

## ΑΝΑΠΤΥΞΗ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗ ΨΗΦΙΑΚΩΝ ΣΥΛΛΟΓΩΝ ΕΦΗΜΕΡΙΔΩΝ

Ελένη Μαμμά

Υπουργείο Εξωτερικών, Ζαλοκώστα 2–3, 100 27, Αθήνα  
e\_mamma@mfa.gr

**Περίληψη:** Σε μία εποχή ραγδαίων αλλαγών στο χώρο της πληροφόρησης, οι νέες τεχνολογίες εμφανίζονται και δρουν με σκοπό την ανάπτυξη διαδικασιών, όπως η υποστήριξη, αξιοποίηση και διάθεση του περιεχομένου των βιβλιοθηκών και των αρχείων. Η δημιουργία Ψηφιακών Βιβλιοθηκών συνδράμει προς τη συγκεκριμένη κατεύθυνση. Στόχος της ακόλουθης μελέτης είναι να παρουσιάσει τη δομή και τη δημιουργία ψηφιακών βιβλιοθηκών και ιδιαίτερα αυτών με συλλογές εφημερίδων. Κατ' αρχήν, θα πραγματοποιηθεί μία εκτενής αναφορά στη σημαντικότητα του υλικού των εφημερίδων στις υπηρεσίες πληροφόρησης, καθώς και στους λόγους για τους οποίους χρειάζονται οι ψηφιακές βιβλιοθήκες εφημερίδων. Δίνεται η σχέση βιβλιοθηκών και εφημερίδων στις μέρες μας και γνωστοποιούνται τα βασικά στάδια και οι τρόποι υλοποίησης μίας συγκεκριμένης ψηφιακής βιβλιοθήκης. Επιπροσθέτως, παρουσιάζονται μοντέλα ψηφιακών βιβλιοθηκών, τα οποία αποτελούν τρέχουσες εφαρμογές και χρησιμοποιούνται για την ανάπτυξη ψηφιακού περιεχομένου.

Τέλος, αναπτύσσεται εκτενώς ένα διεθνώς αναγνωρισμένο πρόγραμμα για ψηφιακές βιβλιοθήκες εφημερίδων του Πανεπιστημίου της Utah των Ηνωμένων Πολιτειών της Αμερικής. Παρουσιάζεται το συγκεκριμένο πρόγραμμα, οι διαδικασίες δημιουργίας της συγκεκριμένης ψηφιακής βιβλιοθήκης από το Πανεπιστήμιο σε συνεργασία με εμπορικές εταιρίες, οι οποίες παρείχαν τεχνολογική υποστήριξη για την πραγμάτωση αυτού του έργου. Γίνεται αναφορά στο ιστορικό του προγράμματος, το είδος του λογισμικού που χρησιμοποιήθηκε, στα βήματα που ακολουθήθηκαν για τη δημιουργία της ψηφιακής βιβλιοθήκης, στην οργάνωση της ιστοσελίδας του Πανεπιστημίου για να καταστεί δυνατή η on–line πρόσβαση στις ψηφιακές συλλογές των εφημερίδων, καθώς και στο μέλλον του συγκεκριμένου προγράμματος. Ακολούθως, παρατίθενται συμπεράσματα και προβληματισμοί σχετικά με την εφαρμογή του προγράμματος, τα οποία μπορούν να αποτελέσουν τη βάση ενός γόνιμου προβληματισμού για τις υπηρεσίες πληροφόρησης, δημοσίου και ιδιωτικού χαρακτήρα, έτσι ώστε να κατανοήσουν την αναγκαιότητα διαχείρισης, οργάνωσης ή και αναδιοργάνωσης του υλικού τους μέσα από τη δημιουργία ψηφιακών βιβλιοθηκών.

**Λέξεις κλειδιά:** Ψηφιακές βιβλιοθήκες, νέες τεχνολογίες, ψηφιακό περιεχόμενο εφημερίδων

# DEVELOPMENT AND MANAGEMENT OF NEWSPAPERS' DIGITAL COLLECTIONS

Eleni Mamma

*Hellenic Ministry of Foreign Affairs, 2–3 Zalokosta str., 100 27, Athens, Greece  
e\_mamma@mfa.gr*

1.04

**Abstract:** The latest rapid changes in the information environment have brought out the outstanding existence of new technologies, whose main aim is the support, the exploitation and the distribution of libraries and archives content. The creation of digital libraries contributes to this undeniable fact. The main goal of this article is to present the structure and the development of digital libraries and especially those which contain newspapers collections. First of all, the importance of newspapers in the information society will be broadly reported. Moreover, the reasons for the necessity of newspaper digital libraries and the relationships between the libraries and the newspapers will be given. The basic steps and the materialization approaches of a specific newspapers digital library will be pointed out. Furthermore, current applications and usages for the development of digital content will be broadly mentioned.

Finally, an international recognized programme for newspaper digital libraries in the Utah University in USA will be introduced. Some interesting procedures of this programme that will be pointed out are the creation steps of the certain digital library taken from the University in collaboration with computing enterprises, which offered their technological support for the accomplishment of the project. In addition, the history of the programme, the kind of the software that was used, the organization of the University site for the online access to the digital newspaper collections, and the future of the programme will come up. Consequently, some conclusions concerning the application of the programme will be set out, which compose the basis of a fruitful speculation for the information community. These conclusions can lead both the public and the private libraries, to foresee the necessity of the organization and the management of their content through the creation of digital libraries.

**Keywords:** Digital libraries, new technologies, newspapers digital content

## 1. ΕΙΣΑΓΩΓΗ

Οι εποχές σηματοδοτούνται από αλλαγές σε όλους τους τομείς της ζωής. Τα ποικίλα δρώμενα, με κοινωνικό, πολιτικό, οικονομικό, πολιτιστικό και αθλητικό ενδιαφέρον, αποτυπώνονται σε διάφορα μέσα, έντυπα και ψηφιακά, τα οποία έχουν ως σκοπό να παρουσιάσουν το κάθε γεγονός στην καθημερινή του διάσταση και με συγκεκριμένο τρόπο. Ένα τέτοιο μέσο αποτελούν και οι εφημερίδες, οι οποίες συνιστούν μία σημαντική πηγή πληροφοριών για όλα τα τρέχοντα γεγονότα της εποχής τους, αλλά και για την καταγραφή του ιστορικού γίνεσθαι.

Στα πλαίσια της ακόλουθης ανακοίνωσης θα πραγματοποιηθεί η προσπάθεια να κατανοήσουμε τη σημαντικότητα του υλικού των εφημερίδων και την ανάγκη δημιουργίας και χρήσης ψηφιακών βιβλιοθηκών εφημερίδων. Επιπροσθέτως, θα δούμε τη σχέση βιβλιοθηκών και εφημερίδων, θα γνωρίσουμε όλα τα στάδια και τις τεχνολογίες δημιουργίας μίας ψηφιακής βιβλιοθήκης για εφημερίδες, τις τρέχουσες εφαρμογές και χρήσεις ψηφιακών βιβλιοθηκών, μοντέλα ψηφιακών βιβλιοθηκών για εφημερίδες, καθώς και λογισμικά, τα οποία διατίθενται για τη δημιουργία τους.

Τέλος, θα γνωρίσουμε το πρόγραμμα ψηφιακών βιβλιοθηκών για εφημερίδες του πανεπιστημίου της Γιούτα στις Η.Π.Α. Θα προσπαθήσουμε να δούμε και να κατανοήσουμε όλες τις διαδικασίες, που πραγματοποιήθηκαν για να δημιουργηθεί η συγκεκριμένη ψηφιακή βιβλιοθήκη από το πανεπιστήμιο σε συνεργασία με δύο εμπορικές εταιρίες, οι οποίες παρέιχαν τη δοκιμασμένη τεχνολογική τους υποστήριξη για την πραγμάτωση αυτού του έργου. Θα δούμε την ανάπτυξη ενός ιστορικού του προγράμματος, το είδος του λογισμικού που χρησιμοποιήθηκε, τα βήματα που πραγματοποιήθηκαν για τη δημιουργία της ψηφιακής βιβλιοθήκης, το πώς οργανώθηκε η ιστοσελίδα του πανεπιστημίου για να είναι δυνατή η on – line πρόσβαση στις ψηφιακές συλλογές των εφημερίδων, καθώς και το μέλλον του συγκεκριμένου προγράμματος, όπως το βλέπουν οι υπεύθυνοι της όλης αυτής προσπάθειας.

Θα παραθέσουμε συμπεράσματα σχετικά με την ανάπτυξη και την εφαρμογή του συγκεκριμένου προγράμματος και θα θέσουμε προβληματισμούς όσον αφορά την δυνατότητα να κατανοήσουν οι σύγχρονες υπηρεσίες πληροφόρησης, δημοσίου και ιδιωτικού χαρακτήρα, το πόσο σπουδαίο και σημαντικό είναι να δομήσουν και να οργανώσουν το υλικό τους μέσα από τη δημιουργία ψηφιακών βιβλιοθηκών.

### *1.1. Σημαντικότητα εφημερίδων*

Πολύ συχνά η ιστορία των εφημερίδων και οι άνθρωποι, που εργάστηκαν ακούραστα για να παραχθούν, μένουν στη λήθη. Οι εφημερίδες αποτελούν μία πρωταρχικής σημασίας πηγή ιστορικής πληροφόρησης, κομμάτι της πολιτιστικής κληρονομιάς μίας χώρας και συνιστούν σημαντικά εργαλεία για την παραγωγή της γνώσης και την προαγωγή της έρευνας. Χρησιμοποιούνται από ιστορικούς, μελετητές, ερευνητές, ειδικούς και γενικότερα ανθρώπους με ποικίλες δραστηριότητες και αποτελούν ένα τεκμήριο πληροφόρησης με πολλαπλή διάσταση, αφού το περιεχόμενό τους καλύπτει ένα ευρύ φάσμα θεμάτων

και ενδιαφερόντων. Ενδεικτικά αναφέρονται τα κοινωνικά, πολιτικά, οικονομικά, επιστημονικά, τεχνολογικά, ιατρικά, πολιτιστικά και αθλητικά της θέματα. Η πληροφόρηση, που μπορεί να έχει ο κάθε ενδιαφερόμενος χρήστης μέσα από ένα τέτοιο υλικό είναι άμεση, έγκυρη, και καλύπτει πολλά γνωστικά πεδία (Arlitsch, Yarr και Edge 2003).

Πάραυτα, οι εφημερίδες αποτελούν και ένα από τα πιο δύσκολα υλικά έρευνας, τα οποία δεν χρησιμοποιούνται εύκολα από τους ερευνητές. Είναι εργαλεία, τα οποία σπάνια ταξινομούνται και οργανώνονται σε ευρετήρια και καταλόγους και έτσι το περιεχόμενό τους δεν αναζητείται με εύκολο τρόπο (Deegan, Steinvel και King 2002). Επιπροσθέτως, οι εφημερίδες βρίσκονται συνήθως σε ξεπερασμένα μέσα, όπως είναι τα μικροφίλμ, τα οποία τις περισσότερες φορές συνιστούν κακής ποιότητας υλικό για να μπορέσει να διαβαστεί και να παραχθεί από αυτό η επιθυμητή πληροφόρηση. Αν στην καλύτερη των περιπτώσεων κρατηθεί, τότε η πρόσβαση σε αυτό είναι τις περισσότερες φορές ανεπιτυχής λόγω των πολιτικών που ακολουθούν οι διάφοροι οργανισμοί σε σχέση με το υλικό τους.

Έτσι λοιπόν, η χρήση των εφημερίδων περιορίζεται συχνά σε έναν χρήστη τη φορά και σε ένα μέρος και όχι στο ηλεκτρονικό ξεφύλλισμα και στην online αναζήτηση και ανάκτηση πληροφοριών από πολλούς και σε διαφορετικά μέρη χρήστες την ίδια χρονική στιγμή. Κάτω από αυτές τις συνθήκες, η δημιουργία ψηφιακών βιβλιοθηκών για εφημερίδες θεωρείται απώτερος, άμεσος και ιδιαίτερα σημαντικός σκοπός.

## *1.2. Βιβλιοθήκες και εφημερίδες*

Η αξία των εφημερίδων ως ένα κομμάτι ιστορικού, αλλά ταυτόχρονα και ενεργού αρχείου, αναγνωριζόταν πάντοτε από τις βιβλιοθήκες. Στις μέρες μας υπάρχουν εκατομμύρια έντυπου υλικού εφημερίδων σε βιβλιοθήκες σε όλο τον κόσμο. Στο συμβατικό περιβάλλον, οι εφημερίδες παρουσιάζουν ένα πρόβλημα όσον αφορά την συντήρησή τους. Παρατηρείται ότι για πολλές δεκαετίες υπήρχε μεγάλη ανησυχία για τη δυνατότητα επιβίωσης των ιστορικών εφημερίδων, δεδομένου ότι οι περισσότερες από αυτές τυπώνονταν σε όξινο χαρτί. Σπουδαίες βιβλιοθήκες, όπως η Βιβλιοθήκη του Κογκρέσου στις Η.Π.Α. και η Βρετανική Βιβλιοθήκη στο Ηνωμένο Βασίλειο, φωτογραφίζουν σε μορφή μικροφίλμ τις εφημερίδες με απώτερο σκοπό να διατηρήσουν το περιεχόμενό τους (Deegan, Steinvel και King 2002).

Υπάρχει, όμως, πάντα η ανησυχία για την κατάσταση διατήρησης του μικροφίλμ, το οποίο δεν έχει παραχθεί και αποθηκευτεί με βάση διεθνή πρότυπα. Επιπροσθέτως, επισημαίνεται τις περισσότερες φορές η ανυπαρξία δομημένης πληροφορίας, η οποία συνίσταται στην έλλειψη ευρετηρίων φύλλων, καταλόγων άρθρων, καταλόγων συγγραφέων και άλλων παραγόντων, όπως τεχνολογικές προδιαγραφές και εργαλεία. Τελευταίο, αλλά όχι λιγότερο σημαντικό, είναι το γεγονός ότι η απομόνωση του περιεχομένου ενός άρθρου από ένα σύνολο κειμένων, που δεν ενδιαφέρουν τον εκάστοτε χρήστη, αποτελούν ένα δύσκολο βήμα. Έτσι λοιπόν, οι παραπάνω παράγοντες μας οδηγούν ολοένα και περισσότερο στο συμπέρασμα πως η δημιουργία ψηφιακών βιβλιοθηκών για εφημερίδες κρίνεται αναγκαία και απαραίτητη. Γι' αυτό το λόγο, θα επιχειρήσουμε παρακάτω να δούμε

γιατί οδηγούμαστε στη δημιουργία ψηφιακών βιβλιοθηκών και τι πρέπει να γνωρίζουμε γι' αυτές.

### *1.2.1. Ψηφιακή βιβλιοθήκη εφημερίδων*

Δημιουργώντας μία ψηφιακή βιβλιοθήκη για εφημερίδες, έχουμε κατ' αρχήν τη δυνατότητα να διατηρήσουμε την εφημερίδα, αλλά ταυτόχρονα και την πρόσβαση στην πληροφορία μέσω του ψηφιακού αντικειμένου, το οποίο βέβαια είναι η ίδια η εφημερίδα. Καθώς ψηφιοποιούμε τις εφημερίδες και κατόπιν δημιουργούμε ψηφιακές συλλογές με την προσθήκη και των μεταδεδομένων τους, μπορούμε να έχουμε εύκολη, γρήγορη και έγκαιρη πρόσβαση στα θέματα που μας ενδιαφέρουν και άρα και στις πληροφορίες, που επιθυμούμε να εντοπίσουμε. Επιπροσθέτως, επιτυγχάνουμε ν' ανακτήσουμε τις πληροφορίες που μας ενδιαφέρουν μέσα από ένα καλά οργανωμένο ψηφιακό υλικό και με τη βοήθεια ενός καλά λειτουργικού πληροφοριακού συστήματος. Η πρόσβαση δεν αφορά μόνο την πληροφορία, αλλά και το ίδιο το ψηφιακό αντικείμενο με τις μορφές πλήρους κειμένου ή και πολυμέσων (εικόνα, ήχος, βίντεο κ.ά.).

Όλα αυτά συναινούν στο να χρησιμοποιούνται παράλληλα αποτελεσματικοί μέθοδοι έρευνας για την απόκτηση των πληροφοριών, ενώ υπάρχει η δυνατότητα να γίνεται χρήση της ψηφιακής βιβλιοθήκης από πολλούς και διαφορετικούς χρήστες, οι οποίοι μπορούν να έχουν ταυτόχρονη πρόσβαση στην πληροφορία και τη γνώση μέσω της μηχανής αναζήτησης. Έτσι, λοιπόν, ο κάθε χρήστης της ψηφιακής βιβλιοθήκης συλλέγει τις πληροφορίες του με τον πιο εύκολο, γρήγορο και εξ αποστάσεως τρόπο.

Κατανοούμε ότι η δημιουργία ψηφιακών βιβλιοθηκών για εφημερίδες είναι αναγκαία και εξαιρετικά κερδοφόρα, εάν μάλιστα αναλογιστούμε τη δυσκολία άντλησης γνώσης από το υλικό μίας εφημερίδας, όταν αυτό είναι ογκώδες σε αριθμούς και πολυποίκιλο στις βιβλιοθήκες και τα αρχεία μέσα στο συμβατικό περιβάλλον. Επίσης, εντοπίζεται πρόβλημα στη διατήρηση της πρόσβασης στο έντυπο υλικό λόγω της ποιότητας του χαρτιού και της αντοχής του στο πέρασμα των χρόνων, του μεγάλου μεγέθους, της εκτενούς παραγωγής φύλλων ημερησίως και των πολιτικών διάθεσης και πρόσβασης, που ακολουθούν οι διάφορες υπηρεσίες πληροφόρησης. Τέλος, η χρήση μικροφίλμ πολλές φορές δυσχεραίνουν το έργο της ενημέρωσης και το κάνουν περισσότερο χρονοβόρο.

Αναλογιζόμαστε, λοιπόν, την χρησιμότητα της ύπαρξης ψηφιακών βιβλιοθηκών εφημερίδων και συνεπώς είναι σημαντικό να γνωρίζουμε κάποια πράγματα όσον αφορά τη δομή και τις λειτουργίες τους.

### *1.2.2. Τι πρέπει να γνωρίζουμε για τις ψηφιακές βιβλιοθήκες εφημερίδων*

Η δημιουργία μίας ψηφιακής βιβλιοθήκης εφημερίδων συνίσταται σε κάποια στάδια, τα οποία πρέπει ευθύς εξαρχής να γνωρίζουμε. Πρέπει να ξέρουμε τις πηγές του υλικού τις οποίες πρώτον διαθέτουμε και δεύτερον μπορούμε να χρησιμοποιήσουμε. Τέτοιου είδους πηγές άντλησης του υλικού μας αποτελούν η αναλογική μορφή, η πρωτογενής ψηφιακή μορφή, και ενδεχομένως άλλου τύπου μορφώματα. Επιπροσθέτως, πρέπει να έχουμε

κατανοήσει πλήρως όλες τις διαδικασίες και τα βήματα, καθώς και τις τεχνολογίες που χρησιμοποιούνται για τη δημιουργία τους. Τέλος, οφείλουμε να γνωρίζουμε τους τρόπους πρόσβασης στο περιεχόμενο του υλικού της εφημερίδας και να δημιουργήσουμε μία ψηφιακή βιβλιοθήκη, η οποία θα επιτρέπει τη διαλειτουργικότητα δεδομένων και τεχνολογικών πλατφόρμων και θα έχει τη δυνατότητα να συνεργάζεται με άλλες ψηφιακές βιβλιοθήκες με σκοπό την ανταλλαγή γνώσης.

Τα βασικά στάδια δημιουργίας μίας ψηφιακής βιβλιοθήκης συνοψίζονται ακολούθως και το καθένα ξεχωριστά από αυτά έχει τη δική του σημαντικότητα και λειτουργικότητα.

### *1.2.3. Βασικά στάδια δημιουργίας μίας ψηφιακής βιβλιοθήκης εφημερίδων*

Βασικός στόχος είναι η δυνατότητα παροχής πλήρους πρόσβασης σε όλα τα άρθρα της εφημερίδας. Για την επίτευξη αυτού του στόχου, ακολουθείται μία ροή εργασιών, που συνίστανται σε συγκεκριμένα στάδια. Κατά τη διάρκεια του πρώτου σταδίου δημιουργούμε συλλογές εφημερίδων με κριτήρια την “ποσότητα”, αλλά και την “ποιότητα” τους. Όταν λέμε “ποσότητα” εννοούμε τη δημιουργία μιας πλήρους συλλογής σε χρονική έκταση, ενώ όταν μιλάμε για “ποιότητα” εννοούμε την καλή φυσική κατάσταση του εντύπου της εφημερίδας. Ακολουθεί το δεύτερο στάδιο, το οποίο αφορά τη συντήρηση των συλλογών από προβλήματα και φθορές. Στο τρίτο στάδιο έχουμε την ψηφιακή καταγραφή των φύλλων. Σύμφωνα με τη διαδικασία αυτή, καταγράφεται στο σύστημα η πλήρης ταυτότητα των φύλλων της εφημερίδας. Αυτή η ταυτότητα καλύπτει τα εξής ακόλουθα πεδία:

- Τίτλος εφημερίδας
- Ιδιοκτήτης/Εκδότης/Διευθυντής/Αρχισυντάκτης/Συντάκτες
- Λοιποί υπεύθυνοι ύλης του φύλλου της εφημερίδας
- Ημέρα/Μήνας/Έτος του φύλλου
- Αριθμός του φύλλου
- Αριθμός/Διαστάσεις των σελίδων του φύλλου
- Προβλήματα και ελλείψεις σελίδων
- Σημειώσεις/Παρατηρήσεις που αφορούν το φύλλο

Με την ολοκλήρωση του τρίτου σταδίου, οι συλλογές ψηφιοποιούνται μέσα από ειδικό πρόγραμμα σάρωσης εικόνων. Η αναλογική μορφή μετατρέπεται σε ψηφιακή μορφή, ενώ παράλληλα πραγματοποιούνται διαδικασίες ελέγχου ποιότητας της ψηφιοποιημένης εικόνας (Gatos, Gouraros και Mantzaris 2001).

Μετά το στάδιο της ψηφιοποίησης ακολουθεί το στάδιο της καταλογογράφησης των άρθρων με την προσθήκη όλων των απαραίτητων μεταδεδομένων. Ενδεικτικά αναφέρονται ορισμένα μεταδεδομένα που περιγράφουν το άρθρο, όπως η τιτλοφορία, η θεματική κατηγορία, οι συγγραφείς, η περίπτωση λογοκρισίας, η περίληψη, διάφορες παρατηρήσεις κ.ά. (Gatos κ.α. 2000).

Στο επόμενο στάδιο ακολουθεί η επεξεργασία με διάφορες τεχνικές όλων των δεδομένων που προκύπτουν από τις προαναφερθείσες διαδικασίες. Δημιουργούνται αποθετήρια των ψηφιακών αντικειμένων, ευρετήρια φύλλων και άρθρων, βάσεις διαχείρισης δεδομένων

και συστήματα αναζήτησης με απώτερο στόχο τη διάθεση του ψηφιακού υλικού στο Διαδίκτυο (Internet), την πρόσβαση των τελικών χρηστών και την ανάκτηση των δομημένων και οργανωμένων άρθρων, έτσι ώστε το πλήρες κείμενο και τα μεταδεδομένα του να παρέχονται στον τελικό χρήστη έγκαιρα και όσο το δυνατόν με μεγαλύτερη ταχύτητα, ακρίβεια και πληρότητα.

#### 1.2.4. Τρέχουσες εφαρμογές και χρήσεις

Μερικές τρέχουσες εφαρμογές και χρήσεις προγραμμάτων ψηφιακών βιβλιοθηκών για εφημερίδες, οι οποίες κατέχουν μία σημαντική θέση στον χώρο τους είναι:

- Το πρόγραμμα ψηφιακών εφημερίδων που πραγματοποιείται από τη βιβλιοθήκη του Πανεπιστημίου της Γιούτα των Ηνωμένων Πολιτειών της Αμερικής (<http://www.lib.utah.edu/digital/unews/about.html>),
- το πρόγραμμα για ψηφιακές συλλογές και διατήρησης εφημερίδων, το οποίο γίνεται από τον οργανισμό OCLC (<http://www.oclc.org>),
- το πρόγραμμα για ψηφιακές βιβλιοθήκες εφημερίδων “E-PRESERVE”, το οποίο αναπτύχθηκε και λειτουργεί στο Αρχείο του Δημοσιογραφικού Οργανισμού Λαμπράκη (<http://www.e-preserve.gr/e-preserve.pdf>),
- το πιλοτικό πρόγραμμα για ψηφιακές συλλογές εφημερίδων, το οποίο πραγματοποιείται στη Βρετανική Βιβλιοθήκη (<http://www.uk.olivesoftware.com>), και τέλος
- Το μεγαλόπνοο πρόγραμμα για δημιουργία ψηφιακών συλλογών εφημερίδων από τη βιβλιοθήκη του Πανεπιστημίου της Ουάσινγκτον ([http://www.lib.washington.edu/mcnews/digital\\_projects.html](http://www.lib.washington.edu/mcnews/digital_projects.html)).

Επιπροσθέτως, παραθέτουμε αξιόλογα και λειτουργικά λογισμικά με βασικό κριτήριο τα αποτελέσματα της ευρείας χρήσης και εφαρμογής τους, όπως:

- Το λογισμικό CONTENTdm του OCLC για δημιουργία και διατήρηση ψηφιακών συλλογών (<http://www.oclc.org/contentdm/default.htm>),
- το λογισμικό ψηφιακών τεχνολογιών Olive (<http://www.olivesoftware.com>),
- το λογισμικό E-PRESERVE για ψηφιακές βιβλιοθήκες εφημερίδων του Αρχείου του Δημοσιογραφικού Οργανισμού Λαμπράκη (<http://www.e-preserve.gr/e-preserve.pdf>).

## 2. ΠΡΟΓΡΑΜΜΑ ΓΙΑ ΨΗΦΙΑΚΕΣ ΒΙΒΛΙΟΘΗΚΕΣ ΤΟΥ ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΤΗΣ ΓΙΟΥΤΑ (UTAH DIGITAL NEWSPAPERS PROJECT)

### 2.1. Ιστορικό

Η ανάπτυξη του συγκεκριμένου προγράμματος ξεκίνησε κατά τη διάρκεια της δεκαετίας του '80, όταν οι βιβλιοθήκες αναζητούσαν τρόπους για να διατηρήσουν το υλικό των εφημερίδων τους και να κάνουν τα κείμενα των εφημερίδων προσβάσιμα και ανακτήσιμα για τους τελικούς χρήστες. Αυτό το πρόγραμμα συνεχίζεται έως και τις μέρες μας (2006),

περνώντας από διάφορες φάσεις, χρηματοδοτήσεις και διαδικασίες. Περιλαμβάνει ψηφιακές συλλογές εφημερίδων της πολιτείας της Γιούτα της περιόδου 1889–1922. Οι τίτλοι των εφημερίδων ξεπερνούν τους 15 και πραγματοποιήθηκαν διάφορες διεργασίες, έτσι ώστε να συλλεχθούν, να δημιουργηθούν οι συλλογές και να ψηφιοποιηθούν. Η δημιουργία της ψηφιακής βιβλιοθήκης έγινε από την βιβλιοθήκη J. Willard Marriott του Πανεπιστημίου της Γιούτα. Αρχικά, πραγματοποίησε μία έρευνα και στη συνέχεια παρουσίασε ένα πρόγραμμα για ψηφιακές βιβλιοθήκες εφημερίδων. Πολύτιμος αρωγός στην προσπάθειά της ήταν η συνεργασία της με δύο εμπορικές εταιρείες με σκοπό την επίτευξη του έργου της (Herbert και Arlitsch 2003).

Οι δύο εμπορικοί συνεργάτες, οι οποίοι είναι καθιερωμένοι στον χώρο τους σε παγκόσμια κλίμακα, προσέφεραν ο καθένας χωριστά τις δικές του τεχνολογίες. Η εταιρεία iArchives Inc. παρέχει το κόψιμο των άρθρων, την περιγραφή τους και συγκομιδή των μεταδεδομένων, τη διαδικασία της οπτικής αναγνώρισης χαρακτήρων (optical character recognition—OCR), τη διαδικασία του XML tagging και τέλος την επαναληκτρολόγηση της τιτλοφορίας. Η εταιρεία DiMeMa Inc. με τη σειρά της παρέχει το λογισμικό CONTENTdm, το οποίο είναι ανοικτής αρχιτεκτονικής και μπορεί να λειτουργήσει σε περιβάλλοντα Windows, Linux και Unix (OCLC Organization 2006).

Στη συνέχεια θα δούμε τον τρόπο με τον οποίο έγινε η διαδικασία της δημιουργίας της ψηφιακής βιβλιοθήκης των συλλογών των εφημερίδων, ξεκινώντας από την ψηφιοποίηση του υλικού και τα βήματα, που ακολουθήθηκαν από την αρχή έως το τέλος.

## *2.2. Διαδικασία Δημιουργίας Ψηφιακής Βιβλιοθήκης Εφημερίδων του Πανεπιστημίου της Γιούτα*

Η διαδικασία αυτή ξεκίνησε με την ψηφιοποίηση του υλικού των εφημερίδων. Περιέχει επτά υψηλού επιπέδου βήματα, τα οποία διαδραματίζονται μέσα από έναν έλεγχο ποιότητας και ασφάλειας των δεδομένων. Τα βήματα είναι τα ακόλουθα:

### **— Βήμα 1: Σκανάρισμα**

Κατά τη διάρκεια του πρώτου βήματος πραγματοποιείται η δημιουργία της ψηφιακής συλλογής τόσο από το έντυπο υλικό, όσο και από το μικροφίλμ. Παράγονται αρχεία εικόνων στα μορφότυπα TIFF και JPEG μέσα από ψηφιακή ανάλυση τουλάχιστον 300 dpi. Στη συνέχεια οι εικόνες αυτές αντιγράφονται και αρχειοθετούνται ως παραγόμενο προϊόν της συγκεκριμένης διαδικασίας.

### **— Βήμα 2: Κόψιμο άρθρων**

Εδώ πραγματοποιείται η δημιουργία κομμένων άρθρων (article clipping) σε ξεχωριστές εικόνες με σκοπό να ανακτώνται μεμονωμένα από όλη την σελίδα του φύλλου της εφημερίδας.

### **— Βήμα 3: Μεταδεδομένα**

Σε αυτό το βήμα έχουμε την περιγραφή των άρθρων και τη συλλογή των μεταδεδομένων τους.



#### — Βήμα 4: Οπτική Αναγνώριση Χαρακτήρων (OCR)

Το “έξυπνο” λογισμικό της οπτικής αναγνώρισης χαρακτήρων “τρέχει” πάνω στις εικόνες και παράγει πλήρες κείμενο για κάθε άρθρο.

#### — Βήμα 5: Αρχεία PDF

Το παραγόμενο πλήρες κείμενο από τη διαδικασία της οπτικής αναγνώρισης χαρακτήρων συνδυάζεται με την εικόνα και δημιουργείται ένα αρχείο PDF για κάθε σελίδα και άρθρο. Ενεργοποιούνται παράλληλα τα αρχεία των μεταδεδομένων του άρθρου στη γλώσσα σήμανσης XML.

#### — Βήμα 6: Εισαγωγή αρχείων

Τα αρχεία PDF και XML εισάγονται στο λογισμικό και ενεργοποιείται αυτόματα η σύνδεση των σελίδων με τα απομονωμένα άρθρα των εφημερίδων.

#### — Βήμα 7: Διακομιστής ιστοσελίδας (Website Server)

Στο τελευταίο στάδιο πραγματοποιείται η φόρτωση των αρχείων στον διακομιστή του πανεπιστημίου και η σύνδεσή τους με την ιστοσελίδα της βιβλιοθήκης J. Willard Marriott.

Ακολουθεί η οργάνωση της ιστοσελίδας του πανεπιστημίου, έτσι ώστε οι τελικοί χρήστες να έχουν on-line πρόσβαση στις ψηφιακές συλλογές των εφημερίδων.

##### 2.2.1. Οργάνωση ιστοσελίδας για online πρόσβαση στις συλλογές

Τα κείμενα των εφημερίδων που έχουν “φορτωθεί” μέσα στην ιστοσελίδα της βιβλιοθήκης του πανεπιστημίου είναι επαρκώς αναζητήσιμα μέσω λέξεων-κλειδιών, ημερομηνίας, τίτλο και τύπο άρθρου, και θέματος. Όλες οι μέθοδοι αναζήτησης, είτε με βάση τα προαναφερθέντα, είτε με βάση το φυλλομέτρημα (browsing), καταλήγουν σε συγκεκριμένα άρθρα ή σε πλήρεις σελίδες, που συνδέονται και με άλλα άρθρα και σελίδες του ίδιου θέματος. Μία δεύτερη διαθέσιμη αναζήτηση γίνεται μέσω του Acrobat toolbar, επειδή οι εικόνες σε PDF περιέχουν “κρυμμένο” κείμενο.

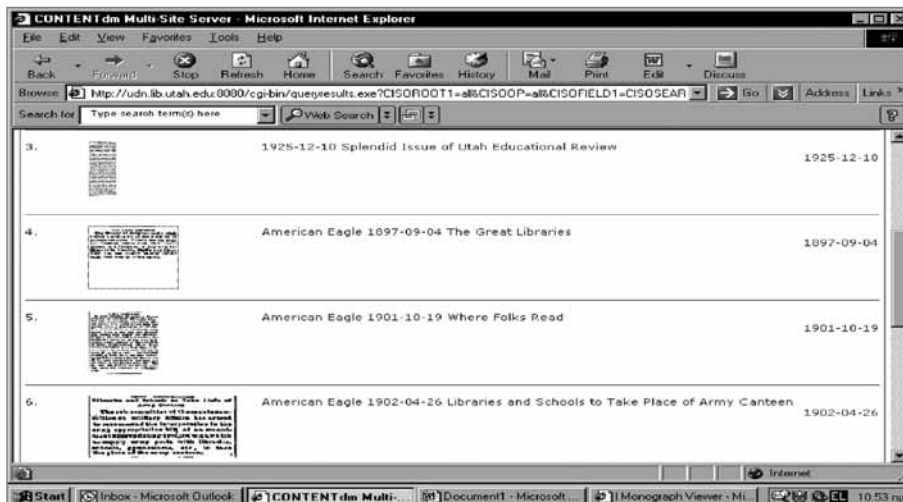
Η κύρια σελίδα προσφέρει μία εισαγωγή για τα χαρακτηριστικά της ιστοσελίδας μαζί με εναλλασσόμενες σελίδες, οι οποίες προκαλούν την περιέργεια του αναγνώστη-χρήστη. Εδώ, ο χρήστης διαθέτει 3 βασικές επιλογές για να δει μία συλλογή.

Η πρώτη επιλογή ενεργοποιεί μία λέξη-κλειδί, η οποία ψάχνει σε ολόκληρη τη συλλογή της εφημερίδας, ενώ υπάρχει και η πρόσθετη επιλογή της αναζήτησης με μία ακριβή φράση (Εικόνα 1).



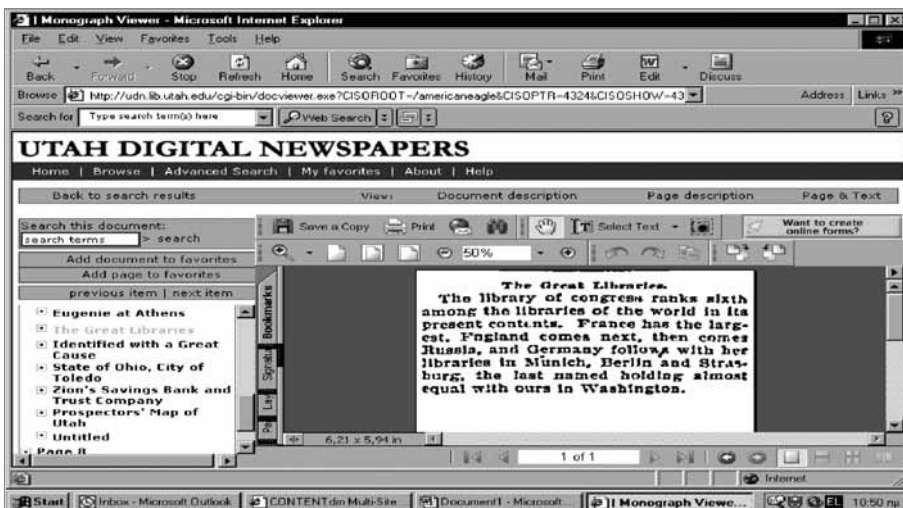
Εικόνα 1: Διαδικασία αναζήτησης άρθρων

Τα αποτελέσματα της αναζήτησης δίνονται σε μία ξεχωριστή σελίδα αποτελεσμάτων αναζήτησης (Εικόνα 2).



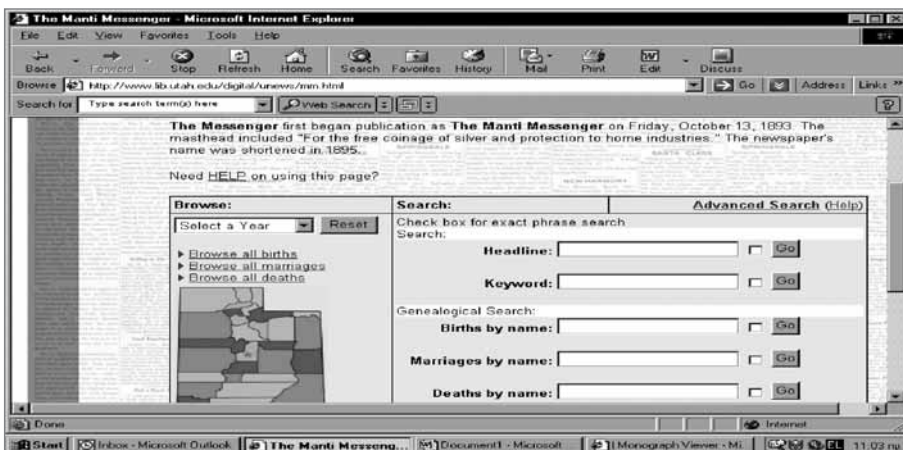
Εικόνα 2: Αποτελέσματα της διαδικασίας της αναζήτησης

Επιλέγοντας ένα συγκεκριμένο άρθρο από τα παραπάνω αποτελέσματα αναζήτησης, ανοίγει ένα παράθυρο, το οποίο δείχνει την εικόνα του κειμένου στην δεξιά πλευρά και παραθέτει ολόκληρο τον πίνακα των περιεχομένων στην αριστερή πλευρά. Ο χρήστης, επιλέγοντας κάτι από τον πίνακα των περιεχομένων μπορεί να βρει εύκολα ένα θέμα, βλέποντας ολόκληρη τη σελίδα ή τα μεμονωμένα άρθρα (Εικόνα 3).



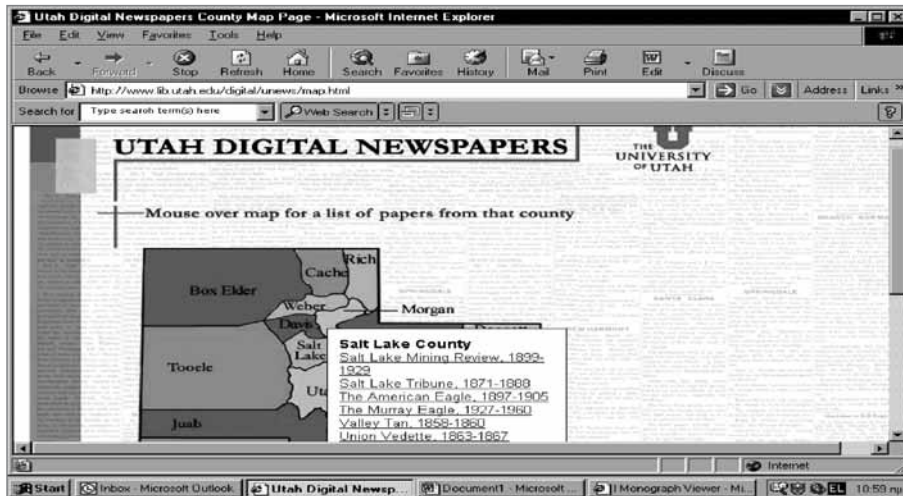
Εικόνα 3: Άρθρα και περιεχόμενα των εφημερίδων

Η δεύτερη επιλογή, που έχει ο χρήστης, είναι να επιλέξει μία συγκεκριμένη εφημερίδα από το μενού των εφημερίδων στην αρχική σελίδα. Διαλέγοντας ένα τίτλο εφημερίδας, εμφανίζεται μία δεύτερη σελίδα όπου εδώ μπορεί να κάνει την αναζήτησή του με δύο τρόπους. Ο ένας είναι να επιλέξει ένα συγκεκριμένο φύλλο και, μέσω του φυλλομετρητή, να ξεφυλλίζει την εφημερίδα, ενώ ο άλλος τρόπος είναι να πάει στο πεδίο “Αναζήτηση” και μέσω λέξεων-κλειδιών, τίτλο άρθρου και άλλων παραμέτρων να βρει τα θέματα, που τον ενδιαφέρουν (Εικόνα 4).



Εικόνα 4: Τρόποι αναζήτησης και εντοπισμού των άρθρων στην εφημερίδα

Η τρίτη επιλογή του χρήστη είναι το ξεφύλλισμα ενός χάρτη της πολιτείας της Γιούτα, ο οποίος δείχνει ποιές επαρχίες έχουν εφημερίδες, που συμπεριλαμβάνονται στο project της Βιβλιοθήκης του Πανεπιστημίου. Πάνω από κάθε επαρχία, εμφανίζονται οι τίτλοι μαζί με τις χρονολογίες των εφημερίδων, που συμπεριλαμβάνονται στη συλλογή (Εικόνα 5).



Εικόνα 5: Χάρτης επαρχιών της Γιούτα με ψηφιοποιημένες συλλογές εφημερίδων

Το σύστημα αναζήτησης και πρόσβασης στις ψηφιακές συλλογές είναι ιδιαίτερα εύχρηστο και φιλικό προς τον χρήστη. Παρέχει τα αποτελέσματα με ταχύτητα και ακρίβεια. Η δυνατότητα πλοήγησης είναι εύκολη, αποτελεσματική και μη χρονοβόρα.

Κατόπιν όλων αυτών, θεωρείται βέβαιο το μέλλον και η περαιτέρω ανάπτυξη και επέκταση του προγράμματος του πανεπιστημίου της Γιούτα, ένα μέλλον που προδιαγράφεται παραγωγικό, λειτουργικό και ευρέως εφαρμόσιμο.

### 2.2.2. Μελλοντικές κατευθύνσεις

Οι βασικοί στόχοι των συνεργαζομένων και συνυπεύθυνων για το πρόγραμμα της ψηφιακής Βιβλιοθήκης του Πανεπιστημίου της Γιούτα είναι υψηλοί, δεδομένης της πολύ καλής λειτουργίας του προγράμματος, της χρηματοδότησης και της συνέχειάς του. Μερικοί από αυτούς τους στόχους του προγράμματος ακολουθούν και είναι οι εξής:

- Επέκταση του αριθμού των συλλογών με την πρόσθεση εκατομμυρίων σελίδων.
- Χρήση νέων τεχνολογιών.
- Συγκομιδή των μεταδεδομένων από τέσσερα διαφορετικά κέντρα, που αφορούν το συγκεκριμένο πρόγραμμα.

- Συνδυασμός των συλλογών για ταυτόχρονες αναζητήσεις σε όλες τις συλλογές, ασχέτως με την τοποθεσία στην οποία βρίσκεται ο διακομιστής με τα δεδομένα των εφημερίδων.
- Εκπαίδευση του προσωπικού ακαδημαϊκών και ιστορικών ιδρυμάτων για την ανάπτυξη προγραμμάτων ψηφιακών βιβλιοθηκών εφημερίδων.

### 2.2.3. Συμπεράσματα

- Το πρόγραμμα χρησιμοποιεί μία αποτελεσματική μέθοδο ψηφιοποίησης του υλικού των εφημερίδων.
- Κάνει χρήση ανοικτών πρωτοκόλλων και πεδίων μεταδεδομένων.
- Το λογισμικό του αποτελεί ένα λογισμικό ευρείας αποδοχής και χρήσης.
- Το σύστημα αναζήτησης και πρόσβασης στις ψηφιακές συλλογές είναι αποτελεσματικό, ταχύ και φιλικό προς το χρήστη.
- Αποτελεί ένα πρόγραμμα ευρέως χρησιμοποιούμενο στις Ηνωμένες Πολιτείες της Αμερικής για τη δημιουργία, οργάνωση και διαχείριση ψηφιακών συλλογών.

Κλείνοντας, θεωρείται αναγκαίο και γόνιμο να τεθούν δύο σημαντικοί προβληματισμοί, οι οποίοι συνοψίζονται στα εξής:

- Κατά πόσο είναι εφικτό οι υπηρεσίες πληροφόρησης, δημοσίου και ιδιωτικού χαρακτήρα, να πειθούν για να δομήσουν και να οργανώσουν το υλικό τους μέσα από τη δημιουργία ψηφιακών βιβλιοθηκών και αρχείων, δεδομένου του υψηλού κόστους του όλου εγχειρήματος και πολλών παραγόντων (προσωπικό, τεχνολογίες, εξοπλισμοί κ.ά.).
- Σε επίπεδο επιλογής μοντέλου δημιουργίας ψηφιακών βιβλιοθηκών με βάση ποια κριτήρια καταλήγουμε σε ένα ήδη υπάρχον ή σε δικό μας μοντέλο ψηφιακής βιβλιοθήκης.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

- Arlitsch, K., L. Yapp και K. Edge 2003. The Utah Digital Newspapers Project. D—Lib Magazine 9, (3), <http://www.dlib.org/dlib/march03/arlitsch/03arlitsch.htm> (πρόσβαση στις 19 Μαΐου 2006).
- Deegan, M., E. Steinvel και E. King 2002. Digitizing historic newspapers: progress and prospects. RLG DigiNews 6, (4), [http://www.rlg.org/preserv/diginews/v6\\_n4\\_feature2.html](http://www.rlg.org/preserv/diginews/v6_n4_feature2.html) (πρόσβαση στις 4 Απριλίου 2006).
- Gatos, B., N. Gouraros και S.L. Mantzaris 2001. An integrated system for creating a Digital Library from Newspaper Archives. Στο Digital Resources for the Humanities (DRH2001) Conference, London, U.K., [http://iit.demokritos.gr/~bgat/DRH\\_v2.pdf](http://iit.demokritos.gr/~bgat/DRH_v2.pdf) (πρόσβαση στις 6 Απριλίου 2006).
- Gatos, B., N. Gouraros, S.L. Mantzaris και P. Tzavelis 2000. A Digital Library from Newspaper Archives. Εργασία στο The Fifth ACM Conference on Digital Libraries (DL2000), San Antonio, Texas, <http://www.lpa.gr/acm/pi.htm> (πρόσβαση στις 10 Απριλίου 2006).
- Herbert, J. και K. Arlitsch 2003. The Digital Newspapers Program at the University of Utah. University of Utah, Marriott Library 47, (1,2), [http://www.lib.utah.edu/digital/unews/serials\\_librarian.html](http://www.lib.utah.edu/digital/unews/serials_librarian.html) (πρόσβαση στις 10 Μαΐου 2006)
- Lambrakis Press Archives x.x. <http://www.lpa.gr> (πρόσβαση στις 7 Ιουνίου 2006).
- OCLC Organization 2006, CONTENTdm Digital Collection Management Software, <http://www.oclc.org/contentdm/default.htm> (πρόσβαση στις 25 Μαΐου 2006).