

ΑΝΑΠΑΡΑΣΤΑΣΗ ΓΝΩΣΗΣ ΣΤΟΝ ΠΑΓΚΟΣΜΙΟ ΙΣΤΟ

XML (e**X**tended **M**arkup
Language)

Ι. Χατζηλυγερούδης

ΕΙΣΑΓΩΓΗ

- 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

`<H2> Nonmonotonic Reasoning: Context-Dependent Reasoning </H2>`

`<i>by V. Marek and M. Truszczyński </i>`

`
`

`Springer 1993
`

`ISBN 0387976892`

`<book>`

`<title> Nonmonotonic Reasoning: Context-Dependent Reasoning </title>`

`<author>V. Marek</author>`

`<author>M. Truszczyński</author>`

`<publisher>Springer</publisher>`

`<year>1993</year>`

`<ISBN>0387976892</ISBN>`

`</book>`

XML vs HTML

- 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

■ Διαφορές (συν.)

- Οι ετικέτες στην HTML είναι προκαθορισμένες (αφορούν μορφοποίηση κειμένου, στόχος: παρουσίαση πληροφοριών)
- Η XML ξεχωρίζει το περιεχόμενο από την μορφοποίηση (στόχος: η παρουσίαση-χρήση πληροφοριών με διάφορους τρόπους)
- 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)

Γλώσσα XML

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

Γλώσσα XML

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

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

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

Στοιχεία (1)

- Αναφέρονται στις κύριες έννοιες/οντότητες ενός εγγράφου XML
- Στοιχείο: `<tag-name>content</tag-name>`
Π.χ. `<lecturer>David Hamilton</lecturer>`
- Τα ονόματα ετικετών (tag-names) ορίζονται από τον χρήστη
 - Πρώτος χαρακτήρας γράμμα/underscore/"':"
 - Δεν μπορεί να αρχίζει με "xml" ή παραλλαγές του

Στοιχεία (2)

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

```
<lecturer>
```

```
  <name>David Hamilton</name>
```

```
  <phone>+302610994555</phone>
```

```
</lecturer>
```

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

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

Γλώσσα XML

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

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="David Hamilton", phone="+302610994555"`

- Παραδείγματα:

`<lecturer name="David Hamilton", phone="+302610994555"/>`

Κενό
στοιχείο

`<order orderNo="23456" customer="John Smith"
date="October 15, 2002">`

`<item itemNo="a528" quantity="1"/>`

`<item itemNo="c817" quantity="3"/>`

`</order>`

Μη κενό
στοιχείο

Χαρακτηριστικά (2)

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

```
<order>
  <orderNo>23456</orderNo>
  <customer>John Smith</customer>
  <date>March 24, 2009</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

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

Σχόλια

- Είναι κομμάτι κειμένου που αγνοείται από τον συντακτικό αναλυτή.
<!-- Αυτό είναι ένα σχόλιο -->

Γλώσσα XML

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

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>Lee Hong</author></name>
```

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

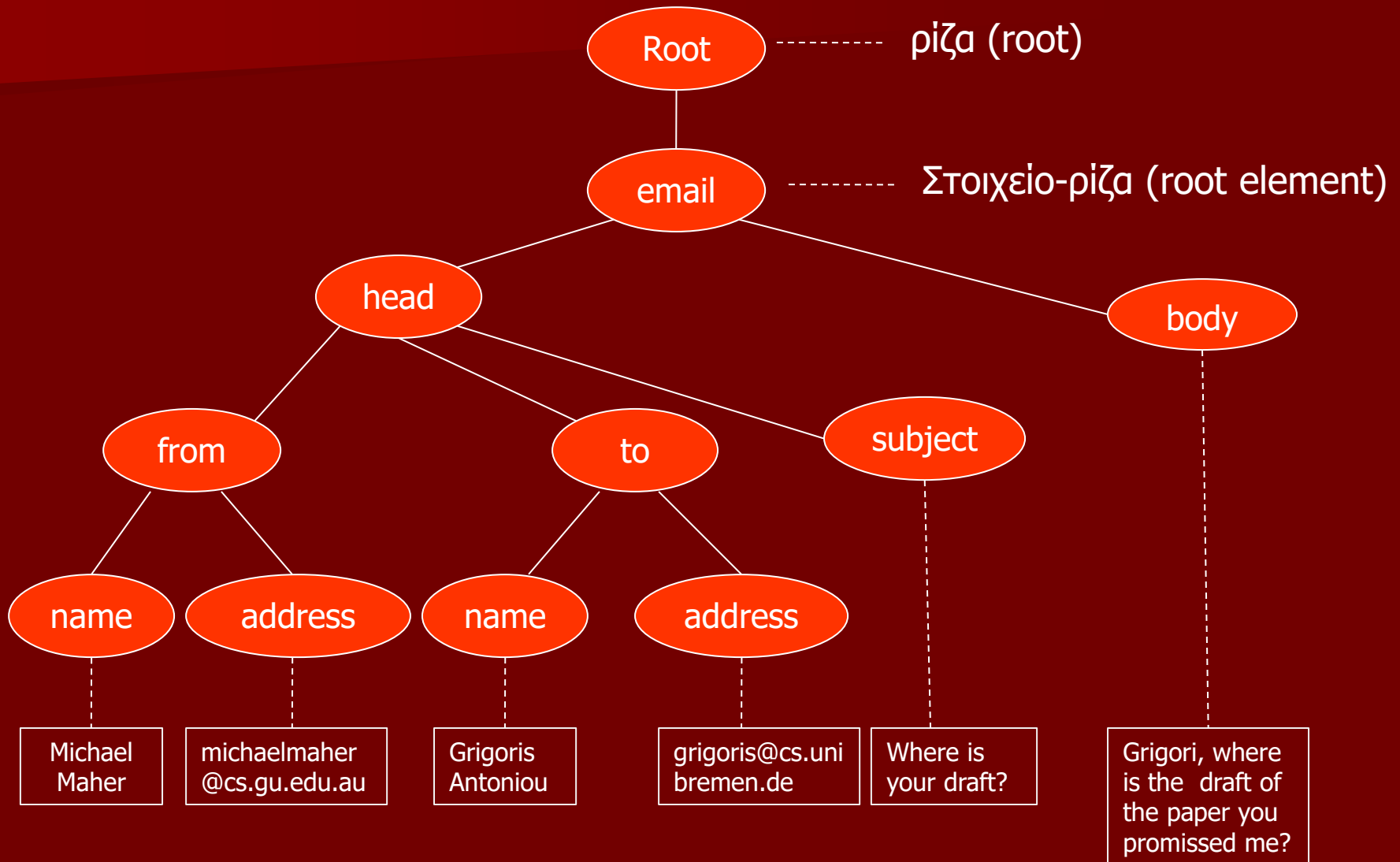
Δενδρικό Μοντέλο Δόμησης (1)

- 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="Michael Maher"
          address=michaelmaher@cs.gu.edu.au"/>
    <to name="Grigoris Antoniou"
        address=grigoris@cs.unibremen.de"/>
    <subject>where is the draft?</subject>
  </head>
  <body>
    Grigori, where is the draft paper you promised me?
  </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="Woo" firstname="Jason"/>`
`<person firstname="Jason" lastname="Woo" />`
είναι ισοδύναμα.

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

Ορισμός Δόμησης

- Ανάγκη για χρήση κοινού «λεξιλογίου» στοιχείων και χαρακτηριστικών για «επικοινωνία» διαφορετικών εφαρμογών.

Που σημαίνει

- Ορισμός ονομάτων στοιχείων και χαρακτηριστικών που θα χρησιμοποιηθούν
- Ορισμός δομής των στοιχείων και χαρακτηριστικών (π.χ. ποιες οι επιτρεπόμενες τιμές των χαρακτηριστικών, ποια στοιχεία μπορούν να εμφανίζονται μέσα σε άλλα στοιχεία κλπ)

Έγκυρο Έγγραφο XML

- Ένα έγγραφο XML είναι έγκυρο αν είναι σωστά δομημένο, χρησιμοποιεί δομικές πληροφορίες και τις εφαρμόζει.

Τρόποι Ορισμού Δόμησης

- DTD (Document Type Definition)
 - Παλαιότερος
 - Πιο περιορισμένος
- XML Schema
 - Νεώτερος
 - Πλουσιότερος

Τα συστατικά ενός DTD μπορούν να οριστούν σε ξεχωριστό αρχείο (εξωτερικό DTD) ή μέσα στο ίδιο έγγραφο XML (εσωτερικό DTD). Προτιμότερο το εξωτερικό.

DTD-Στοιχεία (1)

στοιχείο
(του lecturer)

```
<lecturer>  
  <name>David Hamilton</name>  
  <phone>+302610994555</phone>  
</lecturer>
```

τύπος στοιχείου
(lecturer)

Αντίστοιχο DTD

```
<!ELEMENT lecturer (name, phone)>  
<!ELEMENT name (#PCDATA)>  
<!ELEMENT phone (#PCDATA)>
```

DTD-Στοιχεία (2)

- `<!ELEMENT lecturer (name, phone)>`
(ο τύπος στοιχείου lecturer περιέχει ένα στοιχείο name και ένα στοιχείο phone μ' αυτή τη σειρά)
- `<!ELEMENT lecturer (name|phone)>`
(ο τύπος στοιχείου lecturer περιέχει είτε ένα στοιχείο name είτε ένα στοιχείο phone)
- `<!ELEMENT lecturer ((name, phone) | (phone, name))>`
(ο τύπος στοιχείου lecturer περιέχει ένα στοιχείο name και ένα στοιχείο phone με οποιαδήποτε σειρά)-
προβληματικό για πολλά στοιχεία

DTD-Χαρακτηριστικά

```
<order orderNo="23456" customer="John Smith"
  date="October 15, 2002">
  <item itemNo="a528" quantity="1"/>
  <item itemNo="c817" quantity="3"/>
</order>
```

Ο ορισμός χαρακτηριστικού γίνεται μέσω μιας *λίστας χαρακτηριστικών*

τελεστής πληθικότητας
(cardinality operator)

<!ELEMENT order (item⁺)>

Τύπος στοιχείων που μπορεί να χρησιμοποιηθεί η λίστα

<!ATTLIST order

λίστα χαρακτηριστικών
(attribute list)

| | | |
|----------|-------|------------|
| orderNo | ID | #REQUIRED |
| customer | CDATA | #REQUIRED |
| date | CDATA | #REQUIRED> |

<!ELEMENT item EMPTY>

ονόματα χαρακτηριστικών

<!ATTLIST item

τύποι χαρακτηριστικών

τύποι τιμών

| | | |
|----------|-------|-----------|
| itemNo | ID | #REQUIRED |
| quantity | CDATA | #REQUIRED |
| comments | CDATA | #IMPLIED> |

DTD-Χαρακτηριστικά

- Τελεστές πληθικότητας
 - ? : εμφάνιση καμία ή μία
 - * : εμφάνιση καμία ή περισσότερες
 - + : εμφάνιση μία ή περισσότερες
 - : έλλειψη τελεστή σημαίνει ακριβώς μια εμφάνιση
- Τύποι χαρακτηριστικών
 - (P)CDATA : αλφαριθμητικό ((με) χωρίς parsing)
 - ID : μοναδικό όνομα
 - IDREF/IDREFS : αναφορά/ες σε άλλο/α στοιχείο/α με χαρακτηριστικό ID
 - (v1|...|vn) : απαρίθμηση τιμών

DTD-Χαρακτηριστικά

■ Τύποι τιμών

- **#REQUIRED** : το χαρακτηριστικό αυτό πρέπει να υπάρχει σε κάθε εμφάνιση του τύπου
- **#IMPLIED** : προαιρετική εμφάνιση χαρακτηριστικού (π.χ. σχόλια)
- **#FIXED "value"** : Κάθε στοιχείο πρέπει να έχει αυτό το χαρακτηριστικό, που έχει πάντα την τιμή "value" (προκαθορισμένη τιμή)
- **"value"** : Προεπιλεγμένη τιμή χαρακτηριστικού (μπορεί να επικαλυφθεί από τιμή του χρήστη)

DTD-Χαρακτηριστικά

- Αναφορές (αφορούν τα IDREF, IDREFS)

DTD αρχείο

```
<!ELEMENT family (person*)>
```

```
<!ELEMENT person (name)>
```

```
<!ELEMENT name (#PCDATA)>
```

```
<!ATTLIST person
```

| | | |
|----|----|-----------|
| id | ID | #REQUIRED |
|----|----|-----------|

| | | |
|--------|-------|----------|
| mother | IDREF | #IMPLIED |
|--------|-------|----------|

| | | |
|--------|-------|----------|
| father | IDREF | #IMPLIED |
|--------|-------|----------|

| | | |
|----------|--------|-----------|
| children | IDREFS | #IMPLIED> |
|----------|--------|-----------|

DTD-Χαρακτηριστικά

XML αρχείο

```
<family>
  <person id="bob" mother="mary" father="peter">
    <name>Bob Marley</name>
  </person>
  <person id="bridget" mother="mary">
    <name>Bridget Marley</name>
  </person>
  <person id="mary" children="bob bridget">
    <name>Mary Poppins</name>
  </person>
  <person id="peter" children="bob bridget">
    <name>Peter Marley</name>
  </person>
</family>
```

DTD-Χαρακτηριστικά

■ Οντότητες XML

– Παίζουν διάφορους ρόλους

- Δεσμευτικό θέσης επαναλαμβανόμενων χαρακτήρων
- Τμήμα εξωτερικών δεδομένων
- Μέρος δήλωσης στοιχείων
- **Τυπική χρήση: σαν σταθερές**

– Σύνταξη

<!ENTITY <entity-name> <entity-value>>

Παράδειγμα: **<!ENTITY thisyear "2009">** ← &thisyear;

DTD-Χαρακτηριστικά

Τελικό παράδειγμα

```
<!ELEMENT email (head, body)>
<!ELEMENT head (from, to+, cc*, subject)>
<!ELEMENT from EMPTY>
<!ATTLIST from
    name      CDATA  #IMPLIED
    address   CDATA  #REQUIRED>
<!ELEMENT to EMPTY>
<!ATTLIST to
    name      CDATA  #IMPLIED
    address   CDATA  #REQUIRED>
```

DTD-Χαρακτηριστικά

```
<!ELEMENT cc EMPTY>
```

```
<!ATTLIST cc
```

```
    name      CDATA  #IMPLIED
```

```
    address   CDATA  #REQUIRED>
```

```
<!ELEMENT subject (#PCDATA)>
```

```
<!ELEMENT body (text, attachment*)>
```

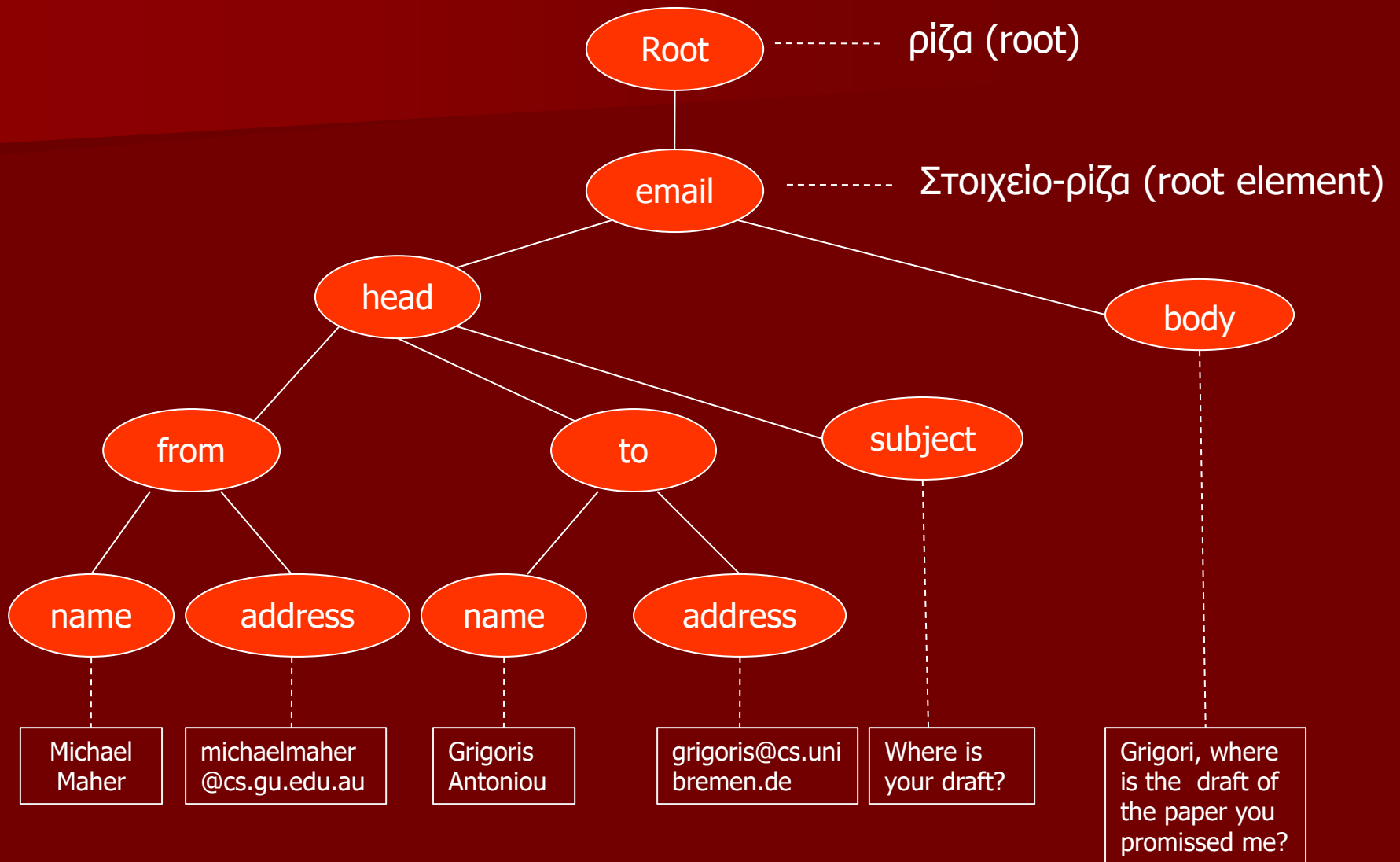
```
<!ELEMENT text (#PCDATA)>
```

```
<!ELEMENT attachment EMPTY>
```

```
<!ATTLIST attachment
```

```
    encoding  (mime|binhex) "mime"
```

```
    file      CDATA  #REQUIRED>
```



DTD-Χαρακτηριστικά

- Χρήση ως σημειογραφία BNF
`<!ELEMENT email (head, body)>` → `email ::= head body`
- Δυνατότητα αναδρομικών ορισμών
`<!ELEMENT bintree ((bintree root bintree)|emptytree)>`