Automaadid, keeled, translaatorid

Varmo Vene

Aivar Annamaa Vesal Vojdani

Arvutiteaduse Instituut
Tartu Ülikool

Administratiivinfo

Üldinfo

- MTAT.05.085 6 ainepunkti.
  - Loengud (24 tundi)
  - Praktikumid (32 tundi)
  - Iseseisev töö (100 tundi)
- Toimumised:
  - E 16.15–18.00, Liivi 2 - 405 (reeglina loeng)
  - N 12.15–14.00, Liivi 2 - 405 (reeglina praktikum)
- Kursuse veebileht:
  https://courses.cs.ut.ee/2015/AKT/spring
- Mitteavalikud materjalid *Moodle*'i keskkonnas:
  http://moodle.ut.ee/

Administratiivinfo

Eesmärgid

Anda ülevaade programmeerimiskeelte klassikalistest transleerimise meetodidest, enamlevinud algoritmidest ja tehnikatest.

Administratiivinfo

Õpiväljundid

Kursuse läbinu omandab:

- ülevaate translaatori erinevate osade ülesannetest ning nende realiseerimismeetodidest;
- oskuse kirjeldada keele lekseeme regulaaravaldistega ja konstrueerida regulaaravaldisele vastava lõpliku automaadi;
- oskuse spetsifitseerida keele süntaksit kontekstivabade grammatikate abil;
- praktilise kogemuse lihtsa programmeerimiskeele translaatori realiseerimisel.

Administratiivinfo

Hinde kujunemine

- Kodutööd (20 punkti)
  - kokku 10 kodutööd
  - iga kodutöö annab 10 punkti
  - tuleb koguda vähemalt 50 punkti
  - summa jagatakse 5-ga
- 1. vahetest (20 punkti; 16. märts)
- 2. vahetest (20 punkti; 20. aprill)
- Eksam (40 punkti)
  - võib asendada projektiga kui kodutöödest on saadud vähemalt 91 punkti.

Administratiivinfo

Hindeskaala

	Hinne	Punkte
A	suurepärane	≥ 91
B	väga hea	$81 - 90$
C	hea	$71 - 80$
D	rahuldav	$61 - 70$
E	kasin	$51 - 60$
F	puudulik	$0 - 50$

Administratiivinfo

2014K tagasiside (1)

- ...
- Raske aine, soovitan kindlalt kodutöid teha.
- Tehke kodutöid, mahukad võivad nad olla, aga tegelikult pole üldse rasked. Kodutööde tegemine aitab 2. vahetestist ja eksamist väga kergelt läbi saada.
- Iseseisva töö maht on suur aga seda väärt. Soovitan kõiki ülesandeid vähemalt proovida lahendada ja kindlasti praktikumides kohal käia. Samuti tasub sirvida courses lehel olevat õpikut, mis seletab pikemalt lahti asju mis loengus võisid segaseks jääda.
- Tegemist on üsna keerulise ainega, kus tuleks eduka soorituse jaoks järjepidevalt kõvasti tööd teha.

Administratiivinfo

2014K tagasiside (2)

- Nõuab aega ja tasub loengutes/praksides aktiivselt osaleda. Hiljem ise õppida suhteliselt keerukas.
- Raskeimad progemiskodutööd siiani. Loodan et oskate hästi javat. Õnneks nad annavad ainult 1p/each.
- Tee kodutöid, aitab kõvasti projekti tegemisel ja kontrolltööde edukal sooritamisel.
- Uus formaat on väga erinev varasemast. Kodutööd aitavad õppida aga annavad vähe punkte (pole hullu kui tegemata jäävad). Eksam oli 1. + 2. vahetest kokku pandud.
- Tegelikkuses ei ole kodutööd nii keerulised kui nad tunduvad. Kontrolltöödeks tasub õppida, sest eksam on väga sarnane ja võimalus samade teadmistega punkte saada on hea!
- ...

Kirjandus

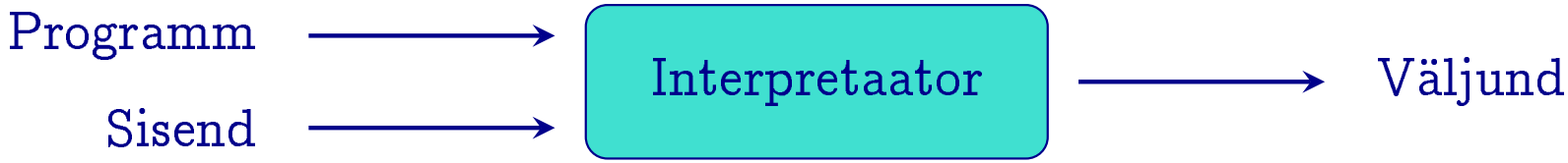
Põhiõpik

- Torben Mogensen. *Introduction to Compiler Design.* Springer, 2011.

Teisi raamatuid

- R. Wilhelm, H. Seidl, S. Hack. *Compiler Design: Syntactic and Semantic Analysis.* Springer, 2013.
- A. Aho, M. Lam, R. Sethi, J. Ullman. *Compilers: Principles, Techniques, and Tools.* Pearson Education, 2006.

Interpretaator



Eelised: Ei toimu programmiteksti eeltöötlemist

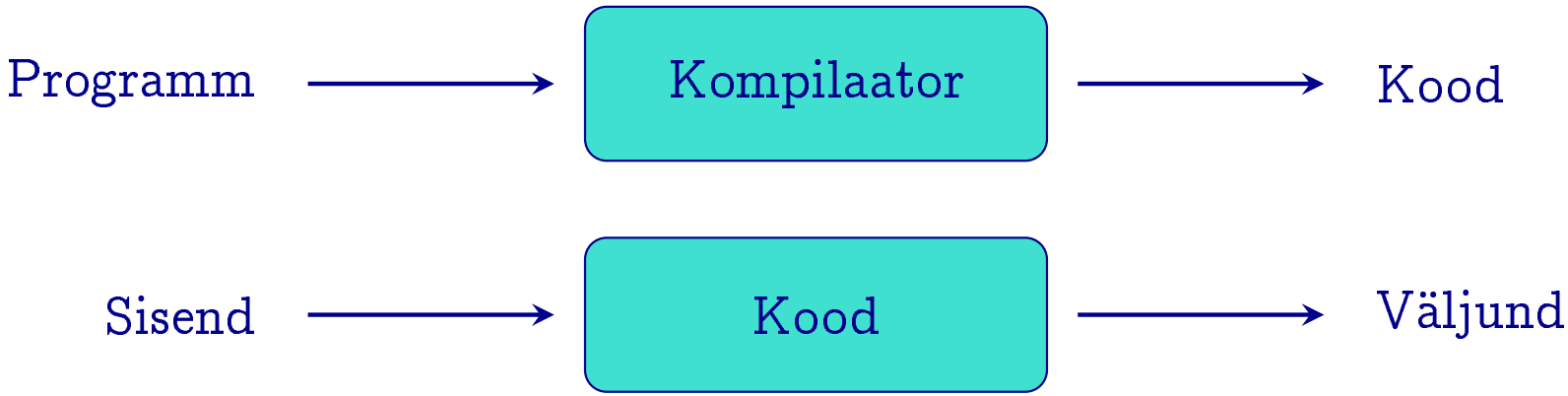
⇒ algkäivituseks kulub vähe aega;

⇒ soodustab interaktiivset programmide kirjutamist.

Puudused: Programmi osasid analüüsitakse täitmisajal korduvalt

⇒ programmide täitmine suhteliselt aeglane.

Kompilaator

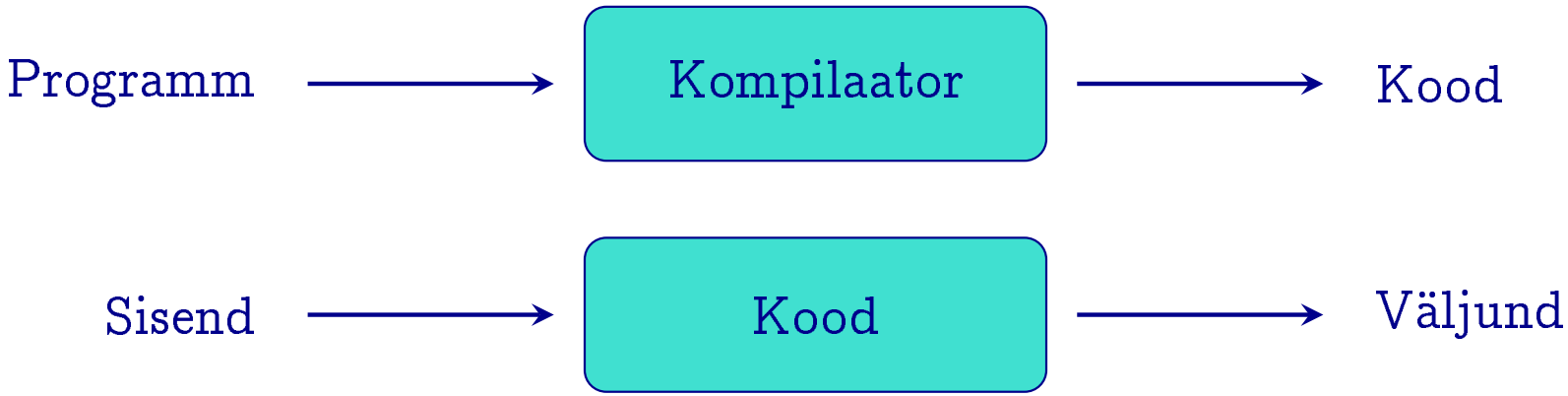


Kaks faasi:

Transleerimisaeg: lähteprogramm transleeritakse täidetavaks (masin-)koodiks.

Täitmisaeg: koodi täitmine antud sisendandmetel.

Kompilaator



Eelised: Programmi analüüsitakse ainult üks kord

$\Rightarrow$ võimaldab programmide (globaalset) optimeerimist;

$\Rightarrow$ programmide täitmine on kiirem.

Puudused: Kompileerimine võtab aega

$\Rightarrow$ kompileerimine tasub ära pikalt töötavate ja tihti kasutatavate programmide korral.

Süntaksjuhitav kompileerimine

- Alates Algol60st, mis oli esimene formaalselt defineeritud süntaksiga keel, juhindub kompilaatori poolt teostatav transleerimisprotsess lähteprogrammi süntaktilisest struktuurist.
- Programmeerimiskeele definitsioon koosneb:

  leksika: defineerib keeles legaalsete sõnade, lekseemide, moodustamise reeglid;

  süntaks: defineerib programmide grammatilise struktuuri;

  semantika: kirjeldab kontekstist sõltuvaid tingimusi (näit. tüüpimisreeglid) ja programmi tähendust.

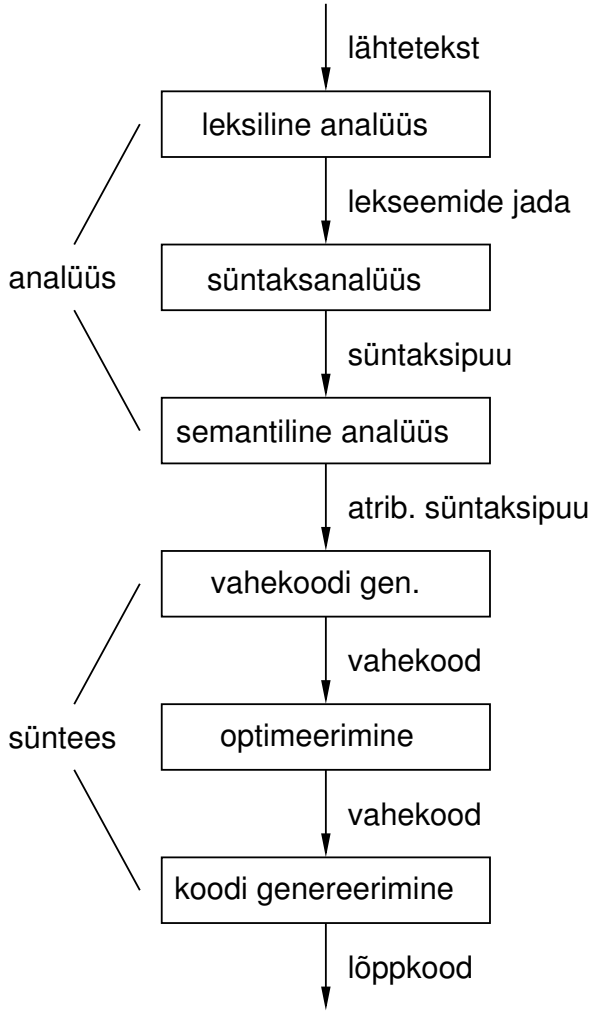
Kompilaatori struktuur

- Kompileerimisprotsess koosneb suures plaanis kahest faasist:

  analüüs: lähteteksti struktuuri äratundmine vastavalt grammatikale ja kontekstuaalsete sõltuvuste kontrollimine;

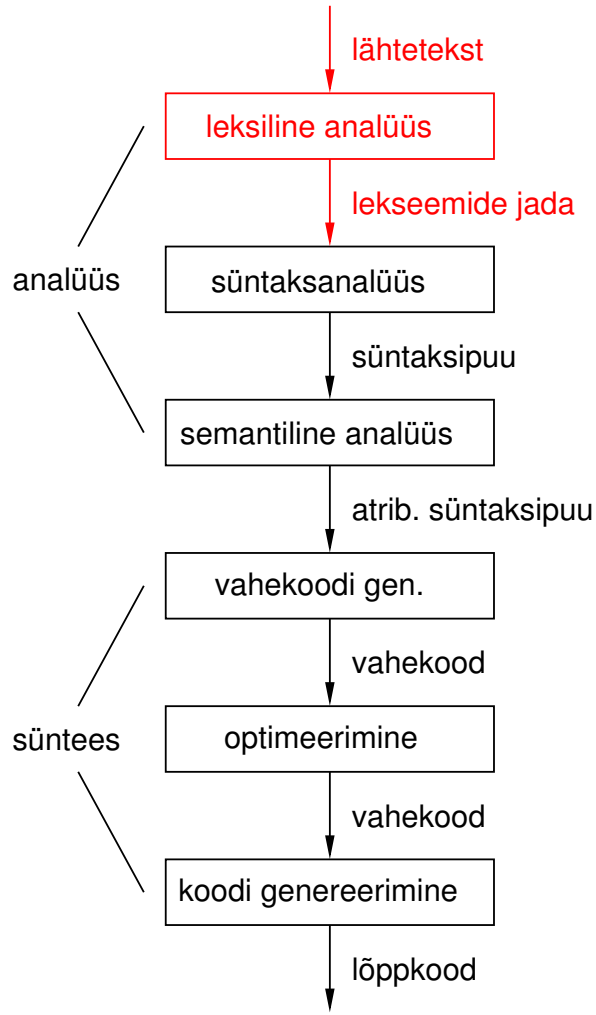
  süntees: koodi genereerimine ja optimeerimine.

- Need omakorda jagunevad mitmeteks alamfaasideks.
- NB! Erinevates faasides teostatavad toimingud võivad toimuda paralleelselt.



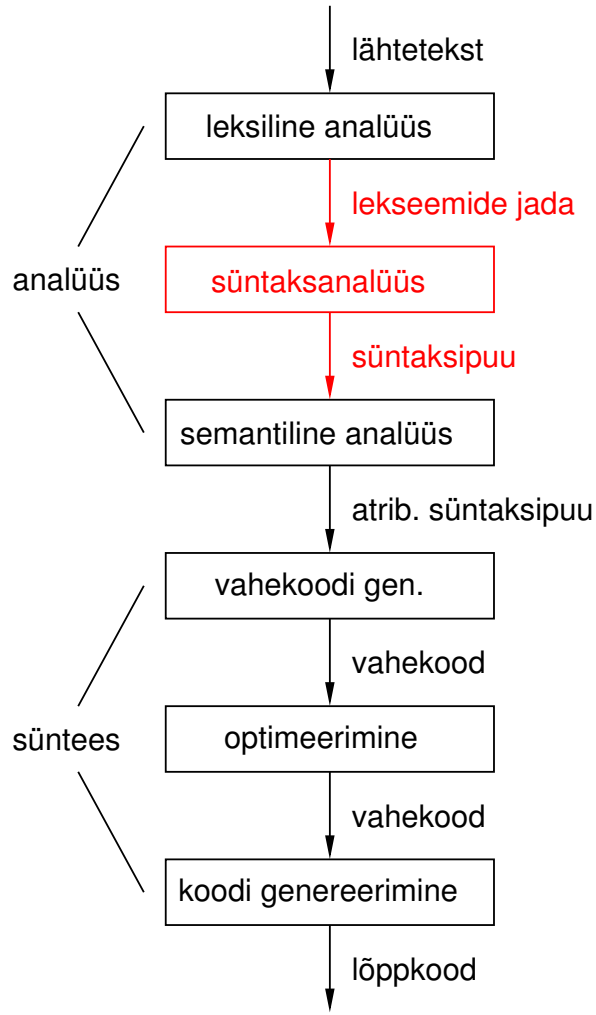
Kompilaatori struktuur

- Leksiline analüüs kontrollib programmi sõnade (literaalsümbolite) vastavust leksilistele reeglitele, eemaldab tühisümbolid ja kommentaarid ning teisendab programmi lekseemide (tokens) jadaks.
- Leksilist analüüsi kutsutakse skaneerimiseks ning vastavat analüsaatorit nimetatakse skanneriks.



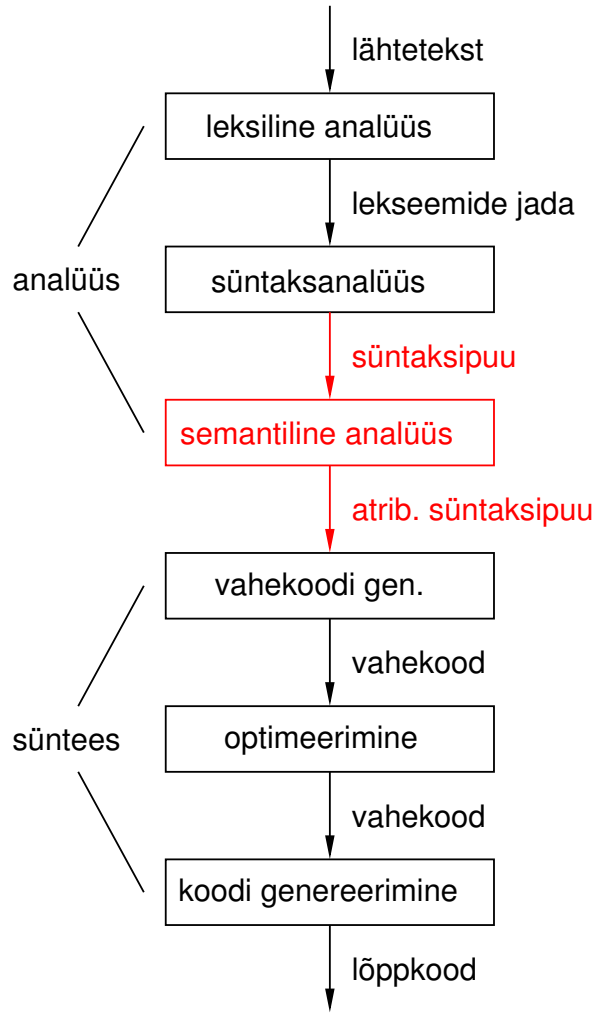
Kompilaatori struktuur

- Süntaksanalüüs kontrollib programmi struktuuri vastavust keele grammatikale ning väljastab programmi esitava (abstraktse) süntaksipuu.
- Süntaksanalüüsi kutsutakse parsimiseks ning vastavat analüsaatorit nimetatakse parseriks.



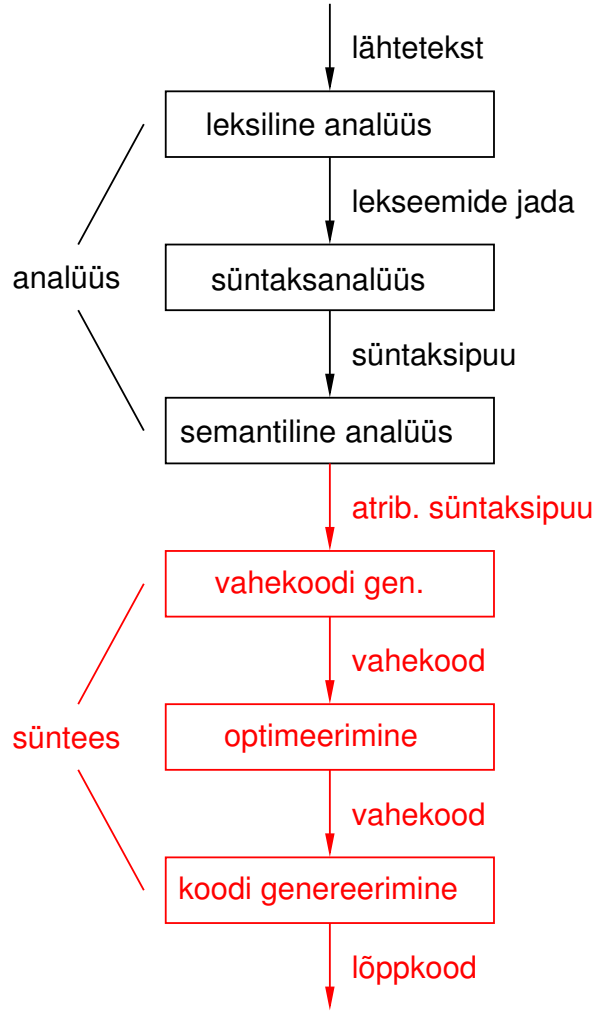
Kompilaatori struktuur

- Semantiline analüüs kontrollib programmi kontekstuaalsete sõltuvuste korrektsust, leiab vastavuse defineerivate ja kasutusesinemiste vahel, leiab esinemiste tüübid ja kontrollib nende vastavust reeglitele, ...
- Süntaksipuu dekoreeritakse tüübi- ja muu kontekstist sõltuva infoga.

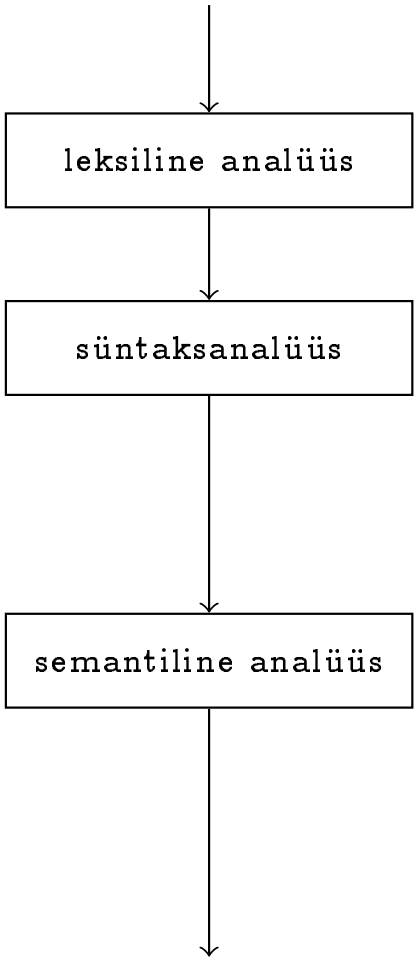


Kompilaatori struktuur

- Sünteesifaasis toimub koodi genereerimine ja optimeerimine.
- Eesmärgiks on riistvara ressursside võimalikult efektiivne ärakasutamine.
- Koodi genereerimine võib toimuda otse peale analüüsi, aga tavaliselt kasutatakse arhitektuurist vähem sõltuvate tegevuste teostamiseks vahekoodi.
- Koodi genereerimisel toimub muuhulgas keelekonstruktsioonide asendamine semantiliselt ekvivalentsete instruktsioonijadadega ning regsitrite määramine.

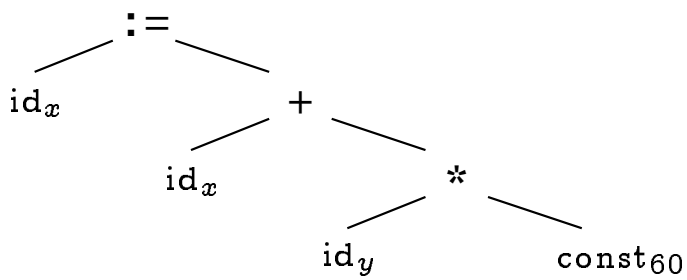


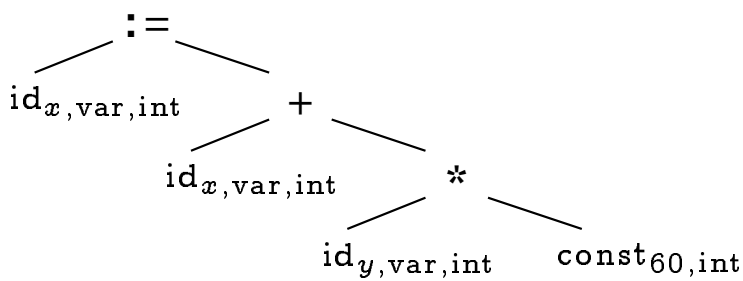
Programmi vaheesitused



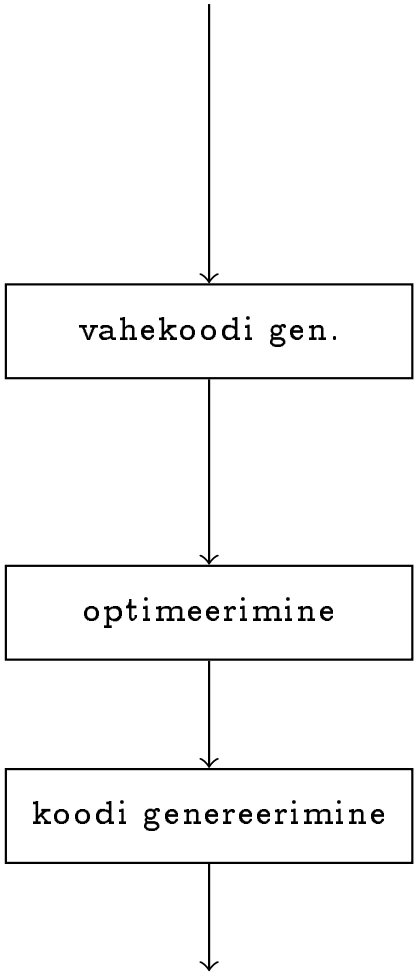

```
x := x + y * 60
```

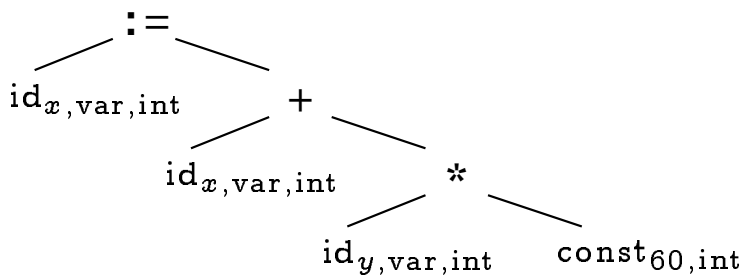
$\mathrm{id}_x\ \mathrm{keyw}_{:=}\ \mathrm{id}_x\ \mathrm{op}_+\ \mathrm{id}_y\ \mathrm{op}_*\ \mathrm{const}_{60}$





Programmi vaheesitused






```
R1 := y
R2 := 60
R3 := R1 * R2
R4 := x
R5 := R3 + R4
x  := R5
```

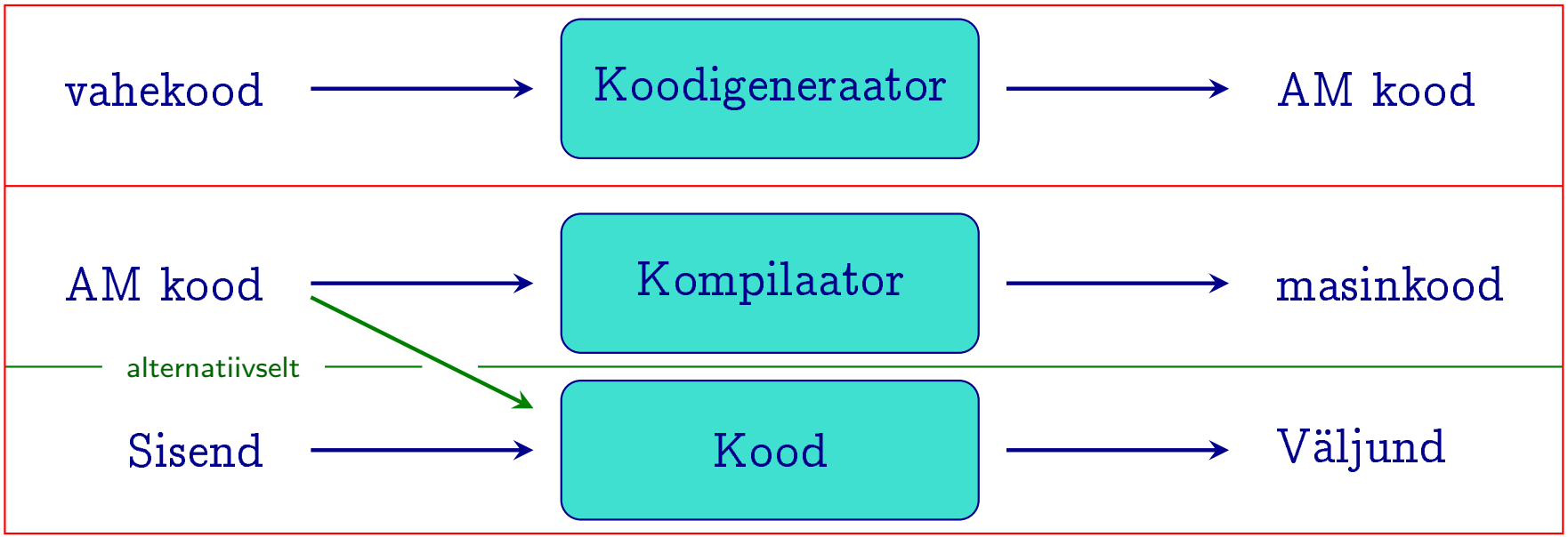
```
R1 := y
R1 := R1 * 60
R1 := R1 + x
x  := R1
```

```
MOV  R1, SP + $8
MULI R1, 60
ADD  R1, $4000
MOV  $4000, R1
```

Abstraktsed masinad

Koodi genereerimine koosneb sageli kahest eraldiseisvast osast:

- programmi transleerimine abstraktse masina koodiks;
- sellest konkreetse koodi genereerimine või ka otse interpreteerimine.



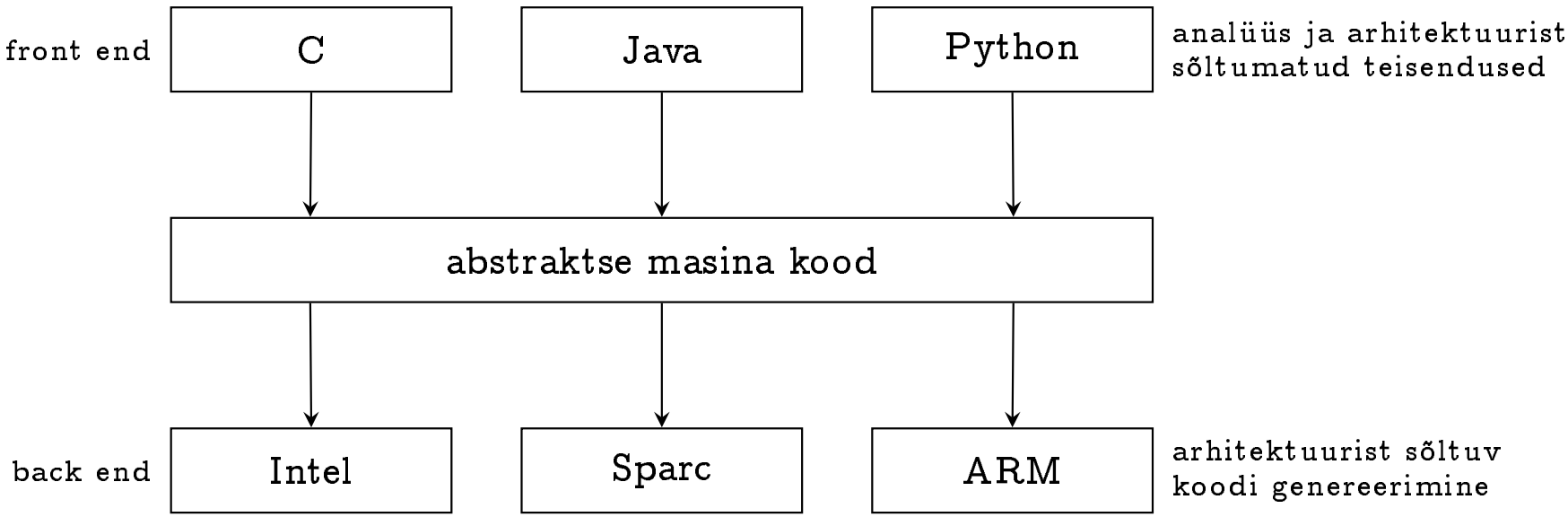
Abstraktsed masinad

Abstraktne masin on idealiseeritud arhitektuuriga masin, mille koodi on lihtne genereerida ja mida on ka lihtne realiseerida erineva ahitektuuriga tegelikel masinatel.

- Keelekonstruktsioonide transleerimine on eraldatud konkreetse riistvara spetsiifilistest omadustest.
- Lihtsustab kompilaatori teisaldamist uue arhitektuuriga masinatele.
- Samuti lihtsustab kompilaatori realiseerimist uute keelte jaoks.

Abstraktsed masinad

Teoreetiliselt, kui on m keelt ja n masinat, siis $n \times m$ kompilaatori jaoks on vaja m front end'i ja n back end'i.



Praktikas töötab see ainult siis, kui keeled ja masinate arhitektuurid on piisavalt lähedased.