

Páttari fyrir útvíkkaðan Fjölni

Arnar Birgisson
Verkefni í *Þýðendum*, haust 2003

Þetta skjal lýsir þáttara fyrir útvíkkaða útgáfu af forritunarmálinu Fjölni. Útvíkkunin felst aðallega í möguleika á að skilgreina innri stef. Útvíkkuninni var hins vegar gerð skil í málýsingu og því verður ekki fjallað um hana hér.

Þáttarinn (og lesgreinir) er skrifaður í tólinu ANTLR (<http://www.antlr.org>). Hér á eftir fylgir lýsing á hverju skrefi þáttunarinnar, ásamt lýsingu á lesgreiningu.

1 Lesgreinir

Lesgreinirinn er tvískiptur, og er notast við s.k. „token stream multiplexing“ í ANTLR. Í upphafi er settur upp gervilesgreinir sem sér um að stýra inntaki forritsins í réttan lesgreini. Gervilesgreinirinn heldur utan um tvö eintök af *alvöru* lesgreinum, annars vegar lesgreini sem vinnur utan einingaskilgreininga sem afmarkast af slaufusvigum, og hins vegar lesgreini sem vinnur innan þeirra. Hér eftir skulum við nefna lesgreinana „forritslesgreini“ (sá sem vinnur utan eininga), og „einingalesgreini“ (sá sem vinnur innan eininga).

Í upphafi byrjar gervilesgreinirinn á að beina öllu inntaki í forritslesgreininn. Þegar forritslesgreinirinn rekst síðan á stafinn { gefur hann út tákni fyrir hann og gefur síðan gervilesgreininum skipun um að skifta yfir í einingalesgreininn. Eftir það fer inntak þýðandans þangað. Þar sem einingar geta innihaldið földuð pör af slaufusvigum heldur einingalesgreinirinn utan um földunardýpt með því að hækka teljara fyrir hvert { sem hann sér í inntakinu og lækka hann fyrir }. Þegar einingarlesgreinirinn rekst síðan á } með földunardýpt 0, þ.e. lokun á slaufusviga sem hann sá ekki opnaðan, sendir hann skipun til gervilesgreinisins um að skipta aftur yfir í þann lesgreini sem á undan fór (þ.e. forritslesgreinisins).

Heildarútkoma úr lesgreiningu er því táknastráumur þar sem gerður er skýr greinarmunur á táknum sem birtast utan eða innan slaufusviga, m.a. er gerður greinarmunur á einingaraðgerðinni * og aðgerðanafninu *.

Inntaksstafróf beggja lesgreina eru stafir með gildin 3 - 255 og er inntakið túlkað skv. ISO-8859-1 (líka þekkt sem *Latin1*) staðli. Forritslesgreinir flytur síðan út táknróf sem er nefnt „FjölurForrit“¹ og inniheldur þau tákni sem sá lesgreini gefur út. Einingalesgreinirinn flytur inn þetta táknróf, bætir við það táknum og flytur aftur út sem „Fjölur“.

Einingalesgreinirinn hefur einnig það hlutverk að skoða aðgerðanöfn og ákveða á grundvelli fyrsta stafs þeirra, til hvaða forgangsflokks þau tilheyra. Forgangsflokkarnir eru sjö og skilar lesgreinirinn mismunandi tákni fyrir hvern flokk.

Lesgreinarnir tveir eru skilgreindir í skránum `FjölurForritLexer.g` og `FjölurEiningLexer.g`.

2 Páttari, fyrsti fasi

Fyrsti fasi þáttunar felst í að taka táknastrauminn frá lesgreiningarfasanum og byggja úr honum máltré (e. *syntax tree*) ef táknastraumurinn (og þar með inntakið) er löglegt Fjölnis-forrit skv. málfræðinni sem hefur verið skilgreind.

Þessi fasi er útfærður með venjulegum ANTLR þáttara í skránni `FjölurParser.g`. Hann flytur inn táknrófið „Fjölur“ og bætir við það nokkrum gervitáknum sem notuð eru til að einfalda uppbyggingu máltrésins, nýtt táknróf er útlutt undir nafninu „FjölurParser“.

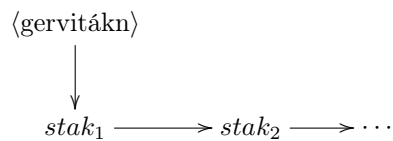
Í þessum fasta þáttarans er inntakið dæmt sem málfræðilega rétt eða rangt, forgangur og tenging aðgerða eru fest og máltréð er smíðað. Flestar reglur í þáttaranum smíða undirtré beint af augum (þ.e. tré sem ekki þarfnast skýringar), en sumar reglur bæta inn gervitáknum í trén og fylgja hér skýringar á þeim helstu.

¹Táknróf x sem flutt er út úr ANTLR forritum er lýst í skrá `xTokenTypes.txt`

ANTLR geymir trjáhnúta í gagnagrind sem inniheldur bendi í fyrsta barn annars vegar, og næsta systkini hins vegar. Myndir af trjám í þessu skjali fylgja þessari gagnagrind þannig að lóðrétt ör sýnir vensl foreldris og fyrsta barns, og lárétt ör sýnir vensl aðliggjandi systkina. Vart þarf að taka fram að foreldri er fyrir ofan barn og röð systkina lesist frá vinstri til hægri.

Aðgerðir Þar sem forgangur og tenging aðgerða kemur ótvírætt fram í smíðuðu tré eru táknið fyrir hina sjö forgangsflokka ekki sett inn í tréð, heldur í staðinn gervitáknið `ADGERD`. Að öðruleyti eru undirtré fyrir aðgerðir eðlileg.

Runur Nafna-, skilgreininga- og segðarunur eru táknaðar með trénu



þar sem $\langle \text{gervitákn} \rangle$ er gervitákn sem einkennir tegund rununnar (runa af nöfnum, skilgreiningum eða segðum), og stökin eru undirtré fyrir viðkomandi tákn.

3 Þáttari, annar fasi

Eftir að fyrsti fasi þáttarans hefur smíðað máltréð tekur annar fasi hans við því og gerir ýmsar einfaldanir. Úttak hans er máltré sem er merkingarfræðilega jafngilt því sem kemur úr fyrsta fasa, en þar sem ýmsum trjám hefur verið útrýmt og/eða breytt. Hér eftir fer yfirlit yfir þær einfaldanir sem þessi fasi framkvæmir.

Þessi fasi er útfærður sem ANTLR „trjáþáttari“ (e. *tree parser*) í skránni `FjölurTransformer.g`. Hann flytur inn táknrófið „FjölurParser“ og flytur það aftur út með viðbótum sem „FjölurTransformer“.

Eftirfarandi einfaldanir og breytingar er framkvæmdar í þessum fasa:

Skilgreiningar Breytuyfirlýsingar (auðkenndar með lykilorðunum `staðvör` og `innflutt`) í upphafi stefja geta í Fjölni komið fyrir í hversu mörgum eintökum sem er og í hvaða röð sem er. Í þessum fasa þáttunar eru slíkar yfirlýsingar teknar saman þannig að í máltré fyrir stefskilgreiningu kemur fyrir nákvæmlega ein runa af nöfnum innfluttra breyta, og ein runa af nöfnum og/eða gildisveitingum fyrir staðværar breytur. Enn fremur eru tómun undirtrjám fyrir þessar skilgreiningar bætt, séu engar skilgreiningar í inntaksforritinu.

Aðgerðir Undirtré fyrir aðgerðabeitingar er í þessum fasa umbreytt í tré sem jafngilda stefköllum. Sem dæmi verður tré fyrir inntakið `x+y` nákvæmlega hið sama og fyrir inntakið `+(;x,y)` þar sem `x` og `y` geta verið hvaða segðir sem er. Þetta er framkvæmt jafnt fyrir tvíundar- og einundaraðgerðir. Að auki er táknið `ADGERD` breytt í táknið `NAFN` svo kóðasmiðurinn sér í raun engan greinarmun á beitingu aðgerða eða stefköllum.

Listar Undirtré fyrir listasegð er umbreytt í tré sem táknar kall í fallið `:. M.ö.o.` kemur út sama tré fyrir inntakið `[x1,x2,...,xN]` og fyrir inntakið `:(;x1,:(;x2,:(;...:(;xN,[])...))`, þar sem `x1, x2, ..., xN` geta verið hvaða segðir sem er.

annars-segðarunur Í `ef-` og `val-`segðum þar sem ekki kemur fram `annars-`hluti með segðarunu er stungið inn jafngildi inntaksins `annars []`.

fyrir-lykkjur `fyrir-lykkjunni fyrir (s1,...,sN, t, r1,...,rM) lykkja q1,...,qK lykkjulok` er breytt í segðarununa sem jafngildir inntakinu `s1,...,sN`, meðan `t lykkja q1,...,qK, r1,...,rM lykkjulok`. Hér geta `s1,...,sN,r1,...,rM,q1,...,qK` og `t` verið hvaða segðir sem er.

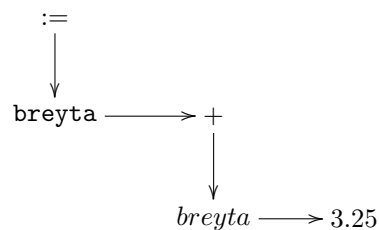
4 Hjálparforrit

Í þáttaranum eru hjálparforrit sem leyfa okkur að keyra þáttaran beint. Í skránni `utils.cpp` er skilgreint fall sem tekur stýristafi í Fjölni og umbreytir í viðeigandi ISO-8859-1 stafi.

Aðalforritið er skilgreint í skránni `prufuthattari.cpp`. Í því er notandanum leyft að keyra þáttarann með skipuninni `falskur_fjölur [-1] [inntaksskrá]`. Ef inntaksskrá er ekki gefin upp er unnið á *stdin*. Ef rofinn -1 er gefinn er þáttarinn ekki keyrður, heldur aðeins lesgreiniingarfasinn. Úttak forritsins í því tilfelli er listi yfir táknið í táknastraum fyrir viðkomandi inntak.

Sé -1 rofinn ekki tekinn fram er þáttarinn keyrður. Úttak hans textaframsetning á máltrénu bæði að loknum fyrsta fasa og að loknum öðrum fasa (auðkennt með strengnum „Eftir fasa 2:“).

Textaframsetningin á máltrénu lítur út eins Lisp/Scheme kóði. Hnútar trésins eru auðkenndir með strengjaparinu `les [heiti tákn]`. Sem dæmi er textaframsetning trésins



á þessa leið:

```

( := [GILDISVEITING]
  breyta [NAFN]
  ( + [ADGERD]
    breyta [NAFN]
    3.25 [FLEYTITALA]
  )
)
  
```