



ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Τεχνολογίες Πολυμέσων

Ενότητα # 5: Κείμενο

Γιώργος Καρυδάκης

Τμήμα Πολιτισμικής Τεχνολογίας και Επικοινωνίας



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αθηνών**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



Σύνοψη προηγούμενης διάλεξης

- Ψηφιακό/Αναλογικό σήμα
- Ψηφιοποίηση
- Δειγματοληψία
- Κβαντισμός
- Κωδικοποίηση
- Συμπίεση

Περιεχόμενα ενότητας

- Δομή κειμένου
- Κωδικοποίηση
- Γραμματοσειρές
- Μορφοποίηση
- Αναγνωσιμότητα/Προσβασιμότητα
- Γλώσσες επισήμανσης/Μεταγλώσσες

Ενότητα # 5

Κείμενο

Κείμενο

- Το θεμελιώδες συστατικό των εφαρμογών πολυμέσων

- Χαρακτηριστικά

- χρώματα

- μορφοποίηση

- γραμματοσειρές

- μέγεθος

- Αλλά και...

-

στοίχιση

- 1. αρίθμηση

- ομαδοποίηση

- διάστημα παραγράφου

- διάστημα χαρακτήρων

Συνεισφορά

- Περιεχόμενο κειμένου
 - ελκυστικό
 - ευανάγνωστο
 - υποβοήθηση προσοχής

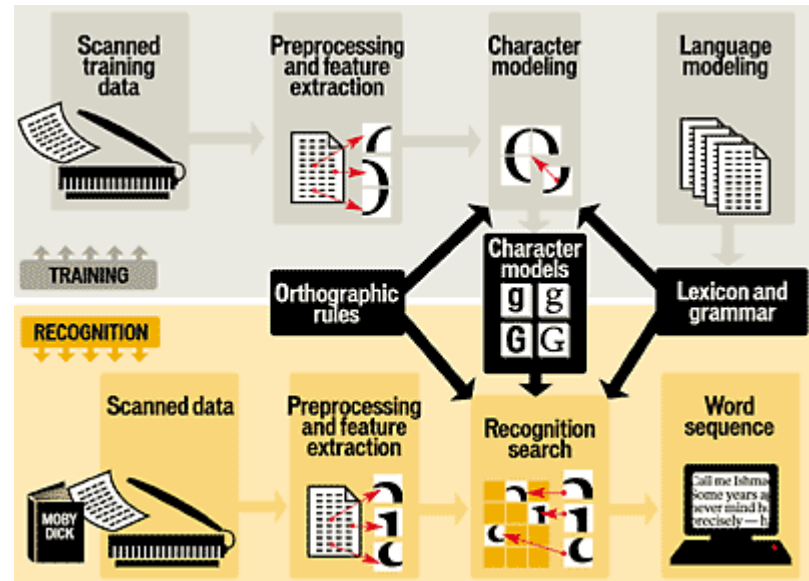
Ιστορική Αναδρομή

- Το κείμενο (text) ήταν ο πρώτος τρόπος απεικόνισης της πληροφορίας σε υπολογιστή και παραμένει μέχρι σήμερα ο κυρίαρχος
- Αρχικά η επικοινωνία ανθρώπου υπολογιστή γινόταν μόνο με κείμενο λόγω της αδυναμίας των αρχικών υπολογιστών να χειριστούν αποτελεσματικά ήχο, εικόνες και video.
- Το κείμενο αποτελεί βασικό φορέα πληροφορίας που δεν πρέπει να παραμελείται σε μια εφαρμογή πολυμέσων.
- Το ζητούμενο δεν είναι η εξάλειψη του κειμένου, αλλά η αποτελεσματική ολοκλήρωση του με τα άλλα είδη πληροφορίας.

Εισαγωγή κειμένου

- Με πληκτρολόγηση
- Με οπτική αναγνώριση χαρακτήρων (Optical Character Recognition - OCR)

Τα OCR συγκρίνουν βήμα-βήμα τη σαρωθείσα επιφάνεια και ψάχνουν για γράμματα. Ότι μοιάζει με γράμμα, το συγκρίνουν με την ενσωματωμένη βάση δεδομένων γραμματοσειρών που διαθέτουν και το αντικαθιστούν στο έγγραφο με την αντίστοιχη τιμή ASCII. Το αποτέλεσμα είναι ένα αρχείο .txt έτοιμο για επεξεργασία



Δομημένο κείμενο

- Το κείμενο μπορεί να θεωρηθεί ότι οργανώνεται σε δενδρική δομή
 - Κεφάλαια
 - Υποκεφάλαια
 - Παραγράφους
 - ...
- Καλύτερη οργάνωση
- Αναζήτηση
- Διάβασμα

Κείμενο

- Πλεονεκτήματα
 - Κυριολεξία
 - Απλότητα
 - Πληροφορίες
- Μειονεκτήματα
 - Κουράζει
 - Εκφραστικότητα
- Με κατάλληλο συνδυασμό με εικόνα, video και ήχο κρατάμε όλα τα πλεονεκτήματα χωρίς να έχουμε μειονεκτήματα

Κωδικοποίηση

- ASCII (American Standard Code for Information Interchange)
 - 7-bit character coding system
 - 128 χαρακτήρες για αλφαριθμητικά και πλεονάζοντα για ειδικά σύμβολα
 - Δεν υπάρχει προτυποποίηση για τα ειδικά σύμβολα
- The Extended Character Set
 - 1-byte character coding system
 - 256 χαρακτήρες
 - επίσης γνωστό ως ISO-Latin-1 character set
 - Υποστηρίζει πολυγλωσσικό σύστημα
 - Χρησιμοποιείται σε HTML σελίδες
- Unicode
 - 16-bit architecture, μέχρι 65.000 χαρακτήρες
 - Version 2.0: υποστηρίζει 25 γλώσσες και 38.885 χαρακτήρες
 - HTML επιτρέπει εισαγωγή Unicode χαρακτήρων
 - `水` παριστάνει τον κινέζικο χαρακτήρα για το νερό

Κείμενο και Τυπογραφία

- Αναπαράσταση κειμένου
 - Αρχικά οι ανάγκες των προγραμμάτων απαιτούσαν ένα περιορισμένο σύνολο κωδικοποιημένων χαρακτήρων όπως αυτό που περιλάμβανε η κωδικοποίηση ASCII
 - Η εξέλιξη όμως των υπολογιστών δημιούργησε την ανάγκη νέων μορφών κωδικοποίησης του κειμένου όπως οι κωδικοσελίδες ISO και Unicode και νέες γλώσσες περιγραφής σελίδας όπως η PostScript22

Κωδικοποίηση χαρακτήρων

- US-ASCII
 - 95 εκτυπώσιμοι χαρακτήρες
 - 33 χαρακτήρες ελέγχου
- ISO-8859-n
 - επέκταση του US-ASCII
 - 256 χαρακτήρες, με κάλυψη και των μη λατινικών αλφαβήτων
- ISO-8859-7
 - Συμπεριλαμβάνει το λατινικό και ελληνικό αλφάβητο σε μονοτονικό σύστημα

ASCII (American Standard Code for Information Interchange) (1)

- ο ASCII είναι ο πιο διαδεδομένος κώδικας αναπαράστασης
- Αναπτύχθηκε από τον Αμερικάνικο οργανισμό Προτυποποίησης και είναι μία απλή κωδικοποίηση χαρακτήρων των 7-bits
- ο κωδικός ASCII περιλαμβάνει **7bit = $2^7 = 128$** χαρακτήρες
- Από αυτούς οι 95 είναι σύμβολα γραφής και οι 33 χαρακτήρες ελέγχου (οι οποίοι χρησιμοποιούνται για να ενεργοποιήσουν ή να σταματήσουν μία ενέργεια που έχει να κάνει με την εγγραφή, την επεξεργασία ή την μετάδοση δεδομένων)

ASCII (2)

- Ένα παράδειγμα ASCII κωδικοποίησης

	H	E	L	L	O
ASCII	48	69	76	76	79
Διαδική απεικόνιση	0011000	1000101	1001100	1001100	1001111

Κωδικοσελίδες ISO (International Standards Organisation)

- Σημαντικό μειονέκτημα της ASCII είναι ότι περιλαμβάνει μόνο λατινικούς χαρακτήρες και δεν υπάρχει πρόνοια για άλλες γλώσσες
- Για αυτό το λόγο δημιουργήθηκαν επεκτάσεις του κώδικα ASCII με 8-bit για τους οποίους μπορούσαν να προστεθούν $2^8=256$ χαρακτήρες
- Οι πρώτες 128 θέσεις είναι ίδιες με την ASCII και οι πλεονάζοντες 128 χαρακτήρες για ειδικά σύμβολα άλλων αλφάβητων
- Γνωστοί κώδικες που χρησιμοποιούνται στην Ελλάδα είναι οι:
 - IBM 851, IBM 869, 737 ή αλλιώς ελληνικός 437 (γιατί προέκυψε από την αμερικάνικη κωδικοσελίδα 437), εΛοΤ 928
 - Η εΛοΤ 928 είναι η πιο σύγχρονη τυποποίηση και καθιερώθηκε αρχικά από τον εΛοΤ και αργότερα έγινε αποδεκτή από τον ISO-Latin-8859-7 ή απλά Latin 7

Unicode

- Αντιμετώπιση των περιορισμών του ISO-8859
 - Πολυγλωσσική επεξεργασία κειμένου
 - Ένας κωδικός αριθμός για κάθε ξεχωριστό χαρακτήρα
 - Κωδικοποίηση ανεξάρτητη από εμφάνιση
 - Συμπεριλαμβάνει πάνω από 30 συστήματα γραφής
- UTF-8 (multibyte Unicode char. encoding)
 - Ένα byte για τους US-ASCII χαρακτήρες (backwards-compatible)
 - Δύο bytes για τους λοιπούς λατινικούς, ελληνικούς κ.α. χαρακτήρες
 - Τρία bytes για τους άλλους βασικούς πολυγλωσσικούς χαρακτήρες
 - Τέσσερα bytes για οποιουσδήποτε άλλους χαρακτήρες

The Unicode Standard

- ISO/IEC 10464-1;1993
- 16 bit κωδικοποίηση -> 65.000 συνδυασμοί
- UTF extension -> 1.000.000 επιπλέον χαρακτήρες

The Unicode actual content

- Latin, Greek, Cyrillic, Armenian, Hebrew, Arabic, Devanagari, Bengali, Gurmuki, Gujarati, Oryia, Tamil, Telugu, Kannada, Malayalam, Thai, Lao, Georgian, Tibetan, Japanese Kana, Korean Hangul, Unified set of Chinese/Japanese/Korean (CJK) ideographs

Κωδικοσελίδες Unicode

- Η ανάπτυξη κωδικοσελίδων ISO έλυσε το πρόβλημα με τα διάφορα αλφάβητα, αλλά λόγω περιορισμού του μεγέθους και πάλι σε καμία κωδικοσελίδα ISO δεν χωρούσαν αρκετοί χαρακτήρες.
- Η Ευρωπαϊκή Ένωση χρειαζόταν πλήθους κωδικοσελίδων για να καλύψει όλες τις χώρες μελών της έτσι δημιουργήθηκε η κωδικοσελίδα Unicode που προτείνει ένα μοναδικό αριθμό για κάθε χαρακτήρα, ανεξάρτητα από το λειτουργικό σύστημα, το λογισμικό και τη γλώσσα.
- Κάθε κωδικοποιημένη λέξη Unicode έχει μήκος **16 bit** επιτυγχάνοντας **$2^{16}=65536$** συνδυασμούς
- Η Unicode παρέχει κωδικοποίηση για τους χαρακτήρες όλων των γλωσσών συμπεριλαμβανομένων των ευρωπαϊκών, της Μέσης Ανατολής και της Ασίας

Μορφές κειμένου στον Η/Υ

- Μη μορφοποιημένο κείμενο (unformatted text):
 - είναι γνωστό και ως καθαρό κείμενο (plaintext) και αποτελείται από σειρές χαρακτήρων του ιδίου μεγέθους. οι χαρακτήρες προέρχονται από ένα πεπερασμένο σύνολο χαρακτήρων
- Μορφοποιημένο κείμενο
 - Είναι γνωστό και ως εμπλουτισμένο κείμενο (rich text). Αποτελείται από σειρές χαρακτήρων, το είδος και το μέγεθος των οποίων ποικίλει, σύμβολα, πίνακες, γραφικά και εικόνες
- Υπερκείμενο
 - είναι η μορφή κειμένου η οποία χρησιμοποιείται κατά κόρον στο Διαδίκτυο και επιτρέπει τη μη γραμμική πλοήγηση ανάμεσα ιστοσελίδες κάθε μια από τις οποίες μπορεί να περιέχει μορφοποιημένο κείμενο.

Αναπαράσταση κειμένου στον Η/Υ

- Το κείμενο μπορεί να περιγραφεί ως ένα σύνολο από χαρακτήρες
- Η επιμέλεια κειμένου περιλαμβάνει:
 - Κωδικοποίηση Χαρακτήρων
 - περιγράφεται μόνο το νοηματικό περιεχόμενο του κειμένου
 - Σε κάθε χαρακτήρα αντιστοιχούμε μια σειρά από bits
 - Για να είναι κατανοητό ένα κείμενο σε 2 διαφορετικούς Η/Υ θα πρέπει να χρησιμοποιούν την ίδια κωδικοποίηση χαρακτήρων
 - Μορφοποίηση Σελίδας
 - καθορίζεται ο τρόπος παρουσίασης του κειμένου

Μορφοποίηση κειμένου

- Η επιβολή κάποιων χαρακτηριστικών εμφάνισης του κειμένου
- ο όρος γραμματοσειρά (font) αναφέρετε σε οικογένεια χαρακτήρων που διαθέτουν κοινά χαρακτηριστικά εμφάνισης
- οικογένεια γραμματοσειρών (typeface) είναι ένα σύνολο γραμματοσειρών με καθορισμένο στύλ εμφάνισης για κάθε χαρακτήρα
- Τυπικές μορφές γραμματοσειρών είναι οι:
 - Γραμματοσειρών με ακρεμόνες (**serif**) Times New Roman
 - Γραμματοσειρών χωρίς ακρεμόνες (**sans-serif**) Arial

Γραμματοσειρές

f f

Γραμματοσειρές

- Μέγεθος
 - 1pt = 1/72 ίντσας

This is ten point Helvetica
This is twelve point
This is fourteen point
This is eighteen point
and this is twenty-four point

- Πλάτος Pitch
 - Προκαθορισμένο
 - Courier
 - Μεταβλητό variable-pitched – some characters wider
 - Times Roman / Palatino

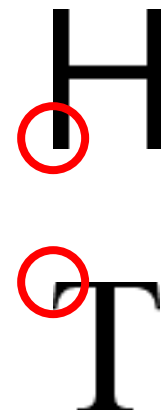
Serif / Sans-serif

- sans-serif – τετραγωνισμένα

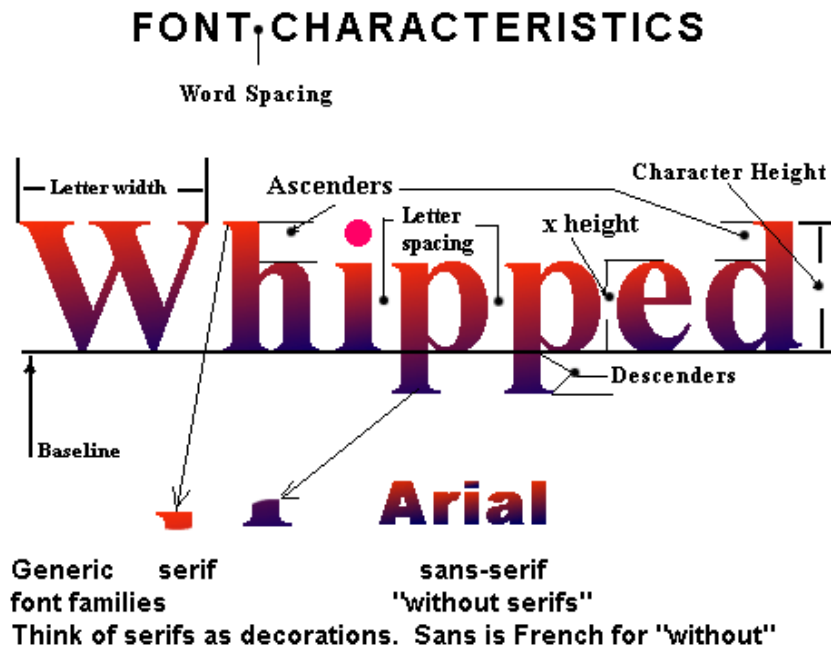
Helvetica

- serif – ομαλές ακμές

Times Roman Palatino



Χαρακτηριστικά γραμματοσειρών



Baseline:

The invisible line on which all text rests

Ascender:

The upper portion of a character that extends above the x-height

Descender

The lower portion of a letter that extends below the baseline

Character width

The precise Width of a character, which will vary from letter to letter unless you are using monospaced letters.

Character Height

Measurement from the baseline to the top of the tallest letter. In HTML when we adjust the line height, this is the measurement that is altered (as measured from the baseline).

Letter Spacing

The spaces between the letters. This can be varied with the kerning control of most graphics programs.

Word Spacing

The spacing between words.

X height

The height of the letter "x"

Κατηγορίες Γραμματοσειρών

- Χαρτογραφική (bitmap fonts - Raster)
 - οι χαρακτήρες περιγράφονται μέσω εικόνων οι οποίες δίνουν την τοποθέτηση των άσπρων και μαύρων pixels για όλα τα σημεία της περιοχής που καταλαμβάνει ένας χαρακτήρας.
 - Για κάθε χαρακτήρα κάθε font υπάρχει μια τέτοια εικόνα.
 - Αν και η επεξεργασία και εκτύπωση αυτών είναι γρήγορη, χάνουν σε:
 - ανεξαρτησία από τη μονάδα εξόδου
 - απαιτήσεις χώρου αποθήκευσης
 - ποιότητα μετά από κάποιο μετασχηματισμό (όπως μεγέθυνση ή περιστροφή).
- Διανυσματική (Vector)
 - η περιγραφή των στοιχείων γίνεται με μαθηματικό τρόπο, γεγονός που διορθώνει τα ελαττώματα των χαρτογραφικών γραμματοσειρών αλλά αυξάνει το χρόνο εκτύπωσης
 - Αυτή τη λογική ακολουθούν οι PostScript Type 1 και TrueType που θα δούμε στη συνέχεια.
 - Για μικρού μεγέθους fonts η χρήση bitmaps είναι αποτελεσματικότερη όσον αφορά την καλύτερη αναπαράσταση των χαρακτήρων

Fonts

- PostScript Type 1:
 - Πρότυπο ψηφιακών χαρακτήρων (International Standards Organization outline font standard, ISO 9541.)
 - Αρχικά αναπτύχθηκε από Adobe Systems για μέσω της γλώσσας περιγραφής σελίδων και fonts PostScript.
 - Στόχος ήταν η δημιουργία εκτυπώσεων υψηλής ανάλυσης (πάνω από 300 dpi) σε εκτυπωτές laser
 - Αναγνωρίζεται από κάθε τύπο Η/Υ (microcomputers - mainframes)
 - Τυπώνεται σε κάθε εκτυπωτή
 - Το Adobe Portable Document Format (PDF) είναι υποσύνολο της γλώσσας PostScript
- True Type
 - Αναπτύχθηκε από την Apple Computers, και μετά δόθηκε άδεια στη Microsoft Corporation.
 - Όπως τα Type 1, το TrueType format είναι διαθέσιμο για δημιουργία νέων fonts

Fonts

- OpenType fonts:
 - Αναπτύχθηκε από κοινού από Adobe και Microsoft.
 - Τα OpenType fonts περιλαμβάνουν είτε PostScript είτε TrueType σε ένα «περιτύλιγμα».
 - Ένα OpenType font είναι ένα αρχείο που ανταποκρίνεται το ίδιο σε Macintosh and Windows platforms.
 - Τα OpenType fonts πλεονεκτούν σε σχέση με τα προηγούμενα επειδή περιέχουν περισσότερους ειδικούς χαρακτήρες και μπορούν να υποστηρίξουν περισσότερες γλώσσες καθώς και επιπλέον τυπογραφικά χαρακτηριστικά όπως small caps, old-style figures, και ligatures (συμπλέγματα γραμμάτων)— όλα σε ένα μοναδικό font

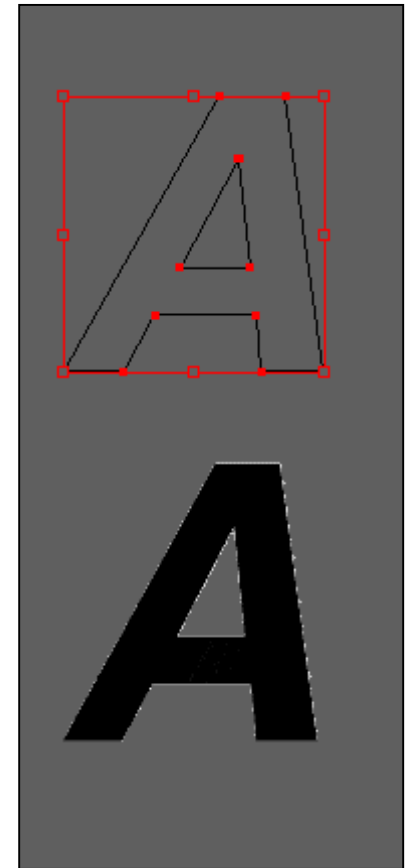
Χαρτογραφικές (bitmap fonts)

- Τα στοιχεία αυτών των γραμματοσειρών περιγράφονται σαν πλέγμα
- Βασικό πλεονέκτημά τους είναι η γρήγορη επεξεργασία και απεικόνιση
- Αντίθετα μειονεκτούν στο ότι έχουν αυξημένες απαιτήσεις χώρου αποθήκευσης και η ποιότητά τους εξαρτάται από την μονάδα εξόδου
- Σημαντικό μειονέκτημά τους είναι η χαμηλή ποιότητα μετά από κάποιο μετασχηματισμό (μεγέθυνση, ή περιστροφή)



Διανυσματικές (bitmap fonts)

- Η περιγραφή αυτών των γραμματοσειρών γίνεται με μαθηματικό τρόπο, γεγονός που διορθώνει τα ελαττώματα των χαρτογραφικών γραμματοσειρών αλλά αυξάνουν το χρόνο εκτύπωσης
- Η εμφάνιση κάθε γράμματος στην οθόνη γίνεται βάση των πληροφοριών σχηματισμού
- Σε αυτήν την κατηγορία ανήκουν οι γραμματοσειρές PostScript Type 1 της Adobe και True Type των Windows και Apple



Αναγνωσιμότητα κειμένου

- Συνήθης ταχύτητα ανάγνωσης από έντυπο υλικό είναι 200-300 λέξεις ανά λεπτό
- Δεν προκύπτει από ομαλή κίνηση των οφθαλμών κατά μήκος του κειμένου, αλλά αντιθέτως παρατηρούνται φαινόμενα παλινδρόμησης, διαδοχικές εστιάσεις και επιταχύνσεις επί του κειμένου από τον αναγνώστη.
- Η αναγνωσιμότητα κειμένου (ταχύτητα ανάγνωσης και αριθμός σφαλμάτων ανάγνωσης) από την οθόνη είναι κατά 30% μειωμένη έναντι ανάγνωσης του ιδίου κειμένου από έντυπο

Αιτίες χαμηλής αναγνωσιμότητας

- Η μειωμένη ποσότητα κειμένου στην οθόνη έναντι του εντύπου
- Γωνία ανάγνωσης περιορίζεται από τη σχετική θέση οθόνης-αναγνώστη
- Μειωμένη συγκέντρωση των χρηστών
- Αριθμός χαρακτήρων και λέξεων ανά γραμμή
- Χρώμα και η χρωματική αντίθεση
- Τυπογραφικά
 - Διαστήματα μεταξύ
 - χαρακτήρων
 - γραμμών
 - Μήκος γραμμών
 - Γραμματοσειρά

Οδηγίες αναγνωσιμότητας

- Λέξεις με κεφαλαία γράμματα είναι κατά 20% πιο δυσανάγνωστες από τις ίδιες λέξεις με μικρά
- Γραμματοσειρές με έμφαση στις ουρές των χαρακτήρων. (serif) είναι καλύτερες από sans serif)
- Η χρήση κενού χώρου σε μια σελίδα κειμένου είναι θετικός παράγων.
- Η απόσταση μεταξύ των χαρακτήρων πρέπει να είναι περίπου 10% της απόστασης μεταξύ των λέξεων.
- Η απόσταση μεταξύ γραμμών τουλάχιστον 50% του ύψους των γραμμών που χρησιμοποιούνται.
- 8-15 λέξεις ανά γραμμή είναι ο βέλτιστος αριθμός για καλή αναγνωσιμότητα.
- Το κείμενο θα πρέπει να χωρίζεται σε παραγράφους μεγέθους 3-5 γραμμών.



κεφαλαία / μικρά
serif / sans serif

χρήση κενού χώρου

απόσταση χαρακτήρων 10% απόστασης λέξεων

απόσταση γραμμών 50% ύψους γραμμάτων

8-15 λέξεις ανά γραμμή

παραγράφους 3-5 γραμμών

Ποιότητα του κειμένου στην οθόνη

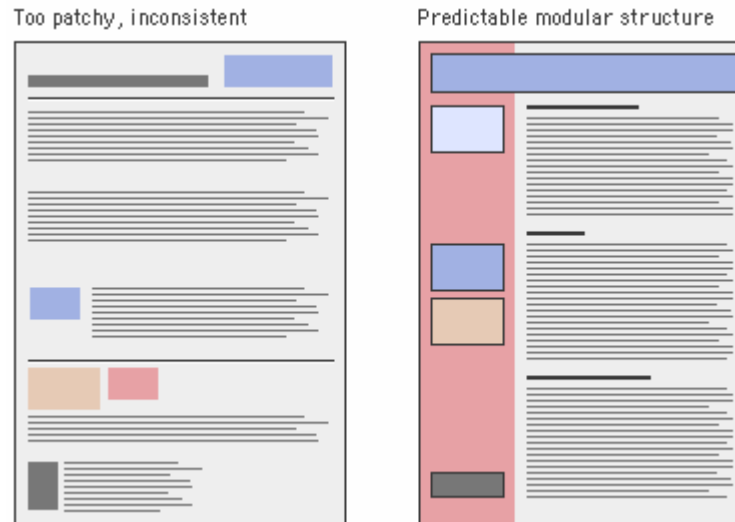
- Ο **άνθρωπος διαβάζει 28% πιο αργά στην οθόνη** από ότι στο χαρτί, γιατί;
- Η αναπαραγωγή του κειμένου στην οθόνη δεν είναι τόσο καλή όσο η απόδοσή του στο χαρτί (screen resolution)
 - τυπωμένο κείμενο σε έντυπα - 1200 dots per inch (dpi) ή μεγαλύτερο
 - computer screens 85 dpi
- Πως μπορούν να αντιμετωπιστούν αυτές οι δυσκολίες;

Ζητήματα τυπογραφίας

- Η καλή τυπογραφία έγκειται στην δημιουργία αντιθέσεων μεταξύ
 - γραμματοσειρών
 - μπλοκ κειμένων
 - τίτλων
 - λευκών περιθωρίων
- Τίποτε δεν έλκει το μάτι και το μυαλό του ανθρώπου περισσότερο από έντονες αντιθέσεις και διακριτικά σχέδια (distinctive patterns) και αυτό μπορεί να επιτευχθεί με τον σχεδιασμό της σελίδας / επιφάνειας διεπαφής της πολυμεσικής εφαρμογής

Χρώμα

- Αν η σελίδα περιέχει πολύ κείμενο
 - χρησιμοποιήστε διακριτικά σχέδια (patterns) με χρώμα
 - τα διακριτικά σχέδια βοηθούν στην δημιουργία της αίσθησης δομής και τάξης
 - αντιθέτως άτακτη χρήση χρωμάτων και μπλοκ κειμένου δημιουργούν την αίσθηση αταξίας και δυσχεραίνουν την ανάγνωση



Στοιχείση (1)

- Τα «λευκά» περιθώρια
 - παρέχουν σημαντική **οπτική ανακούφιση** σε οποιοδήποτε έγγραφο
 - **καθορίζουν την αναγνώσιμη περιοχή** από το κείμενο από το περιβάλλον και από άλλα στοιχεία της σελίδας
 - όταν χρησιμοποιούνται συστηματικά με τον ίδιο τρόπο (consistently) βοηθούν στην **δημιουργία οπτικής δομής**
 - **προσθέτουν επίσης οπτικό ενδιαφέρον** με την αντιπαράθεση του θετικού διαστήματος της οθόνης (κείμενο, graphics) από το αρνητικό (άσπρο) διάστημα.

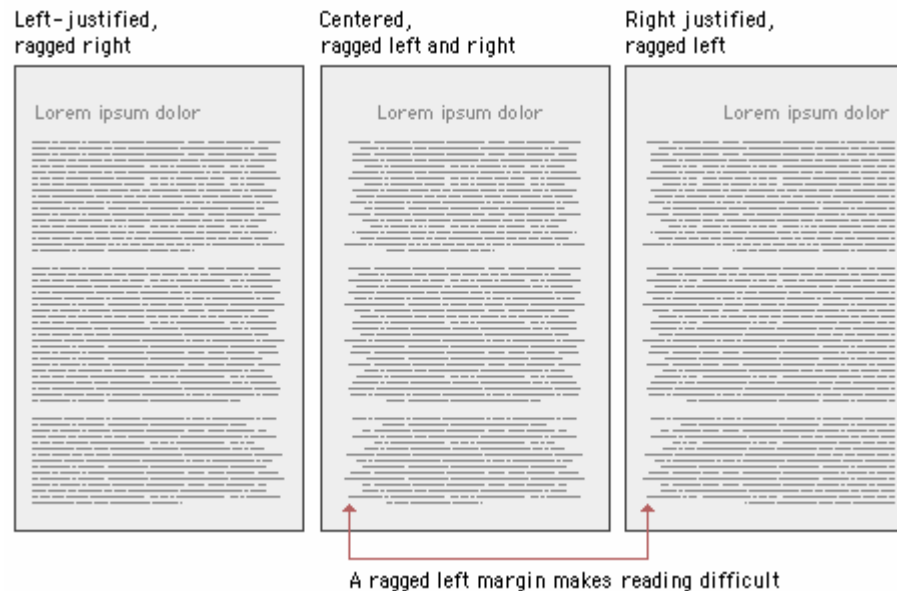
Στοιχίση (2)

- Τα προγράμματα αναπαραγωγής κειμένου (browsers) στην οθόνη δεν έχουν την απαιτούμενη ανάλυση με αποτέλεσμα να δημιουργούνται αυλάκια σε αλληπάλληλες σειρές κειμένου όταν έχουμε πλήρης στοιχίση



Στοιχίση (3)

- Το κείμενο θα πρέπει να είναι **στοιχισμένο** αριστερά ελεύθερο δεξιά με κανονική γραφή (πεζά, κεφαλαία) για μεγαλύτερη αναγνωσιμότητα



Τίτλοι

- οι **τίτλοι** να είναι **στοιχισμένοι αριστερά** και κατά κάτι μεγαλύτεροι από το κυρίως κείμενο ή με διαφορετικό χρώμα
- Κεντραρισμένοι τίτλοι ταιριάζουν σε πλήρη στοίχιση η οποία δεν ενδείκνυται για πολυμεσικές εφαρμογές ιδιαίτερα σχεδιασμένες για το διαδίκτυο



Υπότιτλοι

- οι μεγάλες ποσότητες κειμένου γίνονται δύσκολα κατανοητές στην οθόνη Η/Υ
- Σε περίπτωση που είναι απαραίτητη η παρουσίαση μεγάλων κειμένων είναι προτιμότερο να “**σπάνε**” σε ενότητες σε αρκετές οθόνες, σε αυτό βοηθάει η χρήση τίτλων και υπότιτλων

Γραμματοσειρά

- Κάθε γραμματοσειρά έχει έναν μοναδικό στυλ και επηρεάζει την αρμονία της σύνθεσης με την λεκτικής και οπτικής ροής του περιεχομένου σας (γραφικά)
- Χρησιμοποιείτε γραμματοσειρές που να συμπεριλαμβάνονται στα περισσότερα λειτουργικά συστήματα
- οι **ακρέμονες** γραμματοσειρές λειτουργούν καλύτερα στο **τυπωμένο χαρτί**, ενώ οι **μη-ακρέμονες** γραμματοσειρές λειτουργούν καλύτερα στην **οθόνη** του υπολογιστή γιατί έχουν μεγαλύτερο ύψος σώματος (x-height)
- Καταλληλότερη γραμματοσειρά με την οθόνη είναι η:
 - **Georgia, Verdana, Arial**

Μέγεθος γραμμάτων

- Στις δικτυακές εφαρμογές δεν υπάρχει μεγάλος έλεγχος του μεγέθους της γραμματοσειράς
- Οργάνωση του κειμένου ώστε να διατηρεί λογική αλλά να είναι και αναγνώσιμο
 - Πχ. ο σκοπός του Web ήταν η διασύνδεση των εγγράφων και όχι η παρουσίαση
 - H1, H2, H3, (H4, H5, H6 πολύ μικρά) δεν φτιάχτηκαν με σκοπό να εξυπηρετήσουν γραφιστικές ανάγκες
 - Λύση τα **CCS (Cascading Style Sheets)** – ο διαχωρισμός της δομής και της εμφάνισης

Κεφαλαία ή μικρά; (1)

- ΤΑ ΚΕΦΑΛΑΙΑ ΕΙΝΑΙ ΜΟΝΟΤΟΝΑ
- Η χρήση κεφαλαίων και μικρών είναι πιο ευχάριστη για το μάτι
- Επίσης όταν σαρώνουμε ένα κείμενο για να το διαβάσουμε βασιζόμαστε στο πάνω μέρος του κειμένου
- Η Χρήση Κεφαλαίων Ως Πρώτο Γράμμα Στους Τίτλους εμποδίζει Την εύκολη Ανάγνωση Του Κειμένου, Συμφωνείτε;

Έμφαση (1)

- Μία οθόνη με συμπαγές κείμενο είναι μονότονη
- ο **τονισμός** (highlighting) του κειμένου βοηθά στον έλεγχο της επιλεκτικής αντίληψης και την εστίασης της προσοχής στην τονισμένη πληροφορία
 - **Έντονα** (bold) – χρωματικό κοντράστ
 - *Πλαγιαστά* (italics) – σχηματικό κοντράστ
 - Υπογραμμισμένα (δεν ενδείκνυνται) – βαρύ και δημιουργεί σύγχυση με το σχήμα των γραμμάτων
 - Στο διαδίκτυο η υπογράμμιση παραπέμπει σε υπερσύνδεσμο (link)

Έμφαση (2)

- **Χρώμα** - μπορεί να κατευθύνει την προσοχή του αναγνώστη σε σημαντικά σημεία
- ΧΡΗΣΗ ΚεΦΑΛΑΙΩΝ
- **Αναβόσβημα** (flashing) μπορεί να γίνει ενοχλητικό και δυσκολεύει την ανάγνωση του κειμένου
- Διαστήματα και αύξηση εσοχής (indentation)

Υποδείξεις ευχρηστίας για το κείμενο (1)

- Η κάθε ολοκληρωμένη **ενότητα** να παρουσιάζεται με τον ίδιο τυπογραφικό σχεδιασμό ώστε ο χρήστης να αντιλαμβάνεται την μετάβαση από την μία ενότητα στην επόμενη
- Η **επιλογή χρώματος** θα πρέπει να έχει ως σκοπό την **μεγιστοποίηση της αντίθεσης μεταξύ του κειμένου και του φόντου** γιατί ο βαθμός αναγνωσιμότητας αυξάνεται καθώς αυξάνεται ο βαθμός αντίθεσης
- Το κείμενο θα πρέπει να εμφανίζεται με **σκούρα γράμματα σε ανοιχτό φόντο**
- είναι προτιμότερο να χρησιμοποιείται ένα **ουδέτερο γκρι ή παστέλ φόντο**

Παράδειγμα

Pastel colours as background are easier to read

Pastel colours as background are easier to read

These colours are not inspiring but
at least easily readable

These colours are not good combination for
background and text because there is low contrast
between them which makes it difficult to read

This colour scheme offers contrast, but a
combination of so many colours and the use of
such a complex background makes it very hard to
read

This colour scheme offers contrast and is easier to
read.

Συνέπεια

- Χρώματα
- Λευκά περιθώρια
- Γραμματοσειρές - συνιστάται η χρήση των γραμματοσειρών να είναι σταθερή σε όλη την έκταση της παρουσίασης (**consistency**) – όχι πάνω από 2-3 είδη γραμματοσειρών σε κάθε οθόνη –μάλιστα είναι προτιμότερο να χρησιμοποιείται μια γραμματοσειρά ανά οθόνη και κάθε αλλαγή να είναι για έμφαση
- Στοίχιση
- Έμφαση

Προσβασιμότητα (Accessibility)

- Σχετικά με το κείμενο τα κύρια θέματα πρόσβασης είναι το
 - μέγεθος
 - χρώμακαι απευθυνόμαστε σε άτομα με προβλήματα όρασης και αχρωματοψία
- εξελικτικό κείμενο (scalable text))
- Structural markup
- Έμφαση – χρώμα και αχρωματοψία
- Προσαρμόσιμες επιφάνειες διεπαφής

Μορφές διαχείρισης κειμένου (1)

- Για την διαχείριση κειμένων σε μία εφαρμογή πολυμέσων έχουν καθιερωθεί τρεις μορφές:
 - Μορφή χαρακτήρων
 - Μορφή εικόνας
 - Μορφή φορητού εγγράφου
- **Μορφή χαρακτήρων**
 - Προσοχή σε αυτήν την περίπτωση ο Η/Υ θα πρέπει να διαθέτει την συγκεκριμένη γραμματοσειρά, αλλιώς την αντικαθιστά με άλλη που διαθέτει
 - Για αυτό το λόγο συνιστάται η χρήση ευρέως διαδεδομένων γραμματοσειρών
 - Ή θα πρέπει οι γραμματοσειρές να διατίθενται στον χρήστη

Μορφές διαχείρισης κειμένου (2)

■ Μορφή εικόνας

- Ένα κείμενο που έχει σαρωθεί με την χρήση του σαρωτή μπορεί να αποθηκευτεί ως εικόνα
- Τα περισσότερα προγράμματα συγγραφής δίνουν την δυνατότητα μετατροπής ενός κειμένου σε μορφή εικόνας
- Πλεονεκτήματα – διατηρεί την μορφή του ανεξαρτήτως του συστήματος
- Μειονεκτήματα – μεγαλύτερο αρχείο

■ Μορφή φορητού εγγράφου

- Portable document format (PDF) αποτελεί ένα τρόπο να διατηρηθούν όλα τα αυθεντικά χαρακτηριστικά ενός κειμένου
- Προϊόν της Adobe Acrobat
- Απαραίτητη η χρήση του Acrobat Player

Μορφή κειμένου

- Το τυπωμένο κείμενο μεταφέρει δύο είδη πληροφοριών
 - τη μορφή του κειμένου πάνω στη σελίδα
 - το νοηματικό του περιεχόμενο
- οι απλές κωδικοποιήσεις ASCII, ISO, Unicode κτλ. είναι αρκετές για τη μετάδοση του νοηματικού περιεχομένου, αλλά δεν καθορίζουν τον τρόπο παρουσίασης του νοηματικού περιεχομένου
- Με την ευρεία χρήση του Διαδικτύου έγινε απαραίτητη η ύπαρξη ειδικών κωδικοποιήσεων οι οποίες καθορίζουν και τον τρόπο παρουσίασης της πληροφορίας

Κείμενο και εφαρμογές πολυμέσων

- Σε εφαρμογές πολυμέσων το κείμενο χρησιμοποιείται είτε ως:
 - **κείμενο περιεχομένου** – παρέχει ουσιαστικές πληροφορίες στον χρήστη με την περιγραφή ή παράθεση πληροφοριών
 - **κείμενο περιβάλλοντος** – τίτλοι για τον καθορισμό ενοτήτων, κουμπιά πλοήγησης, μενού επιλογών κτλ.
- Για να είναι αποτελεσματική η παρουσίαση ενός μηνύματος θα πρέπει το κείμενο που χρησιμοποιείται να είναι:
 - σαφές
 - σύντομο
 - αισθητικά ευχάριστο
 - να εμφανίζεται σε κατάλληλη θέση στην οθόνη

Μοντέλα μορφοποίησης κειμένου

- WYSIWYG (What You See is What You Get)
 - Κάθε αλλαγή στο φορμάρισμα εμφανίζεται στην οθόνη ακριβώς όπως θα τυπωθεί στον εκτυπωτή.
- Mark Up Κείμενο
 - Δυο είδη πληροφοριών (semantics + syntactic):
 - η μορφή του κειμένου πάνω στη σελίδα, και το περιεχόμενο.
 - Μίξη του κανονικού κειμένου με οδηγίες παρουσίασης (tags).
 - π.χ. `<center>` Αυτή είναι μια γραμμή με βάση τη γλώσσα HTML `</center>`
 - Η εμφάνιση του φορμαρισμένου κειμένου στην οθόνη γίνεται με τη βοήθεια άλλης εφαρμογής
 - HTML => Browser rendering
 - Latex editor => compile => Postscript πρόγραμμα

Μορφή κειμένου

- Παράδειγμα ειδικής κωδικοποίησης αποτελεί το markup κείμενο το οποίο αποτελεί μίξη του κανονικού κειμένου, με οδηγίες παρουσίασης

Πχ. το σύστημα *troff* στα UNIX

.ce

Αυτή είναι μια γραμμή

Σημαίνει ότι η πρόταση «Αυτή είναι μια γραμμή» πρέπει να κεντραριστεί στη σελίδα

- Η μέθοδος αυτή μπορεί να χρησιμοποιηθεί για τη δημιουργία κειμένων με πολύπλοκη μορφή, όπως επιστημονικά κείμενα με πίνακες, εξισώσεις και ειδικά σύμβολα.
- Άλλα παραδείγματα τέτοιας γλώσσας είναι η SGML, HTML, XML, TEX για UNIX που διευκολύνεται από το σύστημα (LATEX)

Μεταγλώσσα (Markup)

- Μεταγλώσσα ή markup είναι ένα σετ από κωδικούς (**tags**) που χρησιμοποιούνται για να χαρακτηρίσουν την μορφή του κειμένου (πχ. για να δίνουν πληροφορίες στον εκτυπωτή πως να τυπώσει)
- Υπάρχουν τρία είδη markup:
 - **Διαδικαστική μεταγλώσσα** (Procedural mark-up)
 - **Γενική κωδικοποίηση** (Generic coding)
 - **Γενικευμένη μεταγλώσσα** (Generalised mark-up)

Διαδικαστική μεταγλώσσα (procedural markup) (1)

- Χρησιμοποιεί **κώδικα για τον προσδιορισμό της μορφής του κάθε στοιχείου**
- Προσδιορίζει τις διαδικασίες που θα προσδώσουν την επιθυμητή μορφή του κειμένου

πχ. η διαδικαστική μεταγλώσσα προσδιορίζει στο αρχείο ότι η κεφαλή του αρχείου θα είναι *Italics* και οι λίστες θα εμφανίζονται με βούλες:

<ESC-22I> From: J.S. Bach

To: R. Stone

Subject: Supplies for Brandenburgh Ltd. <ESC-22I>

Dear Roll,

Please supply me with the following items:

<ESC-44B> 3 boxes of paper-clips

<ESC-44B> 2 boxes of staples

<ESC-44B> a chair

Yours,

J.S.

Διαδικαστική μεταγλώσσα (2)

- Τα προβλήματα με την διαδικαστική μεταγλώσσα είναι ότι:
 - Διατηρείται μόνο η **μορφή – όχι η δομή**
 - εξαρτάται από το λειτουργικό – platform dependent
 - είναι δύσκολο να απομνημονευτεί

Η γενική κωδικοποίηση (1)

- Η γενική κωδικοποίηση αντικαθιστά τους κωδικούς με macros τα οποία αργότερα ερμηνεύονται από τους κατάλληλους διαδικαστικούς κώδικες (procedural codes)

<italics> From: J.S. Bach

To: R. Stone

Subject: Supplies for Brandenburgh Ltd. <end italics>

Dear Roll,

Please supply me with the following items:

<bullet> 3 boxes of paper-clips

<bullet> 2 boxes of staples

<bullet> a chair

Yours,

J.S.

Η γενική κωδικοποίηση (2)

- Τα πλεονεκτήματα της γενικής κωδικοποίησης είναι ότι:
 - είναι πιο εύκολο να γίνουν αλλαγές
πχ. εύκολα μπορεί να αλλάξει ο ορισμός της βούλας από '•' to '☞'
 - Προσθέτει αυτοματισμό
πχ. βρές όλες τις φορές που το “R. Stone” παρουσιάζεται με Italics
- Μειονεκτήματα:
 - Όπως και στο Procedural markup διατηρείται μόνο η **μορφή – όχι η δομή**

Γενικευμένη μεταγλώσσα (generalised markup)

- Η γενικευμένη μεταγλώσσα (generalised markup) περιγράφει την δομή και όχι απλά την μορφή

<memo>

<header> From: J.S. Bach

To: R. Stone

Subject: Supplies for Brandenburgh Ltd. **<end header>**

<body> Dear Roll,

Please supply me with the following items:

<list><list-item> 3 boxes of paper-clips

<list-item> 2 boxes of staples

<list-item> a chair

<end list>

<signature> Yours,

J.S. **<end signature><end body><end memo>**

Διαφορές τύπων markup (1)

- Η διαφορά μεταξύ **διαδικαστικής μεταγλώσσας** (*procedural markup*) και **δηλωτικής μεταγλώσσας** (*declarative markup*) είναι ότι:
 - η **διαδικαστική** μεταγλώσσα περιγράφει **τι πρέπει να γίνει**
<Display the character '☞' before this item> Item 1
ενώ
 - η **δηλωτική** μεταγλώσσα περιγράφει **τι είναι το περιεχόμενο**
<This is a list item> Item 1
 - Και τα δύο οδηγούν στο ίδιο αποτέλεσμα αλλά με διαφορετική ευελιξία

Διαφορές τύπων markup (2)

- Η δηλωτική μεταγλώσσα (declarative markup) είναι πιο εύκολη στη χρήση της σε σχέση με την διαδικαστική μεταγλώσσα (procedural markup) εφόσον – *“the authors will normally be removed from the printing/display process”*
- ενώ η μεταγλώσσα περιεχομένου είναι πιο δύσκολη για τον Η/Υ αλλά πιο εύκολη στην χρήση της από τον άνθρωπο

SGML (1)

- Η SGML (Standardised General Markup Language) είναι παράδειγμα μεταγλώσσας που αποτελεί μάλιστα και πρότυπο του ISO
- Δημιουργήθηκε από την IBM προκειμένου να λυθεί το πρόβλημα της μη τυποποιημένης εμφάνισης κειμένων σε διάφορα υπολογιστικά συστήματα
- Δεν διαθέτει δικό της περιβάλλον και μπορεί να χρησιμοποιηθεί οποιοσδήποτε επεξεργαστής κειμένου (editor) για να γράψετε SGML
- Αποτελεί την **βάση** της HTML (Hypertext Mark-up Language) and και μητέρα της XML
- είναι **δηλωτική μεταγλώσσα** και όχι διαδικαστική
- είναι **μεταγλώσσα περιεχομένου**

SGML (2)

- Προσδίδοντας την δομή και όχι απλά την μορφή ενός αρχείου:
 - Η πηγή μπορεί να χρησιμοποιηθεί από διάφορες πλατφόρμες
 - Η χρήση νέων συστημάτων δεν επιφέρει υποβιβασμό δεδομένων, καθώς τα νέα συστήματα μπορούν να χρησιμοποιήσουν τις νέες τεχνικές εμφάνισης για να παρουσιάσουν καλύτερα την ίδια δομή του αρχείου
 - Μπορεί να γίνει αναζήτηση βασισμένη στο περιεχόμενο αυξάνοντας την επιτυχία των συστημάτων αναζήτησης
 - Μεγάλοι αριθμοί εγγράφων μπορούν να ρυθμιστούν και να αναλυθούν σαν είναι μια βάση δεδομένων

SGML (3)

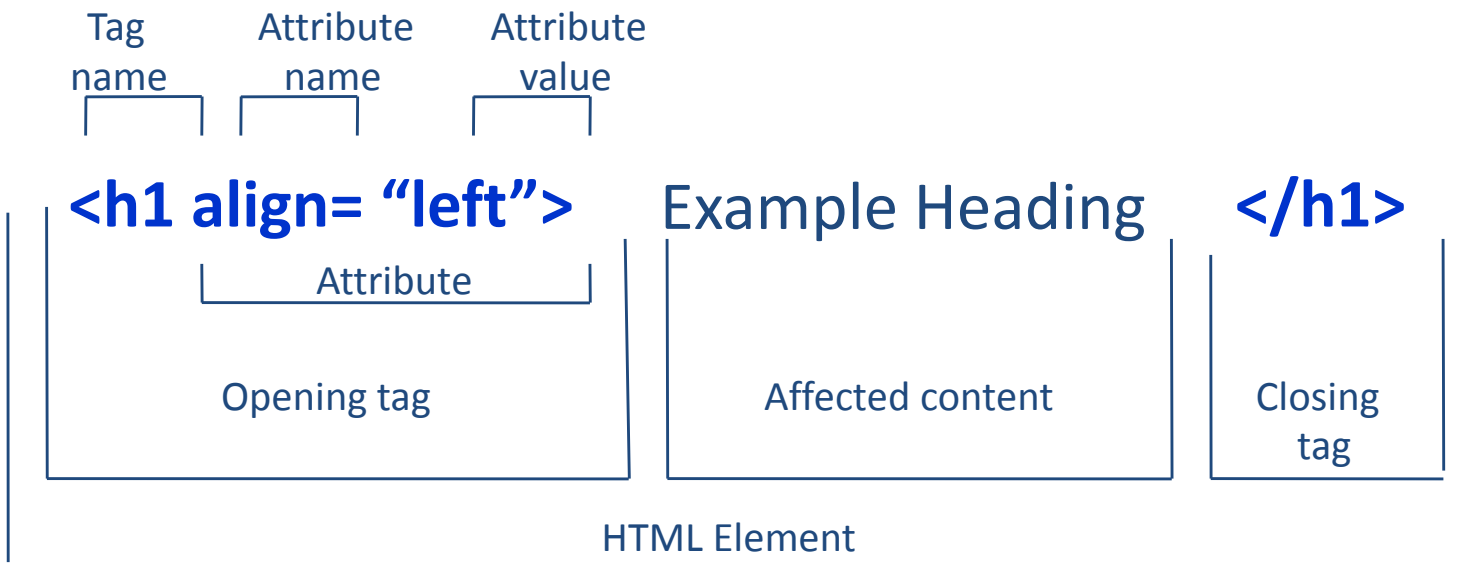
- Η SGML δεν προσφέρει την μεταγλώσσα αυτή καθαυτή
- Η SGML προσφέρει το **περιβάλλον** και τις **τεχνικές** για να γραφτούν οι κανόνες **που υπαγορεύουν την δομή** των αρχείων ή πολύ συχνά μία κάστα αρχείων

HTML (1)

- Η HTML (Hypertext Markup Language) χρησιμοποιείται για την δημιουργία ιστοσελίδων στο Διαδίκτυο
- Η HTML είναι περιγραφική γλώσσα που στηρίζεται στην πρότυπη SGML, στην ουσία είναι ένα DTD της SGML
- Τα αρχεία HTML είναι αρχεία κειμένου που περιέχουν εκτός από το κείμενο και ετικέτες (tag) που μορφοποιούν την παρουσίαση του κειμένου
- Η δημιουργία των αρχείων HTML γίνεται με απλούς συντάκτες κειμένου (notepad) ή με εξειδικευμένα προγράμματα συγγραφής (authoring tools, πχ. Dreamweaver, FrontPage)
- Η HTML χρησιμοποιεί ένα αριθμό από ετικέτες για την μορφοποίηση του κειμένου, την δημιουργία συνδέσμων, εισαγωγή ήχου κτλ.

HTML (2)

- οι ετικέτες της HTML περικλείονται από τα σύμβολα < >
- ο χαρακτήρας / δηλώνει το τέλος της ετικέτας



Παράδειγμα HTML κώδικα

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Translation //EN">  
<html>  
  <head>  
    <title> title content </title>  
  </head>  
  <body>  
    main body content  
  </body>  
</html>
```

XML (1)

- Η XML (eXtensible Markup Language) αναπτύχθηκε από τον World Wide Web Consortium (W3C) για να ξεπεράσει τους περιορισμούς της HTML – που σημαίνει ότι εισάγει νέες δυνατότητες
- Η XML είναι ένα υποσύνολο της SGML και χρησιμοποιεί επίσης ετικέτες

XML (2)

- Η διαφορά της XML με την HTML είναι ότι:
 - Η HTML διευκρινίζει την σημασία της κάθε ετικέτας και προσδιορίζει το πως θα εμφανιστεί σε κάθε πλοηγητή το κείμενο που περιλαμβάνεται σ' αυτήν
 - Η XML χρησιμοποιεί ετικέτες μόνο για να οριοθετήσει κομμάτια δεδομένων και αφήνει την ερμηνεία των δεδομένων στην εφαρμογή που τα διαβάζει
Έτσι ενώ η ετικέτα <p> μπορεί να σημαίνει (p= paragraph) στην HTML, στην XML μπορεί να σημαίνει (p=parameter), (p=price), (p=person)
- Ένα δεύτερο χαρακτηριστικό της XML είναι ότι αυστηρή σε σχέση με τις ετικέτες
 - Αυτό σημαίνει ότι η παράλειψη μιας ετικέτας ή ενός γνωρίσματος δίχως τα σύμβολα (< >) καθιστά άχρηστο το αρχείο XML
 - ενώ η HTML επιτρέπει τέτοιου είδους παραλείψεις

Περιεχόμενα ενότητας

- Δομή κειμένου
- Κωδικοποίηση
- Γραμματοσειρές
- Μορφοποίηση
- Αναγνωσιμότητα/Προσβασιμότητα
- Γλώσσες επισήμανσης/Μεταγλώσσες

Τέλος Ενότητας



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης