

Αρχές Γλωσσών Προγραμματισμού & Μεταφραστών**Τμήμα Μηχανικών Η/Υ & Πληροφορικής****Πανεπιστήμιο Πατρών****Εαρινό Εξάμηνο 2025****Διδάσκοντες: Ι. Γαροφαλάκης, Σ. Σιούτας, Π. Χατζηδούκας****Εργαστηριακή Άσκηση**

Σκοπός της εργαστηριακής άσκησης είναι να εξοικειωθείτε με την περιγραφή μιας γλώσσας σε μορφή BNF, βασικές έννοιες των μεταγλωττιστών και τελικά να υλοποιήσετε δύο από τα τμήματα ενός μεταγλωττιστή, έναν συντακτικό και ένα λεκτικό αναλυτή με τη χρήση των εργαλείων Bison και Flex.

ΔΙΑΔΙΚΑΣΤΙΚΑ**Παραδοτέα**

- ❖ **Γραπτή Αναφορά** που περιλαμβάνει:
 - Τα στοιχεία των μελών της ομάδας (Ονοματεπώνυμο, ΑΜ, Έτος, e-mail)
 - Την περιγραφή της γραμματικής της γλώσσας σε BNF.
 - Τα **ΤΕΛΙΚΑ** αρχεία περιγραφής της γλώσσας, τα οποία δίνονται ως **είσοδος** στα Flex και Bison
 - Screenshots παραδειγμάτων εφαρμογής (τουλάχιστον ένα παράδειγμα για κάθε ερώτημα, για επιτυχημένη και ανεπιτυχή ανάλυση)
 - Σχόλια - Παραδοχές που έγιναν για την ανάπτυξη της εργασίας
- ❖ Συμπιεσμένα σε **ένα αρχείο (zip)** τα:
 - Την παραπάνω γραπτή αναφορά.
 - Τα **ΤΕΛΙΚΑ** αρχεία εισόδου για το flex και το bison.
 - Τον **ΤΕΛΙΚΟ** κώδικα σε C (και το .h αρχείο) που παράγεται από τα δύο προγράμματα και ο εκτελέσιμος κώδικας του parser.
 - Τα **δοκιμαστικά 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.zip), και να υποβληθεί (ΥΠΟΧΡΕΩΤΙΚΑ) στο eclass του μαθήματος και στο μήνυμα να αναφέρονται τα ονοματεπώνυμα, το έτος και οι αντίστοιχοι αριθμοί μητρώου των μελών της ομάδας, καθώς και οι e-mail διευθύνσεις όλων των μελών της ομάδας

Διευκρινίσεις

- ❖ Για τη λειτουργία των εργαλείων Flex και Bison μπορείτε να βρείτε πληροφορίες στο eclass του μαθήματος.
- ❖ Η άσκηση θα γίνει σε ομάδες από 1 έως 3 άτομα.
- ❖ Η συμβολή της στον τελικό βαθμό είναι 20%, βάσει της νέας διαδικασίας αξιολόγησης, ή 30%, βάσει της παλαιάς διαδικασίας αξιολόγησης. **Οι φοιτητές/φοιτήτριες του 2ου έτους ακολουθούν τη νέα διαδικασία αξιολόγησης.**
- ❖ Η εργαστηριακή άσκηση είναι υποχρεωτική.
- ❖ Συμβουλευτείτε τα [Διαδικαστικά](#) του μαθήματος.
- ❖ Οι τελικές ημερομηνίες παράδοσης θα καθοριστούν αναλόγως με τις ημερομηνίες γραπτών εξετάσεων περιόδου Ιουνίου και Σεπτεμβρίου αντίστοιχα. Αναλόγως θα καθοριστεί και η ημερομηνία της προφορικής εξέτασης της εργαστηριακής άσκησης.
- ❖ Τυχόν απορίες ή υποδείξεις αναρτώνται **ΑΠΟΚΛΕΙΣΤΙΚΑ** στο χώρο “Συζητήσεις” στην σελίδα του μαθήματος στο eclass

(<https://eclass.upatras.gr/modules/forum/?course=CEID1091>)

ΠΕΡΙΓΡΑΦΗ ΑΣΚΗΣΗΣ

Η **myHTML** είναι μια ψευδογλώσσα σήμανσης, που ακολουθεί τις προδιαγραφές της HTML, για τη δημιουργία ιστοσελίδων. Η **myHTML** αποτελείται από στοιχεία, τα οποία αποτελούνται από τρία μέρη: την ετικέτα αρχής (start tag), το περιεχόμενο (content), και την ετικέτα τέλους (end tag). Οι ετικέτες αρχής περιέχουν μέσα σε αγκύλες `<>` το όνομα του στοιχείου (`<element_name>`). Οι ετικέτες τέλους είναι παρόμοιας μορφής με τις ετικέτες αρχής, με την προσθήκη μιας καθέτου πριν το όνομα του στοιχείου (`</element_name>`).

Το περιεχόμενο ενός στοιχείου μπορεί να είναι κενό, ή να περιέχει κείμενο, άλλα στοιχεία ή συνδυασμό αυτών. Όταν το περιεχόμενο είναι στοιχείο, τότε έχουμε εμφώλευση στοιχείων.

Προσοχή! Τα στοιχεία πρέπει να είναι σωστά εμφωλευμένα, δηλαδή μια ετικέτα τέλους τερματίζει την ακριβώς προηγούμενη αντίστοιχη ετικέτα αρχής.

Ορισμένοι κανόνες που διέπουν τη **myHTML** είναι οι ακόλουθοι:

- Τα αρχεία myHTML περιέχουν ένα στοιχείο ρίζα το οποίο είναι γονέας όλων των υπόλοιπων στοιχείων.

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

- Εάν δεν αναφέρεται κάτι διαφορετικό, τα στοιχεία έχουν μια ετικέτα τέλους (closing tag), ανεξαρτήτως εάν είναι κενά ή όχι. Κενό είναι ένα στοιχείο το οποίο δεν έχει περιεχόμενο.
- Οι ετικέτες είναι Case Sensitive.
- Τα στοιχεία έχουν συσχετισμένες ιδιότητες οι οποίες καλούνται χαρακτηριστικά (attributes) και εμφανίζονται στην ετικέτα αρχής πριν την αγκύλη `>`. Οι τιμές των χαρακτηριστικών περικλείονται σε μονά ή διπλά εισαγωγικά. Αν το περιεχόμενο της τιμής περιέχει μονά εισαγωγικά, τότε αυτή θα πρέπει να περικλείεται από διπλά εισαγωγικά. Αντίστοιχα, εάν το περιεχόμενο της τιμής περιέχει διπλά εισαγωγικά, αυτή θα πρέπει να περικλείεται σε μονά εισαγωγικά.
- Τα σχόλια στην myHTML είναι της μορφής

<!-- A wild comment appeared --> (σχόλιο μονής γραμμής) ή

<!--

A wild

comment appeared.

--> (σχόλιο πολλαπλών γραμμών)

Σχόλια της μορφής **<!-- A wild comment appeared -->** δεν επιτρέπονται. Τα σχόλια μπορούν να εμφανιστούν ανάμεσα σε στοιχεία ή εντός του περιεχομένου ενός στοιχείου. Δεν επιτρέπεται η εμφάνισή τους εντός μιας ετικέτας αρχής ή μιας ετικέτας τέλους.

Προδιαγραφές γλώσσας

- Κάθε αρχείο myHTML ξεκινά και τελειώνει με το στοιχείο `<MYHTML>`. Αυτό δεν διαθέτει κανένα χαρακτηριστικό.
- Εάν δεν αναφέρεται ρητώς κάτι διαφορετικό, όλα τα στοιχεία διαθέτουν υποχρεωτικά το χαρακτηριστικό (attribute) `id`, η τιμή του οποίου είναι αλφαριθμητικού τύπου. Το χαρακτηριστικό `id` εμφανίζεται ακριβώς μια φορά και, εάν υπάρχουν κι άλλα χαρακτηριστικά εντός της ετικέτας αρχής του στοιχείου, μπορεί να εμφανιστεί σε οποιαδήποτε θέση (πρώτο, τελευταίο ή ενδιάμεσα σε άλλα χαρακτηριστικά).
- Το στοιχείο `<MYHTML>` έχει δυο στοιχεία - παιδιά, τα `<head>` και `<body>`, τα οποία είναι μοναδικά.
- Το στοιχείο `<head>` είναι προαιρετικό και δεν παίρνει κανένα χαρακτηριστικό. Εάν υπάρχει, τότε εμφανίζεται πάντα πριν το στοιχείο `<body>`.
- Τα παιδιά του στοιχείου `<head>` είναι τα στοιχεία `<title>` και `<meta>`. Το στοιχείο `<title>` είναι υποχρεωτικό και εμφανίζεται πρώτο μια φορά. Το στοιχείο `<meta>` είναι προαιρετικό.
- Το στοιχείο `<title>` δεν διαθέτει κανένα χαρακτηριστικό και στο περιεχόμενό του επιτρέπεται μόνο κείμενο.
- Το στοιχείο `<meta>` δεν έχει περιεχόμενο, ούτε ετικέτα τέλους. Για τα επιτρεπόμενα χαρακτηριστικά του στοιχείου υπάρχουν δυο εκδοχές. Είτε το στοιχείο έχει μόνο το χαρακτηριστικό `charset`, η τιμή του οποίου είναι αλφαριθμητικού τύπου, είτε έχει τα χαρακτηριστικά `name` και `content`, τα οποία εμφανίζονται με αυτή τη σειρά και είναι αμφότερα αλφαριθμητικού τύπου. Το στοιχείο `<meta>` μπορεί να εμφανίζεται καμία ή περισσότερες φορές.
- Το στοιχείο `<body>` είναι υποχρεωτικό και δεν παίρνει κανένα χαρακτηριστικό. Το περιεχόμενο του στοιχείου `<body>` μπορεί να είναι κενό.
- Τα παιδιά του στοιχείου `<body>` είναι τα στοιχεία `<p>`, `<a>`, ``, `<form>`, `<div>`. Αυτά μπορούν να εμφανιστούν, σωστά εμφωλευμένα, με οποιαδήποτε σειρά, καμία ή περισσότερες φορές.
- Το στοιχείο `<p>` χρησιμοποιείται για να ορίσει μια παράγραφο. Έχει ως προαιρετικό χαρακτηριστικό το `style`, το οποίο είναι αλφαριθμητικού τύπου, και το περιεχόμενό του στοιχείου μπορεί να είναι κενό ή κείμενο.
- Το στοιχείο `<a>` χρησιμοποιείται για να ορίσει έναν υπερσύνδεσμο εντός μιας σελίδας. Έχει ως επιπλέον υποχρεωτικό χαρακτηριστικό το `href`, το οποίο είναι αλφαριθμητικού τύπου. Το περιεχόμενο του στοιχείου `<a>` μπορεί να είναι κενό ή να περιέχει ακριβώς ένα παιδί - στοιχείο `` ή/και απλό κείμενο.
- Το στοιχείο `` χρησιμοποιείται για να ορίσει μια εικόνα. Διαθέτει δυο επιπλέον υποχρεωτικά χαρακτηριστικά, τα `src` και `alt`, τα οποία είναι τύπου αλφαριθμητικού. Προαιρετικά, διαθέτει τα χαρακτηριστικά `width` και `height`, τα οποία είναι τύπου θετικού ακεραίου. Τα χαρακτηριστικά μπορούν να εμφανίζονται με οποιαδήποτε σειρά. Το στοιχείο αυτό δεν έχει περιεχόμενο, ούτε ετικέτα τέλους.

- Το στοιχείο `<form>` χρησιμοποιείται για να ορίσει μια φόρμα. Τα παιδιά του στοιχείου αυτού είναι ένα ή περισσότερα στοιχεία `<input>` ή στοιχεία `<label>`. Έχει ως προαιρετικό χαρακτηριστικό το `style`, το οποίο είναι αλφαριθμητικού τύπου.
- Ένα στοιχείο `<input>` έχει επιπλέον υποχρεωτικό χαρακτηριστικό `type`, το οποίο καθορίζει τον τύπο του πεδίου. Προαιρετικά ένα στοιχείο `<input>` παίρνει ως χαρακτηριστικό το `value`, που είναι αλφαριθμητικού τύπου, και το `style`, το οποίο είναι αλφαριθμητικού τύπου. Ένα στοιχείο `<input>` δεν έχει περιεχόμενο, ούτε ετικέτα τέλους.
- Ένα στοιχείο `<label>` έχει επιπλέον υποχρεωτικό χαρακτηριστικό `for`, το οποίο είναι αλφαριθμητικού τύπου. Έχει ως προαιρετικό χαρακτηριστικό το `style` το οποίο είναι αλφαριθμητικού τύπου. Το περιεχόμενο ενός στοιχείου `<label>` μπορεί να είναι κενό ή κείμενο.
- Το στοιχείο `<div>` χρησιμοποιείται για να ορίσει μια περιοχή εντός της σελίδας, στην οποία μπορεί να περιέχεται οποιοδήποτε περιεχόμενο. Παιδιά του στοιχείου `<div>` μπορεί να είναι οποιοδήποτε από τα στοιχεία `<p>`, `<a>`, ``, `<form>`, ή να είναι κενού περιεχομένου. Το στοιχείο `<div>` έχει προαιρετικό χαρακτηριστικό το `style` το οποίο είναι τύπου αλφαριθμητικού.

Ερωτήματα

1. (60%)

- a. Δώστε σε BNF το συντακτικό ορισμό της γραμματικής της γλώσσας.
- b. Χρησιμοποιώντας τα προγράμματα Flex και Bison, υλοποιήστε έναν λεξικό και συντακτικό αναλυτή, ο οποίος θα παίρνει ως είσοδο ένα αρχείο γραμμένο στη ψευδογλώσσα που περιγράφηκε παραπάνω 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_name

και θα επιστρέφει το ίδιο το πρόγραμμα στην οθόνη και ένα διαγνωστικό μήνυμα για το αν ήταν ορθώς γραμμένο, ή κατάλληλο μήνυμα σφάλματος (πρέπει να φαίνεται η γραμμή όπου υπάρχει το σφάλμα και η ανάλυση να διακόπτεται εκεί).

2. (30%)

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

- a. Το κείμενο στο περιεχόμενο του στοιχείου `<title>` να είναι μέχρι 60 χαρακτήρες.
- b. Τα `id` των στοιχείων πρέπει να είναι μοναδικά.
- c. Οι τιμές που δίνονται στο χαρακτηριστικό `href` ακολουθούν τη μορφή μιας απόλυτης (absolute) ή σχετικής (relative) URL. Επιπλέον, οι τιμές που δίνονται στο χαρακτηριστικό `href` μπορούν να είναι της μορφής `"#element_id"`, όπου `element_id` είναι το `id` κάποιου στοιχείου εντός της σελίδας.
- d. Οι τιμές που δίνονται στο χαρακτηριστικό `src` ακολουθούν τη μορφή μιας απόλυτης (absolute) ή σχετικής (relative) URL.
- e. Οι επιτρεπόμενες τιμές για το χαρακτηριστικό `type` είναι οι `"text"`, `"checkbox"`, `"radio"` και `"submit"`. Εάν υπάρχει στοιχείο `<input>` τύπου `submit`, εξασφαλίστε ότι θα είναι μοναδικό και θα εμφανίζεται τελευταίο.
- f. Η τιμή που λαμβάνει το χαρακτηριστικό `for` ενός `<label>` στοιχείου θα πρέπει να αντιστοιχεί στο `id` ενός στοιχείου `<input>`. Ένα στοιχείο `<label>` συνδέεται μόνο με ένα στοιχείο `<input>`.

- g. Οι τιμές που δίνονται στο χαρακτηριστικό **style** είναι ζεύγη της μορφής "property:value". Εάν ορίζονται περισσότερες από μια τιμές για το χαρακτηριστικό, τότε 1) αυτές είναι μοναδικές, 2) εμφανίζονται με οποιαδήποτε σειρά και 3) χωρίζονται με **;**, "property1:value1; property2:value2;κ.ο.κ". Οι επιτρεπτές τιμές για τα ζεύγη είναι οι εξής:

| property | value |
|------------------|---|
| background_color | τιμή αλφαριθμητικού τύπου |
| color | τιμή αλφαριθμητικού τύπου |
| font_family | τιμή αλφαριθμητικού τύπου |
| font_size | τιμή θετικού ακεραίου, ακολουθούμενη είτε από "px" είτε από "%". Π.χ. 90px, 160%. |

3. (10%)

Επεκτείνετε τον αναλυτή σας ορίζοντας ένα νέο, δικό σας, χαρακτηριστικό για το στοιχείο **<form>**, το οποίο:

- θα είναι θετικού ακεραίου τύπου και
- θα καθορίζει το ακριβές πλήθος των στοιχείων **<input>** τύπου **checkbox** που ανήκουν στο στοιχείο γονέα.

Στη συνέχεια, υλοποιήστε κατάλληλους ελέγχους, ώστε 1) να εμφανίζεται κατάλληλο διαγνωστικό μήνυμα εάν το πλήθος των στοιχείων checkbox διαφέρει από την τιμή που έχει λάβει το χαρακτηριστικό, 2) να μην επιτρέπεται η χρήση του χαρακτηριστικού αυτού εάν δεν υπάρχει τουλάχιστον ένα στοιχείο checkbox εντός της φόρμας, 3) να μην επηρεάζεται από άλλου τύπου στοιχεία input (π.χ. *text*) που μπορεί να περιέχονται εντός της ίδιας φόρμας.

Σημείωση: Οι ζητούμενοι έλεγχοι για τα ερωτήματα 2 και 3 θα πρέπει να υλοποιηθούν **αποκλειστικά** με κώδικα C/C++.

Ενδεικτικό παράδειγμα συντακτικά ορθού αρχείου myHTML:

```
<MYHTML>
<head>
  <title>Principles of Programming Languages & Compilers</title>
  <meta name="description" content="Flex Bison Tutorial">
  <meta charset="UTF-8">
  <meta name="author" content="Input your name here">
</head>
<body>

  <p id="par1">This is my project.</p>

  <div id="div1" style="color:blue">
    <a id="hyper" href="https://arxes.ceid.gr">
      Here comes an image.
      
    </a>
  </div>

  <form id="form1">
    <label for = "name" id="label1">Name:</label>
    <input type="text" id="name">
    <label id="label2" for="lastname">Last Name:</label>
    <input type="text" id = "lastname">
    <input id="submit" type="submit" value='Submit "it" '>
  </form>

</body>
</MYHTML>
```