Tapaus 10

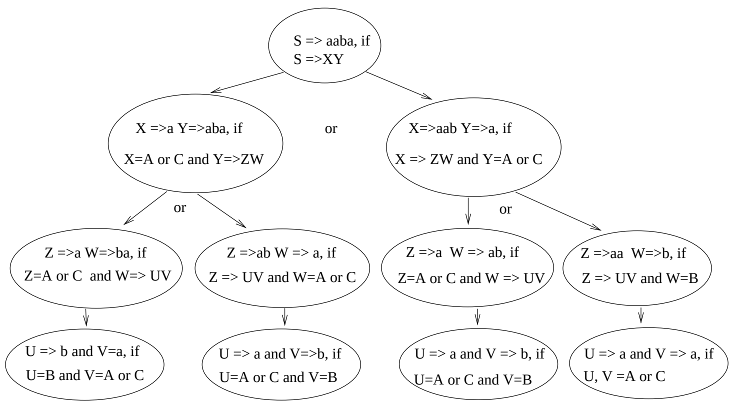
Tällä viikolla on vain yksi ongelma, josta voit saada yhteensä **6 pistettä**.

Pomosi on tuonut sinulle hirveän kasan kontekstittomia kieliä, joille sinun pitäisi laatia tehokkaat jäsentäjät viikossa! Päästyäsi pahimman epätoivon yli huomaat, että kaikki annettujen kielioppien säännöt voidaan esittää seuraavassa muodossa:
$A \to BC$ tai $A \to a$
ja korkeintaan lähtösymboli S on tyhjentyvä (ts. sille voi olla sääntö $S \to \epsilon$).

Mietit ensin jonkinlaista rekursiivista jäsentäjää, joka tutkisi kaikkia mahdollisia merkkijonon johtoja. Esimerkiksi kieliopissa

$S \to AB|BA$
$A \to CC|a$
$B \to BC|b$
$C \to AB|a$

merkkijono *aaba* kuuluu kieleen, jos pätee:



Tajuat kuitenkin, että tällainen ylhäältä alaspäin tenevä jäsentäminen on tuskallisen hidasta ja siinä tehdään paljon turhaa työtä. Voisiko jäsennyspuita rakentaa alhaalta ylöspäin, ts. miettiä ensin, mistä lehdet (päätemerkit) on johdettu, sitten mistä nämä välikesymbolit on johdettu jne. kunnes pääset juureen saakka (jonka pitäisi olla S)?

a) Laadi yleinen jäsennysalgoritmi kaikille tämän muotoisille (ns. *Chomskyn normaalimuodossa*) oleville kieliopeille. Testaa ideaasi ainakin em. kieliopilla ja merkkijonoilla *aababa* ja *abbba*.

b) Entä jos annettu kielioppi ei olekaan tässä muodossa? Voitko jotenkin muuntaa sen ohjelmasi haluamaan muotoon?