

1. Viikkoraportti

Tietorakenteiden harjoitustyö

Joosua Laakso

22. joulukuuta 2013

Työn tekeminen oli aluksi vähän nihkeää koska en ole aiemmin tehnyt mitään tämän kaltaista. Aluksi yritin jos ei-determinististen tila-automaattien muodostaminen olisi ollut mahdollista suoraan säännöllisen lausekkeen merkkijonoesityksestä. Yritys oli tietenkin tuohon tuomittu. Siksi päätin että on viisainta varmaan ensiksi kasata jonkinlainen parseri joka muodostaa merkkijonoesityksestä jäsennykspuun. Myöskin luovuin ei-deterministisistä tila-automaateista ja päätin että käytän kielen tunnistamisessa Brzozowskin derivaattoja.¹

Luulen että olen opiskellut asiaa käsittelevää materiaalia ajassa mitattuna enemmän kuin mitä olen käyttänyt varsinaiseen ohjelmointiin tällä viikolla. Arvioisin että ohjelmointiin olen käyttänyt noin 9-12 tuntia. En ole aivan varma koska monesti ohjelmoidessani menetän ajantajun. Olen listannut joitain lähteitä määrittelydokumentissa. Kaikkia en ole välttämättä listannut ja kaikkia en välttämättä muista, mutta monissa lähteissä oli myös paljon päällekkäisyyttä. Tärkeimpiä ovat säännöllisten lausekkeiden derivaattoja koskevat lähteet, koska niillä on keskeinen asema työssä.

Tällä viikolla opin paljon asioita. Koska en ole vielä käynyt Laskennan mallitkurssia, säännöllisiin lausekkeisiin, säännöllisiin kieliin ja tila-automaatteihin liittyvä teoria on minulle osittain uutta. Olen ainakin kuullut näistä kaikista aiemmin mutta en ole syventynyt niihin. Jos yritän listata kaikki asiat jotka olen oppinut tällä viikolla, niin välttämättäkin unohdan jotain. Yksi tarkempi esimerkki siitä mitä opin, on että on eri tapoja toteuttaa säännöllisten lausekkeiden toiminta: säännöllisestä lausekkeesta voi muodostaa ei-deterministisen tila-automaatin tai voi käyttää Brzozowskin artikkelissa kuvattua derivointimenetelmää. Epädeterministisestä tila-automaatista voi myös muodostaa deterministisen tila-automaatin, samoin kuin derivointimenetelmää voi käyttää deterministisen tila-automaatin muodostukseen. Opin myös että merkkijonoesitys kannattaa muuttaa ensin jäsennykspuuksi ennen kuin alkaa tekemään sille minkäänlaisia operaa-

¹dl.acm.org/citation.cfm?id=321249

tioita. Yksi asia jonka opin vielä on se, että valitsemani työn aihe ei taida olla sieltä helpoimmasta päästä, vaan pelkästään se, että ohjelman saa toimimaan on jo haaste itsessään, puhumattakaan siitä että sen pitäisi toimia tehokkaasti.

Paljon on epäselvää. Koska tämä on uusi asia minulle, minun on vaikea tietää miten minun kannattaa rakentaa koodini, koska en tiedä etukäteen välttämättä mitä tulen tarvitsemaan. Suurin epäselvyys liittyy siis varsinaiseen koodin tuottamiseen itseensä.

En osaa sanoa kuinka hyvin ohjelma on edistynyt, koska en ole seurannut muiden edistymistä jotta voisin verrata, mutta jos aikataulu sanoo että tässä vaiheessa ydintoiminnallisuuden tekeminen pitäisi olla aloitettu, niin voin rehellisesti sanoa että olen varmaan edennyt ihan hyvin. Olen saanut parserin toimimaan kaikilla perusoperaatioilla: tähdellä, katenaatiolla ja unionilla. En kuitenkaan ole vielä aloittanut varsinaista kielentunnistus-toiminnallisuuden toteutusta, olen keskittynyt parseriin. Parseri siis muodostaa merkkijonoesityksestä jäsennyyspuun, jota on helpompi käsitellä kuin pelkkiä merkkijonoja. Aion kuitenkin tehdä tätä työtä myös jouluviikolla, joten uskon että työ edistyy hyvin.

Pyrin siihen että viikon kuluttua ohjelma tukee ainakin perustoiminnallisuutta, eli se pystyy tunnistamaan kielen joka on kuvattu perusoperaattoreilla. Aion siis saada derivointitoiminnallisuuden kuntoon, jonka jälkeen tavoitteen pitäisi olla saavutettu. Vastaani ei ole tullut toistaiseksi ylitsepääsemättömiä ongelmia, joten tavoitteen pitäisi olla saavutettavissa. Valitsemastani aiheesta on kirjoitettu kuitenkin saatavilla lähdemateriaalia yli puolen vuosisadan ajalta, joten suurimmaksi ongelmaksi saattaa muodostua se että onko jo kirjoittamani koodi tarpeeksi laadukasta.