	ΘΕΩΡΙΑ ΥΠΟΛΟΓΙΣΜΟΥ Διδάσκων: Σ. Σκιαδόπουλος	
Άνοιξη 2003	**ΠΡΟΑΙΡΕΤΙΚΟ ΘΕΜΑ**	Ημερ. Παράδοσης: Εξέταση

Μετατροπή αρχείων κειμένου LaTeX σε HTML

Το θέμα με μια ματιά

Σκοπός	Υλοποίηση προγράμματος μετατροπής αρχείων LaTeX (αρχεία με επέκταση .tex) σε αρχεία HTML (αρχεία με επέκταση .html) με ελεγχόμενους κανόνες μετατροπής.
Πλατφόρμα υλοποίησης	gcc ή perl.
Είσοδος	Ένα αρχείο .tex προς μετατροπή και ένα αρχείο .rules με τους κανόνες μετατροπής.
Έξοδος	Ένα αρχείο .html.
Ομάδες εργασίας	Αυστηρά ομάδες 3 ατόμων.

Παραδείγματα εκτέλεσης προγράμματος

Υποθέτουμε ότι το επιθυμητό πρόγραμμα ονομάζεται mylatex2html.

1. **Απλή εκτέλεση**

   ```
   mylatex2html file.tex
   ```

   Το πρόγραμμα διαβάζει το αρχείο file.tex και παράγει το αρχείο file.html με βάση τους εξ΄ορισμού κανόνες.

2. **Εκτέλεση με χρήση επιπρόσθετων κανόνων**

   ```
   mylatex2html file.tex -rules my.rules
   ```

   Το πρόγραμμα διαβάζει το αρχείο file.tex και παράγει το αρχείο file.html με βάση τους εξ΄ορισμού κανόνες και τους κανόνες που υπάρχουν στο αρχείο my.rules.

Γενικές πληροφορίες

Πληροφορίες για τα αρχεία LaTeX αλλά και γενικότερη πληροφόρηση για την επεξεργασία κειμένου με LaTeX μπορείτε να βρείτε στην σελίδα http://www.tug.org/. Μπορείτε ακόμα να κατεβάσετε μια υλοποίηση του LaTeX για Microsoft Windows από την σελίδα http://www.miktex.org/.

Πληροφορίες για τα αρχεία HTML μπορείτε να βρείτε στην σελίδα http://www.tug.org/. Δείτε ακόμα και την σελίδα http://www.htmlhelp.com/.

Βήματα υλοποίησης

Αρχικά πρέπει να εντοπίσετε τα *περιβάλλοντα* (δομικά στοιχεία) των αρχείων .tex και .html. Περιβάλλοντα ενός κειμένου αποτελούν για παράδειγμα: (α) οι επικεφαλίδες τμημάτων του κειμένου (ενότητες, υποενότητες, εδάφια κ.λπ.), (β) οι μαθηματικές εκφράσεις (γ) κείμενα με ιδιαίτερα χαρακτηριστικά (υπογραμμισμένο, έντονο κ.λπ.), (δ) ο πίνακας περιεχομένων και σχημάτων, (ε) οι βιβλιογραφικές αναφορές, (στ) πίνακες ή εικόνες κ.ά.

Στην συνέχεια πρέπει να βρείτε τρόπους μετάβασης από τα δομικά στοιχεία των αρχείων .tex σε αυτά των αρχείων .html. Για παράδειγμα μια εντολή

```
\section{Intoduction}
```

σε ένα αρχείο .tex μπορεί να μετατραπεί σε μια εντολή

```
<H1>Intoduction</H1>
```

σε ένα αρχείο .html.

Δεν θα δυσκολευτείτε να βρείτε ένα σύνολο από βασικούς κανόνες μετατροπής. Οι κανόνες αυτοί πρέπει να ενσωματωθούν στο πρόγραμμά σας ως εξ΄ορισμού λειτουργίες. Θα υπάρχει όμως η δυνατότητα στο χρήστη να δηλώσει και τους δικούς του κανόνες μετατροπής. Σε περίπτωση αμφισημίας, θα εκτελούνται οι κανόνες του χρήστη. Οι κανόνες που θέλει να προσθέσει ή αλλάξει ο χρήστης θα εμπεριέχονται σε ένα αρχείο κειμένου το οποίο θα δίνετε σαν παράμετρος στο πρόγραμμα (με την επιλογή -rules). Κάθε κανόνας θα γράφεται σε μια γραμμή του κειμένου και θα έχει την μορφή:

```
latex environment, corresponding html environment
```

Για παράδειγμα:

```
\emph{#1}, <B>#1</B>
```

Ο παραπάνω κανόνας χρησιμοποιεί την μεταβλητή #1 και μετατρέπει τα περιβάλλοντα \emph{κείμενο} σε <B>κείμενο</B> (αντί του πιο λογικού <I>κείμενο</I>). Κανόνες με περισσότερες παραμέτρους μπορούν να κωδικοποιηθούν με τις μεταβλητές #2, #3 κ.λπ.

Ας σημειωθεί τέλος ότι σε κείμενα LaTeX υπάρχουν πολλοί τρόποι για να οριστεί ένα περιβάλλον. Σας παρουσιάζουμε μερικά παραδείγματα.

1. \environmentName{κείμενο}
2. {\environmentName κείμενο}
3. \begin{environmentName}κείμενο\end{environmentName}
4. \[κείμενο\]
5. $κείμενο$
6. $$κείμενο$$

Σημεία ενδιαφέροντος

1. Το πρόγραμμά σας πρέπει να υποστηρίζει και εντολές include, input και bibliography στα αρχεία .tex. Για παράδειγμα, αν επεξεργάζεστε το αρχείο file.tex και αυτό εμπεριέχει την εντολή

   ```
   \include{anotherlatexfile.tex}
   ```

   το πρόγραμμά σας συνεχίζει και επεξεργάζεται το αρχείο anotherlatexfile.tex. Όταν το αρχείο anotherlatexfile.tex τελειώσει η μετάφραση συνεχίζεται στο αρχείο file.tex.

2. Πρέπει να δώσετε ιδιαίτερη προσοχή στην εμβέλεια των περιβαλλόντων στα αρχεία .tex γιατί μπορεί να είναι φωλιασμένα. Για παράδειγμα το φωλιασμένο περιβάλλον

   ```
   \subsubsection{A subsubsection with a text in \texttt{courier}}
   ```

   σε ένα κείμενο LaTeX πρέπει να μεταφραστεί ως

   ```
   <H3>A subsubsection with a text in <TT>courier</TT></H3>
   ```

   στο κείμενο HTML.

3. Το πρόγραμμα μπορεί να σας μοιάζει με απλή ακολουθία από αναζητήσεις και αντικαταστάσεις. Στην πραγματικότητα είναι κάτι πολύ πιο δύσκολο και γενικότερο. Δείτε για παράδειγμα τις περιπτώσεις των φωλιασμένων περιβαλλόντων. Θα πρέπει ακόμα να δώσετε προσοχή στην πολυπλοκότητα του αλγορίθμου που θα προτείνεται. Μπορείτε να δώσετε μια λύση που να διαβάζει μόνο μια φορά το αρχείο εισόδου;

4. Πρέπει να φροντίσετε το παραγόμενο κείμενο HTML να είναι ευκολοδιάβαστο ώστε ο χρήστης αν θέλει να μπορεί εύκολα να το διορθώσει.

Σχετικές εργασίες

Δείτε το `latex2html` στην σελίδα `http://www.latex2html.org/` για ένα αντίστοιχο πρόγραμμα. Στην σελίδα αυτή μπορείτε να βρείτε πολλά παραδείγματα αρχείων `.tex` και `.html`, το πρόγραμμα μετατροπής `latex2html` με τον πηγαίο κώδικά του. Πειραματιστείτε αρκετά με το `latex2html` πριν αρχίσετε να γράφετε το δικό σας πρόγραμμα. Στην παραπάνω διανομή υπάρχουν πολλά παραδείγματα.

Πρόσθετη βαθμολογία

Πρόσθετη βαθμολογία θα δοθεί σε εργασίες που παρέχουν επιπρόσθετη λειτουργικότητα, όπως μετατροπή (α) πινάκων (β) εικόνων και (γ) παραπομπών και αναφορών σε σχήματα, εικόνες και βιβλιογραφία (περιβάλλοντα `\label`, `\bibitem`, `\cite` και `\ref`).

Τι πρέπει να παραδώσετε

1. Αναλυτικό κείμενο με την λύση – αλγόριθμο που προτείνετε. Στο κείμενο πρέπει να αναφέρετε

   (α′) Λεπτομερή περιγραφή του αλγορίθμου που προτείνεται. Αναφερθείτε αναλυτικά στα περιβάλλοντα που υποστηρίζετε καθώς και σε αυτά που *δεν* υποστηρίζετε.

   (β′) Παρουσίαση αποτελεσμάτων του προγράμματός σας για τουλάχιστον 3 διαφορετικές εισόδους.

   (γ′) Αναλυτική και ασυμπτωτική μελέτη της πολυπλοκότητας του αλγορίθμου.

2. Εκτελέσιμο και πηγαίο κώδικα σε δισκέτα ή `CD`. Ο πηγαίος κώδικας πρέπει να περιλαμβάνει αναλυτικά σχόλια που θα παραπέμπουν στην αναφορά και στα παραδείγματά της.

Τα παραπάνω πρέπει αν παραδοθούν ιδιοχείρως (και όχι με `email`) την ημέρα της εξέτασης. Σημειώστε ότι δεν θα βαθμολογηθούν εργασίες που

- δεν συνοδεύονται από αναλυτικό κείμενο ή
- περιέχουν προγράμματα που δεν μεταγλωττίζονται σε `gcc` ή `perl`.

Παραρτήματα

Α′ Κείμενα LaTeX

Ένα κείμενο σε LaTeX έχει στην ακόλουθη δομή:

```
\documentclass[10pt,a4paper]{article}
\usepackage{epsfig}

% this is a comment
\begin{document}

% A section
\section{This is a section}

% A subsection
\subsection{This is a subsection}

% Read and process another tex file
\input{anotherlatexfile.tex}

$x_2+3y^{2+3}\leq 4$ % some maths
\end{document}
```

Β′ Κείμενα HTML

Το αντίστοιχο κείμενο του Παραρτήματος Α′ θα μπορούσε να είναι το παρακάτω.

```
<HTML>
<HEAD>
   <META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=iso-8859-7">
   <META NAME="Author" CONTENT="my latex to html translator">
</HEAD>

<!-- this is a comment -->
<BODY>

<!-- A section -->
<H1>This is a section</H1>

<P>The first paragraph of our text.

<P>The second paragraph of our text.

<!-- A subsection -->
<H2>This is a subsection</H2>

<!-- Read and process another tex file -->
... text that corresponds to anotherlatexfile.tex ...

<P><VAR>x</VAR><SUB>2</SUB>+3<VAR>y</VAR><SUP>2+<VAR>z</VAR></SUP><= 4 <!-- some maths -->

</BODY>
</HTML>
```