

# Linguaggi e Traduttori

Compito dell'11 ottobre 1997

Si costruisca, utilizzando la coppia di programmi LEX e YACC (o Flex e Bison), un traduttore guidato dalla sintassi in grado di inserire un indice in un documento HTML.

## Linguaggio di ingresso

Un documento HTML è costituito da un testo ASCII annotato mediante opportune parole chiave. Tutte le parole chiave sono racchiuse tra i simboli "<" e ">". All'interno di tali simboli possono essere presenti anche eventuali modificatori e parametri delle parole chiave. Per le parole chiave ed i parametri è indifferente l'uso di lettere maiuscole o minuscole. I due caratteri delimitatori con il loro contenuto costituiscono un **tag**.

Un documento è composto da due sezioni: **intestazione** e **corpo**. L'intero documento è compreso tra le parole chiave "html" e "/html"; l'intestazione è racchiusa tra le parole chiave "head" e "/head"; il corpo è racchiuso tra le parole chiave "body" e "/body".

All'interno dell'intestazione può comparire il titolo del documento racchiuso tra le parole chiave "title" e "/title".

All'interno del corpo compare il testo del documento intercalato da varie parole chiave. Le parole chiave possono iniziare con un carattere alfabetico o con il carattere "/".

I titoli sono racchiusi tra le parole chiave "Hn" e "/Hn" con  $n$  in  $\{1,2,3,4\}$ .

Il traduttore deve riconoscere i commenti che seguono la sintassi dell'HTML, ovvero iniziano con la serie di caratteri "<!--" e terminano con i caratteri "-->".

## Scopo del programma.

Il traduttore deve ripetere sullo standard output il proprio input eliminando i commenti ed aggiungendo nella parte finale del corpo l'indice costruito in base ai titoli presenti nel documento.

L'indice deve essere iniziato con il titolo di primo livello "Indice" che non sarà inserito nell'indice stesso.

L'indice deve essere organizzato gerarchicamente, tramite l'uso di liste non numerate annidate. Gli elementi di una lista non numerata sono racchiusi tra le parole chiave "ul" e "/ul". Ogni elemento della lista è preceduto dalla parola chiave "li".

Gli elementi dell'indice sono i titoli presenti nel documento. Essi sono riferimenti ipertestuali alle posizioni dei titoli all'interno del documento. È possibile ottenere tale effetto nel seguente modo:

- racchiudere i titoli nel documento tra la parola chiave "a" con parametro "name" e la parola chiave "/a"; il valore del parametro name deve essere un identificatore unico assegnato al titolo;
- racchiudere i titoli riportati nell'indice tra la parola chiave "a" con parametro "href" e la parola chiave "/a"; il valore assegnato al parametro href deve essere l'identificatore unico assegnato al titolo relativo;

Quando il programma incontra un errore, sia sintattico che semantico, deve segnalarlo e terminare l'esecuzione: non è richiesta la gestione ed il recupero degli errori.

## Esempio

Dato il seguente documento in ingresso:

```

<HTML><HEAD><TITLE>Prova</TITLE></HEAD>
<BODY>
<!-- .... -->
<H1>Titolo_1</h1>
<h2>Titolo_1_1</H2>
Testo <b>vario</b>
<H1>Titolo_2</h1>
<h2>Titolo_2_1</H2>
Vario testo
<h2>Titolo_2_2</H2>
<ul>
<li>pippo
<li>pluto
</ul>
<hr>
<!-- Fine dell'esempio -->
</body></HTML>

```

Il traduttore dovrebbe generare un documento simile al seguente:

```

<HTML><HEAD><TITLE>Prova</TITLE></HEAD>
<BODY>
<a name="H_1">
<H1>Titolo_1</H1>
</a>
<a name="H_1_1">
<H2>Titolo_1_1</H2>
</a>
Testo <b>vario</b>.
<a name="H_2">
<H1>Titolo_2</h1>
</a>
<a name="H_2_1">
<h2>Titolo_2_1</H2>
</a>
Vario testo
<a name="H_2_2">
<h2>Titolo_2_2</H2>
</a>
<ul>
<li>pippo
<li>pluto
</ul>
<hr>
<H1>Indice</H1>
<ul>
<li><a href="#H_1">Titolo_1</a>
<ul>
<li><a href="#H_1_1">Titolo_1_1</a>
</ul>
<li><a href="#H_2">Titolo_2</a>
<ul>
<li><a href="#H_2_1">Titolo_2_1</a>
<li><a href="#H_2_2">Titolo_2_2</a>
</ul>
</li>
</ul>
</BODY></HTML>

```