

Linguaggi e Traduttori – Laboratorio

Compito del 26/01/2004

Si costruisca, utilizzando la coppia di programmi LEX e YACC (o Flex e Bison), un traduttore guidato dalla sintassi in grado di analizzare un sottoinsieme del linguaggio LaTeX.

Linguaggio di ingresso

Un documento LaTeX è costituito da testo ASCII annotato tramite opportuni comandi. Un comando è costituito da un backslash ('\') seguito da una parola chiave (costituita da caratteri alfanumerici). Una parola chiave è seguita da un attributo, racchiuso tra '{' e '}'. Un attributo è costituito da una serie di parole, costituite da caratteri alfanumerici, e separate da spazi. LaTeX permette inoltre di definire degli ambienti: tutto il testo che compare tra le parole chiave `\begin{<nome_ambiente>}` e `\end{<nome_ambiente>}` appartiene ad un ambiente e viene formattato secondo le regole di quell'ambiente.

Un documento LaTeX è formato da un preambolo e dal corpo del documento. Il preambolo è costituito dal comando `\documentstyle{<classe>}`.

Il corpo del documento è racchiuso tra i comandi `\begin{document}` e `\end{document}`. All'interno del corpo compare il testo intercalato da comandi e da definizioni di ambienti. E' necessario quindi che ad ogni comando `\begin{<nome_ambiente>}` sia associato un comando `\end{<nome_ambiente>}`. Il corpo del documento può poi essere suddiviso in sezioni e sottosezioni tramite i comandi `\section{<nome_sezione>}` e `\subsection{<nome_sezione>}`, che iniziano rispettivamente una nuova sezione o sottosezione.

Il traduttore deve anche riconoscere i commenti, che sono composti da una serie di caratteri iniziata da % e terminata dal new-line.

Scopo del programma.

Il traduttore deve analizzare un file LaTeX e stampare su stdout alcune statistiche.

In particolare deve visualizzare:

- la classe del documento
- l'elenco degli ambienti definiti in tutto il documento
- l'elenco (numerato) delle sezioni e delle sotto-sezioni definite nel documento

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. Ciò significa che il programma deve verificare che il documento segua le regole del linguaggio d'ingresso (come specificato al punto precedente).

Esempio

Dato il seguente documento in ingresso:

```
\documentstyle{article}
%Preambolo.
\begin{document}
\section{Introduzione a LaTeX}
```

Un esempio di documento scritto con LaTeX.

```
\subsection{Suddivisione del documento}
```

Lo stile article prevede una suddivisione in sezioni e sottosezioni.

```
\subsection{Paragrafi}
```

Il testo di un paragrafo termina quando nel sorgente viene incontrata una riga vuota.

```
\subsection{Gli ambienti}    % definizione di un'altra sottosezione
```

LaTeX utilizza gli ambienti per definire dei comportamenti circoscritti a zone particolari del testo.

```
\begin{center}
esempio di testo centrato.
\end{center}
```

```
\section{Figure}
\subsection{Introduzione}
```

```
Esempio figura.
\begin{figure}
\caption{Esempio di sottotitolo}
\end{figure}
% Fine del documento.
```

```
\end{document}
```

Il traduttore dovrebbe visualizzare su stdout:

Classe del documento: article

Elenco degli ambienti utilizzati:

document
center
figure

Elenco delle sezioni:

1. Introduzione a LaTeX
1.1 Suddivisione del documento
1.2 Paragrafi
1.3 Gli ambienti
2. figure
2.1 Introduzione