

Ηλεκτρονική Δημοσίευση Ψηφιοποίηση κειμένου και εικόνων

Σπύρος Βερονίκης
Ηλεκ/γος Μηχ/κός & Μηχ/κός Η/Υ
Δρ. Τμήματος Αρχειονομίας, Βιβλιοθηκονομίας & Μουσειολογίας
spver@ionio.gr

Νοέμβριος 2020

- 1 Εισαγωγή στην ψηφιοποίηση
- 2 Λογική και στρατηγικές έργων ψηφιοποίησης
 - Ψηφιοποίηση για πρόσβαση και διατήρηση
 - Στρατηγικές ψηφιοποίησης και βιωσιμότητα
- 3 Διαδικασία ψηφιοποίησης
 - Βήματα ψηφιοποίησης
 - Γενικές οδηγίες ψηφιοποίησης
- 4 Ψηφιοποίηση αντικειμένων λεκτικού και οπτικού περιεχομένου
 - Τεχνικά θέματα
- 5 Συστάσεις για την ψηφιοποίηση κειμένου και εικόνων

Για σήμερα...

- Ψηφιοποίηση αντικειμένων Βιβλιοθηκών, Αρχείων και Μουσείων
- Κίνητρα και στρατηγικές ανάπτυξης έργων ψηφιοποίησης
- Επισκόπηση διαδικασίας ψηφιοποίησης
- Τεχνικοί παράγοντες που επηρεάζουν την ψηφιοποίηση
- Συστάσεις για δράσεις και έργα ψηφιοποίησης

Ορισμός

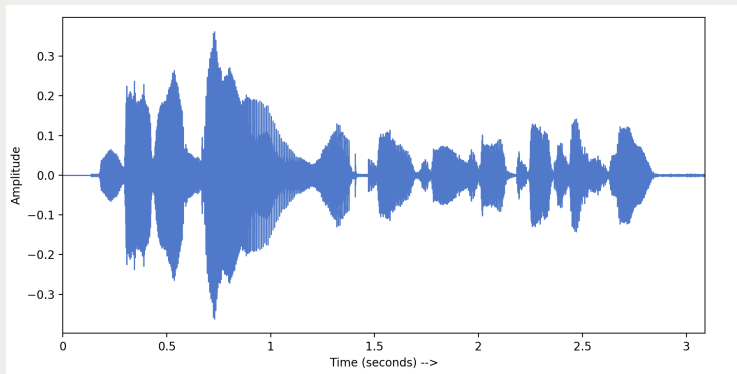
Ψηφιοποίηση είναι η διαδικασία μετατροπής και αποθήκευσης ενός αναλογικού σήματος σε ψηφιακό σήμα το οποίο αναπαριστά το ίδιο περιεχόμενο.

Όμως, τι είναι αναλογικό σήμα;

Ως αναλογικό σήμα νοείται κάθε μορφή συνεχούς, ηλεκτρικού σήματος του οποία τα χρονικά χαρακτηριστικά (π.χ. ένταση, συχνότητα, περίοδος, κ.α.) είναι αναπαραστάσεις των χρονικών χαρακτηριστικών κάποιας άλλης φυσικής ποσότητας ή φαινομένου.

Παραδείγματα αναλογικών σημάτων συνιστούν οι καταγραφές θερμοκρασίας, ηχητικής έντασης, φωτεινότητας, κ.α.

Παράδειγμα ηχητικού σήματος



Πηγή: <https://towardsdatascience.com>

Στο παραπάνω σήμα αποτυπώνεται η ένταση φωνής σε σχέση με τον χρόνο.

Ψηφιοποίηση στις Βιβλιοθήκες, Αρχεία και Μουσεία

Μιλώντας για έργα ψηφιοποίησης στις Βιβλιοθήκες, τα Αρχεία και τα Μουσεία εννοούμε το έργο της δημιουργίας ψηφιακών αναπαραστάσεων, των αντικειμένων πληροφόρησης που διαθέτουν (έντυπο υλικό (βιβλία), ηχογραφήσεις, βιντεοσκοπήσεις, ζωγραφικούς πίνακες, μακέτες, αγάλματα, αγγεία, κοσμήματα κ.α.)

- Το περισσότερο υλικό πληροφόρησης που διαθέτουμε είναι τυπωμένο σε χαρτί. Άλλοι τύποι είναι τα micro-films, slides, audio/video tapes, κ.α.
- Ο κύριος όγκος υλικού που περιλαμβάνεται σε ΨΒ προέρχεται από έργα ψηφιοποίησης (Europeana, Digital Public Library of America, Hathi Trust DL,)
- Χάρη σ' αυτά τα έργα (JSTOR, Google Books Project, Open Content Alliance Initiative, etc) δημιουργήθηκαν οι ΨΒ.

Παράδειγμα ψηφιοποίησης κειμένου

Κατά την διαδικασία αυτή λαμβάνονται στιγμιότυπα απεικόνισης του αντικειμένου και δημιουργούνται κωδικοσειρές συμβόλων 0 και 1 οι οποίες αποτελούν αναπαραστάσεις αυτών των στιγμιτύπων.

Ένα κείμενο απαρτίζεται από σύμβολα ενός αλφαβήτου τα οποία δημιουργούν φθόγγους, λέξεις, προτάσεις, παραγράφους, κλπ.

Για την ψηφιοποίησή του πρέπει πρώτα αυτό το πληροφοριακό περιεχόμενο να περιγραφεί (κωδικοποιηθεί) μ' έναν ψηφιακό τρόπο, δηλαδή χρησιμοποιώντας μόνον δύο σύμβολα: 0 (μηδέν) και 1 (ένα).

Ο πίνακας ASCII ήταν ένας από τους πρώτους τρόπους που χρησιμοποιήθηκαν για την κωδικοποίηση αγγλικών χαρακτήρων. Αποτελεί μια αντιστοίχιση των λατινικών γραμμάτων σε μια σειρά οκτάδας συμβόλων 0 και 1.

Για παράδειγμα:

- Η: αντιστοιχεί σε 0100.1000
- e: αντιστοιχεί σε 0110.0101
- l: αντιστοιχεί σε 0110.1100
- ο: αντιστοιχεί σε 0110.1111

οπότε η λέξη *Hello* αντιστοιχεί στην κωδικοσειρά:

0100.1000 0110.0101 0110.1100 0110.1100 0110.1111

Ψηφιοποίηση του κειμένου είναι η διαδικασία δημιουργίας αυτής της κωδικοσειράς (αναπαράστασης) και της αποθήκευσής της σ' ένα μέσο αποθήκευσης ώστε να είναι διαθέσιμη όποτε χρειαστεί.

Η ψηφιοποίηση δεν πρέπει να συγχέεται με την σάρωση (σκανάρισμα), καθώς είναι έννοια ευρύτερη και αναφέρεται στον τρόπο δημιουργίας μιας ψηφιακής αναπαράστασης ενός αντικειμένου του φυσικού μας κόσμου.

Η σάρωση είναι μια από τις πολλές διαθέσιμες τεχνικές ψηφιοποίησης (η φωτογράφιση με μια ψηφιακή μηχανή είναι μια άλλη εναλλακτική).

Η ψηφιοποίηση μπορεί να εφαρμόζεται σε λεκτικά τεκμήρια, σε εικόνες, σε ηχογραφήσεις, σε φιλμ, σε 3D αντικείμενα, κ.α.

Συχνά, κατά την ψηφιοποίηση πέραν της προσπάθειας αναπαράστασης του περιεχομένου του αντικειμένου, καταβάλλεται προσπάθεια καταγραφής και συνοδών πληροφοριών, όπως των τυχόν ιδιαίτερων χαρακτηριστικών του αντικειμένου.

Η ψηφιοποίηση αποσκοπεί και σε άλλους στόχους πέραν της δημιουργίας μιας ψηφιακής αναπαράστασης ενός αντικειμένου πληροφόρησης.

Συχνά μας ενδιαφέρει:

- η δυνατότητα διατήρησης του περιεχομένου
- η επικοινωνία του περιεχομένου με άλλα άτομα σε διαφορετικό τόπο ή χρόνο από τον δικό μας
- ο χειρισμός του περιεχομένου με τρόπο που δεν είναι δυνατός ενεργώντας στο φυσικό αντικείμενο.

Για παράδειγμα, ψηφιοποιώντας έναν σπάνιο ζωγραφικό πίνακα, μπορούμε :

- να μεταφέρουμε εύκολα την εικόνα του σε άλλα άτομα ανά τον πλανήτη,
- να αποκαταστήσουμε (οπτικά) φθαρμένα τμήματά του,
- να δημιουργήσουμε πολλαπλά αντίγραφα (και παραλλαγές τους)

Στην περίπτωση ενός παλαιού, ασπρόμαυρου κινηματογραφικού φιλμ, μπορούμε να αποκαταστήσουμε τα χρώματά του, αποδίδοντας καλύτερα το περιεχόμενό του.

Σε μια θορυβώδη ηχογράφιση, μπορούμε να απομακρύνουμε θόρυβο από το υπόβαθρο της σκηνης ηχογράφησης, ώστε να ακούγεται καλύτερα και καθαρότερα ο ομιλητής. Παράλληλα, μπορούν να ενσωματωθούν στοιχεία όπως audio scripts για σχολιασμό, μετάφραση, ερμηνεία, κ.α.

Η ψηφιοποίηση ενός κτιρίου μας επιτρέπει να δημιουργήσουμε ένα μοντέλο του κτιρίου και να έχουμε όψεις και θεάσεις του οι οποίες δεν ήταν μέχρι πρότινος εύκολα διαθέσιμες (π.χ. τομές κτιρίου, απεικονίσεις στις οποίες υπερτίθενται άλλες πληροφορίες ή διαγράμματα).

Παραδείγματα αποτελούν οι ψηφιακές αποτυπώσεις εμβληματικών κτιρίων, ιστορικού ενδιαφέροντος όπως ο Παρθενώνας και οι πυραμίδες της Αιγύπτου.

Διαχωρισμός χρονομεταβαλλόμενων και μη αντικειμένων

Κατά την ψηφιοποίηση αντικειμένων πληροφόρησης, συχνά τα διαχωρίζουμε σε δύο μεγάλες κατηγορίες:

- 1** Χρονικά αμετάβλητα: είναι αυτά των οποίων το περιεχόμενο είναι στατικό σε σχέση με τον χρόνο. Π.χ. κείμενα, εικόνες, χάρτες, εφημερίδες, περιοδικά, πίνακες, παρτιτούρες, κ.α.
- 2** Χρονικά μεταβαλλόμενα: είναι αυτά των οποίων το πληροφοριακό περιεχόμενο μεταβάλλεται σε σχέση με τον χρόνο. Π.χ. ηχογραφήσεις, μαγνητοσκοπήσεις και animations

Διαχωρισμός χρονομεταβαλλόμενων και μη αντικειμένων

Η ψηφιοποίηση των χρονικά μεταβαλλόμενων αντικειμένων είναι διαδικασία περισσότερο σύνθετη, καθώς πρέπει να αποτυπωθεί και η χρονική διάσταση της πληροφορίας.

Η ψηφιοποίηση των χρονικά αμετάβλητων αντικειμένων, γίνεται συνήθως με μια διαδικασία φωτογραφικής αποτύπωσης.

Στην περίπτωση της ψηφιοποίησης 3D αντικειμένων, συνήθως λαμβάνεται ένα μεγάλο πλήθος φωτογραφικών αποτυπώσεων από ελαφρώς διαφορετικές γωνίες λήψης και ακολουθεί η συρραφή τους σε μια ενιαία φωτογραφία.

Εξέλιξη της τεχνολογίας ψηφιοποίησης

Χάρη στην έρευνα κατά την δεκαετία του 1980 και την ευρεία διάχυση των προσωπικών Η/Υ, η επόμενη δεκαετία γνώρισε μια έκρηξη έργων ψηφιοποίησης καθώς αναπτύχθηκε ένα μεγάλο πλήθος απαραίτητων εργαλείων (σαρωτές, λογισμικά, κ.α.)

Αυτό ώθησε περισσότερο και την ανάγκη δημιουργίας αποθηκευτικών μέσων μεγαλύτερης χωρητικότητας. Μια συσκευή αποθήκευσης με ικανότητα 40MB το 1970 είχε τις διαστάσεις ενός κομοδίνου και το 1980 ενός κουτιού τσιγάρων. Την δεκαετία του 1990, η ικανότητα αποθήκευσης ενός σκληρού δίσκου έφτανε μερικές εκατοντάδες MB και την επόμενη δεκαετία μερικές εκατοντάδες GB.

Παρομοίως, οι επίπεδοι σαρωτές έγιναν καλύτεροι, με ικανότητα αποτύπωσης ολοένα και μικρότερης επιφάνειας του φυσικού αντικειμένου καταγράφοντας έτσι λεπτομέρεις δυσδιάκριτες στο γυμνό μάτι.

Το πλήθος των χρωμάτων που μπορούσαν να αποτυπωθούν ολοένα και αυξάνονταν.

Παράλληλα, έγιναν ταχύτεροι, μειώνοντας δραστικά τον χρόνο που απαιτείται για την ολοκλήρωση μιας σάρωσης, επιτρέποντας έτσι την ψηφιοποίηση περισσότερου υλικού.

Πλέον, διατίθενται σαρωτές τεχνολογίας laser οι οποίοι μπορούν να χρησιμοποιηθούν για την σάρωση και 3D αντικειμένων.

Ψηφιοποίηση στους φορείς πληροφόρησης

- Φορείς όπως οι βιβλιοθήκες, τα Αρχεία και τα Μουσεία ξεκίνησαν να συμμετέχουν σε έργα ψηφιοποίησης στα μέσα του 1990.
- Ως το τέλος της δεκαετίας, οι μισές βιβλιοθήκες παγκοσμίως συμμετείχαν σε κάποιο έργο ψηφιοποίησης. Μικρότεροι φορείς, ακολούθησαν κατά την επόμενη πενταετία.

Η χρηματοδότηση, το απαραίτητο προσωπικό και η εξειδίκευση παραμένουν ακόμη και σήμερα προκλήσεις που χρήζουν διαχείρισης.

Πλέον φορείς πληροφόρησης αναλαμβάνουν ή συμμετέχουν σε έργα ψηφιοποίησης :

- για να ανταπεξέλθουν στις σύγχρονες ανάγκες συμπεριφοράς αναζήτησης του κοινού που εξυπηρετούν
- για να φέρουν στον ψηφιακό κόσμο τα τεκμήρια που διαθέτουν, ώστε να υποστηρίξουν τις διαδικασίες της μάθησης και έρευνας.

Πλέον, το μη ψηφιοποιημένο υλικό περνά στην αφάνεια και το ψηφιοποιημένο υλικό που δεν έχει αντίκτυπο εγκαταλείπεται.

Περιεχόμενα

- 1 Εισαγωγή στην ψηφιοποίηση
- 2 Λογική και στρατηγικές έργων ψηφιοποίησης
 - Ψηφιοποίηση για πρόσβαση και διατήρηση
 - Στρατηγικές ψηφιοποίησης και βιωσιμότητα
- 3 Διαδικασία ψηφιοποίησης
 - Βήματα ψηφιοποίησης
 - Γενικές οδηγίες ψηφιοποίησης
- 4 Ψηφιοποίηση αντικειμένων λεκτικού και οπτικού περιεχομένου
 - Τεχνικά θέματα
- 5 Συστάσεις για την ψηφιοποίηση κειμένου και εικόνων

Γιατί να ψηφιοποιήσουμε;

Τα δύο βασικά κίνητρα ψηφιοποίησης στους φορείς πληροφόρησης είναι:

- η ενίσχυση της πρόσβασης και προσπέλασης στο διαθέσιμο περιεχόμενο
- η διατήρηση υλικού που είναι πολύτιμο, ευαίσθητο και από υλικό που φθείρεται

Άλλα κίνητρα είναι ο περιορισμός φθοράς από την χρήση του απτών, φυσικών τεκμηρίων, η δυνατότητα εξυπηρέτησης περισσότερων χρηστών ταυτόχρονα και η συγκέντρωση προσοχής γύρω από τον φορέα.

Βελτίωση προσβασιμότητας

Η ψηφιοποίηση βελτίωσε την πρόσβαση των χρηστών στο υλικό πληροφόρησης, καθώς αυτό είναι διαθέσιμο χωρίς χρονικούς ή χωρικούς περιορισμούς. Επιπλέον, διατίθεται περισσότερο υλικό, ιδίως σε άτομα που δεν βρίσκονται κοντά στον φορέα πληροφόρησης.

Το περισσότερο διαθέσιμο υλικό επιτρέπει την έρευνα και τη συνεργασία μεταξύ ατόμων από διαφορετικά πεδία και χώρους, καταρρίπτοντας έτσι στεγανά μεταξύ των διαφορετικών ειδών επιστημών (κοινωνικές, ανθρωπιστικές, θετικές, καλλιτεχνικές, κ.α.)

Πρόσβαση σε ευαίσθητο υλικό

Η ψηφιοποίηση απομάκρυνε περιορισμούς πρόσβασης, όπως η απόσταση του ενδιαφερομένου από τον φορέα πληροφόρησης και η ανάγκη διαφύλαξης πολύτιμων, μοναδικών χειρογράφων/εικόνων/χαρτών.

Επιπλέον, σε ορισμένες περιπτώσεις βελτίωσε την ποιότητα αναπαράστασης περιεχομένου, αποκαθιστώντας (ψηφιακά) τμήματα του αντικειμένου που είχαν καταστραφεί (π.χ. ξεθωριασμένα χρώματα, σκισμένες ή τρύπιες σελίδες, κ.α.)

Έτσι, ερευνητές, εκπαιδευτικοί, μαθητές, το γενικό κοινό, κ.α. αποκτούν πρόσβαση σε περιεχόμενο (πολλές φορές ιστορικής αξίας) το οποίο μέχρι πρότινος ήταν δύσκολα ή καθόλου διαθέσιμο.

Παραδείγματα ψηφιοποίησης ευαίσθητου υλικού

Παραδείγματα ευαίσθητων εγγράφων αποτελούν για παράδειγμα τα ημερολόγια του Νεύτωνα, έγγραφα από την διοίκηση των Ενετών στο Ιστορικό Αρχείο της Βενετίας, η Διακήρυξη της Ανεξαρτησίας των ΗΠΑ, χειρόγραφα και κώδικες αιώνων τα οποία φυλάσσονται σε μοναστήρια, κ.α.

Παραδείγματα συλλογών:

<http://bav.bodleian.ox.ac.uk/>

<http://www.bl.uk/manuscripts/Default.aspx>

Η ανάγκη διατήρησης του διαθέσιμου υλικού πληροφόρησης

Η ψηφιοποίηση, στοχεύει:

- 1** στην διατήρηση του υλικού, ως ένας τρόπος διάσωσης του πολύτιμου και ευαίσθητου υλικού
- 2** ως μια στρατηγική για την διατήρηση και ενίσχυση της διάθεσης του υλικού, από τους φορείς διαχείρισής του.

Στην πρώτη περίπτωση, δημιουργούνται υψηλής πιστότητας αντίγραφα και μοντέλα αναπαράστασης των φυσικών τεκμηρίων, τα οποία μπορούν εφ' εξής να χρησιμοποιούνται. Δίνεται επίσης η δυνατότητα να χρησιμοποιούνται και παράγωγα (ή παραλλαγές αυτών).

Ψηφιοποίηση για την μετάβαση στην ψηφιακή εποχή

Στην δεύτερη περίπτωση, η ψηφιοποίηση είναι αμφιλεγόμενη. Έχει κανείς να αντιμετωπίσει:

- την μεγάλη ποικιλία διαφορετικών μορφότυπων, μέσων και τεχνικών ψηφιοποίησης,
- το συμβιβασμό μεταξύ απαιτούμενων/διαθέσιμων πόρων και διατηρητέας πληροφορίας,
- την ακεραιότητα και διατήρηση αυθεντικότητας του περιεχομένου που ψηφιοποιείται,
- την δυναμική και μεταβλητότητα που χαρακτηρίζει τα ψηφιακά μέσα και πρότυπα αποθήκευσης ψηφιακής πληροφορίας, κ.α.

Πόσο εύκολο είναι σήμερα να δείτε το περιεχόμενο μιας έκθεσης, αποθηκευμένης σε floppy disk?

Εύθραυστα αναλογικά μέσα αποθήκευσης

Το περιεχόμενο των μη ψηφιοποιημένων τεκμηρίων των φορέων πληροφόρησης συχνά αποτυπώνονται σε υλικό το οποίο είναι ευαίσθητο και παρουσιάζει έντονη φθορά στο πέρασμα του χρόνου.

Το χαρτί είναι ιδιαίτερα ανθεκτικό αλλά απαιτεί ιδιαίτερη προσοχή στον χειρισμό του καθώς είναι ευαίσθητο. Επίσης, εικόνες και φιλμς συχνά αποτυπώνονται σε ευαίσθητα και ασταθή μέσα, όπως γυάλινες ή κεραμικές επιφάνειες, φιλμ με βάση την ακετόνη ή την νιτρική κυτταρίνη (cellulose nitrate).

Περιεχόμενα

- 1 Εισαγωγή στην ψηφιοποίηση
- 2 Λογική και στρατηγικές έργων ψηφιοποίησης
 - Ψηφιοποίηση για πρόσβαση και διατήρηση
 - Στρατηγικές ψηφιοποίησης και βιωσιμότητα
- 3 Διαδικασία ψηφιοποίησης
 - Βήματα ψηφιοποίησης
 - Γενικές οδηγίες ψηφιοποίησης
- 4 Ψηφιοποίηση αντικειμένων λεκτικού και οπτικού περιεχομένου
 - Τεχνικά θέματα
- 5 Συστάσεις για την ψηφιοποίηση κειμένου και εικόνων

Η βελτίωση προσβασιμότητας και διατήρηση του υπάρχοντος αναλογικού υλικού είναι τα κύρια κίνητρα και στόχοι μιας στρατηγικής ψηφιοποίησης, αλλά οι φορείς δεν περιορίζονται μόνον σ' αυτά.

Συχνά, έργα ψηφιοποίησης ξεκινούν και για άλλους λόγους, όπως :

- υποστήριξη εκπαιδευτικών προγραμμάτων
- συμμετοχή της τοπικής κοινότητας χρηστών
- κάλυψη των αναγκών των χρηστών, κ.α.

Κάθε έργο ψηφιοποίησης, διαφοροποιείται από κάθε άλλο λόγω των χαρακτηριστικών του υλικού προς ψηφιοποίηση, των διαφορετικών αιτιολογήσεων επιλογής υλικού, τα τεχνικά πρότυπα, το επίπεδο και την ποιότητα σύλληψης πληροφορίας, κ.α.

Σε γενικές γραμμές, η στρατηγική ψηφιοποίησης καθορίζεται:

- βάσει της μορφής και των χαρακτηριστικών του υλικού
- βάσει των στόχων των προγραμμάτων/έργων ψηφιοποίησης
- βάσει της τρέχουσας και εν δυνάμει χρήσης του υλικού
- βάσει του κοινού στο οποίο θα διατεθούν τα αποτελέσματα

Απαιτητική ψηφιοποίηση

Όταν η ψηφιοποίηση στοχεύει κυρίως στην διατήρηση του υλικού, τότε λαμβάνεται μέριμνα ώστε αυτή να γίνει με τον καλύτερο δυνατό τρόπο. Εξαντλούνται περιθώρια και διαθέσιμοι πόροι προκειμένου να επιτευχθεί το καλύτερο δυνατό αποτέλεσμα.

Από τα ψηφιακά αρχεία, μπορούν μεταγενέστερα να δημιουργηθούν παραλλαγές στις οποίες αποτυπώνεται όση πληροφορία χρειάζεται, για τον σκοπό που προορίζονται (π.χ. χαμηλότερης ανάλυσης εικόνες, μικρότερο βάθος χρώματος, συμπίεση πληροφορίας, κ.α.).

Μαζική ψηφιοποίηση

Όταν η ψηφιοποίηση στοχεύει στην βελτίωση προσβασιμότητας στο διαθέσιμο υλικό, κατά κανόνα επιλέγεται χαμηλότερο επίπεδο ποιότητας, το οποίο εξοικονομεί πόρους και χρόνο, δίνοντας τη δυνατότητα να ψηφιοποιηθεί περισσότερο υλικό (π.χ. Google Books Project).

Έτσι, διαμορφώνεται ένα **ισοζύγιο μεταξύ ποιότητας και ποσότητας** στο οποίο οι έχοντες την ευθύνη διοίκησης του έργου ψηφιοποίησης πρέπει να ισορροπήσουν.

Βιωσιμότητα ενός έργου ψηφιοποίησης

Ένα έργο ψηφιοποίησης είναι έργο επίπονο, χρονοβόρο και δαπανηρό. Προκειμένου να ολοκληρωθεί και να μην εγκαταλειφθεί θα πρέπει να έχει απτά οφέλη για τον φορέα που το αναλαμβάνει. Η ψηφιοποίηση πρέπει

- να ενσωματώνεται και να αξιοποιείται από τις παρεχόμενες υπηρεσίες του φορέα
- να εστιάζει στους κύριους στόχους της αποστολής του φορέα
- να βασίζεται σε συστηματικές και όχι ευκαιριακές χρηματικές εισροές
- να περιλαμβάνει ένα πλάνο για την μακροχρόνια διατήρηση του ψηφιοποιημένου υλικού

Η ψηφιοποίηση είναι μια διαδικασία η οποία εκτελείται σε διάφορα σταδιακά βήματα. Δεν αφορά μόνον την σάρωση της αναλογικής πληροφορίας αλλά περιλαμβάνει και την επεξεργασία και την τεκμηρίωση (περιγραφή) ώστε να μπορούν αυτά να αξιοποιηθούν με τον καλύτερο δυνατό τρόπο.

Περιεχόμενα

- 1 Εισαγωγή στην ψηφιοποίηση
- 2 Λογική και στρατηγικές έργων ψηφιοποίησης
 - Ψηφιοποίηση για πρόσβαση και διατήρηση
 - Στρατηγικές ψηφιοποίησης και βιωσιμότητα
- 3 Διαδικασία ψηφιοποίησης
 - Βήματα ψηφιοποίησης
 - Γενικές οδηγίες ψηφιοποίησης
- 4 Ψηφιοποίηση αντικειμένων λεκτικού και οπτικού περιεχομένου
 - Τεχνικά θέματα
- 5 Συστάσεις για την ψηφιοποίηση κειμένου και εικόνων

Βήματα ψηφιοποίησης

Ένα έργο ψηφιοποίησης διαχωρίζεται σε φάσεις, και σε κάθε φάση λαμβάνουν χώρα ορισμένες ενέργειες. Τυπικά, οι φάσεις είναι οι εξής:

- 1 Προγραμματισμός έργου και επιλογή υλικού προς ψηφιοποίηση
- 2 Μετατροπή αναλογικού σήματος σε ψηφιακό
- 3 Επεξεργασία ψηφιακού σήματος (για βελτίωση) και δημιουργία παράγωγων αρχείων
- 4 Καταγραφή μεταδεδομένων ψηφιοποιημένου αντικειμένου
- 5 Ενσωμάτωση των ψηφιακών αντικειμένων και των μεταδεδομένων τους σε μια ΨΒ
- 6 Διατήρηση των ψηφιοποιημένων αντικειμένων και των περιγραφών τους

Προγραμματισμός έργου και επιλογή υλικού προς ψηφιοποίηση

- Είναι η πρώτη και πιο σημαντική φάση του έργου ψηφιοποίησης.
- Καθορίζεται ο σκοπός του έργου,
- οι στόχοι του,
- ο προϋπολογισμός,
- το χρονοδιάγραμμα,
- το προσωπικό,
- οι ρόλοι και αρμοδιότητες κάθε εμπλεκόμενου

Επιπλέον, γίνονται οι προετοιμασίες του υλικού, καθορίζονται οι τεχνικές προδιαγραφές και επιλέγεται ο εξοπλισμός ψηφιοποίησης που θα χρησιμοποιηθεί.

Παράγοντες που πρέπει να λαμβάνονται υπόψη είναι :

- τυχόν περιορισμοί πνευματικής ιδιοκτησίας
- η μορφή, το είδος και το μέγεθος/έκταση των τεκμηρίων
- η τρέχουσα κατάστασή τους και τα τυχόν μοναδικά τους χαρακτηριστικά
- οι απαιτήσεις χειρισμού των ευαίσθητων τεκμηρίων

Μετατροπή αναλογικού σήματος σε ψηφιακό

Η δεύτερη φάση είναι αυτή της σύλληψης πληροφορίας και συχνά γίνεται λαμβάνοντας μία ή περισσότερες φωτογραφίες του τεκμηρίου.

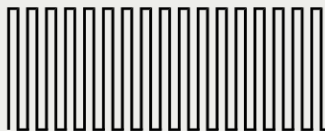
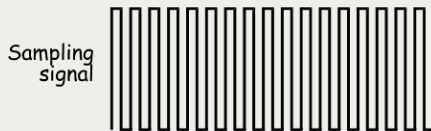
Η πληροφορία που αποτυπώνεται στο αναλογικό τεκμήριο δειγματοληπτείται (sampled) με συσκευές όπως σαρωτές, ψηφιακές κάμερες ή άλλοι μετατροπείς αναλογικού-σε-ψηφιακό σήμα.

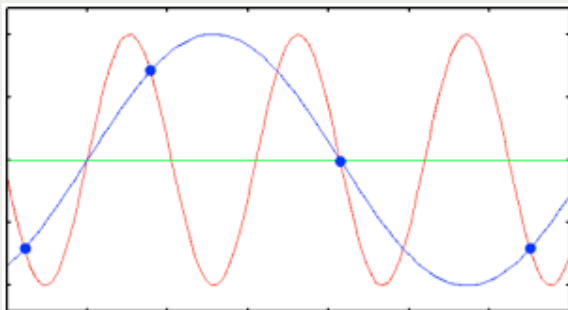
Σε απαιτητικά έργα, δημιουργούνται τα master-files που θα χρησιμοποιηθούν για την διατήρηση του υλικού: 2 αντίγραφα. Το ένα για φύλαξη και το άλλο για χρήση (θα δημιουργηθούν άλλες εκδοχές αναπαραστάσεων).

Δειγματοληψία αναλογικού σήματος

Είναι αδύνατον να καταγράψουμε όλη την πληροφορία, γι' αυτό και λαμβάνουμε διαδοχικά δείγματα αυτής. **Κατά μία έννοια, απορρίπτουμε πληροφορία!**

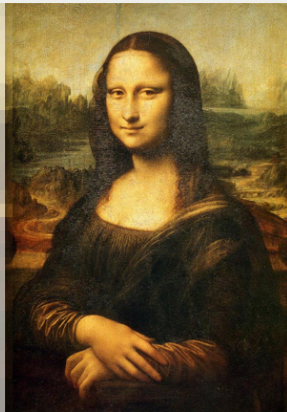
Ωστόσο, διατηρούμε τόση όση χρειάζεται για να έχουμε μια ικανοποιητική αναπαράσταση (αναλόγως των απαιτήσεων ψηφιοποίησης)





Συχνότητα δειγματοληψίας: ορίζει πόσα δείγματα λαμβάνονται σε κάθε δευτερόλεπτο. Πολύ λίγα δείγματα μπορούν να είναι παραπλανητικά.

Παράδειγμα μη επαρκούς δειγματοληψίας



Επεξεργασία ψηφιακού σήματος

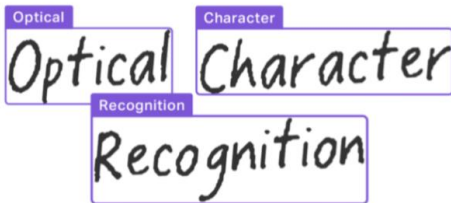
Σ' αυτή τη φάση δημιουργούνται τα αρχεία στα οποία αποθηκεύεται το ψηφιακό σήμα και λαμβάνει χώρα η επεξεργασία βελτίωσής του (βελτίωση ποιότητας, πρόσθετες λειτουργίες, κ.α.).

Στα master-files αποθηκεύεται η ανεπεξέργαστη πληροφορία (raw data).

Η βελτίωση/διόρθωση μπορεί να αφορά σε απομάκρυνση σημάτων ξένων προς το περιεχόμενο (θορυβώδη σήματα), βελτίωση τονικών αποχρώσεων, αντίθεσης, εστίασης, συγχρονισμό σημάτων ήχου/βίντεο, κ.α.

Δημιουργία παράγωγων αρχείων

Συχνά δημιουργούνται και παράγωγα αρχεία, όπως μεταγραφές (scripts). Για παράδειγμα, απί την ψηφιοποίηση ενός κειμένου, κατά την επεξεργασία OCR αναπαράγεται το κείμενο σε επεξεργάσιμη μορφή.



Δημιουργία μεταδεδομένων

Αυτή η φάση αφορά:

- στην επιλογή του ενδεδειγμένου σχήματος μεταδεδομένων
- στην διαμόρφωση/δημιουργία καρτέλας μεταδεδομένων (metadata template)
- στην επιλογή λεξικού όρων για την καταχώρηση τιμών στα πεδία μεταδεδομένων
- στη συγκέντρωση οδηγών και καλών πρακτικών τεκμηρίωσης
- στη δημιουργία περιγραφικής τεκμηρίωσης για κάθε αντικείμενο που ψηφιοποιείται

Κατά την τεκμηρίωση, πέραν της καταγραφής περιγραφικών δεδομένων, καταχωρούνται και δομικά, διαχειριστικά και τεχνικά μεταδεδομένα. Στα απαιτητικά έργα, ακολουθεί μια επισκόπηση (review) της φάσης τεκμηρίωσης.

Παράδειγμα μεταδεδομένων εικόνας

Metadata fields

View and configure collection and administrative fields.

Collection field properties

View, add, edit and delete fields. Enable full text searching and controlled vocabulary. After you have added, changed, or deleted fields, index the collection to update changes.

	Field name	DC map	Data type	Large	Search	Hide	Required	Vocab		add field
1	Title	Title	Text	No	Yes	No	Yes	No	move to ▾	edit delete
2	Subject	Subject	Text	No	Yes	No	No	No	move to ▾	edit delete
3	Description	Description	Text	Yes	Yes	No	No	No	move to ▾	edit delete
4	Creator	Creator	Text	No	Yes	No	No	No	move to ▾	edit delete
5	Publisher	Publisher	Text	No	Yes	No	No	No	move to ▾	edit delete
6	Date	Date	Date	No	Yes	No	No	No	move to ▾	edit delete
7	Type	Type	Text	No	Yes	No	No	No	move to ▾	edit delete
8	Format	Format	Text	No	Yes	No	No	No	move to ▾	edit delete
9	Contributors	Contributors	Text	No	No	No	No	No	move to ▾	edit delete
10	Identifier	Identifier	Text	No	No	No	No	No	move to ▾	edit delete
11	Source	Source	Text	No	No	No	No	No	move to ▾	edit delete
12	Language	Language	Text	No	No	No	No	No	move to ▾	edit delete
13	Relation	Relation	Text	No	No	No	No	No	move to ▾	edit delete
14	Coverage	Coverage	Text	No	No	No	No	No	move to ▾	edit delete
15	Rights	Rights	Text	No	No	No	No	No	move to ▾	edit delete
16	Audience	Audience	Text	No	No	No	No	No	move to ▾	edit delete
17	Transcript	None	Full Text Search	Yes	Yes	No	No	No	move to ▾	edit delete
	Field name	DC map	Data type	Large	Search	Hide	Required	Vocab		add field

Ενσωμάτωση στην ΨΒ

Αυτή είναι η φάση που ολοκληρώνει την ψηφιοποίηση. Τα ψηφιοποιημένα αντικείμενα και οι περιγραφές τους καταχωρούνται και συσχετίζονται σε μια πλατφόρμα ΨΒ ώστε να μπορούν να ευρετηριαστούν και να είναι διαθέσιμα για μελλοντική χρήση.

Υπάρχουν ελεύθερα αλλά και εμπορικά διαθέσιμες πλατφόρμες ΨΒ για την καταχώρηση των ψηφιοποιημένων αντικειμένων, με διαφορετικές διαχειριστικές δυνατότητες και ευελιξία παραμετροποίησης (π.χ. Collective Open Access, Greenstone, Omeka, DSpace, CONTENTdm, LUNA, ...)

Μέσω της πλατφόρμας μπορούν να ορίζονται δικαιώματα πρόσβασης, σημειώματα περιγραφής των όρων χρήσης (περί πνευματικής ιδιοκτησίας), κ.α.

Διατήρηση ψηφιοποιημένων αρχείων

Η φάση αυτή περιλαμβάνει τις απαιτούμενες ενέργειες για την μακροχρόνια φύλαξη και διατήρηση προσπελασιμότητας των ψηφιοποιημένων αρχείων.

Είναι σημαντικό να θυμόμαστε ότι πληροφορία η οποία απορρίπτεται κατά την ψηφιοποίηση είναι συχνά μη ανακτήσιμη, οπότε τα master-files είναι αυτά τα οποία έχει πλέον κανείς και πρέπει να διατηρεί.

Διατήρηση ψηφιοποιημένων αρχείων

Για την διαφύλαξη του υλικού πρέπει να δημιουργηθεί μια *Πολιτική Μακροχρόνιας Διατήρησης* και να δημιουργηθεί η κατάλληλη υποδομή και στρατηγική που θα εξασφαλίζει:

- την ταυτοποίηση
- την αρχειακή αποθήκευση
- τον έλεγχο της ακεραιότητας και αυθεντικότητας
- την τακτική λήψη αντιγράφων ασφαλείας
- την ανανέωση
- την μεταφορά των αρχείων σε νέες δομές και μορφότυπους (migration)
- την χρηματοδότηση των ανωτέρω

Περιεχόμενα

- 1 Εισαγωγή στην ψηφιοποίηση
- 2 Λογική και στρατηγικές έργων ψηφιοποίησης
 - Ψηφιοποίηση για πρόσβαση και διατήρηση
 - Στρατηγικές ψηφιοποίησης και βιωσιμότητα
- 3 Διαδικασία ψηφιοποίησης
 - Βήματα ψηφιοποίησης
 - Γενικές οδηγίες ψηφιοποίησης
- 4 Ψηφιοποίηση αντικειμένων λεκτικού και οπτικού περιεχομένου
 - Τεχνικά θέματα
- 5 Συστάσεις για την ψηφιοποίηση κειμένου και εικόνων

Γενικές οδηγίες ψηφιοποίησης

Η εμπειρία πολλών ετών έργων και προγραμμάτων ψηφιοποίησης οδήγησε στην δημιουργία οδηγών ψηφιοποίησης, ώστε:

- να διασφαλίζεται η υψηλή ποιότητα και η δυνατότητα διατήρησης των αρχείων ψηφιοποίησης
- να υποστηρίζεται η τρέχουσα και μελλοντική χρήση αυτών
- να ενισχύεται η διαλειτουργικότητα και ανταλλαγή ψηφιοποιημένων δεδομένων/πληροφοριών μεταξύ φορέων

Παραδείγματα αποτελούν:

- *Framework of Guidance for Building Good Digital Collections* (NISO)
- *Moving Theory into Practice* (Cornell University)
- *CDL Guidelines for Digital Images* (California university Digital Library), κ.α

Ελάχιστες απαιτήσεις ψηφιοποίησης

- 1** Ψηφιοποίηση στην μεγαλύτερη δυνατή ανάλυση που ενδείκνυται για το συγκεκριμένο αντικείμενο
- 2** Δημιουργία και διατήρηση πηγαίων αρχείων master-files που μπορούν να χρησιμοποιηθούν για ασφάλεια και δημιουργία παράγωγων αρχείων
- 3** Δημιουργία ψηφιακών αρχείων τα οποία είναι σε μορφή που μπορεί να ανταλλάσσεται με άλλους φορείς και εξοπλισμό
- 4** Διασφάλιση συστηματικής ποιότητας ψηφιοποιήσεων

Ελάχιστες απαιτήσεις ψηφιοποίησης

- 5 Ψηφιοποίηση σε επαρκή επίπεδα ποιότητας ώστε να αποφεύγονται οι επαναλήψεις ψηφιοποίησης των ίδιων αντικειμένων για την αποτύπωση περισσότερης πληροφορίας
- 6 Ψηφιοποίηση του ίδιου του πρωτότυπου αντικειμένου (όποτε είναι δυνατόν) και όχι αντιγράφων του (που ίσως υστερούν σε ποιότητα/πληροφορία)
- 7 Διάθεση αρχειακού αποθηκευτικού χώρου & υποδομών και διαχείριση εργασιών διατήρησης

Digital-master files (πηγαία αρχεία)

Κατά την ψηφιοποίηση αντικειμένων δημιουργούνται:

- τα (*digital*) *master-files* (or *raw-data files*)
- παράγωγα αρχεία (*derivative files*)

Ο σκοπός χρήσης τους είναι διαφορετικός και διαφέρουν και σε άλλα στοιχεία όπως το μέγεθος, η έκταση, ο βαθμός συμπίεσης ψηφιακής πληροφορίας, ο μορφότυπος, κ.α.

Στα πηγαία αρχεία καταγράφεται η πληροφορία των φυσικών αντικειμένων, ακατέργαστη και με την καλύτερη δυνατή ποιότητα.



Iconic image of Lena Soderberg (ca. 1980)

Η αρχική εικόνα (master file) μπορεί να χρησιμοποιηθεί για να δημιουργηθούν παραλλαγές της, όπως οι παραπάνω. Ένας κοινός μορφότυπος εικόνων για μακρά διατήρηση είναι ο TIFF.

Συνήθη γνωρίσματα των πηγαίων αρχείων

Για την δημιουργία των πηγαίων αρχείων (master-files) συχνά εφαρμόζονται οι εξής συστάσεις :

- δεδομένων των διαθέσιμων πόρων, γίνεται ψηφιοποίηση στην καλύτερη δυνατή ποιότητα
- αποτυπώνεται η μέγιστη δυνατή ποσότητα πληροφορίας, ώστε να μην χρειαστεί να επαναληφθεί η ψηφιοποίηση.
- η πληροφορία που φέρουν είναι ασυμπιεστη, έτσι ώστε να μην απορρίπτονται δεδομένα

Συνήθη γνωρίσματα των πηγαίων αρχείων

- χρησιμοποιούνται καλά τεκμηριωμένοι μορφότυποι αρχείων, μη εμπορικοί (patented) ώστε να είναι δυνατή η ανοικτή πρόσβαση στο περιεχόμενό τους.
- δεν αποθηκεύονται ως αρχείο διατήρησης τυχόν τροποποιήσεις που έγιναν κατόπιν της ψηφιοποίησης
- αποδίδεται στο αρχείο μια ονομασία βάσει ενός καθιερωμένου σχήματος ονοματοδοσίας.

Παράγωγα αρχεία

Τα *παράγωγα αρχεία* (*derivative files*) δημιουργούνται από τα πηγαία αρχεία και χρησιμοποιούνται για να εξυπηρετήσουν ιδιαίτερους σκοπούς, αναλόγως των αναγκών που εξυπηρετούν (π.χ. παρουσίαση, εκτύπωση, ενσωμάτωση σε πολυμεσικό περιεχόμενο, κ.α.)

- φέρουν λιγότερη πληροφορία απ' ότι τα πηγαία.
- δεν δημιουργούνται κατ' ανάγκη κατά τον χρόνο ψηφιοποίησης. Μπορούν να δημιουργούνται και μεταγενέστερα, κατ' απαίτηση (ad-hoc)

Παράγωγα αρχεία

Για τη δημιουργία τους, ενδείκνυται:

- να διατηρείται μικρό το μέγεθός τους, ώστε να μεταφέρεται γρήγορα η πληροφορία που περιέχουν
- όταν αναπαριστούν οπτικό περιεχόμενο, να αποθηκεύονται σε τυποποιημένους, συμπιεσμένους μορφότυπους όπως JPEG
- οι μορφότυποι που επιλέγονται να υποστηρίζονται από τους σύγχρονους φυλλομετρητές.

Ενδειγμένοι μορφώτυποι αρχείων ψηφιοποίησης

Ανάλογα με το είδος του φυσικού μέσου στο οποίο εμπεριέχεται η πληροφορία προς ψηφιοποίηση, επιλέγεται ο κατάλληλος μορφώτυπος για τα πηγαία αρχεία και τα παράγωγα αρχεία. Για παράδειγμα:

- **Κείμενα:** *master files:* TIFF, *derivatives:* JPEG, PDF
- **Φωτογραφικό υλικό:** *master files:* TIFF, *derivatives:* JPEG, JPEG2000
- **Ηχογραφήσεις:** *master files:* WAV/BWF, *derivatives:* MP3
- **Μαγνητοσκοπήσεις:** *master files:* JPEG2000/MXF, *derivatives:* MPEG-4

Γενικά, επιλέγονται μορφώτυποι οι οποίοι να υποστηρίζονται από όλα τα είδη υπολογιστών (platform-independent και να είναι ανοικτοί (non-patented).

Συστάσεις για την ονοματοδοσία αρχείων ψηφιοποίησης

Η ονοματοδοσία των αρχείων ψηφιοποίησης καθορίζεται πριν την έναρξη της διαδικασίας.

Είναι σημαντικό να γίνεται σωστά, ώστε να μπορούν τα αρχεία να αναγνωρίζονται εύκολα. Επιπλέον, διασφαλίζεται διαλειτουργικότητα μεταξύ διαφορετικών συστημάτων.

Δεν υπάρχει λόγος και δεν ενδείκνυται να καταγράφεται πολλή πληροφορία στην ονομασία των αρχείων. **Έχουμε τα μεταδεδομένα γι' αυτό!**

Η επιλογή του τρόπου ονοματοδοσίας πρέπει να γίνεται έτσι ώστε να εξυπηρετεί τους σκοπούς του έργου.

Γενικές συστάσεις είναι οι εξής :

- να αποδίδονται μοναδικά ονόματα σε κάθε αρχείο ψηφιοποίησης
- να χρησιμοποιούνται αλφαριθμητικοί χαρακτήρας (λατινικά γράμματα και οι αριθμοί 0-9)
- να περιλαμβάνεται ο αναγνωριστικός κωδικός του φορέα (όταν υπάρχει)
- να αριθμούνται σειριακά τα αρχεία που ψηφιοποιούνται (με ίδιο πλήθος ψηφίων)
- να χρησιμοποιούνται έγκυρες επεκτάσεις αρχείων (δηλωτικές του μορφότυπου)
- να περιορίζεται η ονομασία του αρχείου στους 31 χαρακτήρες (μαζί με τα γράμματα μορφότυπου)

Για παράδειγμα : IU-01-IMAGE00001.jpg

Ψηφιοποίηση αντικειμένων λεκτικού και οπτικού περιεχομένου

Η πληροφορία η οποία δεν μεταβάλλεται χρονικά, χαρακτηρίζεται ως *στατική*. Τέτοιου είδους πληροφορία συναντάμε σε χειρόγραφα και άλλα κείμενα καθώς και σε εικόνες (π.χ. φωτογραφίες, slides, ζωγραφικούς πίνακες, κ.α).

Για την ψηφιοποίηση της στατικής πληροφορίας συχνά χρησιμοποιούνται σαρωτές εικόνων και ψηφιακές φωτογραφικές μηχανές. Αυτές οι συσκευές, δημιουργούν ψηφιδωτά αρχεία (raster files or **bitmaps**). Κάθε τέτοιο αρχείο είναι μια ορθογώνια διάταξη από ψηφίδες (pixels) και σε κάθε ψηφίδα αντιστοιχείται μια ακέραια τιμή ή ένα σύνολο ακεραίων τιμών, ενδεικτικών της χρωματικής έντασης (φωτεινότητας) της αντίστοιχης περιοχής.

Παραδείγματα ψηφιδωτών αρχείων



Για τις ασπρόμαυρες εικόνες η ακέραια τιμή αντιστοιχείται σε μια τονική απόχρωση του γκρι. Συνήθως το μηδέν αντιστοιχεί στο μαύρο, ο μεγαλύτερος ακέραιος στο άσπρο και οι ενδιάμεσοι αντιστοιχούνται στους γκρι τόνους.

Στις έγχρωμες εικόνες, συνήθως αποθηκεύονται 3 ακέραιοι για κάθε pixel, ένας για κάθε βασικό χρώμα της παλέτας χρωμάτων (colorspace) που χρησιμοποιείται.

Χαρακτηριστικά ψηφιδωτών αρχείων

Τα ψηφιδωτά αρχεία εικόνων (raster files) φέρουν ένα πλήθος χαρακτηριστικών, όπως :

- το πλήθος χρωματικών συνιστωσών που αποτυπώνονται στην εικόνα (π.χ. binary/grayscale, 3-channel color)
- το πλήθος χρωμάτων που αποτυπώνονται (color depth)
- το πλήθος ψηφίδων που χρησιμοποιούνται κατά πλάτος και κατά ύψος της εικόνας (resolution)
- ο χώρος αποθήκευσης που καταλαμβάνει το ψηφιακό αρχείο (αποτιμάται σε kB, MB,)

Πέραν αυτών, τα ψηφιακά αρχεία μπορούν να συνοδεύονται και από άλλες πληροφορίες (μεταδεδομένα) όπως την τοποθεσία λήψης της φωτογραφίας, την ημερομηνία λήψης, το είδος τυχόν συμπίεσης, τα χαρακτηριστικά της συσκευής λήψης της φωτογραφίας, κ.α.

Περιεχόμενα

- 1 Εισαγωγή στην ψηφιοποίηση
- 2 Λογική και στρατηγικές έργων ψηφιοποίησης
 - Ψηφιοποίηση για πρόσβαση και διατήρηση
 - Στρατηγικές ψηφιοποίησης και βιωσιμότητα
- 3 Διαδικασία ψηφιοποίησης
 - Βήματα ψηφιοποίησης
 - Γενικές οδηγίες ψηφιοποίησης
- 4 Ψηφιοποίηση αντικειμένων λεκτικού και οπτικού περιεχομένου
 - Τεχνικά θέματα
- 5 Συστάσεις για την ψηφιοποίηση κειμένου και εικόνων

Τεχνικά χαρακτηριστικά ψηφιοποιημένων εικόνων

Η ποιότητα σύλληψης και απόδοσης πληροφορίας μιας ψηφιοποιημένης εικόνας εξαρτάται από ένα πλήθος τεχνικών χαρακτηριστικών αυτής όπως :

- το πλήθος των χρωματικών συνιστωσών (colorspace)
- το βάθος χρώματος
- η ανάλυση σάρωσης/φωτογράφισης
- η συμπίεση δεδομένων
- ο μορφότυπος αποθήκευσης του ψηφιακού αρχείου

Αυτοί οι παράγοντες πρέπει να εξετάζονται σε σχέση με τον σκοπό ψηφιοποίησης και να δίνεται ιδιαίτερη προσοχή στις περιπτώσεις δημιουργίας των πηγαίων αρχείων (master-files).

Ανάλυση ψηφιοποιημένης εικόνας (resolution)

Η ανάλυση τις εικόνας καθορίζει το πλήθος δειγμάτων που λήφθηκαν κατά την διαδικασία της ψηφιοποίησης.

Προκειμένου να αποτυπωθεί η απεικονιζόμενη πληροφορία, η επιφάνεια του αντικειμένου διαχωρίζεται σε μικρές τετράγωνα περιοχές (pixels) και για κάθε μία εξ' αυτών αποτυπώνεται η χρωματική ένταση.

Όσο περισσότερες οι περιοχές (πλήθος δειγμάτων), τόσο μεγαλύτερη η ανάλυση ή η διακριτική ικανότητα.

Κατά τη διαδικασία της ψηφιοποίησης, πριν τη λήψη του στιγμιότυπου, καθορίζεται το πλήθος των περιοχών (ψηφίδων) ανά τετραγωνική ίντσα που πρόκειται να καταγραφούν. Αποτιμάται σε **Dots-Per-Inch** ή **Pixels-Per-Inch**.

Παράδειγμα λήψης εικόνας με χαμηλή και υψηλή ανάλυση



Σε μια χαμηλής ανάλυσης εικόνας, το μέγεθος της περιοχής δειγματοληψίας είναι μεγάλο και γίνεται εύκολα αντιληπτό από το μάτι. Έτσι, οι ακμές δεν αποδίδονται σωστά, παρά ως σύνολα κλιμακωτών ευθυγράμμων τμημάτων.

Resolution trade-offs

Όσο μεγαλύτερη η ανάλυση της ψηφιοποιημένης εικόνας, τόσο περισσότερη πρωτογενής πληροφορία καταγράφεται.

Μικρές εικόνες οδηγούν σε μικρό πλήθος ψηφίδων και αντιστρόφως. Ωστόσο, η αναλογία διατηρείται. Ωστόσο, οι μεγάλες αναλύσεις απαιτούν

περισσότερο χρόνο για την σύλληψη πληροφορίας και περισσότερο χώρο στο ψηφιακό μέσο αποθήκευσης.

Η επεξεργασία υψηλής ανάλυσης εικόνων, προϋποθέτει και την διάθεση εξοπλισμού υψηλών προδιαγραφών.

Η διακίνηση μεγάλων ψηφιοποιημένων αρχείων δεσμεύει πολύτιμους πόρους (π.χ. εύρος) και μειώνει την ταχύτητα μετάδοσης.

Κατά κανόνα, οι υψηλές αναλύσεις επιλέγονται για τη δημιουργία των πηγαίων αρχείων.

Για την ψηφιοποίηση κειμένων, συχνά η ελάχιστη πυκνότητα ψηφίδων που χρησιμοποιείται είναι 300DPI (περίπου 17×17px).

Εάν πρόκειται να ακολουθήσει διαδικασία αυτοματοποιημένης μεταγραφής του κειμένου (π.χ. OCR) τότε επιλέγονται ακόμη υψηλότερες αναλύσεις (π.χ. 600 DPI).

Για την σάρωση εικόνων διαστάσεων περίπου 20cm × 25cm συνήθως επιλέγεται ανάλυση πυκνότητας 600 DPI.

Αν και η αύξηση της πυκνότητας δειγμάτων οδηγεί στην καταγραφή περισσότερης πληροφορίας, υπάρχει ένα όριο πέραν του οποίου το κόστος υπερβαίνει το όφελος.

Πλήθος τονικών αποχρώσεων (color depth)

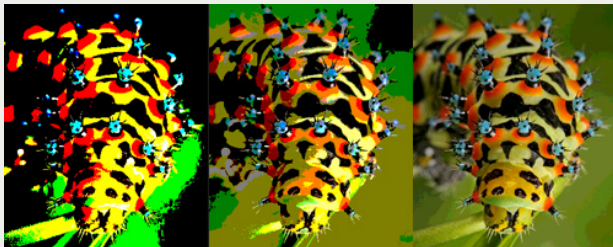
Το πλήθος των χρωμάτων που καταγράφονται κατά την ψηφιοποίηση μιας εικόνας επιδρά σημαντικά στην ποιότητα αναπαράστασης της ψηφιοποιημένης πληροφορίας.

Μπορούμε να θεωρήσουμε ότι στον φυσικό κόσμο υπάρχουν άπειρες τονικές αποχρώσεις. Ωστόσο, το ανθρώπινο μάτι μπορεί να διακρίνει μόνον ένα πεπερασμένο πλήθος τονικών διαβαθμίσεων του ορατού φάσματος φωτός.

Το πλήθος χρωμάτων που θα καταγραφούν επιλέγεται πριν την έναρξη της ψηφιοποίησης. Μπορεί να είναι μόλις δύο χρώματα (άσπρο/μαύρο), τονικές αποχρώσεις του γκρι ή μερικά εκατομμύρια έγχρωμων αποχρώσεων.

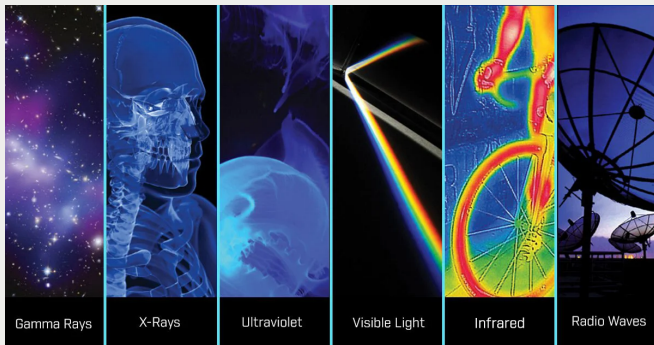
Συνήθως, το πλήθος των χρωμάτων επιλέγεται να είναι πολλαπλάσιο των αριθμητικών δυνάμεων του 2, δηλαδή 2^N όπου $N = 1, 2, \dots$

Παράδειγμα σύλληψης με διαφορετικό πλήθος χρωμάτων



Παράδειγμα αναπαράστασης χρησιμοποιώντας 8, 32, και 256 διαφορετικά χρώματα. Όσο αυξάνεται το πλήθος των χρωμάτων που δύνανται να αποτυπωθούν, τόσο πιο αληθοφανής γίνεται η αναπαράσταση της ψηφιοποιημένης πληροφορίας.

Οι συμβατικοί scanners συλλαμβάνουν μόνον τις αντανάκλασεις του ορατού φωτός. Για την αντανάκλαση μη ορατού φωτός (π.χ. υπεριώδη/υπέρυθρη) απαιτείται εξειδικευμένος εξοπλισμός, ο οποίος σπανίως χρησιμοποιείται.



Τέτοιος εξοπλισμός μπορεί να συλλάβει και να παρέχει πληροφορία σχετικά με στοιχεία τα οποία δεν γίνονται άμεσα αντιληπτά στο μάτι. Για παράδειγμα, στους ζωγραφικούς πίνακες μπορεί να αποκαλύψει στρώσεις μπογιάς που έχουν επικαλυφθεί μεταγενέστερα ή ακόμη και την χημική σύσταση των χρωμάτων που χρησιμοποιήθηκαν.

TIPs:

- οι αισθητήρες των σύγχρονων ψηφιακών φωτογραφικών μηχανών μπορούν να συλλάβουν το υπέρυθρο φως που ανακλάται από τα διάφορα αντικείμενα. Ωστόσο, συχνά φέρουν ενσωματωμένο φίλτρο για την απόρριψη αυτής της πληροφορίας, πριν καν αυτή να φτάσει στον αισθητήρα.
- μια συσκευή ακτινογραφίας μπορεί επίσης να θεωρηθεί ως ένας σαρωτής!

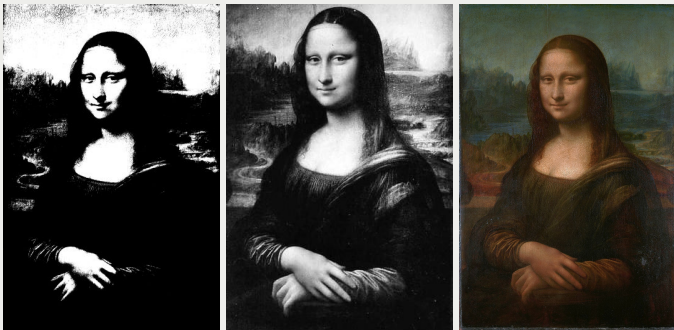
Color mode

Το πλήθος των χρωματικών συνιστωσών (color mode) καθορίζει το πλήθος των βασικών χρωμάτων που καταγράφονται στο αρχείο ψηφιοποίησης.

Αυτό το πλήθος χρωμάτων (συνήθως 1 ή 3) αντιστοιχεί στο πλήθος των χρωματικών συνιστωσών στις οποίες αναλύεται κάθε απόχρωση ορατού φωτός που αντανακλάται από το φυσικό (αναλογικό) αντικείμενο.

Έτσι, για την ασπρόμαυρη απόδοση/αναπαράσταση ενός αντικειμένου καταγράφεται 1 μόνον συνιστώσα, η οποία μπορεί να λαμβάνει τονικές αποχρώσεις της ασπρόμαυρης κλίμακας.

Εναλλακτικά, μπορεί να καταγράφονται 3 βασικές χρωματικές συνιστώσες (π.χ. Red-Green-Blue) οι οποίες συνδυαζόμενες σε κατάλληλη ένταση να δημιουργούν το αρχικό χρώμα. Το πλήθος των τονικών αποχρώσεων κάθε βασικής συνιστώσας, καθορίζεται από το *βάθος χρώματος*.



Τρεις διαφορετικοί τρόποι χρωματικής απόδοσης του ίδιου αντικειμένου (bitonal, grayscale, colored modes).

Συμπίεση

Συμπίεση είναι η διαδικασία απόρριψης πληροφορίας που καταγράφηκε κατά την ψηφιοποίηση, σε βαθμό όμως τέτοιο που η εναπομείνουσα πληροφορία να επαρκεί για την ικανοποιητική αναπαράσταση της πληροφορίας που καταγράφηκε.

Στις περισσότερες περιπτώσεις, η απόρριψη πληροφορίας είναι μια διαδικασία μη-αναστρέψιμη (*lossy compression*). Ωστόσο, υπάρχουν και περιπτώσεις κατά τις οποίες η διαδικασία συμπίεσης μπορεί να αντιστραφεί και να ανακτηθεί η αρχικώς καταγεγραμμένη πληροφορία (*lossless compression*).



Η συμπίεση επιτυγχάνεται απορρίπτοντας πληροφορία (π.χ. πλήθος τονικών αποχρώσεων) και σ' αυτή την περίπτωση αυτή η πληροφορία είναι αδύνατον να ανακτηθεί.

Loseless compression

Για να επιτευχθεί συμπίεση χωρίς απώλεια πληροφορίας, πρέπει η πληροφορία που απορρίπτεται να είναι πλεονάζουσα.

Για παράδειγμα, μπορεί να εφαρμοστεί σε περιπτώσεις όπου το παρασκήνιο μιας εικόνας είναι σαφώς περιγεγραμμένο και έχει σε όλη του την έκταση το ίδιο ακριβώς χρώμα.

Εικόνες οι οποίες αποθηκεύονται κατά το πρότυπο TIFF και JPEG 2000 υποστηρίζουν αυτό το είδος συμπίεσης και μπορούν να χρησιμοποιηθούν για την δημιουργία των πηγαίων αρχείων που θα εξυπηρετούν σκοπούς δημιουργίας παραγώγων αρχείων (και όχι αποθήκευσης).

Μορφότυποι αρχείων ψηφιοποίησης

Οι μορφότυποι αποθήκευσης (*formats*) συνιστούν έναν τρόπο κωδικοποίησης και οργάνωσης της ψηφιακής πληροφορίας που παράγεται κατά την διαδικασία της ψηφιοποίησης. Στην συντριπτική πλειοψηφία των περιπτώσεων ψηφιοποίησης, το αρχείο που παράγεται είναι μια ψηφιδωτή εικόνα.

Για την αποθήκευση της πληροφορίας που περιέχεται στην εικόνα έχουν προταθεί διάφοροι μορφότυποι, καθ' ένας με τα ισχυρά και αδύναμα σημεία του, αναλόγως του σκοπού δημιουργίας του.

Οι συχνότερα χρησιμοποιούμενοι μορφότυποι αποθήκευσης ψηφιοποιημένων εικόνων είναι οι εξής: TIFF, JPEG, JPEG 2000, PDF, PDF/A, PNG

TIFF format

Ο μορφότυπος TIFF (Tagged Image File Format) είναι ευρέως διαδεδομένος και συχνά χρησιμοποιούμενος για την αποθήκευση των πηγαίων αρχείων.

- Μπορεί να χρησιμοποιηθεί με όλους τους χρωματικούς χώρους
- Υποστηρίζει συμπίεση με ή χωρίς απώλεια πληροφορίας
- συνδυάζει πληροφορία από ψηφιδωτές εικόνες και ένα σύνολο επισημειώσεων (tags)
- είναι μορφότυπος ανοικτός και καλά τεκμηριωμένος

JPEG file format

Ο μορφότυπος JPEG (Joint Photographic Experts Group) σχεδιάστηκε στοχεύοντας στην συμπίεση.

- Δημοσιεύθηκε το 1992 και είναι χρήσιμος για εφαρμογή όπου ο διαθέσιμος χώρος είναι μικρός και υπάρχει ανάγκη για πολλαπλές ψηφιοποιήσεις (π.χ. φορητές φωτογραφικές μηχανές).
- Χρησιμοποιείται κυρίως για τη δημιουργία των παράγωγων αρχείων
- Ο βαθμός συμπίεσης μπορεί να επιλέγεται από τον χρήστη (τυπικά 10:1)
- Υποστηρίζεται από όλους τους φυλλομετρητές.

JPEG 2000 file format

Το JPEG 2000 δημιουργήθηκε από το Joint Photographic Experts Group προς αντικατάσταση του παλαιότερου JPEG.

- Δημιουργήθηκε το 2000 και χρησιμοποιεί αποδοτικότερο αλγόριθμο συμπίεσης
- Υποστηρίζει συμπίεση με απώλεια και χωρίς απώλεια πληροφορίας
- Μπορεί να χειρίζεται εύκολα μεγάλα αρχεία, πλούσια σε περιεχόμενο, με δυνατότητα μεγέθυνσης και μετατόπισης. Γι' αυτό χρησιμοποιείται συχνά σε μεγάλες εικόνες χαρτών, εφημερίδων, σχεδίων, κ.α. μεγάλης έκτασης εικόνων
- Δεν υποστηρίζεται εγγενώς από τους σύγχρονους φυλλομετρητές, παρά απαιτεί την ενσωμάτωση ενός προγράμματος προβολής (viewer).

PDF file format

Ο μορφότυπος PDF (Portable Document Format) δημιουργήθηκε το 1993 και χρησιμοποιείται κυρίως στις περιπτώσεις εγγράφων.

- Είναι ανοικτός μορφότυπος από το 2008
- Διατηρεί σταθερή δομή (fixed layout) της καταγεγραμμένης πληροφορίας
- Χρησιμοποιείται κυρίως για τα παράγωγα αρχεία (π.χ. OCR)

PDF/A file format

Ο μορφότυπος PDF/A επεκτείνει τις δυνατότητες του PDF και σχεδιάστηκε έχοντας κατά νου την μακροχρόνια αποθήκευση και διατήρηση αρχείων.

- Καθιερώθηκε ως ο μορφότυπος για την αρχειακή αποθήκευση
- ενσωματώνει ένα μηχανισμό απεικόνισης του περιεχομένου και της διάρθρωσής του (layout) ανεξάρτητο από τα εργαλεία και το σύστημα που χρησιμοποιήθηκε για την δημιουργία του και την προβολή του
- όλες οι γραμματοσειρές και τα μεταδεδομένα ενσωματώνονται στο αρχείο ώστε να είναι διαθέσιμα όποτε χρειαστούν.

PNG file format

Ο μορφότυπος PNG (Portable Network Graphics) δημιουργήθηκε για την αντικατάσταση του παλαιότερου μορφότυπου GIF.

- Υποστηρίζει την αποθήκευση μονόχρωμων και έγχρωμων εικόνων
- είναι ανοικτός μορφότυπος
- χρησιμοποιείται ευρέως στο διαδίκτυο
- υποστηρίζει διαφάνεια χρωμάτων
- χρησιμοποιείται για τα παράγωγα αρχεία

Συνοψίζοντας ...

- Οι μορφότυποι κωδικοποίησης εικόνων χρησιμοποιούνται ήδη για μεγάλο χρονικό διάστημα και έχουν πλέον "ωριμάσει" αρκετά.
- Δεν συμβαίνει το ίδιο και με τους μορφότυπους αποθήκευσης πληροφορίας σε video (time-changing content) όπου υπάρχει μια συνεχής και έντονη εξέλιξη, σε μια προσπάθεια αντιμετώπισης των νέων προκλήσεων που εγείρονται από τις τεχνολογικές εξελίξεις.

Συστάσεις για την ψηφιοποίηση κειμένου και εικόνων


Είτε πρόκειται για ψηφιοποίηση εγγράφων είτε για ψηφιοποίηση εικόνων, η διαδικασία είναι παρόμοια. Το αποτέλεσμα είναι συνήθως η δημιουργία μιας ψηφιδωτής, ψηφιακής εικόνας στην οποία αποτυπώνεται η εικόνα του φυσικού (απτού) τεκμηρίου.

Ωστόσο, τεχνικές παράμετροι όπως η ανάλυση σάρωσης και το βάθος χρώματος διαφοροποιούνται σημαντικά.

Επιπλέον, τα κείμενα συχνά υπόκεινται σε περαιτέρω επεξεργασία ώστε η εικόνα κειμένου που καταγράφεται να μετατραπεί σε σύνολο λέξεων, προτάσεων, παραγράφων, κλπ στο οποίο μπορεί κανείς να αναζητήσει.

Ψηφιοποίηση κειμένων

Κατά την ψηφιοποίηση κειμένων δεν αρκεί μόνον η λήψη ενός στιγμιότυπου της εικόνας τους αλλά είναι επιθυμητό αυτή η εικόνα να μετατραπεί σε κείμενο το οποίο να μπορεί να αναγνωρισθεί και επεξεργαστεί από κατάλληλο λογισμικό.

Έτσι, κάθε εικόνα συμβόλου πρέπει να αποκτήσει και μια σημασιολογική σημασία και ταυτότητα. Για παράδειγμα, το  δεν θα είναι απλώς ένας κύκλος αλλά θα είναι ένα *ό μικρον*, το οποίο:

- μπορεί να είναι ένα άρθρο
- μπορεί να συνδυάζεται με ένα "i" για να δημιουργήσουν ένα σύνθετο φωνήεν
- μπορεί να βρίσκεται στο τέλος μιας λέξης, δηλώνοντας ότι αυτή αναφέρεται σε ένα ουδέτερο αντικείμενο, κλπ.

Ανάλογα με τα χαρακτηριστικά του έντυπου εγγράφου, καθορίζονται και οι διάφορες παράμετροι ψηφιοποίησης.

- σε έγγραφα με μικρό μέγεθος χαρακτήρων, επιλέγεται υψηλή ανάλυση ψηφιοποίησης (πολλαπλά δείγματα)
- σε έγγραφα στα οποία χρησιμοποιούνται μελάνια διαφορετικών χρωμάτων, επιλέγεται η έγχρωμη ψηφιοποίηση (και όπου απαιτείται) με υψηλό πλήθος τονικών αποχρώσεων

Άλλα χαρακτηριστικά των έντυπων τεκμηρίων που επηρεάζουν τις παραμέτρους ψηφιοποίησης είναι η τρέχουσα κατάστασή τους, εάν πρόκειται για χειρόγραφα ή τυπωμένα από μηχανή κείμενα, εάν περιέχουν εικόνες (ασπρόμαυρες/έγχρωμες) ή όχι, κ.α.

Προτεινόμενες τιμές παραμέτρων για βιβλία

Για την ψηφιοποίηση βιβλίων, αναλόγως του τύπου τους προτείνονται συνήθως τα εξής (ως ελάχιστες απαιτήσεις):

- Λεκτικά κείμενα χωρίς φωτογραφίες: ανάλυση 300DPI, βάθος χρώματος 8bit, μονόχρωμη σάρωση (grayscale)
- Λεκτικά κείμενα με φωτογραφίες: ανάλυση 400DPI, βάθος χρώματος 8 ή 24bit, μονόχρωμη ή έγχρωμη σάρωση
- Σπάνια βιβλία: ανάλυση 400DPI, βάθος χρώματος 24bit, έγχρωμη σάρωση
- Χειρόγραφα: ανάλυση 400DPI, βάθος χρώματος 24bit, έγχρωμη σάρωση.

Στις περιπτώσεις που πρόκειται να ακολουθήσει και η επεξεργασία του κειμένου σε μορφή αναγνωρίσιμη από λογισμικά, τότε η ανάλυση σάρωσης μπορεί να αυξάνεται έως και τα 400DPI. Ωστόσο, το πλήθος χρωμάτων σπανίως χρειάζεται να είναι μεγαλύτερο από 2 έως 16 διαφορετικών τονικών αποχρώσεων.

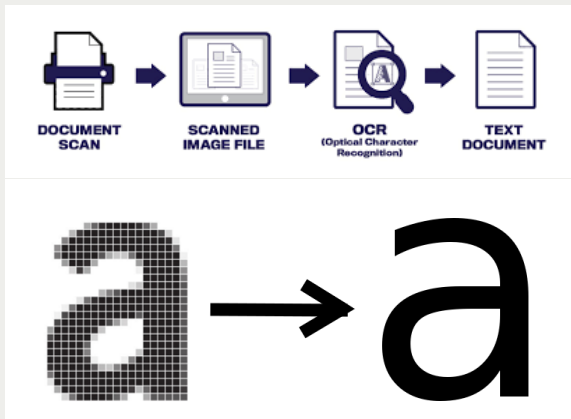
Μετατροπή εικόνων σε κείμενο (OCR)

Η διαδικασία μετατροπής μιας εικόνας συμβόλων ενός αλφαβήτου σε εννοιολογικά σύμβολα (αναγνώσιμο και επεξεργάσιμο κείμενο) ονομάζεται *Οπτική Αναγνώριση Χαρακτήρων* (Optical Character Recognition - OCR)

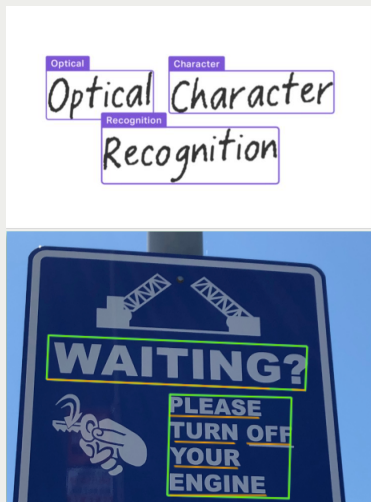
Χάρη σ' αυτή, από την εικόνα ενός κειμένου ένα λογισμικό μπορεί να αναγνωρίσει τα σύμβολα που χρησιμοποιούνται και να εκτελέσει μια σειρά ελέγχων επί αυτών. Για παράδειγμα, το λογισμικό:

- μπορεί να ελέγξει την ορθογραφία της κάθε λέξης
- μπορεί να αναζητήσει συνώνυμα της κάθε λέξης
- μπορεί να αναζητήσει στο διαδίκτυο (ή άλλη πηγή) για άλλα κείμενα με αναφορά σε κάποια από τις λέξεις, κ.α.

Η διαδικασία ψηφιοποίησης ενός κειμένου και μετατροπής του σε αναγνώσιμη μορφή περιγράφεται αφαιρετικά στο παρακάτω σχήμα.



Παραδείγματα real-time OCR



- Χάρη στην τεχνική, μπορούμε να έχουμε ανάγνωση και κατανόηση του περιεχομένου (NLP) ενός τυπωμένου κειμένου.
- Η αναγνώριση μικρών κειμένων μπορεί να εκτελείτε σε πραγματικό χρόνο, με φορητές συσκευές όπως τηλέφωνα και tablets, δημιουργώντας on-line μεταφραστές.

Παραδείγματα τέτοιων λογισμικών είναι τα ABBY FineReader, OmniPage, Readiris Pro. Το Tesseract είναι αντίστοιχο λογισμικό, ανοικτού κώδικα, το οποίο δημιουργήθηκε στα εργαστήρια της Hewlett-Packard.

Η μετατροπή εικόνας σε κείμενο υποστηρίζεται για πολλά είδη αλφαβήτων (λατινικό, ελληνικό, κ.α.). Είναι μια διαδικασία με υψηλό ποσοστό ορθής αναγνώρισης χαρακτήρων (π.χ. 95%), ιδίως για την αγγλική γλώσσα, καθώς έχει υπάρξει περισσότερη έρευνα λόγω της μεγαλύτερης συγκέντρωσης ενδιαφέροντος.

Εκτός της αναγνώρισης των συμβόλων, ορισμένα από τα προγράμματα μπορούν να αναγνωρίσουν και να αναπαραστήσουν και την δομή/διάταξη (layout) του εγγράφου.

Σημεία προς προσοχή:

Παράγοντες οι οποίοι μπορούν να επηρεάσουν την αποτελεσματικότητα της μετατροπής είναι οι εξής:

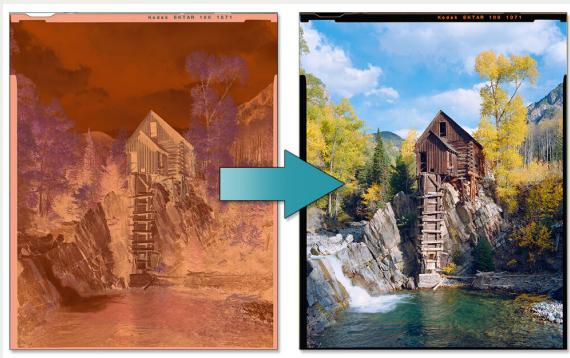
- η ποιότητα των ψηφιοποιημένων εικόνων
- το κατά πόσο είναι ευανάγνωστο το αποτυπωμένο κείμενο. Σε χειρόγραφα, όπου δεν υπάρχει μεγάλος βαθμός τυποποίησης του τρόπου γραφής των συμβόλων, τα προγράμματα OCR συχνά αποτυγχάνουν.
- ο βαθμός αντίθεσης του κειμένου έναντι του υποβάθρου της εικόνας (contrast)
- η γλώσσα στην οποία είναι γραμμένο το αρχικό κείμενο

Συστάσεις Ψηφιοποίησης Εικόνων

Οι εικόνες είναι ο τύπος τεκμηρίων που ψηφιοποιείται συχνότερα. Ωστόσο, η διαδικασία παρουσιάζει σημαντικές δυσκολίες, που οφείλονται στην ποικιλομορφία των χαρακτηριστικών του έντυπου υλικού, όπως:

- το είδος του μέσου αποθήκευσης (χαρτί, καμβάς, αρνητικό φιλμ, κ.α.)
- οι διαστάσεις των έντυπων εικόνων (από πολύ μικρές έως πολύ μεγάλες)
- την εύθραυστη κατάσταση του αναλογικού μέσου αποτύπωσης της εικόνας
- τις τυχόν αλλοιώσεις που φέρει η εικόνα (κυρίως οι παλιές εικόνες)

Παραδείγματα εικόνων



Οι εικόνες δεν βρίσκονται πάντα τυπωμένες σε χαρτί. Ενδέχεται να διατίθενται υπό μορφή φιλμ, το οποίο έχει μικρές διαστάσεις και τα χρώματα δεν είναι άμεσα εμφανή.

Παραδείγματα εικόνων



Οι διαστάσεις της εικόνας προς ψηφιοποίηση δεν είναι πάντα μικρές, γεγονός που εγείρει προκλήσεις κατά την ψηφιοποίησή τους.

Παραδείγματα εικόνων



Ορισμένες εικόνες είναι διαθέσιμες σε εύθραυστο αποθηκευτικό μέσο και απαιτούν ιδιαίτερη προσοχή στον χειρισμό τους. Παραπάνω απεικονίζεται η αποκατάσταση της αρχικής εικόνας η οποία έγινε στο ψηφιοποιημένο αρχείο.

Όλοι αυτοί οι παράγοντες βαρύνουν στην απόφαση της επιλογής της μεθόδου, της τεχνικής και του εξοπλισμού που θα χρησιμοποιηθεί για την ψηφιοποίηση των εικόνων.

Οι παράμετροι ψηφιοποίησης των εικόνων, επηρεάζονται σε μεγάλο βαθμό από το μέγεθος της αρχικής εικόνας, το μέσο εκτύπωσης και τον σκοπό ψηφιοποίησης.

Για τις περιπτώσεις διάσωσης και αρχειακής αποθήκευσης του περιεχομένου που απεικονίζεται σε μια εικόνα, έχει κυριαρχήσει η αποθήκευση σε μορφότυπο TIFF.

Ενδειγμένες τιμές

Οι ψηφιοποιημένες εικόνες που δημιουργούνται έχουν πάντοτε ορθογώνια μορφή και οι διαστάσεις τους αποτιμώνται σε pixels.

- Εικόνες διαστάσεων 8×10": ανάλυση 400ppi, 8 or 24bits, grayscale/color. Δημιουργούν αρχεία διαστάσεων 3200×4000px
- Εικόνες διαστάσεων 5×7": ανάλυση 625ppi, 8 or 24bits, grayscale/color. Δημιουργούν αρχεία διαστάσεων 3125×4375px
- Εικόνες διαστάσεων 4×5": ανάλυση 800ppi, 8 or 24bits, grayscale/color. Δημιουργούν αρχεία διαστάσεων 3200×4000px
- Εικόνες διαστάσεων 4×2.5": ανάλυση 1200ppi, 8 or 24bits, color. Δημιουργούν αρχεία διαστάσεων 4800×3000px

Image size Pixels (Virtual Size of Scans)	Megapixel rating	Print size (inches) at 200ppi	Print size (inches) at 300ppi
640 x 480	0.3	3.2 x 2.4	2.1 x 1.6
1,024 x 768	0.8	5.1 x 3.8	3.4 x 2.5
1,280 x 960	1.2	6.4 x 4.8	4.2 x 3.2
1,504 x 1,000	1.5	7.5 x 5.0	5.0 x 3.3
1,632 x 1,224	2.0	3.3 x 6.1	5.4 x 4.1
2,000 x 1,312	2.6	10.0 x 6.6	6.7 x 4.4
2,240 x 1,488	3.3	11.2 x 7.4	7.5 x 5.0
2,275 x 1,520	3.5	11.4 x 7.6	7.6 x 5.1
2,272 x 1,704	3.9	11.4 x 8.5	7.6 x 5.7
2,590 x 1,920	5.0	13.0 x 9.6	8.6 x 6.4
3,008 x 2,000	6.0	15.0 x 10.0	10.0 x 6.7
4,256 x 2,848	12.1	21.3 x 14.2	14.2 x 9.5
4,536 x 3,024	13.7	22.7 x 15.1	15.1 x 10.1
5,782 x 3,946	22.8	28.9 x 19.7	19.3 x 13.2

Ρυθμίσεις ανάλυσης και πλήθους χρωμάτων για αρνητικά φιλμ

Στην περίπτωση που οι εικόνες είναι αποθηκευμένες σε κάποιο μικροφίλμ, οι αναλύσεις που χρησιμοποιούνται είναι κατά κανόνα υψηλότερες καθώς λόγω της μικρής επιφάνειας λαμβάνονται περισσότερα δείγματα.

- διαστάσεις 8×10": ανάλυση 800ppi, 8 or 24bits, grayscale/color. Δημιουργούν αρχεία διαστάσεων 6400x8000px
- διαστάσεις 4×5": ανάλυση 1200ppi, 8 or 24bits, grayscale/color. Δημιουργούν αρχεία διαστάσεων 4800x6000px
- φιλμ 35mm: ανάλυση 4000ppi, 8 or 24bits, grayscale/color. Δημιουργούν αρχεία διαστάσεων 5480px κατά την μεγάλη διάσταση του στιγμιότυπου.

Πηγές / Παραπομπές



Iris Xie and Krystyna Matusiak, (2016) "*Discover Digital Libraries: Theory and Practice*", Amsterdam: Elsevier. ISBN: 9780124171121