

TÖL304G

Forritunarmál

Vikublað 1

Snorri Agnarsson

16. ágúst 2024

Efnisyfirlit

1	Inngangur	1
2	Hnit kennara	1
3	Lágmörk á verkefnaskilum	2
4	Miðmísserispróf	2
5	Dæmaskil og einkunnir	2
6	Gradescope	3
7	Piazza	3
8	Upptökur fyrirlestra	3
9	Hæfnisviðmið	3
10	Helstu forritunarmál	4
11	Áhersluatriði	4
12	Bækur og annað lesefni	5
13	Áætlun	5

14 Zoom fundir	6
15 Efni vikunnar	6
16 Hvað er „mál“?	6
17 Samhengisfrjáls mál og BNF	6
18 Útleiðslur	7
19 Regluleg mál	9
20 Reglulegar segðir	9
21 Endanlegar stöðuvélar	10
22 Extended BNF	11
23 Dæmatímar og dæmahópar	12

1 Inngangur

Í þessu námskeiði verða kynntar grunnhugmyndir sem liggja að baki flestra forritunarmála. Tilgangurinn er meðal annars sá að nemendur geri sér grein fyrir mismuni forritunarmála, notkunargildi hinna mismunandi mála og þeirra hugmynda, sem að baki liggja, og að nemendur verði dómbærir á kosti og galla hinna ýmsu forritunarmála.

Áhersluatriði eru betur skilgreind hér að neðan.

Bók til hliðsjónar er *Programming Languages* eftir Robert W. Sebesta.

Bókin er ætluð sem ítarefni en við munum ekki fylgja henni í neinum smáatriðum. Þeir sem hafa aðgang að öðrum bókum um forritunarmál geta örugglega notað þær í staðinn til hliðsjónar.

2 Hnit kennara

Nafn: Snorri Agnarsson

Aðsetur: Gróska.

Sími: 861 3270

Tölvupóstfang: snorri@hi.is

3 Lágmark á verkefnaskilum

Samkvæmt reglum námsbrautarinnar á að skilgreina lágmark fyrir þau verkefni sem skilað hefur verið um miðbik misserisins. Þeir nemendur sem ekki ná því lágmarki á tilsettum tíma verða sjálfkrafa skráðir úr námskeiðinu og fá ekki að þreyta próf. Ef viðkomandi nemandi hefur gilda afsökun, svo sem veikindi, og getur sannfært kennarann um að hann eða hún hafi möguleika á að klára námskeiðið, þá má veita undanþágu frá þessari reglu. Ef þið lendið í vandræðum með þessar reglur þá ættuð þið að hafa samband við mig, helst áður en ég skrái ykkur úr námskeiðinu.

Í þessu námskeiði þarf að skila að minnsta kosti fjórum einstaklingsverkefnum (Gradescope), fjórum hópverkefnum (Gradescope¹) fyrir lok áttundu viku námskeiðsins. Í flestum vikum, en ekki öllum, verður möguleiki á að skila einu einstaklingsverkefni og einu hópverkefni.

4 Miðmisserispróf

Um miðbik misserisins, sennilega í lok áttundu viku, verður miðmisserispróf sem gildir aðeins til hækkunar á lokaeinkunn. Sé einkunnin úr miðmisserisprófinu hærri en lokaprófseinkunnin gildir miðmisseriseinkunnin 30% á móti lokaprófseinkunninni, sem gildir þá 70% af endanlegri prófseinkunn.

5 Dæmaskil og einkunnir

Í námskeiðinu verða oftast tvenn vikuleg verkefni, eitt einstaklingsverkefni og eitt hópverkefni. Hópverkefni má vinna í dæmatíma. Engin takmörk eru sett á fjölda nemenda í hópi sem vinnur hópverkefni nema hvað nemandi má ekki vera í fleiri en einum skilahópi fyrir sama verkefni. Samræður milli hópa eru í fínu lagi. Hópaskipting má vel vera mismunandi í mismunandi vikum og þið ráðið sjálf hópaskiptingu.

Hópverkefnum og einstaklingsverkefnum skal skila fyrir miðnætti á fimmtudegi í skilaviku verkefnanna. Engum einstaklingsverkefnum eða hópverkefnum skal skila í fyrstu viku.

Fyrir hópverkefni mun einhver í hópnum þurfa að skila verkefninu í Gradescope og sá eða sú skal þá tilgreina hverjir aðrir nemendur eru í hópnum.

Meðaltal 7 bestu einkunna einstaklingsverkefna mynda 25% af einkunn námskeiðsins.

Meðaltal 7 bestu einkunna hópverkefna mynda 25% af einkunn námskeiðsins.

¹<https://gradescope.com>

Ekki er tekið á móti verkefnum eftir að skilafrestur er útrunninn.

Einkunn úr lokaprófi gildir 50% af lokaeinkunn, en athugið að það þarf að ná prófi. Ef prófseinkunnin er falleinkunn munu verkefniseinkunnir verða hunsaðar og prófseinkunnin verður lokaeinkunn. Einkunn úr miðmisserisprófi er vegin inn áður en athugað er hvort prófinu hefur verið náð.

Afritun lausna einstaklingsverkefna er forboðin. Þó er í lagi og mælt er með að nemendur ræði saman sín á milli um dæmin og lausnarhugmyndir tengdar þeim. En hver og einn nemandi skal skrifa (og prófa og sýna prófanir, ef hægt er) sínar eigin lausnir og skila þeim.

6 Gradescope

Gradescope kerfið hefur gefið góða raun og við munum nota það í þessu námskeiði bæði fyrir einstaklingsverkefni og hópverkefni.

Lausnum á einstaklingsverkefnum og hópverkefnum skal skila í Gradescope í gegnum vefviðmót Gradescope.

7 Piazza

Við munum nota spjallkerfið Piazza² til að skiptast á skoðunum um efni námskeiðsins. Allar almennar fyrirspurnir um námskeiðið eiga því heima þar. Ekki senda slíka fyrirspurnir í tölvupósti heldur beinið þeim til kennara, nemenda eða allra gegnum Piazza.

8 Upptökur fyrirlestra

Fyrirlestrar verða teknir upp sem myndskaið í Panopto kerfinu sem verða aðgengileg gegnum Ugluna og Canvas³.

9 Hæfnisviðmið

Við lok námskeiðsins mun nemandi

- skilja hvernig málfræði forritunarmála er lýst, geta komist að því hvort tiltekið forrit sé málfræðilega rétt fyrir gefna mállýsingu, geta skrifað mállýsingar fyrir einföld mál

²<https://www.piazza.com>

³<https://ugla.hi.is>

- skilja hvað rökstudd forritun er og geta forritað á rökstuddan hátt
- skilja og geta útskýrt hvernig umdæmi breytu virkar í ýmsum forritunarmálum ásamt því hvernig minnissvæðum og vakningarfærslum fyrir breytur er úthlutað og tengd saman í ýmsum forritunarmálum
- skilja hinar ýmsu gerðir viðfangaflutninga og geta notað þær
- skilja fallsforritun, listavinnslu, endurkvæmni og halaendurkvæmni og geta notað þær
- skilja hvað ruslasöfnun er og geta lýst nokkrum aðferðum fyrir ruslasöfnun
- skilja hvað fjölnota einingar eru og geta forritað fjölnota einingar í nokkrum forritunarmálum

10 Helstu forritunarmál

Forritunarmálin sem notuð verða eru trúlega Java, C++, C#, Scheme, Haskell, Morpho og CAML. Scheme er bálkmótað, einfalt en öflugt afbrigði af LISP. CAML er afbrigði af fallsforritunarmálinu ML. Haskell er hreint fallsforritunarmál, náskylt CAML, sem er í vaxandi notkun. Morpho er forritunarmál sem undirritaður hefur hannað og þróað með Hallgrími H. Gunnarssyni. Java þekkja flestir nú orðið. C# frá Microsoft er svipað og Java.

Einnig munum við eilítið nota forritunarmálið Dafny⁴, en einungis til að hnykkja á hugmyndum um rökstudda forritun.

11 Áhersluatriði

Áhersla verður lögð á eftirfarandi atriði.

- Málfræði forritunarmála, BNF, EBNF og málrit. Einnig smávegis um regluleg mál, þ.e. reglulegar segðir og endanlegar stöðuvélar.
- Innviðir og hönnunargrundvöllur bálkmótaðra forritunarmála s.s. Ödu, Pascal, Morpho og Scheme.
- Viðfangaflutningar í forritunarmálum: Gildisviðföng, afritsviðföng, tilvísunarviðföng, nafnviðföng og löt viðföng.
- Grunnhugmyndir í fallsforritun, löt og ströng gildun, hrein fallsforritun, kostir hennar og gallar.

⁴<https://tio.run/#dafny>

- Einingaforritun í Morpho, Jövu og kannski C# og/eða C++. Sérstök áhersla verður lögð á fjölnota einingar (t.d. *template* í C++ og *generic* í Jövu).
- Skjölun forrita og sannprófun, einkum m.t.t. einingaforritunar og forritunar stórra kerfa.
- Listavinnsla í Scheme, Morpho, CAML og Haskell.
- Hlutbundin forritun í Java, C# og Scheme, áhersla er lögð á innviði hlutbundinnar forritunar, þ.e. hvernig boð vinna og samband þeirrar virkni við arfgengi.
- Fallsforritun í Scheme, CAML, Morpho og Haskell.
- Ruslasöfnunaraferðir, sem notaðar eru í málum s.s. Scheme, Morpho, Java, C#, CAML og Haskell.
- Líklega verður farið í samskeiða forritun í Morpho og Java.

Þau atriði, sem mest áhersla er lögð á að nemendur tileinki sér eru innviðir bálkmótaðra forritunarmála, einingaforritun, listavinnsla/fallsforritun (það tvennt hangir mjög saman), ruslasöfnun og hlutbundin forritun. Ekki er ætlast til að nemendur leggi á minnið smáatriði í málfræði forritunarmála. Áherslan er á merkingu forritunarmálanna og mismunandi eðli þeirra.

12 Bækur og annað lesefni

Bókin eftir Sebesta er til hliðsjónar í námskeiðinu og vísað verður í seinni vikublöðum á ýmsar hjálparskrár og vefsíður, til dæmis handbók fyrir forritunarmálið Morpho, auk rita um Scheme, Haskell, CAML og fleira.

Til dæmis er gagnlegt að líta á eftirfarandi vefsíður:

1. EBNF⁵
2. Veforðabók⁶

13 Áætlun

Líkleg efnisröð í námskeiðinu er í skjalinu `fmal_plan.pdf` í Uglunni/Canvas.

⁵https://en.wikipedia.org/wiki/Extended_Backus-Naur_form

⁶<http://foldoc.org>

14 Zoom fundir

Ef nemendur þurfa viðtal við mig er þægilegast að nota Zoom⁷. Þanta má Zoom viðtal með því að hringja í mig. Oftast er ég laus í hádeginu, nema á föstudögum.

15 Efni vikunnar

Í þessari og næstu viku förum við í mállýsingar forritunarmála og byrjum að kynna okkur innviði bálkmótaðra forritunarmála.

Í langflestum tilfellum er málfræði nútíma forritunarmála lýst með samhengisfrjáls-um mállýsingum, þ.e. BNF.

Oftast er málfræðinni lýst í tveimur skrefum, eitt skref lýsir frumeiningum málsins, þ.e. lykilorðum, sértáknum s.s. svigum, kommu, o.s.frv., og lesföstum s.s. heiltöluföstum og strengföstum. Í sama skrefi er oftast lýst því sem koma má milli frumeininga málsins, t.d. bilstafir og athugasemdir. Frumeiningum nútíma forritunarmála má langoftast lýsa sem reglulegum málum. Til dæmis er mál löglegra fleytitölufasta í C++ reglulegt mál.

Eftir að frumeiningum málsins hefur verið lýst er heildarmálfræði málsins lýst á BNF sniði. Þá er reiknað með því að frumeiningarnar, sem lýst var í fyrra skrefi séu tákni innan þess stafrófs, sem unnið er með.

16 Hvað er „mál“?

Mál er einfaldlega mengi strengja. Strengur er endanleg runa tákna úr einhverju mengi, sem við þá köllum táknróf eða stafróf (alphabet) málsins.

17 Samhengisfrjáls mál og BNF

BNF⁸ er aðferð, svokallað meta-mál, til að skilgreina mál. Með BNF er í grundvallaratriðum átt við það sama og átt er við þegar talað er um samhengisfrjálsar mállýsingar (context-free grammar).

BNF stendur nú fyrir Backus-Naur Form. BNF stóð einu sinni fyrir Backus Normal Form, því John Backus, sá sem fann upp FORTRAN forritunarmálið, átti mikinn þátt í þróun þess. En Peter Naur, sem ritstjóri skilgreiningarinnar á ALGOL-60 forritunarmálinu átti einnig mikinn þátt í að koma BNF í almenna notkun innan tölvunarfræðinnar, og þess vegna er BNF nú almennt látið standa fyrir Backus-Naur Form.

⁷<https://eu01web.zoom.us/j/5528462515>

⁸https://en.wikipedia.org/wiki/Backus-Naur_form

En BNF á sér einnig rætur í málvísindum, því málvísindamaðurinn Noam Chomsky skilgreindi samhengisfrjálsar mállysingar áður en Backus og Naur skilgreindu BNF. Chomsky skilgreindi reyndar fleiri tegundir mállysinga, m.a. fyrir regluleg mál, sem rætt er um hér að neðan.

Dæmi um BNF skilgreiningu:

```
<expr> ::= <num> | ( <expr> ) | <expr> <op> <expr>
<op> ::= + | - | * | /
<num> ::= <digit> | <digit> <num>
<digit> ::= 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9
```

eða, betur sniðsett:

```
⟨expr⟩ ::= ⟨num⟩
          | ( ⟨expr⟩ )
          | ⟨expr⟩ ⟨op⟩ ⟨expr⟩
⟨op⟩ ::= +
        | -
        | *
        | /
⟨num⟩ ::= ⟨digit⟩
          | ⟨digit⟩ ⟨num⟩
⟨digit⟩ ::= 0
           | 1
           | 2
           | 3
           | 4
           | 5
           | 6
           | 7
           | 8
           | 9
```

Þessi BNF skilgreining lýsir máli sem inniheldur strengi sem eru segðir („formúlur“) með heiltölugildum og venjulegum reikniaðgerðum.

18 Útleiðslur

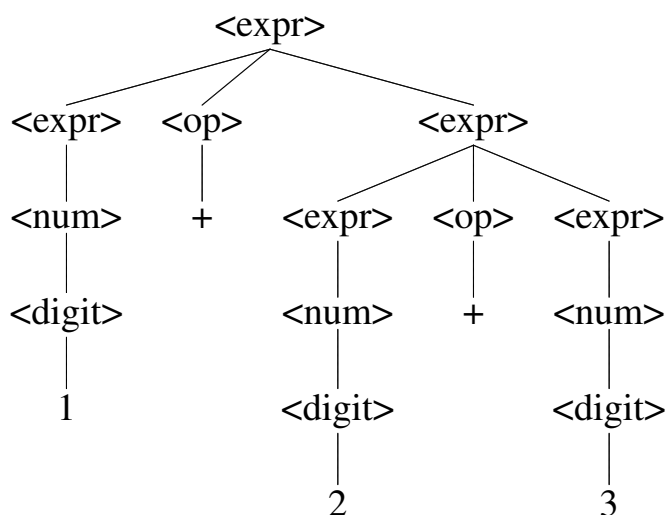
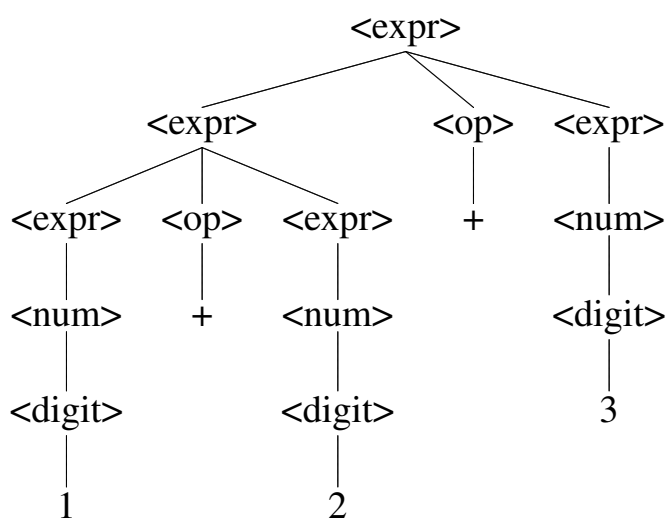
Þegar gefin er BNF skilgreining eins og að ofan má yfirleitt *leiða út* ótakmarkaðan fjölda strengja í málinu. Dæmi:


```

<expr>  =>
<expr> <op> <expr> =>
<expr> <op> <expr> <op> <expr>  =>
<num> <op> <expr> <op> <expr>  =>
<digit> <op> <expr> <op> <expr>  =>
1 <op> <expr> <op> <expr>  =>
1 + <expr> <op> <expr>  =>
...  =>
1 + 2 + 3

```

Takið eftir að strenginn $1 + 2 + 3$ má leiða út á fleiri en einn hátt. Mállýsing þessi er því *margræð* (*ambiguous*). Margræðni (*ambiguity*) er oft talin óheppileg ef um forritunarmál er að ræða. Margræðnin sést vel ef við notum *útleiðslutré*:



Ætlast er til að þið getið búið til slík útleiðslutré fyrir gefinn streng í máli sem skilgreint er með tiltekinni BNF skilgreiningu.

Ef unnt er að lýsa tilteknu máli með BNF (þ.e. með samhengisfrjálsri mállyssingu) þá segjum við að *málið* sé samhengisfrjálst (ekki aðeins mállyssingin, sem augljóslega er samhengisfrjáls samkvæmt skilgreiningu).

Þegar málfræði forritunarmála er lýst er nú til dags næstum alltaf notuð einhver aðferð sem er jafngild BNF. Dæmi um slíkar aðferðir, aðrar en BNF sjálft, eru EBNF (Extended BNF) og málrít (syntax diagrams). Ætlast er til að þið getið lesið og skilið allar þessar þrjár aðferðir.

19 Regluleg mál

Frumeiningar forritunarmála, svo sem lykilorð, strengfastar, heiltölufastar, fleytitölufastar, o.s.frv., eru yfirleitt regluleg mál (regular language). Til dæmis er mál fleytitölufasta í Jövu reglulegt mál og mál strengfasta er annað reglulegt mál. Athugið að þetta eru að sjálfsögði tvö mismunandi regluleg mál, sem aftur eru notuð í skilgreiningu þriðja málsins (þ.e. Java).

Regluleg mál eru undirmengi samhengisfrjálsra mála, þ.e. öll regluleg mál eru samhengisfrjáls, en ekki öfugt. Regluleg mál eru einfaldari en önnur samhengisfrjáls mál, og til eru einfaldari aðferðir en BNF til að lýsa reglulegum málum. Ætlast er við að þið getið lesið og notað endanlegar stöðuvélar⁹ (finite state automaton, finite state machine) og reglulegar segðir (regular expressions) til að lýsa einföldum reglulegum málum.

Regluleg mál eru mikið notuð í ýmsum tölvuverkefnum, til dæmis í skeljum og í leitarforritum svo sem grep.

20 Reglulegar segðir

Ein aðferð til að skilgreina regluleg mál er *reglulegar segðir*. Reglulegar segðir yfir stafróf Σ má skilgreinda á eftirfarandi hátt:

- \emptyset er regluleg segð sem skilgreinir málið $M(\emptyset) = \emptyset = \{\}$
- ϵ er regluleg segð sem skilgreinir málið $M(\epsilon) = \{\epsilon\}$
- Ef $x \in \Sigma$ þá er x regluleg segð sem skilgreinir málið $M(x) = \{x\}$, sem er eins staks mengi sem inniheldur eins stafs streng.
- Ef x og y eru reglulegar segðir þá er xy regluleg segð sem skilgreinir málið $M(xy) = \{uv \mid u \in M(x) \wedge v \in M(y)\}$.
- Ef x og y eru reglulegar segðir þá er $x \mid y$ regluleg segð sem skilgreinir málið $M(x \mid y) = \{w \mid w \in M(x) \vee w \in M(y)\}$.

⁹https://en.wikipedia.org/wiki/Finite-state_machine

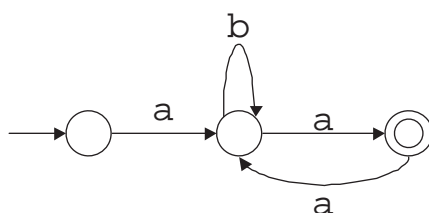
- Ef x er regluleg segð þá er $x^?$ regluleg segð sem skilgreinir málið $M(x^?) = M(x) \cup \{\epsilon\}$.
- Ef x er regluleg segð þá er x^* regluleg segð sem skilgreinir málið $M(x^*) = \bigcup_{n=0}^{\infty} M(x^n)$.
- Ef x er regluleg segð þá er x^+ regluleg segð sem skilgreinir málið $M(x^+) = \bigcup_{n=1}^{\infty} M(x^n)$.
- Ef x er regluleg segð þá er (x) regluleg segð sem skilgreinir málið $M((x)) = M(x)$.

Í ofangreindum skilgreiningum gerum við ráð fyrir að séu u og v eru strengir þá sé strengurinn uv skilgreindur sem samskeyting u og v . Svipað gildir um samskeytingu fleiri en tveggja strengja. Einnig gerum við ráð fyrir að veldishafning strengja sé skilgreind með eftirfarandi:

- $x^0 = \epsilon$
- $x^{n+1} = xx^n$ (skeytt er saman x og x^n).

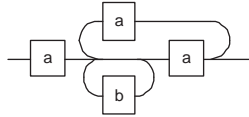
21 Endanlegar stöðuvélar

Eins og fram hefur komið eru endanlegar stöðuvélar enn ein aðferð til að skilgreina regluleg mál. Þessi mynd sýnir dæmi um endanlega stöðuvél sem ber kennsl á strengi yfir stafrófið $\{a, b\}$. Í málinu eru m.a. strengirnir aa , aba , $abba$, o.s.frv., en einnig má stinga aa einhvers staðar inn í löglegan streng og fá þannig út annan löglegan streng í málinu.



Strengi í málinu sem svona stöðuvél skilgreinir má fá með því að byrja í byrjunarstöðunni, sem er hringurinn sem örin lengst til vinstri bendir á, eða almennt sem sú ör bendir á sem ekki byrjar í einhverri stöðu, og fara síðan gegnum núll eða fleiri stikur og enda í lokastöðu. Lokastöður eru þær sem teiknaðar eru með tvöföldum hring. Ef síðan er skeytt saman þeim stöfum sem stikurnar eru merktar með þá fæst strengur í málinu. Aðrir strengir en þeir sem framleiða má með þessum hætti eru ekki í málinu.

Þessa endanlegu stöðuvél má einnig teikna sem málrit:



Takið eftir að í málritinu eru kassar fyrir lokatáknin sem samsvara stikum í stöðuvélinni. Dæmi um strengi í þessu máli eru aa, aba, abba, aaaa, en ekki t.d. aaa eða aabaa.

Ætlast er til að nemendur geti skilið svona endanlegar stöðuvélar og málrit, og geti borið kennsl á hvort tilteknir strengir séu í málinu sem slíkar vélar og rit skilgreina. Fleiri dæmi um málrit má finna í handbókinni fyrir forritunarmálið Morpho, sem mun verða sett í Ugluna/Canvas þegar þar að kemur.

22 Extended BNF

Extended BNF¹⁰ eða EBNF er nú orðið algeng aðferð til að skilgreina málfræði forritunarmála. EBNF er svipað og BNF, en þar hefur verið bætt við hugmyndum úr reglulegum segðum. Ýmis afbrigði eru til af EBNF, eins og af BNF, en til er nú alþjóðlegur staðall¹¹ fyrir EBNF, sem líklegt er að flestir notendur EBNF fari nálægt að fara eftir. Sjá einnig grein eftir R.S. Scowen¹².

Í EBNF eru lokatáknin merkt sérstaklega með því að setja gæsalappir utan um þau (andstætt venjunni í BNF, þar sem millitáknin eru merkt með <...>), en millitáknin eru venjuleg ómerkt orð eða jafnvel orðasambönd.

Í EBNF er síðan bætt við eftirfarandi mál fyrirbærum í hægri hlutum reglna, sem eiga rætur að rekja til reglulegra segða:

- Nota má slaufusviga til að tákna endurtekningar: $\{X\}$ er látið standa fyrir núll eða fleiri X .
- Nota má hornklofa til að tákna einstakan valkost: $[X]$ er látið standa fyrir núll eða eitt X .
- Nota má sviga til að safna saman, og nota má $|$ til að aðskilja valkosti: Til dæmis er $(X | Y), Z$ jafngilt $X, Z | Y, Z$.

Samkvæmt staðaltillögunni sem liggur fyrir er samskeyting í EBNF táknuð með aðgerðinni , (koma), sem er ólíkt BNF, þar sem samskeyting hefur ekkert aðferðartákn.

Dæmi um einfalda EBNF mállýsingu er eftirfarandi:

```
expression =
    term, { op, term } ;
```

¹⁰Sjá t.d. bls. 131-134 í Sebesta.

¹¹<http://www.cl.cam.ac.uk/~mgk25/iso-ebnf.html>

¹²<http://www.cl.cam.ac.uk/~mgk25/iso-14977-paper.pdf>

```

term =
    number | '(' , expression , ')' ;
op =
    '+' | '-' | '*' | '/' ;
number =
    digit , { digit } ;
digit =
    '0' | '1' | '2' | '3' | '4' |
    '5' | '6' | '7' | '8' | '9' ;

```

Mállýsing þessi skilgreinir sama mál og BNF mállýsingin frammar í þessu vikublaði. Einnig mætti lýsa sama máli með:

```

expression =
    ( number | '(' , expression , ')' ) ,
    { ( '+' | '-' | '*' | '/' ) ,
      ( number | '(' , expression , ')' ) } ;
number =
    ( '0' | '1' | '2' | '3' | '4' |
      '5' | '6' | '7' | '8' | '9' ) ,
    { '0' | '1' | '2' | '3' | '4' |
      '5' | '6' | '7' | '8' | '9' } ;

```

23 Dæmatímar og dæmahópar

Dæmatímar munu hefjast í viku 2.

Þið ráðið því sjálf í hvaða dæmatíma þið mætið í hverri viku. Ef það veldur því að of margir hrúgast saman þá munum við reyna að taka á því.