

Exp	:	Id	{...}
		$Exp' + Exp$	{...}
		$Exp' * Exp$	{...}
		$'(Exp)'$	{...}
		;	
Id	:	'a'	{...}
		'b'	{...}
		Id 'a'	{...}
		Id 'b'	{...}
		Id '0'	{...}
		Id '1'	{...}
		;	

missä {...} sisältää jonkin toiminnon C-koodina.

Dokumentin rakenteen kuvaus

Dokumentin rakenteen kuvaukseen käytettävät kielet kuten HTML, XML ja Latex voidaan esittää kontekstittomina kielioppeina. Tällaisen kielen merkkijonot ovat dokumentteja, joissa esiintyy tiettyjä kielen semantiikkaa kuvaavia merkkejä (*tags*). Esimerkiksi HTML-dokumentissa $\langle OL \rangle$ ja $\langle /OL \rangle$ määrittelevät järjestetyn listan. HTML-jäsennin tarkistaa, että dokumentti on HTML:n syntaksin mukainen ja ohjaa dokumentin prosessointia. XML:llä puolestaan voidaan kuvata tekstin semantiikkaa – mitä tietyt merkkijonot merkitsevät. Esimerkiksi $\langle ADDR \rangle$ Perhospolu 17 C 3, 33000 Kukkamäki $\langle /ADDR \rangle$ kertoo, että osajono on osoite.

5.13 Harjoituksia kontekstittomista kielistä

1. Olkoon kontekstiton kielioppi $G = (V, \Sigma, P, S)$, missä $V = \{A, B, C, D, S\} \cup \Sigma$, $\Sigma = \{Jim, big, green, cheese, ate\}$ ja säännöt P ovat:
 $A \rightarrow B|CA$
 $S \rightarrow ADA$
 $C \rightarrow big|green$
 $B \rightarrow cheese|Jim$
 $D \rightarrow ate$

Millaisia merkkijonoja kuuluu kieliopin G kuvaamaan kieleen kuuluu? Anna joitain esimerkkejä! Kuuluvatko seuraavat lauseet kieleen $L(G)$: "big big green cheese", "ate Jim big cheese"?

2. Laadi kontekstittomat kieliopit seuraaville aakkoston $\Sigma = \{a, b\}$ kielille:
- $L_1\{ww^R \mid \text{missä } w^R = a_n a_{n-1} \dots a_0 \text{ kun } w = a_0 a_1 \dots a_n\}$
 - $L_2 = \{w c w^R \mid c \in \Sigma\}$
 - $L_3 = \{w w^C \mid \text{missä } w^C \text{ on saatu } w\text{:stä muuttamalla kaikki } a:tb\text{:ksi ja päinvastoin}\}$

3. Laadi seuraavan kieliopin kuvaaman kielen tunnistava äärellinen automaatti. Mikä on vastaava säännöllinen lauseke?

$$\begin{aligned} S &\rightarrow aA|bB \\ A &\rightarrow aS|bA \\ B &\rightarrow bB|\epsilon \end{aligned}$$

4. Laadi oikealle lineaarinen kielioppi kielen $\{w \in \{a, b\}^* \mid w \text{ ei sisällä merkkijonoa } abaa \text{ kuvaamiseen}\}$

5. Ulkoavaruuden olioiden kielessä kaikki sanat noudattavat Blurbsin muotoa. Blurb on Whoozit, jota seuraa yksi tai useampi Whatzit. Whoozit on kirjain 'x', jota voi seurata haluttu määrä 'y'-kirjaimia (myös nolla (0) kappaletta). Whatzit on kirjain 'q', jota seuraa joko kirjain 'z' tai 'd', jota seuraa Whoozit.

Anna Blurbsin kielen kuvaava kontekstiton kielioppi. Osaatko nyt laatia vastaavan säännöllisen lausekkeen? (Vihje: piirrä ensin automaatti kieliopista!)

6. Säännöllisille lausekkeille pätee seuraava päättelysääntö:

Jos $r = rs \cup t$, niin $r = ts^*$, kun $\epsilon \notin L(s)$.

Osoita tämä sääntö todeksi kontekstittomien kielioppien avulla!

7. Laadi kontekstiton kielioppi, joka tuottaa seuraavan esimerkin tapaiset, rajattoman monista sisäkkäisistä for-silmukoista ja alkeisoperaatioista a koostuvat ohjelmointikielen lauseet:

```
for (i = N; i < N; i++) {
  for (j=N; j<N; j++) {
    a;
  }
}
```

missä N on kokonaisluku.

8. Kuvaa seuraavanlaiset C-kielen ohjelmat kontekstittomana kielioppina.

- funktio on muotoa

$$\text{fktyyppi fuktnimi}(\text{partyppi par } (, \text{partyppi par})^*)$$

$$\{$$

$$\quad \text{fktrunko}$$

$$\}$$

missä fktyyppi voi olla void tai int ja partyppi on int

- fktrunko on muotoa iflause tai sijoitus
- sijoitus on muotoa $\text{muuttuja}=\text{arvo}$;
- par ja muuttuja ovat 1-kirjaimisia muuttujan nimiä
- arvo on kokonaisluku
- iflause on muotoa
 $\text{if (ehto) iflause sijoitus}$; tai
 $\text{if (ehto) iflause sijoitus; else iflause sijoitus}$; tai
 tyhjä merkkijono
- ehto on muotoa $(\text{muuttuja}==\text{arvo})$

9. Piirrä oheinen pinoautomaatti M :

$$M = (\{q_0, q_1, q_2\}, \{a, b\}, \{A\}, \delta, q_0, \{q_1, q_2\}), \text{ missä}$$

$$\delta(q_0, a, \epsilon) = \{(q_0, A)\}$$

$$\delta(q_0, \epsilon, \epsilon) = \{(q_1, \epsilon)\}$$

$$\delta(q_0, b, A) = \{(q_2, \epsilon)\}$$

$$\delta(q_1, \epsilon, A) = \{(q_1, \epsilon)\}$$

$$\delta(q_2, b, A) = \{(q_2, \epsilon)\}$$

$$\delta(q_2, \epsilon, A) = \{(q_2, \epsilon)\}$$

Millaisen kielen M tunnistaa? Tutki kaikki laskentapolut syöteillä aab , abb ja aba ! Hyväksykö M merkkijonot $aabb$ ja $aaab$?

10. Merkitään $w^R = \text{merkkijono } w \text{ takaperin kirjoitettuna}$ (so. jos $w = a_1a_2 \dots a_n$, niin $w^R = a_n \dots a_2a_1$). Merkkijono on palindromi, jos $w = w^R$ (esimerkiksi "isorikassikakissakirosi"). Tarkastellaan aakkoston $\{a, b\}$ palindromien muodostamaa kieltä $\text{PAL} = \{w \in \{a, b\} \mid w = w^R\}$.
- a) Laadi kielen tuottava kontekstiton kielioppi.
 b) Muodosta kielen tunnistava pinoautomaatti.
11. Laadi kontekstiton kielioppi, joka tuottaa lukua 90 pienemmät roomalaiset numerot. Lukumerkit: L=50, X=10, V=5, I=1. Esim. XII=12, XIV=14, XL=40, XLIX=49. Yksinkertaisuuden vuoksi voit olettaa, että tyhjä merkkijono esittää lukua 0 ja merkkijono LXL lukua 90.

12. Muodosta pinoautomaatti, joka tunnistaa edellisen tehtävänannon mukaiset lukua 90 pienemmät roomalaiset numerot!

13. Kissastanian logiikkakoulussa on edetty kontekstittomiin kieliin. Tunnin aiheena on seuraava kieli:

Naukaisu voi alkaa yhdellä tai useammalla miu :lla ja lopussa on yhtä monta mau :ta. Välissä voi olla 0 tai useampia u -kirjaimia. Tai naukaisu on maukaisu, jota seuraa yksi tai useampi au . Maukaisu puolestaan alkaa $miau$:lla tai mau :lla, jota seuraa 0 tai useampia u -kirjaimia.

- Anna kielen tuottava kielioppi.
- Anna joitain esimerkkejä kielen sanoista.
- Laadi pinoautomaatti, joka tunnistaa kielen.

14. Millaisen kielen oheinen pinoautomaatti M tunnistaa?

$$M = (\{q_0, q_1\}, \{a, b\}, \{\underline{A}, \underline{A}_1, \underline{B}, \underline{B}_1, A, A_1, B, B_1\}, \delta, q_0, \{q_0\})$$

$$\delta(q_0, a, \epsilon) = \{(q_1, \underline{A}_1)\}$$

$$\delta(q_0, b, \epsilon) = \{(q_1, \underline{B})\}$$

$$\delta(q_1, a, \epsilon) = \{(q_1, \underline{A}_1)\}$$

$$\delta(q_1, a, B) = \{(q_1, \underline{B}_1)\}$$

$$\delta(q_1, a, B_1) = \{(q_1, \epsilon)\}$$

$$\delta(q_1, a, \underline{B}) = \{(q_1, \underline{B}_1)\}$$

$$\delta(q_1, a, A_1) = \{(q_1, A)\}$$

$$\delta(q_1, a, \underline{A}_1) = \{(q_1, \underline{A})\}$$

$$\delta(q_1, b, \epsilon) = \{(q_1, B)\}$$

$$\delta(q_1, b, A) = \{(q_1, \epsilon)\}$$

$$\delta(q_1, b, A_1) = \{(q_1, B_1)\}$$

$$\delta(q_1, b, \underline{A}_1) = \{(q_1, \underline{B}_1)\}$$

15. Osoita, että seuraavat kieliopit ovat moniselitteisiä! Osaatko laatia saman kielen kuvaavan yksiselitteisen kieliopin?

$$(a) S \rightarrow aSb|SS|\epsilon$$

$$(b) S \rightarrow aSb|aaSb|\epsilon$$

16. Kieli $L = \{a^i b^j c^k \mid i = j \text{ tai } j = k\}$ tiedetään luonnostaan moniselitteiseksi. Anna kielen kuvaava kielioppi. Perustele miksi kielelle ei voi antaa yksiselitteistä kielioppia!

17. Laadi kielioppi, joka tuottaa kielen $\{a^m b^n c^{m+n} \mid m, n \geq 0\}$. Onko laatimasi kielioppi yksi- vai moniselitteinen?

18. Mitä haittaa on siitä, jos kieli on luonnostaan moniselitteinen?

19. Osoita, että seuraavat kielet ovat kontekstittomia!

$$\text{a) } \{a^m b^n \mid m \geq n\}$$

$$\text{b) } \{a^m b^n c^p d^q \mid m + n = p + q\}$$

$$\text{c) } \{uawb \mid u, w \in \{a, b\}^*, |u| = |w|\}$$

$$\text{d) } \{a^m b^n \mid n \leq m \leq 2n\}$$

20. Olkoon aakkosto $\Sigma = \{m, i, u, a\}$. Tarkastellaan seuraavaa kielioppia:

$$S \rightarrow miuNmauS \mid SmiauA \mid \epsilon$$

$$N \rightarrow miuNmau \mid U$$

$$U \rightarrow uU \mid \epsilon$$

$$A \rightarrow Aau \mid \epsilon$$

Anna merkkijonon *miumiuuumaumaumiaumiauuu* vasen ja oikea johto sekä jäsenyspuu! Onko kielioppi yksi- vai moniselitteinen?

21. Seuraava kielioppi on melkein $LL(1)$ -muodossa. Muunna se $LL(1)$ -muotoon!

$$S \rightarrow miuNmau \mid SnauU$$

$$N \rightarrow miuNmau \mid U$$

$$U \rightarrow Uu \mid \epsilon$$

22. Täyttääkö seuraava kielioppi $LL(1)$ -ehdon?

$$S \rightarrow (L) \mid p \mid q$$

$$L \rightarrow LandS \mid LorS \mid S$$

Millaisen kielen kielioppi kuvaa?

23. Laadi (pseudokoodina) rekursiivinen jäsentäjä kieliopin kuvaamalle kielelle (Huom! Muunna ensin tarvittaessa $LL(1)$:een!).

24. Olkoon L kieli, jonka sanat koostuvat mistä tahansa tekstistä sekä laillisista sulkurakenteista. Esimerkiksi lause ”(suuri) kissaeläin (joko leijona tai tiikeri { jotka harvinaisia } tai sapelihammastiikeri { joka sukupuuttoon kuollut [ks. tarkemmin Kurten] – harmi –} tai leopardi) on viehättävä (mutta vaarallinen!) tuttavuus.” kuuluu kieleen. Voit olettaa yksinkertaisuuden vuoksi, että tekstiosuudet sisältävät vain pieniä kirjaimia a..z sekä välilyöntejä.

Anna kielen kuvaava $LL(1)$ -kielioppi ja suunnittele sille rekursiivinen jäsentäjä!

25. Tarkastellaan yksinkertaista ohjelmointikieltä, jonka kaikki ohjelmat ovat muotoa *begin* LAUSE; *end*. LAUSE voi olla yksinkertainen sijoitus MUUTTUJA:=TEKIJA;; MUUTTUJA:=TEKIJA+TEKIJA; tai MUUTTUJA:=TEKIJA-TEKIJA;; missä TEKIJÄ on joko muuttuja a, \dots, z tai etumerkitön kokonaisluku. Tai LAUSE on *while*-rakenne: *while* (EHTO) *do begin* LAUSE *end*. EHTO puolestaan vertailee kahta tekijää operaatioilla $<, <=, >, >=, ==, <>$.

Laadi kielelle kielioppi ja kuvaa, miten se jäsennetään rekursiivisesti!

26. Laadi kontekstiton kielioppi, joka kuvaa yhden muuttujan x polynomeja. Yksinkertaisuuden vuoksi voit olettaa, että termien kertoimet ja eksponentit ovat yksinumeroisia kokonaislukuja ja ensimmäinen termi on etumerkitön. Termien ei tarvitse olla missään tietyssä järjestyksessä ja samanasteisia termejä voi polynomissa olla useita. Anna seuraavien merkkijonojen jäsennykset kieliopisasi: $2 * x^2 - 2 * x + 1$, $x - x^2 - 1 + 2 * x^2 - 3x$, x , $5!$ Osaatko laatia kielelle rekursiivisen jäsentimen?
27. Laadi ohjelma, joka muuntaa HTML-listarakenteet vastaaviksi latex-listoiksi. HTML-listat ovat muotoa `(text)` ja `(text)` ja latexissa niitä vastaavat listat `\begin{itemize}(\item text)\end{itemize}` ja `\begin{enumerate} (\item text)\end{enumerate}`. Huomaa, että sisäkkäisiä listoja voi olla useita!
28. Osoita, että seuraavat kielet ovat deterministisiä:
- $\{a^m b^n \mid m \neq n\}$
 - $\{w c w^R \mid w \in \{a, b\}^*\}$
 - $\{a^m c b^m\} \cup \{a^m d b^{2m}\}$

29. Muunna kielioppi

$$\begin{aligned} S &\rightarrow (S) | A \\ A &\rightarrow SS | \epsilon \end{aligned}$$

Chomskyn normaalimuotoon. Esitä myös välivaiheet (ϵ -produktioiden poisto, yksikköproduktioiden poisto)!

30. Muunna kielioppi

$$\begin{aligned} S &\rightarrow ABC | a \\ A &\rightarrow aAaa | \epsilon \\ B &\rightarrow bBbb | \epsilon \\ C &\rightarrow cCa | c \end{aligned}$$

Chomskyn normaalimuotoon. Esitä myös välivaiheet!

31. Simuloi CYK-algoritmin toimintaa sen ratkaistessa, kuuluvatko merkkijonot bbaab, ababab ja aabba kieliopin

$$S \rightarrow AS|b$$

$$A \rightarrow SA|a$$

tuottamaan kieleen. Myönteisissä tapauksissa esitä merkkijonoille kieliopin mukaiset jäsennesspuut.

32. Tarkastellaan kieltä $L = \{a^i b^j c^k \mid i = j \text{ tai } j = k\}$.
- Laadi kielelle kielioppi, joka on Chomskyn normaalimuodossa.
 - Simuloi CYK-algoritmia syötteellä *aabbcc*.
 - Lue kaikki merkkijonon jäsennesspuut CYK-*taulukosta*.

33. Toteuta CYK-algoritmi tietokoneohjelmana kieliopille

$$S \rightarrow AB|BC$$

$$A \rightarrow BA|a$$

$$B \rightarrow CC|b$$

$$C \rightarrow AB|a$$

Testaa ohjelmaasi erilaisilla merkkijonoilla!

34. Millä tavalla jäsentäisit seuraavat kielet? Perusstrategia riittää.
- HTML
 - Propositilogiikan lauseet, jotka koostuvat atomilauseista A, B, \dots, Z ja operaatioista $\neg, \wedge, \vee, \Rightarrow, \Leftrightarrow$, sekä suluista.
 - C- tai Pascal-ohjelmakoodi
 - SQL-kysely
35. Olkoon $L_1 = \{a^n b^{2n} c^m \mid n, m \geq 0\}$ ja $L_2 = \{a^n b^m c^{2m} \mid n, m \geq 0\}$. Onko $L_1 \cap L_2$ kontekstiton kieli? Perustele vastauksesi!
36. Laadi tehokas algoritmi, joka tutkii, onko välikesymboli tyhjentyvä (ts. voiko se tuottaa ϵ :in jossain lausejohdoksessa)!
37. Osoita, että kontekstittomat kielet ovat suljettuja yhdisteen, katenaation ja sulkeuman suhteen. Ts. jos L_1 ja L_2 ovat kontekstittomia, niin
- $L_1 \cup L_2$
 - $L_1 L_2$
 - L_1^*
- ovat kontekstittomia. (Vihje: oletta, että on olemassa pinoautomaatit $M(L_1)$ ja $M(L_2)$ ja laadi sulkeumakielille pinoautomaatit.)

38. Kieltä $\{a^n b^n c^n | n \geq 0\}$ ei voida tunnistaa tavallisella pinoautomaatilla. Voisiko sen tunnistaa, jos käytössä olisi kaksipinoinen automaatti? Jos voi, niin piirrä ko. automaatin siirtymäkaavio ja simuloi sen toimintaa! Jos ei, niin perustele, miksi ei!
39. Osoita kontekstittomien kielten pumppauslemmalla, että kieli $\{ww \mid w \in \{a, b\}^*\}$ ei ole kontekstiton!
40. Osoita kontekstittomien kielten pumppauslemmalla, että kieli $\{a^p \mid p \text{ on alkuluku}\}$ ei ole kontekstiton!
41. Kertaustehtävä: Selosta mitä seuraavat kieliluokat tarkoittavat ja kuinka jäsenyysongelma ratkaistaan tehokkaimmin kussakin luokassa!
- a) Lineaariset kielet
 - b) Deterministiset kielet
 - c) LL(1)-kielet
 - d) Yksiselitteiset kielet
 - e) Luonnostaan moniselitteiset kielet