SGML (Standard Generalized Markup Language)

Γλώσσα για την περιγραφή της δομης και του περιεχομένου ηλεκτρονικών κειμένων

Πλήρης – Σύνθετη – Δύσκολη στην εκμάθηση και την χρήση

↓

HTML (HyperText Markup Language)

Υποσύνολο της SGML – Χρήση στο WEB

Έμφαση στην παρουσίαση των δεδομένων και όχι στο περιεχόμενο – Σταθερό σύνολο ετικέτων (tag set)

– Καθορισμένη σημασία ετικετών – Αδυναμία επεκτασιμότητας (Ορισμός νέων ετικετών)

↓

XML (eXtensible Markup Language)

Συνδυαζει την δυναμικότητα της SGML με την απλοτητα της HTML

Δεν υπαρχουν ετικετες ορισμενες από την γλωσσα - Ο χρήστης οριζει τις δικες του ετικετες (πλήρως επεκτάσιμη)

Οι ετικέτες δεν έχουν ερμηνεία από μόνες τους – Ο χρήστης τους δίνει σημασία ανάλογα με την εφαρμογή

Οι ετικέτες περιγραφουν τα δεδομένα τα οποία περικλείουν (μεταδεδομένα – μεταγλώσσα)

Δενδρική δομή των ετικετών στο κείμενο (tag nesting) – Οι ετικέτες κατάλληλα φωλιασμένες

Υποστηρίζει το σύνολο χαρακτήρων UNICODE

Παράδειγμα XML αρχείου

```
<?xml version="1.0" standalone="yes" encoding="Windows-1253"?>
<!- Example of a simple XML file -->
<example>
<student code="p3960134">
     <name>
        <firstname>Ιωάννης</firstname>
        <lastname>Κάποιος</lastname>
     </name>
     <department έτος="4">Πληροφορική</department>
     <address>
        <street>Πατησίων 238</street>
        <city zipcode="12356">Αθήνα</city>
     </address>
</student>
</example>
```

DTD (Document Type Definition)

Ορίζει την δομή(τον τύπο) ενός αρχείου XML, περιέχει δηλ. κάποιες δηλώσεις αναφορικά με:

•Το root-element (Κάθε XML αρχείο έχει ένα root-element)

•Ποια elements επιτρέπονται να εμφανίζονται και τον τύπο τους (element – mixed content)

•Σχέση μεταξύ των elements (element nesting)

•Τα attributes του κάθε element και τον τύπο τους

Το DTD μπορεί να είναι :

•Internal DTD στην αρχή του αρχείου (standalone="yes")

•External DTD (standalone="no")

Δεν είναι απαραίτητο -- Είναι όμως σημαντικό :

•για την τεκμηρίωση του εγγράφου

•για την ανταλλαγή δεδομένων μεταξύ βάσεων δεδομένων, και

•για την παροχή μετά-δεδομένων (meta-data) στον XML parser

Εγκυρότητα αρχείων XML

Well-formed -- το αρχείο ακολουθεί τους κανόνες σύνταξης της γλώσσας (εκφράζονται σε μορφή EBNF)

Γενικά αυστηρό συντακτικό

Πρέπει για κάθε στοιχείο να υπάρχει start-tag και end-tag

Τα στοιχεία (elements) φωλιάζουν σωστά το ένα μέσα στο άλλο και δεν υπάρχει επικάλυψη

Οι τιμές των attributes περικλείονται μέσα σε " "

Valid -- ενα well-formed αρχείο το οποίο έχει ένα κατάλληλο DTD του οποίου ακολουθεί τις δηλώσεις και τους περιορισμούς

XML Parsers

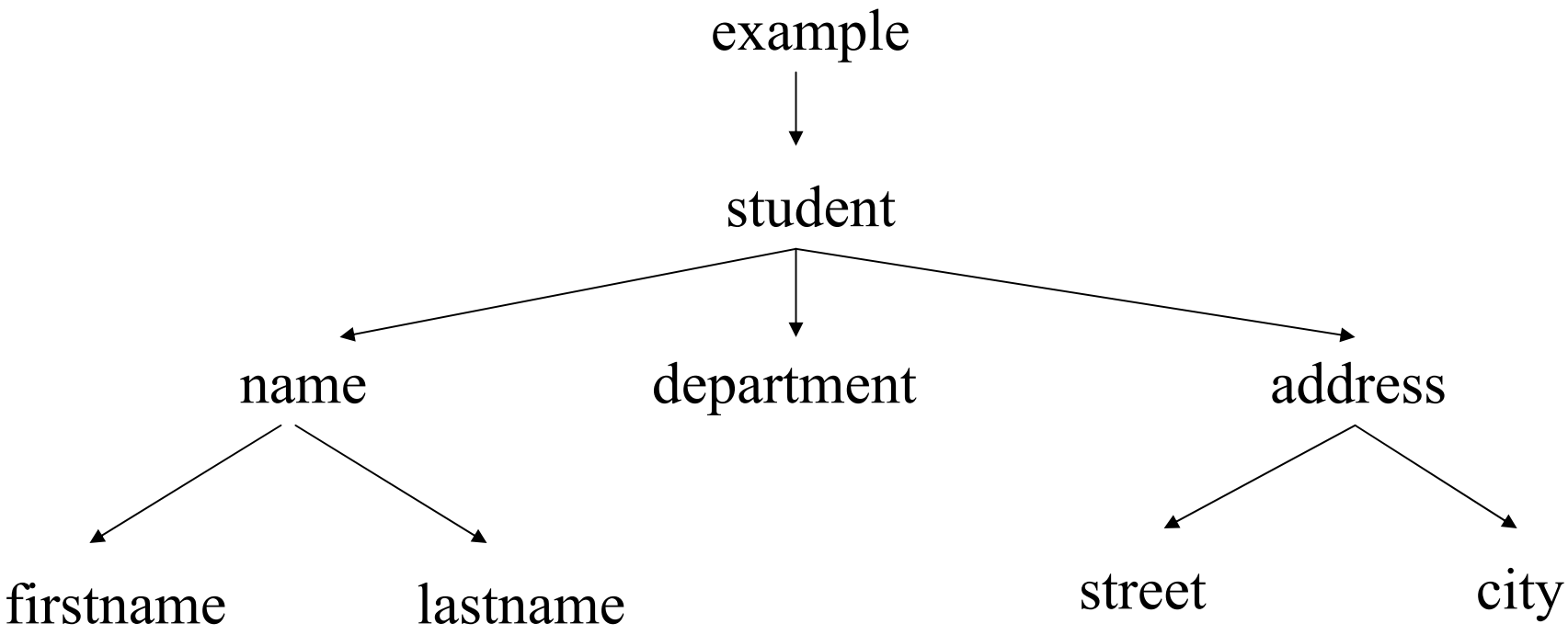
Εργαλεία τα οπόια δέχονται ως είσοδο το XML έγγραφο μαζί με το τυχόν DTD και ελέγχουν την εγκυρότητά του σύμφωνα με τους κανόνες της γλώσσας και τις δηλώσεις που περιέχονται στο DTD

Validating Ελέγχει εάν το αρχείο είναι well-formed και valid εξετάζοντας όλα τα DTD (internal-external)

Non-validating Εξετάζει μόνο το internal DTD και ελέγχει μόνο εάν είναι well-formed

Χρησιμοποιούν τις πληροφορίες του DTD για να κατασκευάσουν το δεντρικό μοντέλο ανάλυσης των δεδομένων ενός κειμένου XML (parse-tree) και μεταβιβάζουν τα δεδομένα του εγγράφου XML στην εφαρμογή του χρήστη για επεξεργασία

Το Δέντρο ανάλυσης των δεδομένων για το προηγούμενο παράδειγμα :



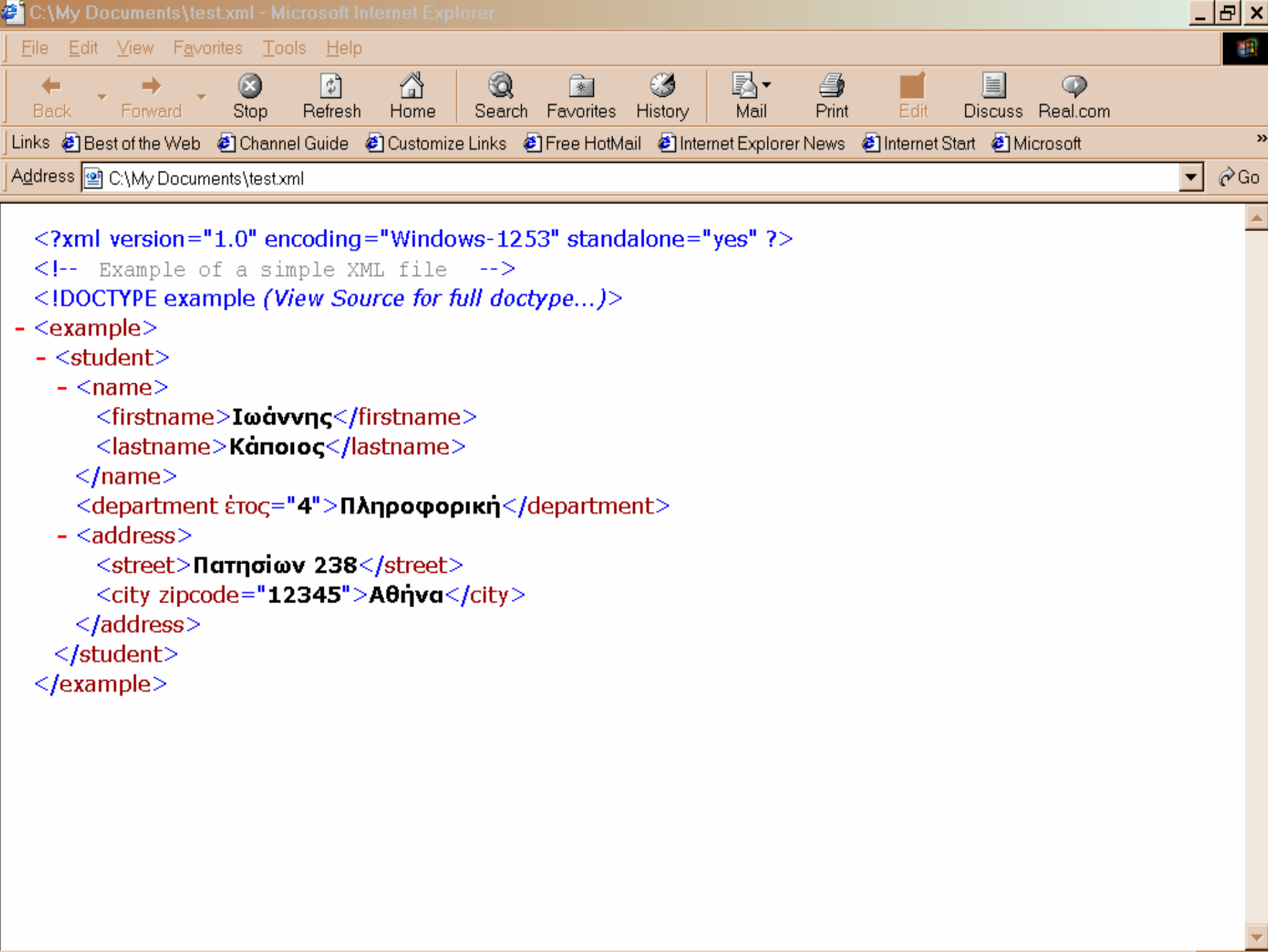
Schema : Επέκταση του DTD (document-centric)

•Προσφέρει δυνατότητες ορισμού τύπων δεδομένων

•Συνεισφέρει στην ανταλλαγή δεδομένων μεταξύ οργανισμών που έχουν υιοθετήσει το ίδιο Schema για τις συναλλαγές τους.

Το DTD για το παράδειγμά μας :

```
<!DOCTYPE example [
<!ELEMENT example (student)*>
<!ELEMENT student (name, department, address)>
<!ATTLIST student code ID #REQUIRED>
<!ELEMENT name (firstname, lastname)>
<!ELEMENT firstname (#PCDATA)>
<!ELEMENT lastname (#PCDATA)>
<!ELEMENT department (#PCDATA)>
<!ATTLIST department έτος CDATA>
<!ELEMENT address (street, city)>
<!ELEMENT street (#PCDATA)>
<!ELEMENT city (#PCDATA)>
<!ATTLIST city zipcode CDATA>
]>
```


C:\My Documents\test.xml - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Favorites History Mail Print Edit Discuss Real.com
Links Best of the Web Channel Guide Customize Links Free HotMail Internet Explorer News Internet Start Microsoft
Address C:\My Documents\test.xml
Go
<?xml version="1.0" encoding="Windows-1253" standalone="yes" ?>
<!-- Example of a simple XML file -->
<!DOCTYPE example (View Source for full doctype...)>
- <example>
- <student>
- <name>
<firstname>Ιωάννης</firstname>
<lastname>Κάποιος</lastname>
</name>
<department έτος="4">Πληροφορική</department>
- <address>
<street>Πατησίων 238</street>
<city zipcode="12345">Αθήνα</city>
</address>
</student>
</example>

APIs (Application Programming Interfaces) για XML Parsers

•**DOM (**Document Object Model) : tree – based

Κατασκευάζει ένα δέντρο που αντανακλά την δομή του εγγράφου XML και του οποίου οι κόμβοι είναι τα στοιχεία (elements) του εγγράφου. Το δέντρο μπορεί να προσπελαστεί με οποιαδήποτε γλώσσα (Java, C++, Perl κτλ)

•**SAX** (Simple API for Xml) : event - driven

Απλώς αναγνωρίζει events, όπως ετικέτες αρχής και τέλους, σχόλια κτλ. Ο χρήστης πρέπει προγραμματιστικά να ερμηνεύσει αυτά τα events και να κατασκευάσει το δικό του μοντέλο δεδομένων

Το SAX είναι πιο γρήγορο και λιγότερο απαιτητικό σε μνήμη , απαιτεί όμως περισσότερη δουλειά από τον χρήστη.

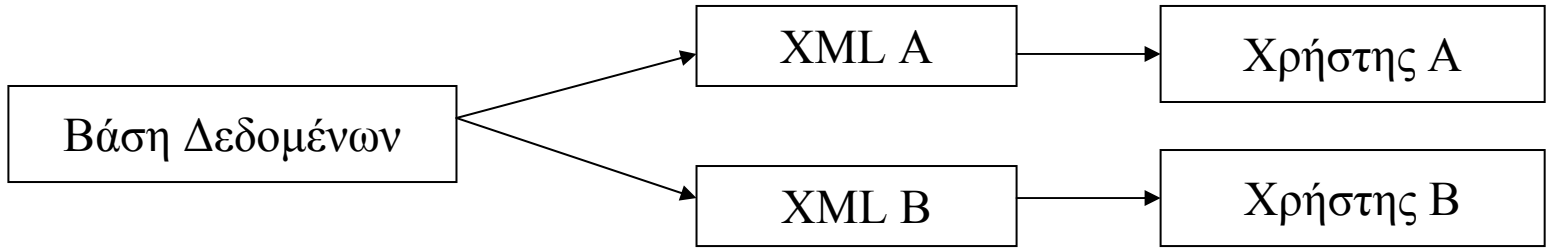
Πότε χρησιμοποιείται το καθένα;

DOM : Έγγραφα document-centric για χρήση από ανθρώπους (human-consumption), όπου μας ενδιαφέρει η δομή του εγγράφου και η σειρά που εμφανίζονται τα στοιχεία στο κείμενο.

SAX : Έγγραφα data-centric που έχουν παραχθεί ή θα χρησιμοποιηθούν από μηχανές (machine-readable), όπως κείμενα βασισμένα σε αποτέλεσμα κάποιας query .

Παραδείγματα Εφαρμογών της XML

1. Ενδίαμεσο επίπεδο για την ανταλλαγή δεδομένων μεταξύ ετερογενών βάσεων δεδομένων
2. Παρουσιάση των ίδιων δεδομένων με διαφορετική όψη σε διαφορετικούς χρήστες με βάση κάποιο κριτήριο.



3. Υλοποίηση ευφυών Web agents για την συλλόγη πληροφοριών στο Διαδίκτυο σχετικά με διάφορους χρήστες-επισκέπτες ιστοσελίδων με έναν τρόπο πιο προσωπικό χάρη στην ύπαρξη μετά-δεδομένων

Υπερδεσμοί (Hyperlinks) στην XML

- **Simple Link** : Παρόμοιος με τον σύνδεσμο <A> της HTML

```
<link xml:link="simple" href="locator">Link Text</link>
```

- **Extended Link (XLink) :** Εκφράζει σχέση μεταξύ περισσότερων από 2 πόρους (resources) , πιο ευέλικτος & ισχυρός από τον Simple Link

```
<elink xml:link="extended" role="annotation>
  <locator xml:link="locator" href="text.loc">The Text</locator>
  <locator xml:link="locator" href="annot.loc">Annotations</locator>
</elink>
```

- **Extended Pointer (XPointer) :** Αυθαίρετος εντοπισμός πόρων στο έγγραφο μέσω διάσχισης του ιεραρχικού δέντρου των δεδομένων του.

XML και Βάσεις Δεδομένων

Μεταφορά δεδομένων από την Βάση σε XML έγγραφο

2 τρόποι για την αντιστοίχιση (mapping) της δομής της βάσης στο αρχείο

1. **Template-driven** : Δεν υπάρχει προκαθορισμένη αντιστοίχιση, απλά εισάγονται εντολές SELECT σε ένα πρότυπο(template) και παράγεται το αρχείο. Χρησιμοποιείται για μεταφορά δεδομένων από σχεσιακές Β.Δ.

```
<FlightInnfo>
    <Intro>The following seats have available seats:</Intro>
    <SelStmt>SELECT Airline, FlNum, Depart FROM Flights</SelStmt>
    <Conclude>We hope one of these meets your needs</Conclude>
</FlightInfo>
```

Η εντολή SELECT θα αντικατασταθεί από τα αποτελέσματα:

```
<Flights>
<Row>
    <Airline>O.A.</Airline>
    <FlNum>123</FlNum>
    <Depart>May 25, 2000 15:45</Depart>
</Row> ...
</Flights>
```

XML και Βάσεις Δεδομένων

2. **Model-driven** : Ένα μοντέλο δεδομένων επιβάλλεται στην δομή του XML αρχείου σύμφωνα με την δομή της βάσης. Για σχεσιακές Β.Δ. :

```
<database>
  <table>
      <row>
         <column1>… </column1>
         <column2>… </column2> …
      </row> …
  </table> …
</database>
```

Για αντικειμενοστραφεις Β.Δ. χρησιμοποιειται ως μοντελο ένα δεντρο αντικειμενων στο οποίο τα elements αντιστοιχουν σε αντικείμενα και τα attributes και τα PCDATA στις ιδιότητες τους (properties).

Αξίζει να σημειωθεί ότι υπαρχουν επίσης και προϊόντα τα οποία χρησιμοποιούν αλγορίθμους για την δημιουργία DTD από το Σχήμα της (σχεσιακής) Βάσης και το αντίστροφο.