

# Määrittelydokumentti

## Tietorakenteiden harjoitustyö

Joosua Laakso

22. joulukuuta 2013

### Työn aihe

Olen valinnut toteutettavan työn aiheeksi säännöllisten lausekkeiden tulkin. Säännölliset lausekkeet ovat olennainen osa teoreettista tietojenkäsittelytiedettä. Säännöllinen lauseke kuvaa äärellistä tila-automaattia jolle syötetään merkkijono jonka automaatti joko hyväksyy tai hylkää riippuen siitä, kuuluuko merkkijono säännöllisen lausekkeen määrittelemään säännölliseen kieleen.

### Ohjelman käyttö

Ohjelmalle annetaan syötteenä säännöllisen lausekkeen määrittelevä merkkijono, sekä merkkijono jonka säännöllinen lauseke joko hylkää tai hyväksyy.

### Toteutus

Annetusta säännöllisen lausekkeen merkkijonoesityksestä muodostetaan ensin sitä vastaava jäsennyspuu. Varsinaisen toiminnan toteutukseen aijon käyttää Brzozowskin säännöllisten lausekkeiden derivaattoja.<sup>1</sup> Ideana säännöllisten lausekkeiden derivaatoissa on se, että lauseke derivoidaan jonkin kyseessä olevan säännöllisen kielen kirjaimen suhteen, jonka tuloksena on uusi säännöllinen lauseke, joka hyväksyy kaikki merkkijonot jotka alkoivat derivoidulla kirjaimella, ilman kyseistä kirjainta. Esimerkiksi jos säännöllinen lauseke  $(ab)|c$  derivoidaan kirjaimen  $a$  suhteen, tuloksena on uusi säännöllinen lauseke  $b$ . Derivaattoja voidaan käyttää testaamaan, kuuluuko jokin merkkijono johonkin säännölliseen kieleen siten, että muodostetaan säännöllisen kielen hyväksyvä säännöllinen lauseke ja derivoidaan säännöllinen lauseke järjestyksessä jokaisen merkkijonon kirjaimen suhteen. Jos tuloksena on säännöllinen lauseke joka hyväksyy tyhjän merkkijonon, alkuperäinen merkkijono kuuluu säännöllisen lausekkeen määrittelemään säännölliseen kieleen.<sup>2</sup>

### Tila-automaatin muodostus

Ohjelman toimintaa on mahdollista tehostaa siten, että säännöllisestä lausekkeesta muodostetaan kielen tunnistava deterministinen äärellinen tila-automaatti.

<sup>1</sup>[dl.acm.org/citation.cfm?id=321249](http://dl.acm.org/citation.cfm?id=321249)

<sup>2</sup>[maniagnosis.crsr.net/2012/02/practical-regular-expression.html](http://maniagnosis.crsr.net/2012/02/practical-regular-expression.html)

On olemassa algoritmi, joka muodostaa tehokkaan tila-automaatin säännöllisestä lausekkeesta hyödyntäen Brzozowskin derivaattoja automaatin muodostuksessa.<sup>3</sup> Erityisesti suurilla syötteillä tällainen menettelytapa saattaa olla ajallisesti mitattuna edullinen. Säännöllisistä lausekkeista muodostettuja tila-automaatteja voisi myös jollain menetelmällä tallentaa välimuistiin, vähentäen tila-automaatiksi kääntämisestä aiheutuvaa tehokkuuden menetystä.

## Työn tavoitteet

Työn ensisijaisena tavoitteena on tuottaa toimiva toteutus algoritmista, joka tunnistaa säännöllisen lausekkeen määrittelemän säännöllisen kielen. Minimitalavoite on tuottaa ohjelma joka tukee ainakin säännöllisten lausekkeiden perusoperaatioita eli tähteä, katenaatiota ja unionia. Lisätoiminnallisuutena voi lisätä muita suosittuja operaatioita, kuten plus-operaattorin ja määrällisen toiston.

## Tehokkuus

Työn päätavoite on saada aikaan toimiva toteutus. Tehokkuuden kannalta toivon kuitenkin että saisin aikaan toteutuksen jonka käyttämä aika ei kasvaisi räjähdysmäisesti, vaan olisi aikavaativuudeltaan polynominen syötteen ja säännöllisen lausekkeen pituuksien kasvaessa, jonka pitäisi olla teoriassa mahdollista.<sup>4</sup>

---

<sup>3</sup>[www.mpi-sws.org/~turon/re-deriv.pdf](http://www.mpi-sws.org/~turon/re-deriv.pdf)

<sup>4</sup>[perl.plover.com/NPC/](http://perl.plover.com/NPC/)