



Γλώσσες Σήμανσης – Markup Languages

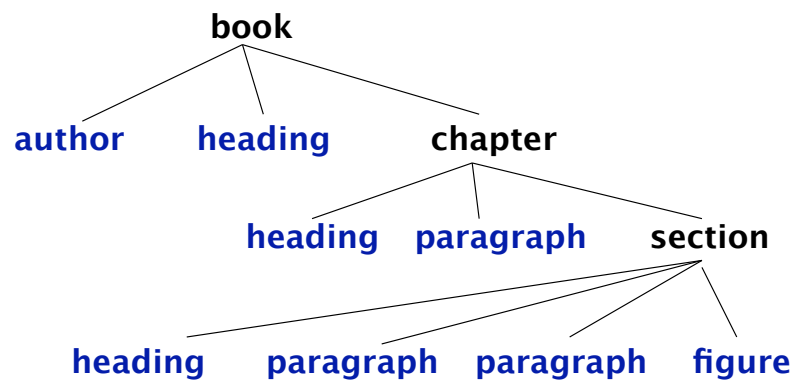
Στοιχεία SGML



- SGML (Standard Generalized Markup Language): Προδιαγραφή Γενικών Γλωσσών Σήμανσης.
- Πρότυπο ISO 8879.
- Οι γλώσσες «σήμανσης» καθορίζουν σημαντήρες (tags) και κανόνες για εισαγωγή των σε έγγραφα (documents).
- Έγγραφο (document):
 - Περιεχόμενο (content) - πληροφορίες
 - Λογική δομή (Ιεραρχική-δενδρική): σύνολο από λογικά στοιχεία (elements, logical units), οργανωμένα σε ένα «δένδρο εγγράφου» (document tree).

2

Δένδρο Εγγράφου



3

Έγγραφα στην SGML (SGML Documents)



- Η αναπαράσταση της δομής ενός εγγράφου γίνεται με την οριοθέτηση τμημάτων του περιεχομένου που συνιστούν ενιαία λογικά στοιχεία, με τους αντίστοιχους σημαντήρες, σύμφωνα με τα πρότυπα της SGML. Π.χ.:

```
<book> <author>Erik Wilde</author>
<heading>Wilde's WWW</heading>
<chapter id="sgml"> <heading>SGML</heading>
<paragraph>....
<section> <heading>Content & presentation</heading>....
</section>
```
- Με την εισαγωγή τους σε ένα έγγραφο οι σημαντήρες:
 - Αναπαριστάνουν ρητά (επισημαίνουν) τη λογική δομή του εγγράφου.
 - Επιτρέπουν τον διαχωρισμό του περιεχομένου (content) από την παρουσίαση (formatting) του εγγράφου.

4

Περιεχόμενο και Παρουσίαση Εγγράφων



- Περιεχόμενο Εγγράφου: περιλαμβάνει την λογική δομή ενός εγγράφου, το πληροφοριακό περιεχόμενο των δομικών στοιχείων του. Η λογική δομή του εγγράφου καθορίζεται από το είδος του εγγράφου (βιβλίο, άρθρο, περιοδικό κλπ)
- Παρουσίαση Εγγράφου: ο τρόπος της φυσικής παρουσίασης ενός εγγράφου, σε σελίδες χαρτιού, στην οθόνη ενός υπολογιστή, ή με οποιονδήποτε άλλο τρόπο. Συνήθως υπάρχουν κανόνες για το πώς θα παρουσιαστεί ένα έγγραφο, ανεξάρτητοι από το περιεχόμενό του. Π.χ.:
 - αρίθμηση παραγράφων
 - μορφοτύπηση παραγράφων
 - θέση υποσημειώσεων
- Παράλληλα με την SGML, η χρήση της Document Style Semantics and Specification Language μπορεί να καθορίσει την «παρουσίαση» του περιεχομένου ενός εγγράφου.

5

Δόμηση δια της Σημάνσεως



- Σήμανση εγγράφων (markup): η ανάμειξη περιεχομένου με πληροφορίες για το περιεχόμενο μέσα στο ίδιο έγγραφο.
- Η σήμανση γίνεται κατά τρόπον ώστε ο αναγνώστης ενός εγγράφου (άνθρωπος ή μηχανή) να μπορεί να ξεχωρίσει χωρίς αμφιβολία το κείμενο από τον «σχολιασμό» του.
- Η δομική πληροφορία καθορίζεται υπό μορφή στοιχείων (elements), τα οποία έχουν μοναδικό όνομα (π.χ. chapter ή heading).
- Για τον διαχωρισμό του στοιχείων σήμανσης από το κείμενο-περιεχόμενο, τα στοιχεία τοποθετούνται μεταξύ οριοθετών σήμανσης (markup delimiters).
- Ένα στοιχείο σήμανσης πλαισιωμένο από οριοθέτες σήμανσης λέγεται **σημαντήρας** (tag).

6

Παράδειγμα SGML



```
<book><author>Erik Wilde</author>  
<heading>Wilde's WWW</heading>  
<chapter id="sgml"><heading>SGML</heading>  
<paragraph>....  
<section><heading>Content & presentation</heading>....  
</section>
```

- Οι σημαντήρες βοηθούν στην αναγνώριση της δομής του εγγράφου.
- Κάθε στοιχείο σήμανσης έχει διαφορετική δομή.
- Επιπρόσθετη πληροφορία μπορεί να ενταχθεί σε ένα στοιχείο σήμανσης με ορισμό **κατηγορημάτων** (**attributes**) για τον εναρκτήριο σημαντήρα του.

7

Document Type Definition (DTD)

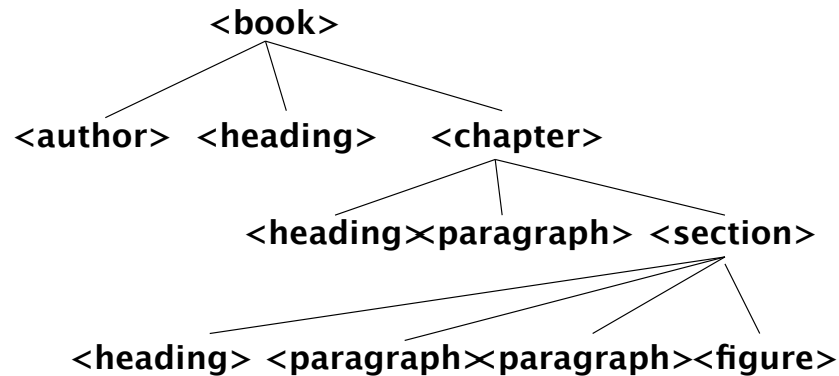


- Το DTD καθορίζει τους κανόνες σύμφωνα με τους οποίους δομείται ένα έγγραφο κωδικοποιημένο σύμφωνα με την SGML.
- Ομοειδή έγγραφα της SGML, κωδικοποιούνται σύμφωνα με ένα κοινό DTD, πράγμα που διευκολύνει την ηλεκτρονική τους επεξεργασία (αυτόματος έλεγχος ορθότητας, επίλυσης κανόνων παρουσίασης κοκ)
- Π.χ. DTD βιβλίου:

```
<!ELEMENT book (author, heading,chapter+) >  
<!ELEMENT chapter (heading, (paragraph|figure)*,section*)>  
<!ELEMENT section (heading, (paragraph|figure)*) >  
<!ATTLIST (chapter|section)  
id CDATA #IMPLIED>  
<!ELEMENT author (#PCDATA) >
```

8

Ιεραρχική Δομή SGML Εγγράφων



9

Χρησιμότητα των DTD



- Δημιουργία Εγγράφων (με σωστή δόμηση)
- Επικύρωση Εγγράφων (validation)

10

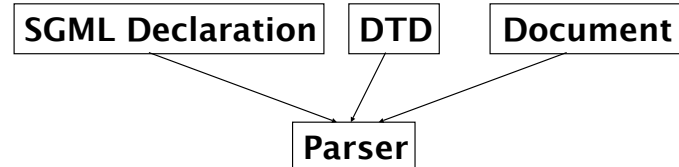
Παρουσίαση Εγγράφου



1. Επικύρωση Εγγράφου σύμφωνα με το DTD του.
2. Υποκατάσταση των στοιχείων σήμανσης με εντολές μορφοτύπησης (για το χρησιμοποιούμενο σύστημα παρουσίασης).
3. Έλεγχος συγκειμένου (αλλιώς εμφανίζεται ένα heading μέσα σε section κι αλλιώς μέσα σε chapter).
4. Επιπρόσθετες εντολές επεξεργασίας (π.χ. Από τα heading δημιουργείται ευρετήριο).

11

Σύστημα Επεξεργασίας SGML: Γενική Μορφή



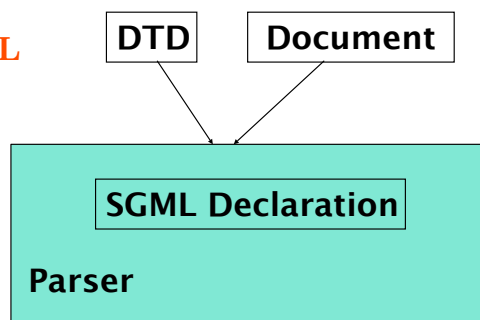
- **SGML Declaration:** βασικοί ορισμοί και συμβάσεις πάνω στις οποίες στηρίζεται η συγγραφή ενός έγκυρου DTD (π.χ. Λεπτομέρειες για την αναπαράσταση των σημαντήρων).
- Χρησιμοποιείται από τον parser για την ανάλυση του DTD και του εγγράφου SGML.

12

Εξειδικευμένος Συντακτικός Αναλυτής SGML



Π.χ.: XML

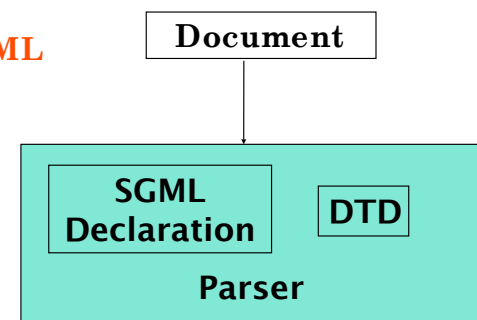


13

Εξειδικευμένος Συντακτικός Αναλυτής SGML



Π.χ.: HTML



14

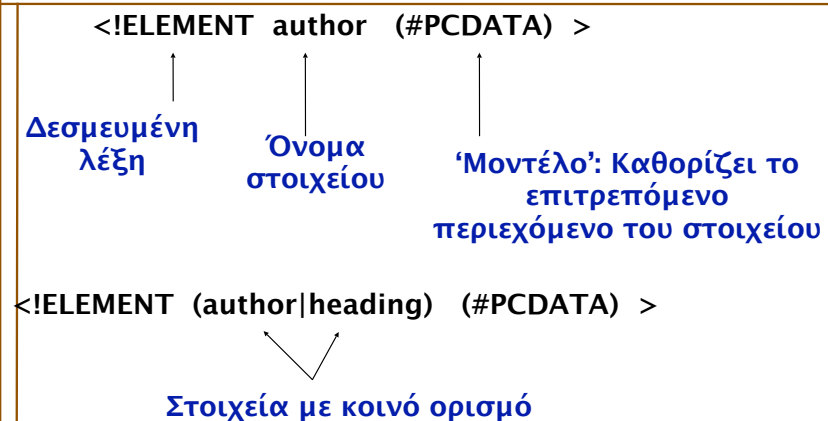
Η γραμματική των DTD



- Καθορίζει μια κατηγορία (κλάση) εγγράφων που ακολουθούν μια κοινή γραμματική.
- Προσδιορίζει:
 - Τα λογικά στοιχεία μια κλάσης εγγράφων: ένα έγγραφο σε SGML είναι ένα σύνολο από στοιχεία (elements), καθένα εκ των οποίων ανήκει σε μια συγκεκριμένη κλάση. Στο DTD ορίζονται επίσης τα κατηγορήματα (attributes) τα οποία αντιστοιχούν σε κάθε στοιχείο.
 - Οι κανόνες βάση των οποίων συνδυάζονται τα στοιχεία, καθορίζοντας μια δενδρική ιεραρχία.

15

Ορισμός Στοιχείων SGML



16

Μοντέλα Ομάδων (Model Groups)



- Στον ορισμό ενός στοιχείου SGML, έπονται του ονόματος του στοιχείου και καθορίζουν το περιεχόμενο το οποίο επιτρέπεται να περικλεισθεί από τους σημαντήρες του εν λόγω στοιχείου.
- Connectors (“ , ” , “&” , “|”):
<!ELEMENT book (front,body,trailer)>
<!ELEMENT front (author & title & date)>
<!ELEMENT author (editor|writer)>
- Occurrence Indicators:
<!ELEMENT body (chapter+)>
<!ELEMENT trailer (appendix*, glossary?, index)>
- Στοιχεία κειμένου: <!ELEMENT emphasis (#PCDATA)>

17

Μοντέλα Ομάδων



<!ELEMENT book (front,body,trailer)>

- Και τα τρία στοιχεία, με την καθορισμένη σειρά

<!ELEMENT front (author & title & date)>

- Και τα τρία στοιχεία με οποιαδήποτε σειρά

<!ELEMENT author (editor|writer)>

- Η' το ένα ή το άλλο στοιχείο

18

Εξαιρέσεις (exceptions)



- Η SGML προσφέρει επιπρόσθετους μηχανισμούς προσδιορισμού στοιχείων που συμπεριλαμβάνονται ή αποκλείονται από το περιεχόμενο ενός στοιχείου.
- **Συμπερίληψη (inclusion):** στοιχεία που συμπεριλαμβάνονται μπορούν να χρησιμοποιηθούν σε οποιοδήποτε σημείο σε μια ομάδα μοντέλων (είτε στο στοιχείο είτε σε οποιοδήποτε στοιχείο συμπεριλαμβάνεται στα περιεχόμενά του):
<!ELEMENT chapter (heading,(paragraph|table)*
+(emphasis) >
- **Αποκλεισμός (exclusion):**
<!ELEMENT emphasis (#PCDATA) -(emphasis) >

19

Κατηγορήματα (attributes)



<!ATTLIST **reference** **id** CDATA **#REQUIRED**
type (section|page) **section**>

Όνομα στοιχείου για το οποίο καθορίζουμε τα κατηγορήματα

Δεσμευμένη λέξη

Όνόματα κατηγορημάτων

Default value

Υποχρεωτικό κατηγορημα

- **Ορισμός κατηγορηματος:**
Attribute-name declared-value default-value
- **Declared-value:**
 - Λίστα από τιμές που μπορεί να λάβει το κατηγορημα
 - Reserved name, ενδεικτικό του τύπου της τιμής που επιτρέπεται να λάβει το κατηγορημα

20

Reserved Names



- **CDATA**: προσδιορίζει κατηγορηματικά που δέχεται ως τιμή συμβολοσειρά χαρακτήρων SGML.
- **ID**: προσδιορίζει ότι στο αντίστοιχο στοιχείο δίνεται μία μοναδική ταυτότητα – κάθε χρήση του στοιχείου σε ένα έγγραφο, μπορεί να ταυτισθεί κατά μοναδικό τρόπο.
- **IDREF**: κατηγορηματικά αυτής της μορφής παραπέμπει σε στοιχείο του ίδιου εγγράφου μέσω της ταυτότητας του στοιχείου.
- **IDREFS**: κατηγορηματικά που αποτελείται από λίστα IDREF.
- **NAME**: έγκυρο όνομα SGML
- **NUMBER**: προσδιορίζει ότι το κατηγορηματικά δέχεται αριθμητικές τιμές μόνο.

21

Reserved Names



- **#IMPLIED**: αν στο έγγραφο δεν καθορίζεται τιμή για το συγκεκριμένο κατηγορηματικά, τότε ο αναλυτής (parser) μπορεί να αναθέσει στο κατηγορηματικά μια υπονοούμενη, προκαθορισμένη τιμή.
- **#REQUIRED**: καθορίζει ότι είναι υποχρεωτική η εισαγωγή τιμής για το συγκεκριμένο κατηγορηματικά, στον εναρκτήριο σημαντήρα του αντίστοιχου στοιχείου.
- **#FIXED**: χρησιμοποιείται όταν η τιμή του κατηγορηματικού είναι σταθερή.

22

Οντότητες (entities)



- Μια συλλογή από χαρακτήρες, στην οποία μπορεί να γίνει αναφορά σαν αδιάσπαστο σύνολο.
- Δήλωση οντοτήτων:

Entity Reference

```
<!ENTITY Sigma CDATA “&#931;”>
```

Entity name optional

Replacement text

23

Αναφορά σε οντότητες (entity reference)



- Αναφορές σε γενικές οντότητες: αφορούν οντότητες καθορισμένες στο DTD του εγγράφου. Αποτελούνται από:

&entity1;

Entity name

Entity reference open

Entity reference close

- Αναφορές σε οντότητες χαρακτήρων εκτός πληκτρολογίου (π.χ. Ελληνικών κλπ). Π.χ. Σ

24

Έγγραφα SGML



- Θεωρητικά, η ανταλλαγή ενός εγγράφου SGML πρέπει να συνοδεύεται από την ανταλλαγή του SGML declaration και του DTD.
- Αφού η XML και η HTML είναι γλώσσες σήμανσης με προκαθορισμένη δήλωση SGML, η ανταλλαγή εγγράφων αυτού του τύπου δεν προϋποθέτει και την ανταλλαγή της δήλωσης SGML.
- Η HTML διαθέτει προκαθορισμένο DTD, κι έτσι η ανταλλαγή αρχείων HTML **δεν** απαιτεί ούτε την ανταλλαγή DTD.
- Επειδή ωστόσο στο ΠΠΠ συνυπάρχουν διαφορετικές εκδόσεις της HTML, είναι σκόπιμο στο προίμιο ενός εγγράφου HTML να δηλώνεται η ακολουθούμενη έκδοση.