

Säännöllisten lausekkeiden tulkki

Säännöllinen lauseke tarkoittaa merkkijonoa, jonka avulla pyritään etsimään tekstistä lausekkeen ehdot täyttäviä merkkijonoja. Säännöllisiä lausekkeitä ja työssä tarvittavia epädeterministisiä äärellisiä käsitellään tarkemmin kurssilla laskennan mallit.

Työssä toteutetaan ohjelma, joka osaa rajallisesti tulkita Java 7:n mukaisia säännöllisiä lausekkeitä.^[1.] Tavoitteena on kyetä tulkitsemaan mahdollisimman monta Java 7:n ahneista kvanttoreista ja kaikki loogiset operaattorit. Ohjelmaan lisätään ylimääräistä Java 7:n mukaista toiminnallisuutta, jos aikarajoitteet antavat myöten.

Valitsin aiheen, koska halusin oppia lisää säännöllisten lausekkeiden toiminnasta. Säännöllisiin lausekkeisiin olin törmännyt aiemmin ohjelmoinnin jatkokurssilla (MOOC), jossa niitä käsiteltiin tosin vain pintapuolisesti. Sen lisäksi pääsen jo etukäteen tutustumaan laskennan mallit -kurssin aihepiiriin, josta olen kiinnostunut.

Tulkki perustuu Ken Thompsonin esittelemään algoritmiin^[2.], jonka avulla rakennetaan säännöllisestä lausekkeesta epädeterministinen äärellinen automaatti. Automaatti pystyy päättelemään onko annettu merkkijono säännöllisen lausekkeen mukainen ja palauttaa true tai false tuloksen mukaan.

LÄHTEET:

<http://docs.oracle.com/javase/7/docs/api/java/util/regex/Pattern.html> [1.]

http://en.wikipedia.org/wiki/Thompson's_construction_algorithm [2.]