

HTML-validering i endelige automater og push-down-automater

Vi ønsker å lage en automat som verifiserer eller forkaster HTML-dokumenter. For at automaten ikke skal bli alt for stor, bruker vi følgende forenklinger:

- Antall tagger er redusert
- Taggene skrives, om ønskelig, i en forhåndsdefinert kortform
- Ingen tagger har attributter (som HREF="http://www.example.com" etc)

Tagg	Kortform
<HTML> </HTML>	H /H
<HEAD></HEAD>	O /O
<TITLE> </TITLE>	T /T
<BODY> </BODY>	B /B
<P> </P>	P /P
 	I /I
<A> 	A /A

Følgende regler skal gjelde:

- <HTML> må være første element, </HTML> siste
- <HEAD> </HEAD> må komme før <BODY> men kan utelates
- <TITLE> </TITLE> må komme mellom <HEAD> og </HEAD>
- <HTML>, <HEAD>, <TITLE> og <BODY> kan kun forekomme én gang
- <BODY> </BODY> kan ikke utelates
- <P></P> og <A> må komme mellom <BODY> og </BODY>
- <A> kan kun komme mellom <P> og </P>, men ikke mellom og
- kan kun komme mellom <P> og </P> eller <A> og
- Ingen tagg kan komme mellom to tagger av seg selv. For eksempel er ikke dette lov:
 - o II/I/I (``)
 - o AA/A/A (`<A><A>`)
 - o PP/P/P (`<P><P></P></P>`)

Alle karakterer som kan komme ellers i et HTML-dokument skriver vi som k (karakter), hvor følgende regler gjelder.

- k kan kun komme
 - o mellom <TITLE> og </TITLE>
 - o mellom <P> og </P>
 - o mellom <A> og
- k kan aldri komme mellom og

Her er noen eksempler på lovlige HTML-dokumenter, i følge reglene over:

- HOThei/T/OBPI/I/P/B/H
- HB/B/H
- HO/OBPA/A/PPI/I/P/H

Oppgave 1

Oppgave 1.1

Kan dette HTML-språket uttrykkes i et regulært språk? Hvis ikke, kom med et pumpelemma-bevis.

Oppgave 1.2

Lag en automat som godkjenner språket over.

Oppgave 2

Vi legger til følgende tagger: <DIV></DIV>, som forkortes til D og /D. Følgende regler gjelder for denne:

- <DIV></DIV> kan kun komme mellom <BODY> og </BODY>
- <DIV></DIV> kan komme mellom <DIV> og </DIV> og <P> </P>
- k kan komme mellom <DIV> </DIV>

Oppgave 2.1

Kan dette utvidete språket uttrykkes i et regulært språk? Hvis ikke, kom med et pumpelemma-bevis.

Oppgave 2.2

Lag en automat som godkjenner det utvidete språket.