



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

ΑΝΟΙΚΤΑ ακαδημαϊκά
μαθήματα ΠΠ

Αναπαράσταση Γνώσης στον Παγκόσμιο Ιστό

Ενότητα 2: XML

Δομημένα Έγγραφα Ιστού, Μέρος 1^ο - XML

Ιωάννης Χατζηλυγερούδης

Πολυτεχνική Σχολή

Τμήμα Μηχ/κών Η/Υπολογιστών & Πληροφορικής

Περιεχόμενα ενότητας

Μέρος 1^ο - XML

1. Εισαγωγή
2. Η Γλώσσα XML
3. Δενδρικό Μοντέλο Δόμησης



Εισαγωγή

Εισαγωγή

- SGML (Standard Generalized Markup Language)
 - Διεθνές πρότυπο ορισμού μεθόδων αναπαράστασης πληροφοριών ανεξάρτητα από συσκευή/σύστημα και χρήση/ανάγνωση
 - Εφαρμογές SGML: Γλώσσες που υπακούουν στο πρότυπο αυτό. Π.χ.
 - HTML
 - XML



XML vs HTML (1 από 3)

```
<H2> Context-Dependent Reasoning</H2>
  <i>by <b>V. Marek</b> and <b>M. Truszczynski</b></i>
  <br>
  Springer 1993<br>
  ISBN 0387976892
```

```
<book>
  <title>Context-DependentReasoning</title>
  <author>V. Marek</author>
  <author>M. Truszczynski</author>
  <publisher>Springer</publisher>
  <year>1993</year>
  <ISBN>0387976892</ISBN>
</book>
```



XML vs HTML (2 από 3)

■ Ομοιότητες

- Είναι γλώσσες σήμανσης (markup)
- Χρησιμοποιούν ετικέτες (tags)
- Εύκολα αναγνώσιμες από άνθρωπο

■ Διαφορές

- Όλες οι ετικέτες στην XML πρέπει να κλείνουν (στην HTML όχι: π.χ.
)
- Η HTML είναι δύσκολα αναγνώσιμη από μηχανές
- Η HTML δεν περιέχει δομική πληροφορία (οντότητες-σχέσεις)
- Η XML επιτρέπει ορισμό περιορισμών σε τιμές (π.χ. έτος)



XML vs HTML (3 από 3)

- Διαφορές (συν.)
 - Οι ετικέτες στην HTML είναι προκαθορισμένες (αφορούν μορφοποίηση κειμένου)
 - Η XML ξεχωρίζει το περιεχόμενο από την μορφοποίηση
 - Η XML είναι μια μετα-γλώσσα για σήμανση-οι ετικέτες δεν είναι προκαθορισμένες, αλλά ορίζονται από τους χρήστες → **επεκτάσιμη** (MathML, BSML, HRML, AML, NewsML, IRML κλπ)



Η Γλώσσα XML

Γλώσσα XML

- Πρόλογος
- Στοιχεία (XML elements)
- Χαρακτηριστικά (XML attributes)
- Σχόλια
- Οδηγίες επεξεργασίας



Γλώσσα XML - Πρόλογος

- **Δήλωση XML (XML declaration)**

```
<?xml version="1.0" [encoding="UTF-16"?]>
```

```
<?xml version="1.0" [encoding="UTF-16"? standalone  
= "no"]>
```

- **Αναφορά σε εξωτερικά έγγραφα**

```
<!DOCTYPE book SYSTEM "book.dtd">
```

(οι δομικές πληροφορίες βρίσκονται στο τοπικό αρχείο book.dtd)



Γλώσσα XML - Στοιχεία (1)

- Αναφέρονται στις κύριες έννοιες (οντότητες) ενός εγγράφου XML
- Στοιχείο: `<tag-name>content</tag-name>`

Π.χ. `<lecturer> Ioannis Chatzis </lecturer>`

- Τα ονόματα ετικετών (tag-names) ορίζονται από τον χρήστη
 - Πρώτος χαρακτήρας γράμμα/underscore/”:"
 - Δεν μπορεί να αρχίζει με “xml” ή παραλλαγές του



Γλώσσα XML - Στοιχεία (2)

- Το περιεχόμενο (content) ενός στοιχείου μπορεί να είναι κείμενο ή άλλα στοιχεία ή τίποτα. Π.χ.

```
<lecturer>  
  <name>Ioannis Chatzis</name>  
  <phone>+302610994555</phone>  
</lecturer>
```

- Αν δεν υπάρχει περιεχόμενο (content), τότε το στοιχείο αποκαλείται κενό (*empty*). Π.χ.

```
<lecturer></lecturer> ή <lecturer/>
```



Γλώσσα XML - Χαρακτηριστικά (1)

- Ένα κενό στοιχείο μπορεί να περιέχει ορισμένες ιδιότητες με τη μορφή «χαρακτηριστικών»
- Χαρακτηριστικό: `<όνομα>=<τιμή>` στην ετικέτα ανοίγματος ενός στοιχείου (κενού ή μη). Π.χ.

```
name="Ioannis Chatzis", phone="+302610994555"/>
```

Κενό
στοιχείο

- Παραδείγματα:

```
<lecturer name="Ioannis Chatzis", phone="+302610994555"/>
```

```
<order orderNo="23456" customer=" John Doe"  
      date="May 02, 2015">  
  <item itemNo="a528" quantity="1"/>  
  <item itemNo="c817" quantity="3"/>  
</order>
```

Μη κενό
στοιχείο



Γλώσσα XML - Χαρακτηριστικά (2)

- Το προηγούμενο μπορεί να γραφεί και ως εξής (χωρίς χρήση χαρακτηριστικών, με ένθετα στοιχεία):

```
<order>
  <orderNo>23456</orderNo>
  <customer>John Doe</customer>
  <date>March 25, 2015</date>
  <item>
    <itemNo>a528</itemNo>
    <quantity>1</quantity>
  </item>
  <item>
    <itemNo>c817</itemNo>
    <quantity>3</quantity>
  </item>
</order>
```

Τα χαρακτηριστικά δεν μπορούν να είναι ένθετα.



Γλώσσα XML - Σχόλια

- Είναι κομμάτι κειμένου που αγνοείται από τον συντακτικό αναλυτή.

```
<!-- Αυτό είναι ένα σχόλιο -->
```



Γλώσσα XML - Οδηγίες Επεξεργασίας

- Παρέχουν ένα μηχανισμό μεταβίβασης πληροφοριών σε μια εφαρμογή σχετικά με το χειρισμό στοιχείων από αυτή.
- Γενική μορφή:

`<?target instruction ?>`

Π.χ. `<?stylesheet type="text/css" href="mystyle.css" ?>`



Κανόνες Δόμησης Εγγράφου XML

- Μόνο ένα εξωτερικό στοιχείο (στοιχείο ρίζα)
- Κάθε στοιχείο περιέχει μια ετικέτα ανοίγματος και μια αντίστοιχη κλεισίματος
- Δεν υπάρχουν επικαλύψεις ετικετών:

```
<author><name>Ioannis Chatzis</author></name>
```

- Τα χαρακτηριστικά σ' ένα στοιχείο έχουν μοναδικά ονόματα
- Τα ονόματα των στοιχείων και των ετικετών πρέπει να είναι επιτρεπτά

Δενδρικό Μοντέλο Δόμησης (1)

- Είναι εφικτό να αναπαριστούμε σωστά δομημένα έγγραφα XML σαν δένδρα
- Κανόνες:
 - Υπάρχει ακριβώς μια ρίζα
 - Δεν υπάρχουν κύκλοι
 - Κάθε κόμβος έχει ακριβώς ένα γονέα (πλην της ρίζας)
 - Κάθε κόμβος έχει μια ετικέτα
 - Η σειρά των στοιχείων είναι σημαντική (όχι όμως των χαρακτηριστικών)

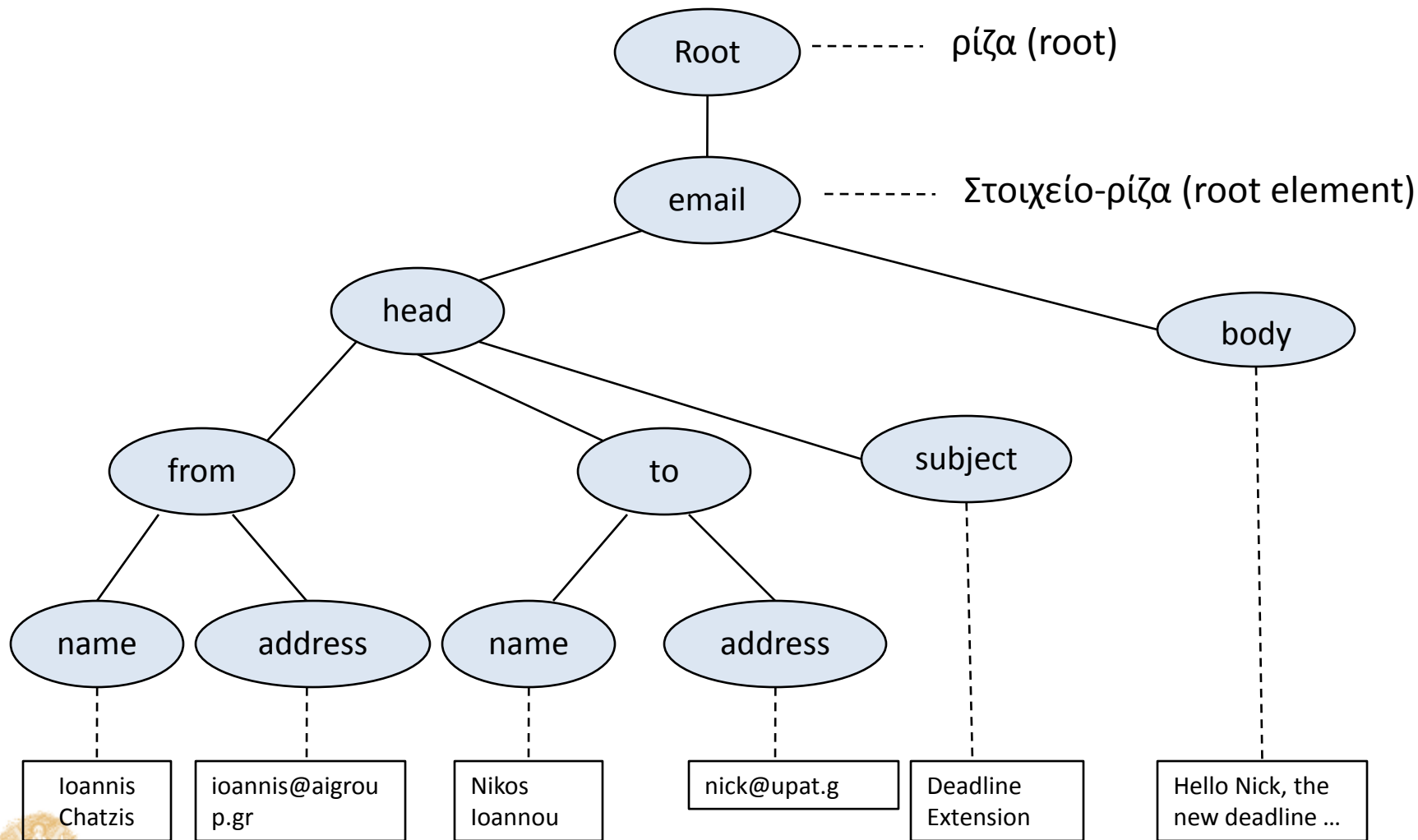


Δενδρικό Μοντέλο Δόμησης (2)

```
<?xml version="1.0" encoding="UTF-16"?>
<!DOCTYPE email SYSTEM "email.dtd">
<email>
  <head>
    <from name="Ioannis Chatzis"
      address="ioannis@aigroup.gr"/>
    <to name="Nikos Ioannou"
      address="nick@upat.gr"/>
    <subject>Deadline Extension</subject>
  </head>
  <body>
    Hello Nick, the new deadline for the project
    submission is 25/12/2022, Cheers
  </body>
</email>
```



Δενδρικό Μοντέλο Δόμησης (3)



Δενδρικό Μοντέλο Δόμησης (4)

- Κάποια προβλήματα

- Δεν αναπαρίσταται το γεγονός ότι η σειρά των στοιχείων είναι σημαντική ενώ των χαρακτηριστικών όχι.

Π.χ. τα `<person lastname= "Doe" firstname= "John"/>`
`<person firstname= "John" lastname= "Doe" />`
είναι ισοδύναμα.

- Δεν υπάρχει διάκριση τύπων κόμβων (στοιχείων, χαρακτηριστικών κλπ).



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αθηνών**» έχει χρηματοδοτήσει μόνο την 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.



Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0.



Σημείωμα Αναφοράς

Copyright: Πανεπιστήμιον Πατρών, Ιωάννης Χατζηλυγερούδης, 2015.
«Αναπαράσταση Γνώσης στον Παγκόσμιο Ιστό. XML». Έκδοση: 1.0. Πάτρα
2015. Διαθέσιμο από τη δικτυακή διεύθυνση:
<https://eclass.upatras.gr/courses/CEID1098/>



Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



[1] <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Ως **Μη Εμπορική** ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

Σημείωμα Χρήσης Έργων Τρίτων

- Οι διαφάνειες είναι κατά μεγάλο μέρος βασισμένες στο βιβλίο "Εισαγωγή στον Σημασιολογικό Ιστό", των Grigoris Antoniou και Frank van Harmelen, Β' Έκδοση, 2009 (Ελληνική Έκδοση).

