

**Αρχές Γλωσσών Προγραμματισμού & Μεταφραστών****Τμήμα Μηχανικών Η/Υ & Πληροφορικής****Πανεπιστήμιο Πατρών****Εαρινό Εξάμηνο 2018****Διδάσκων: Ι. Γαροφαλάκης****Εργαστηριακή Άσκηση**

Η XML αποτελεί συντομογραφία του eXtensible Markup Language. Είναι μια γλώσσα σήμανσης που σχεδιάστηκε με σκοπό να αποθηκεύει και να μεταφέρει δεδομένα με ένα δομημένο τρόπο. Η XML σχεδιάστηκε με σκοπό να είναι αυτοπεριγραφική.

Η XML αποτελείται από *στοιχεία*. Καθένα φέρει 3 μέρη, την ετικέτα αρχής (start tag), το περιεχόμενο (content) και την ετικέτα τέλους (end tag). Το όνομα του *στοιχείου* εμφανίζεται στη ετικέτα αρχής (<element-name>) και την ετικέτα τέλους (</element-name>). Το περιεχόμενο μπορεί να είναι κείμενο, άλλα στοιχεία ή συνδυασμός αυτών, όπως επίσης μπορεί να είναι και κενό. Όταν το περιεχόμενο είναι στοιχείο τότε έχουμε εμφωλευμένα στοιχεία. Το σημαντικό που πρέπει να σημειωθεί σε αυτό το σημείο είναι ότι τα στοιχεία πρέπει να είναι σωστά εμφωλευμένα, δηλαδή μια ετικέτα τέλους, τερματίζει ακριβώς την προηγούμενη αντίστοιχη ετικέτα αρχής.

Τα *στοιχεία* έχουν συσχετισμένες ιδιότητες οι οποίες καλούνται *χαρακτηριστικά*, οι οποίες παίρνουν τιμές και εμφανίζονται στην ετικέτα αρχής πριν το ">". Για παράδειγμα: <title lang="en">Everyday Italian</title>.

Οι κανόνες που διέπουν την γλώσσα XML είναι οι εξής:

- Τα αρχεία XML πρέπει να περιέχουν ένα στοιχείο *ρίζα* το οποίο είναι πατέρας όλων των υπολοίπων στοιχείων:

```
<root>
  <child>
    <subchild>.....</subchild>
  </child>
</root>
```

- Όλα τα XML στοιχεία πρέπει να έχουν ένα *closing tag*
- Ένα κενό στοιχείο μπορεί να γραφτεί και σαν self-closing tag <element/>
- Τα XML tags είναι Case Sensitive.
- Οι τιμές των XML χαρακτηριστικών πρέπει να εισέρχονται με εισαγωγικά.

```
<note date="12/11/2007">
  <to>Tove</to>
  <from>Jani</from>
</note>
```

- Τα σχόλια στην XML

Τα σχόλια στην XML είναι της μορφής <!-- Ένα σχόλιο -->.

Σχόλια της μορφής <!-- Αυτό είναι ένα -- σχόλιο --> δεν επιτρέπονται.

Σε αυτή την εργαστηριακή άσκηση θα υλοποιήσουμε ένα υποσύνολο του XML προτύπου που χρησιμοποιεί το Microsoft Excel. Το πλήρες πρότυπο είναι διαθέσιμο στον σύνδεσμο: [https://msdn.microsoft.com/en-us/library/aa140066\(v=office.10\).aspx](https://msdn.microsoft.com/en-us/library/aa140066(v=office.10).aspx)

Το υποσύνολο που θα χρησιμοποιήσουμε στην εργαστηριακή άσκηση αποτελείται από τα εξής στοιχεία:

<ss:Workbook>, <ss:Styles>, <ss:Style>, <ss:Worksheet>, <ss:Table>, <ss:Column/>, <ss:Row>, <ss:Cell>, <ss>Data>

Οι κανόνες και περιορισμοί που ισχύουν στην εργαστηριακή άσκηση είναι οι εξής:

- Το στοιχείο ρίζα είναι το <ss:Workbook>.
- Το <ss:Workbook> έχει υποχρεωτικά τουλάχιστον ένα στοιχείο <ss:Worksheet> και προαιρετικά στοιχεία <ss:Styles>. Εάν υπάρχουν στοιχεία <ss:Styles>, αυτά εμφανίζονται πριν από το πρώτο στοιχείο <ss:Worksheet>.
- Το <ss:Styles> περιέχει προαιρετικά στοιχεία <ss:Style>.
- Το <ss:Style> έχει κενό περιεχόμενο ενώ έχει υποχρεωτικό χαρακτηριστικό το ss:ID. Το ss:ID δέχεται τιμή τύπου αλφαριθμητικού και χρησιμοποιείται σαν αναγνωριστικό.
- Το <ss:Worksheet> έχει προαιρετικά στοιχεία <ss:Table>, απαιτούμενο χαρακτηριστικό το ss:Name και προαιρετικό χαρακτηριστικό το ss:Protected. Το ss:Name δέχεται τιμή τύπου αλφαριθμητικού, ενώ το ss:Protected δέχεται τιμή τύπου Boolean.
- Το <ss:Table> έχει προαιρετικά στοιχεία <ss:Column/> και <ss:Row>, τα οποία εμφανίζονται με αυτή τη σειρά εάν υπάρχουν και τα δύο, όπως επίσης προαιρετικά χαρακτηριστικά τα ss:ExpandedColumnCount, ss:ExpandedRowCount και ss:StyleID. Τα 2 πρώτα χαρακτηριστικά δέχονται τιμές τύπου Θετικού Ακεραίου, ενώ το ss:StyleID δέχεται τιμές τύπου αλφαριθμητικού.
- Το <ss:Column/> είναι κενό στοιχείο κι έχει προαιρετικά χαρακτηριστικά τα ss:Hidden, ss:Width και ss:StyleID. Το ss:Hidden δέχεται τιμές τύπου Boolean, το ss:Width δέχεται τιμές τύπου Θετικού Ακεραίου, ενώ το ss:StyleID δέχεται τιμές τύπου αλφαριθμητικού.
- Το <ss:Row> έχει προαιρετικά στοιχεία <ss:Cell> και φέρει ως προαιρετικά χαρακτηριστικά τα ss:Height, ss:Hidden και ss:StyleID. Το ss:Hidden δέχεται τιμές τύπου Boolean, το ss:Height δέχεται τιμές τύπου Θετικού Ακεραίου, ενώ το ss:StyleID δέχεται τιμές τύπου αλφαριθμητικού.
- Το <ss:Cell> έχει προαιρετικά στοιχεία <ss>Data> και φέρει ως προαιρετικά χαρακτηριστικά τα ss:MergeAcross και ss:MergeDown, τα οποία δέχονται τιμές τύπου Θετικού Ακεραίου, όπως επίσης και το ss:StyleID, το οποίο δέχεται τιμές τύπου αλφαριθμητικού.
- Το <ss>Data> έχει απαιτούμενο χαρακτηριστικό το ss:Type, το οποίο μπορεί να πάρει τιμές μια εκ των Number, DateTime, Boolean, String, ενώ το περιεχόμενο μπορεί να είναι κείμενο ή κενό.
- Σε οποιοδήποτε σημείο του κώδικα υποστηρίζονται τα σχόλια της XML.

Ένα ενδεικτικό αποδεκτό αρχείο XML είναι το εξής:

```
<ss:Workbook>
  <ss:Styles>
    <ss:Style ss:ID="s123"></ss:Style>
    <ss:Style ss:ID="x123"></ss:Style>
  </ss:Styles>
  <ss:Worksheet ss:Name="sheet1">
    <ss:Table ss:ExpandedColumnCount="2">
      <ss:Column ss:StyleID="s123"/>
      <ss:Column ss:StyleID="s123"/>
      <ss:Row>
        <ss:Cell>
          <ss:Data ss:Type="Number">1234</ss:Data>
        </ss:Cell>
        <ss:Cell>
          <ss:Data ss:Type="String">string data</ss:Data>
        </ss:Cell>
      </ss:Row>
    </ss:Table>
  </ss:Worksheet>
</ss:Workbook>
```

### Ερωτήματα

#### 1. (70%)

- i. Δώστε σε **BNF** το συντακτικό ορισμό της γραμματικής της γλώσσας.
- ii. Χρησιμοποιώντας τα προγράμματα Flex και Bison, υλοποιήστε έναν αναλυτή (parser), ο οποίος θα παίρνει ως είσοδο ένα αρχείο γραμμένο στη γλώσσα που περιγράφηκε πιο πάνω και θα ελέγχει **σε ένα πέραςμα** αν το πρόγραμμα είναι συντακτικά ορθό. Το πρόγραμμά σας θα καλείται από τη γραμμή εντολών ως εξής:

***prompt> myParser.exe file.txt***

και θα επιστρέφει **στην οθόνη** το ίδιο το πρόγραμμα έχοντας εφαρμόσει κατάλληλη στοίχιση (κάθε μπλοκ πάει ένα tab «μέσα») στα διάφορα μπλοκ του προγράμματος, ένα διαγνωστικό μήνυμα για το αν ήταν ορθώς γραμμένο, ή κατάλληλο μήνυμα σφάλματος (πρέπει να φαίνεται η γραμμή όπου υπάρχει το σφάλμα).

#### 2. (20%)

Τροποποιήστε τον αναλυτή σας ώστε να εκτελεί επιπλέον τους εξής ελέγχους και να εμφανίζει σχετικά διαγνωστικά μηνύματα σε περίπτωση σφάλματος

- i. Όταν υπάρχουν στοιχεία ss: Style θα πρέπει να χρησιμοποιούν μοναδικές τιμές στα ss:ID.
- ii. Όταν υπάρχει το χαρακτηριστικό ss:StyleID η τιμή του πρέπει να υπάρχει δηλωμένη νωρίτερα σαν τιμή σε ss:ID.
- iii. Όταν υπάρχει το χαρακτηριστικό ss:ExpandedColumnCount τότε εφόσον υπάρχουν <ss:Column/> ο αριθμός τους πρέπει να συμφωνεί με την τιμή του χαρακτηριστικού.
- iv. Όταν υπάρχει το χαρακτηριστικό ss:ExpandedRowCount τότε εφόσον υπάρχουν <ss:Row/> ο αριθμός τους πρέπει να συμφωνεί με την τιμή του χαρακτηριστικού.

## 3. (10%)

Τροποποιήστε τον αναλυτή σας ώστε στο <ss:Data> να εκτελεί επιπλέον έλεγχο τύπων μεταξύ του τύπου του περιεχομένου του στοιχείου και της τιμής στο χαρακτηριστικό ss:Type.

**Παραδοτέα**

- ❖ **Γραπτή Αναφορά** που περιλαμβάνει
  - ο Την περιγραφή της γραμματικής της γλώσσας σε **BNF**.
  - ο Τα **ΤΕΛΙΚΑ** αρχεία περιγραφής της γλώσσας, τα οποία δίνονται ως **είσοδος** στα Flex και Bison
  - ο Screenshots παραδειγμάτων εφαρμογής (τουλάχιστον ένα παράδειγμα για κάθε ερώτημα)
  - ο Σχόλια
- ❖ Συμπιεσμένα σε **ένα αρχείο (zip)** τα:
  - ο Την πιο πάνω γραπτή αναφορά
  - ο Τα **ΤΕΛΙΚΑ** αρχεία εισόδου για το flex και το bison.
  - ο Τον **ΤΕΛΙΚΟ** κώδικα σε C (και το .h αρχείο) που παράγεται από τα δύο προγράμματα και ο εκτελέσιμος κώδικας του parser.
  - ο Τα αρχεία που δόθηκαν σαν είσοδο στον parser για να ελεγχθεί η σωστή λειτουργία του

**Το αρχείο zip πρέπει να έχει όνομα τους αριθμούς μητρώου των ατόμων της ομάδας διαχωρισμένους με \_ και από τον μικρότερο στο μεγαλύτερο (π.χ. 1000\_1543\_2788\_3972.zip), και να σταλεί (ΥΠΟΧΡΕΩΤΙΚΑ) με email στο [pergantis@ceid.upatras.gr](mailto:pergantis@ceid.upatras.gr) με θέμα ASKISI ARXES GLWSSWN 2018 και στο μήνυμα να αναφέρονται τα ονοματεπώνυμα, το έτος, οι αντίστοιχοι αριθμοί μητρώου και οι e-mail διευθύνσεις των μελών της ομάδας**

**Διευκρινήσεις**

- ❖ Για τη λειτουργία των εργαλείων Flex και Bison μπορείτε να βρείτε πληροφορίες στη σελίδα του μαθήματος.
- ❖ Η άσκηση θα γίνει σε ομάδες έως 3 ατόμων.
- ❖ Η συμβολή της στον τελικό βαθμό είναι 30%.
- ❖ Η άσκηση είναι υποχρεωτική και πρέπει ο βαθμός της να είναι τουλάχιστον 5.
- ❖ Τελικές ημερομηνίες παράδοσης είναι οι ημερομηνίες γραπτών εξετάσεων περιόδου Ιουνίου και Σεπτεμβρίου αντίστοιχα.
- ❖ Για τυχόν απορίες ή υποδείξεις μπορείτε να απευθύνεστε με e-mail στους μεταπτυχιακούς φοιτητές
  - ο Περγαντή Δημήτρη ([pergantis@ceid.upatras.gr](mailto:pergantis@ceid.upatras.gr))
  - ο Κορμπά Αντωνία ([korba@ceid.upatras.gr](mailto:korba@ceid.upatras.gr))

(Θα ήταν προτιμότερο πάντως να γίνεται χρήση του forum του μαθήματος στο myCeid).