύσκολη)
 - Καλό interface με μεγάλες δυνατότητες παραμετροποίησης
 - Μικρό manual
 - Άριστη τεχνική υποστήριξη από τους δημιουργούς
 - Έκδοση 0.3.3 (ουσιαστικά όμως το κυρίως μέρος τους βασίζεται στην 0.0.9)
 - Ανανεώνεται πολύ τακτικά με νεώτερες εκδόσεις
- Τελικά αποφασίστηκε να εφαρμοστεί το λογισμικό CDSware, το οποίο άρχισε να λειτουργεί τον Μάρτιο του 2003.

CDSware (CERN DOCUMENT SERVER SOFTWARE)

Είναι το λογισμικό που αναπτύσσεται και εφαρμόζεται από το CERN Document Server, στο CERN (Ευρωπαϊκός Οργανισμός Πυρηνικής Έρευνας) που εδρεύει στην Ελβετία. Υποστηρίζει τη δημιουργία εξυπηρετητών ηλεκτρονικών προδημοσιεύσεων (electronic preprint servers), δημόσιων καταλόγων βιβλιοθηκών (OPACs) και συστημάτων τεκμηρίων (document system) στο web. Είναι συμβατό με το πρωτόκολλο συγκομιδής μεταδεδομένων OAI-PMH (Open Archive Initiative— Protocol for Metadata Harvesting) και χρησιμοποιεί ως βασικό βιβλιογραφικό πρότυπο το MARC21.

Δημιουργήθηκε για τη διαχείριση πολύ μεγάλων αποθετηρίων (repositories) που περιέχουν ποικίλους τύπους υλικού, όπως περιγραφές μουσειακών αντικειμένων, συλλογές εμπιστευτικών αλλά και δημοσίων τεκμηρίων, κλπ. Στο CERN, το CDSware διαχειρίζεται περισσότερες από

400 συλλογές δεδομένων, περιέχει πάνω από 650.000 βιβλιογραφικές εγγραφές και 320000 τεκμήρια πλήρους κειμένου. Καλύπτει προδημοσιεύσεις, άρθρα, αναφορές, εργασίες, παρουσιάσεις και πρακτικά συνεδρίων, βιβλία, περιοδικά, φωτογραφίες και βίντεο.

Εφαρμόζει τεχνικές πρόσκτησης μεταδεδομένων όπως η υπηρεσία WebSubmit που επιτρέπει στους συγγραφείς ή/και στους βιβλιοθηκονόμους να καταχωρούν απ' ευθείας τα τεκμήρια και τα μεταδεδομένα τους. Είναι ένα λογισμικό που παραχωρείται από το 2002 ως ολοκληρωμένο πακέτο υπό την ονομασία CDSware, έκδοση **0.0.9**. Η νεώτερη έκδοσή του είναι η **0.3.3** που παρουσιάστηκε στις 16 Ιουλίου 2004.

Οι απαιτήσεις του συστήματος εξαρτώνται από το μέγεθος των συλλογών, τον αριθμό των χρηστών, την πλατφόρμα της βάσης δεδομένων, κτλ. Τα βασικά τεχνικά χαρακτηριστικά του είναι τα εξής:

- Λειτουργικό σύστημα: οποιοδήποτε Unix σύστημα
- Γλώσσα προγραμματισμού: Python και PHP
- Βάση δεδομένων: MySQL
- Web server: Apache
- Μηχανή αναζήτησης: CDSware
- Χρησιμοποιεί δική του τεχνολογία ευρετηρίασης (indexing) και μηχανή αναζήτησης που είναι σχεδιασμένη για μεγάλες συλλογές δεδομένων.
- Άλλο: WML (website META language)
- Υποστηριζόμενοι clients: όλοι οι clients HTML 4.0 και νεώτεροι

Παραμετροποίηση λογισμικού

Από τον Δεκέμβριο του 2002 έως τον Μάρτιο του 2003, έγινε η παραμετροποίηση του λογισμικού. Οι βιβλιοθηκονόμοι σε άμεση συνεργασία με τον τεχνικό υποστήριξης, διατύπωσαν τις ανάγκες και τις απαιτήσεις τους, οι οποίες βέβαια διαμορφώθηκαν και εξαρτήθηκαν άμεσα από τις δυνατότητες του ίδιου του λογισμικού. Η παραμετροποίηση αφορούσε τις διεπαφές της καταχώρησης, της επεξεργασίας των μεταδεδομένων και της διεπαφής των χρηστών. Διαμορφώθηκαν έτσι ώστε να είναι λειτουργικά και για τους βιβλιοθηκονόμους αλλά και για τους χρήστες.

Σχεδιασμός διεπαφής και εγγραφή χρηστών

Κατά το σχεδιασμό της δημιουργίας ενός αρχείου ψηφιακών τεκμηρίων, τα θέματα που αφορούν τους χρήστες είναι πρωταρχικής σημασίας. Όταν αναφερόμαστε σε χρήστες δεν πρέπει να ξεχνάμε πως υπάρχουν οι εσωτερικοί χρήστες (οι βιβλιοθηκονόμοι) και οι εξωτερικοί (οι χρήστες που αναζητούν τις βάσεις). Κατά την παραμετροποίηση του λογισμικού ελήφθησαν μέτρα που μπορούν να αποτρέψουν κάθε είδους κατάχρηση, να προστατέψουν τη συλλογή, να διασφαλίσουν την ακεραιότητα των δεδομένων αλλά και τη μελλοντική χρήση της βάσης ως μια αξιόπιστη πηγή πληροφόρησης.

Από τη στιγμή που τα ψηφιακά τεκμήρια γίνονται διαθέσιμα μέσω του Διαδικτύου και το λογισμικό που εφαρμόστηκε δίνει μεγάλη ελευθερία όσον αφορά και την παραμετροποίησή της διαπεφής με τον χρήστη, έγινε προσπάθεια ώστε να σχεδιαστεί μια διεπαφή που να είναι εύχρηστη και εύκολη στην πλοήγηση.

Εκτός όμως από την τροποποίηση της διεπαφής χρηστών όσον αφορά τον τρόπο παρουσίασης των μεταδεδομένων, καθώς και διαφόρων μετατροπών όσον αφορά την αποδοχή και την εμφάνιση των ελληνικών χαρακτήρων, στην εφαρμογή του Α.Π.Θ. έχει αφαιρεθεί από την διεπαφή χρηστών η δυνατότητα αυτόματης δημιουργίας passwords που επιτρέπει την καταχώρηση ή τροποποίηση τεκμηρίων από τους ίδιους τους χρήστες. Την δυνατότητα αυτή την έχει μόνο ο διαχειριστής του συστήματος. Αυτό κρίθηκε αναγκαίο αφού αποφασίστηκε να μην εφαρμοστεί προς το παρόν το μοντέλο της υποβολής των τεκμηρίων από μεμονωμένους χρήστες.

Δημιουργία και καταχώρηση μεταδεδομένων

Το βασικό σχήμα μεταδεδομένων που χρησιμοποιείται είναι το MARC21. Το σύστημα υποστηρίζει επίσης και την δημιουργία μεταδεδομένων με το βασικό (basic) Dublin Core και τα σύνολα χαρακτήρων Unicode UTF-8 και ISO 8859-7.

Τα μεταδεδομένα μπορούν να εμφανιστούν με διάφορους τρόπους. Στην πρώτη οθόνη στην οποία εμφανίζονται τα τεκμήρια που έχει ανακτήσει ο χρήστης, εμφανίζονται με την απλούστερη και πιο κατανοητή μορφή τους: πρόκειται για την εμφάνιση στο HTML format. Η οθόνη αυτή όπως και όλες οι άλλες μπορούν να διαμορφωθούν και να περιέχουν οποιαδήποτε στοιχεία θέλουμε και με οποιαδήποτε μορφή ή σειρά εμφάνισης. Υπάρχουν όμως και πρόσθετες οθόνες όπου τα στοιχεία των βιβλιογραφικών εγγραφών που καταχωρούνται, μετατρέπονται αυτόματα και παρουσιάζονται σε MARC21 HTML (brief και detailed), XML MARC και σε XML Dublin Core. Ουσιαστικά τα βιβλιογραφικά δεδομένα σώζονται σε μορφή XML MARC, αλλά η MARC HTML αφορά μια διαφορετική και πιο κατανοητή για τον κοινό χρήστη μορφή.

Όσον αφορά την εφαρμογή του MARC21 στην περίπτωση των **διδακτορικών διατριβών**, εκτός των θεματικών επικεφαλίδων στα ελληνικά, καταχωρούνται ταυτόχρονα και στα αγγλικά. Επιπρόσθετα ζητείται εγγράφως από τους συγγραφείς τους να ορίζουν λέξεις-κλειδιά (στην ελληνικά και αγγλική γλώσσα) που καταχωρούνται στο πεδίο 653 του MARC21 (uncontrolled terms) καθώς επίσης και περίληψη στα ελληνικά και αγγλικά που καταχωρείται στο πεδίο 520 του MARC21.

Οι καθιερώσεις των όρων που καταχωρούνται σε όλες τις συλλογές, όσον αφορά τα ονόματα των ελληνικών φυσικών προσώπων, βασίζονται στο «Αρχείο καθιερωμένων επικεφαλίδων» της Εθνικής Βιβλιοθήκης της Ελλάδος, έκδοση 2003. Τα ξενόγλωσσα ονόματα καθιερώνονται από το Library of Congress Authorities (<http://authorities.loc.gov/>). Η καθιέρωση των θεματικών όρων γίνεται με βάση το Library of Congress Subject Headings (LCSH) μεταφρασμένων στα ελληνικά. Συγκεκριμένα η απόδοση των θεματικών όρων στα ελληνικά γίνεται από τους καταλογογράφους του Συστήματος Βιβλιοθηκών του Α.Π.Θ., όπως αυτά καταχωρούνται στον ηλεκτρονικό κατάλογο της βιβλιοθήκης. Πρέπει να σημειωθεί πως υπάρχει εξειδικευμένη ομάδα βιβλιοθηκονόμων που συντηρεί και ενημερώνει τα αρχεία καθιερωμένων όρων (authority files) η οποία ελέγχει και αποδίδει όλους τους νέους όρους. Στην περίπτωση των διδακτορικών διατριβών επειδή αυτές αφορούν καινοτόμα και πολύ εξειδικευμένα θέματα κρίθηκε απαραίτητη και η προσωπική επικοινωνία με τους διδάκτορες ώστε να επιτευχθεί η δημιουργία όσο το δυνατόν πιο έγκυρων μεταδεδομένων.

Αναζήτηση και ανάκτηση τεκμηρίων

Οι καλύτερες συλλογές ψηφιακών τεκμηρίων του κόσμου θα ήταν άχρηστες χωρίς τη δυνατότητα εύκολης αναζήτησης και γρήγορης ανάκτησής τους. Για τη σωστή λειτουργία του μηχανισμού αναζήτησης μεγάλη σημασία έχει η επάρκεια των μεταδεδομένων που τεκμηριώνουν το υλικό της συλλογής. Η προσθήκη των μεταδεδομένων εκ των υστέρων είναι δύσκολη. Τα τεκμήρια πρέπει να παρακολουθούνται από την πηγή τους ως τη ψηφιοποίηση, την προσθήκη μεταδεδομένων, την αρχειοθέτηση, τη διόρθωση και την παράδοση. Επομένως είναι σημαντικό να ληφθούν υπόψη εκ των προτέρων οι πιθανές χρήσεις προτού καθοριστούν οι παράμετροι αναζήτησης (π.χ. ποια πεδία στη βάση δεδομένων θα ευρετηριαστούν).

Στο έργο αυτό επιλέχθηκε να δίνεται η δυνατότητα ευρετηριασμού των εξής πεδίων του MARC21:

- αριθμός εισαγωγής εγγραφής (037)
- συγγραφέας (100, 110)
- τίτλος (ή τίτλοι) (240, 245, 246)
- περίληψη (520)
- θεματικές επικεφαλίδες (600, 610, 630, 650, 651)
- λέξεις-κλειδιά (653)
- πηγή (τίτλος εφημερίδας ή περιοδικού) (786)
- επιβλέπων καθηγητής (900), και
- συμβουλευτική επιτροπή (901)

Δυνατότητες έρευνας

Με βάση την παραμετροποίηση του λογισμικού που εφαρμόστηκε, παρέχονται οι εξής δυνατότητες έρευνας:

- Επιλογής simple search και advanced search
- Ταυτόχρονη αναζήτηση σε όλα τα πεδία μεταδεδομένων
- Αναζήτηση σε επιλεγμένα πεδία μεταδεδομένων (π.χ. μόνο στο πεδίο τίτλου ή συγγραφέα)
- Λογική Boolean (Δυαδική λογική)
- Ξεφύλλισμα (Browse)
 - κατά συγγραφέα
 - κατά τίτλο
 - κατά θεματικό όρο
 - κατά αριθμό εισαγωγής

Διατήρηση ψηφιακών αρχείων

Η μακροπρόθεσμη διατήρηση των ψηφιακών τεκμηρίων θα πρέπει να ληφθεί υπόψη πριν την ψηφιοποίηση. Οι αποφάσεις που θα ληφθούν σε κάθε στάδιο του κύκλου της δημιουργίας του αρχείου θα έχουν επίδραση στο πως θα εφαρμοστούν οι στρατηγικές της ψηφιακής διατήρησης. Η ψηφιακή διατήρηση είναι μια διεργασία κατά την οποία οι πολιτικές και οι διαδικασίες εφαρμόζονται έτσι ώστε το ψηφιακό αρχείο να είναι χρησιμοποιήσιμο ανεξάρτητα από

μελλοντικές αλλαγές στην τεχνολογία υλικού και λογισμικού. Η προσέγγιση που ακολουθήθηκε για να επιτευχθεί ο στόχος αυτός είναι η εφαρμογή σύγχρονων, αξιόπιστων, διεθνώς διαδεδομένων προτύπων (XML, MARC, DC, PDF), έτσι ώστε να είναι δυνατή η μετατροπή τους και η μεταφορά τους σε οποιοδήποτε άλλο πρότυπο ή λογισμικό προκύψει στο μέλλον.

Ασφάλεια ψηφιακών αρχείων, αυθεντικοποίηση και παροχή αδειών

Η ασφάλεια αφορά στο να καταστεί το αρχείο ψηφιακών τεκμηρίων όσο το δυνατόν περισσότερο ασφαλές από πληθώρα απειλών και προβλημάτων. Πάντως, είναι απαραίτητη μια ισορροπία μεταξύ των απαιτήσεων ασφάλειας και των αναγκών των χρηστών. Για παράδειγμα, εκτεταμένη ασφάλεια θα μπορούσε να αποθαρρύνει πιθανούς χρήστες που δεν έχουν μεγάλη εμπειρία στη χρήση δικτυακών πηγών.

Η παροχή της χρήσης των τεκμηρίων είναι ένας ακόμη παράγοντας που πρέπει να ληφθεί υπόψη, ιδιαίτερα όταν αυτά βρίσκονται υπό καθεστώς copyright. Η νομοθεσία που αφορά τα πνευματικά δικαιώματα στην χώρα μας αποβαίνει σε βάρος των βιβλιοθηκών και τους δημιουργεί προβλήματα ακόμη και στην διάθεσή τους για προσφορά νέου ψηφιακού υλικού. Για το λόγο, αυτό στο ξεκίνημα της δημιουργίας των ψηφιακών συλλογών και μέχρι που ολοκληρώθηκαν οι προσπάθειες απόκτησης των δικαιωμάτων χρήσης των άρθρων από τις εφημερίδες και τα περιοδικά (Δεκέμβριος 2003), η πρακτική που ακολουθήθηκε ήταν αυτή του περιορισμού χρήσης των τεκμηρίων εκτός πανεπιστημίου (IP recognition). Τελικά έγινε ο έλεγχος ώστε η βάση αυτή να μην περιλαμβάνει τεκμήρια από έντυπα που δεν έδωσαν την άδεια χρήσης των άρθρων τους, και από τον Σεπτέμβριο του 2004 η πρόσβαση και στις δύο συλλογές έγινε ελεύθερη για όλους του χρήστες μέσω του δικτυακού τόπου του Συστήματος Βιβλιοθηκών του Α.Π.Θ. «**Μέντωρ**» (<http://web.lib.auth.gr/portal/>) ή πιο συγκεκριμένα από το τόπο <http://cds.lib.auth.gr>.

Όσον αφορά τις άδειες χρήσης, όπως αναφέρθηκε και προηγουμένως, αποφασίστηκε να μην εφαρμοστεί προς το παρόν το μοντέλο της υποβολής των τεκμηρίων από μεμονωμένους χρήστες, αλλά οι άδειες αυτές να ελέγχονται αποκλειστικά και μόνο από τον διαχειριστή του συστήματος. Τέλος, το σύστημα έχει ενσωματώσει όλες τις τελευταίες αναβαθμίσεις (updates) και προστατεύεται από δικό του σύστημα ελεγχόμενης πρόσβασης (firewall).

Ψηφιακό υδατογράφημα

Οι ιδιοκτήτες των ψηφιακών τεκμηρίων συχνά διστάζουν να τα διαθέσουν σε ένα δικτυακό περιβάλλον λόγω του κινδύνου εύκολης αντιγραφής και παράνομης χρήσης τους. Μια λύση που εφαρμόζεται για την αποφυγή τέτοιων προβλημάτων είναι το ψηφιακό υδατογράφημα (watermark). Για να είναι νομικά αποδεκτό ένα ψηφιακό υδατογράφημα πρέπει να μην επιδέχεται παρερμηνεία. Οπτικά υδατογραφήματα μπορεί να περιλαμβάνουν την τοποθέτηση ενός λογοτύπου σε ένα ουδέτερο σημείο της εικόνας, ώστε να καθίσταται γνωστή η καταγωγή της ή την καταγραφή μιας ειδικής ειδοποίησης που εμφανίζεται κατά το κατέβασμα (downloading) του τεκμηρίου. Οι χρήστες μπορούν μεν να αφαιρέσουν αυτό το λογότυπο, αυτό όμως συνεπάγεται την καταπάτηση του copyright. Πάντως το μεγαλύτερο ποσοστό των χρηστών δεν αφαιρεί το υδατογράφημα όταν αυτό τοποθετηθεί με τρόπο ορθολογικό ο οποίος να μην αλλοιώνει την ποιότητα του ψηφιακού τεκμηρίου.

Στο έργο που πραγματοποιείται στο Α.Π.Θ. έχει ήδη αποφασιστεί το λογότυπο που θα χρησιμοποιηθεί ως ψηφιακό υδατογράφημα σε όλα τα ψηφιακά αντικείμενα που καταχωρούνται στη βάση (αφορά τα αρχικά του ιδρύματος, δηλαδή το ακρωνύμιο Α.Π.Θ που θα εκτείνεται διαγωνίως του τεκμηρίου). Αποφασίστηκε να μην δημιουργείται σε κάθε ένα τεκμήριο ξεχωριστά αλλά να εισαχθεί στο τέλος του έργου ταυτόχρονα στο σύνολο των τεκμηρίων.

Προσωπικό - στατιστικά στοιχεία

Για την δημιουργία της βάσης δεδομένων των διδακτορικών διατριβών απασχολούνται 2 βιβλιοθηκονόμοι (οι οποίοι καταλογογραφούν και καταχωρούν στον ηλεκτρονικό κατάλογο της βιβλιοθήκης/Horizon και την έντυπη μορφή τους) και για την δημιουργία της ψηφιοποίησης της συλλογής άρθρων εφημερίδων και περιοδικών απασχολούνται 4 βιβλιοθηκονόμοι. Πρέπει να σημειωθεί ότι αυτοί οι 6 βιβλιοθηκονόμοι αρχικά ήταν 4, αλλά λόγω των αυξημένων απαιτήσεων του όγκου του υλικού και του πλήθους των εργασιών που απαιτούνται, κρίθηκε αναγκαία η ενίσχυσή τους με δύο ακόμη βιβλιοθηκονόμους. Επιπλέον για το σύνολο του έργου της ψηφιοποίησης απασχολείται ένας τεχνικός υποστήριξης, ο οποίος αρχικά συνεργάστηκε με τους βιβλιοθηκονόμους για την επιλογή του λογισμικού και αργότερα με την παραμετροποίηση του λογισμικού και εν τέλει έχει την τεχνική εποπτεία και υποστήριξη των ψηφιακών βάσεων που δημιουργούνται.

Όσον αφορά τις **διδακτορικές διατριβές** και την διαδικασία παραλαβής, επεξεργασίας, μετατροπής τους σε pdf, καταλογογράφησης και καταχώρησής τους στη βάση, έχει υπολογιστεί πως κατά μέσο όρο γίνονται 0,18 καταχωρήσεις διατριβών ανά ανθρωπόωρα.

Όσο δε αφορά την ψηφιοποίηση και την επεξεργασία της συλλογής **άρθρων εφημερίδων και περιοδικών**, σύμφωνα με την έως τώρα ροή εργασίας έχει διαπιστωθεί πως το σκανάρισμα, η επεξεργασία και η αποθήκευση υλικού σε μορφή jpeg και pdf, αποφέρει κατά μέσο όρο 6,5 σκαναρισμένα άρθρα ανά ανθρωπόωρα και 2,25 καταχωρήσεις άρθρων στο CDSoftware ανά ανθρωπόωρα. Από το στοιχείο αυτό γίνεται φανερό πως η διαδικασία της καταλογογράφησης και δημιουργίας των μεταδεδομένων απαιτεί σχεδόν τον τριπλάσιο χρόνο από αυτόν της ψηφιοποίησης του υλικού.

Μελλοντικός σχεδιασμός

Επιπλέον, στο πλαίσιο της δημιουργίας ψηφιακών βάσεων δεδομένων στο Α.Π.Θ. (Γ' ΚΠΣ, Επιχειρησιακό πρόγραμμα "Κοινωνία της πληροφορίας", του Υπουργείου Πολιτισμού, χρονική περίοδος 2004-2005), έχουν προγραμματιστεί να ψηφιοποιηθούν και να καταχωρηθούν στο ίδιο λογισμικό, με τις ανάλογες βέβαια τροποποιήσεις στην παραμετροποίηση του CDSoftware και οι εξής συλλογές:

- 5000 τεκμήρια που αποτελούνται από παλαιύτυπα, φωτογραφίες, γκραβούρες, υδατογραφίες, πίνακες ζωγραφικής, γελοιογραφίες, χαρακτηριστικά (Δωρεά Τρικόγλου)
- Χειρόγραφα νεοελλήνων λογοτεχνών (περίπου 20000 επιστολές)

Γενικότερος στόχος, είναι η παραγωγή και η διάθεση ψηφιακού υλικού στην ελληνική γλώσσα, ιδιαίτερα δε όσον αφορά στον τομέα των ανθρωπιστικών επιστημών στον οποίο μπορεί να υπάρξει ευρύ πεδίο δράσης και συμμετοχής ιδιαίτερα με υλικό που προέρχεται από ελληνικά εκπαιδευτικά ιδρύματα. Καθήκον των ελληνικών εκπαιδευτικών ιδρυμάτων είναι η παραγωγή

υλικού για την ισότιμη συμμετοχή της ελληνικής γλώσσας στις πηγές του διαδικτύου, καθώς επίσης και η απόκτηση εμπειρίας για την μελλοντική δημιουργία υπηρεσιών ψηφιακής δημοσίευσης μελετών και ερευνών που εκπονούνται στο ίδρυμα.

Βιβλιογραφία

1. CERN Document Server Software, <http://cdsware.cern.ch> [πρόσβαση 25-8-2004]
2. CERN Document Server, <http://cds.cern.ch> [πρόσβαση 25-8-2004]
3. Ellis, Judith (ed.) (2000) *Η διαχείριση των αρχείων*. Ελληνική Αρχαιακή Εταιρεία, Τυπωθήτω
4. Vesely, Martin (2003) “*CERN Document Server Software: document management system for grey literature in networked environment*” [presentation]. GL5, Amsterdam, The Netherlands, December 4-5, <http://cdsware.cern.ch/publications/gl5.pdf> [πρόσβαση 25-8-2004]
5. *Εκσυγχρονισμός του Συστήματος Βιβλιοθηκών του Α.Π.Θ. – Γ’ ΚΠΣ* [Τεχνικό δελτίο] (2000), σσ. 48-57
6. Σίτας, Ανέστης (2002) “Ψηφιοποίηση υλικού και δημιουργία ψηφιακών βάσεων δεδομένων” *Σύγχρονη Βιβλιοθήκη & Υπηρεσίες Πληροφόρησης*, 14, σσ. 19-26