

## Määrittelydokumentaatio: Säännöllisten lauseiden tulkki

### Yleistä:

Olen työssäni tarvinnut jo muutaman kerran säännöllisiä lauseita. Yleensä olen googlettanut, copy-pastennut ja rukoillut että lauseet toimivat ilman että olen ymmärtänyt että mitä oikeasti tapahtuu. Esim. useaan otteeseen olen tarvinnut tietoa html-sivulta ja aina käyttänyt samaa (+.?) säännöllistä lausetta kopioimaan tekstin tietyn tagin väliltä. Toinen säännöllinen lause mihin olen törmännyt on ".\*", ilmeisesti se tarkoittaa että sana saa jatkua miten vaan tms. Säännöllisiin lauseisiin olen törmännyt ainakin QTP-testiautomaatiotyökalussa ja JMeter-kuormitustestaustyökalussa. Olisi varmaan korkea aika ymmärtää mitä nämä säännölliset lauseet ihan oikeasti tarkoittavat.

Mitä algoritmeja ja tietorakenteita toteutat työssäsi

Alan purkamaan säännöllisten lauseiden tulkkia osoitteesta:

<http://www.regular-expressions.info/reference.html>

Aloitin kohdasta Character.

Ongelmaksi tulee tuon escape merkin toteutus, koska se saattaa olla literaali tai sitten sillä voi olla oma merkityksensä. Ongelma on että merkkejä \-merkin jälkeen voi olla vaikka miten pitkä, siellä voi olla aaltosuluissa lisämääreitä esim. `\N{U+263D}`. Joten vaihtoehtoja on paljon, miten tämän saisi tehokkaimmin hoidettua.

Seuraavaksi tulee Character set jne. Varmaan ongelmaksi tulee että miten saa nämä toimimaan yhdessä.

### Yleisesti:

Kaikki säännöllisen lauseiden merkit eivät vielä tuttuja mutta jos sulut/hakasulut tarkoittavat suunnilleen samaa kuin matematiikassa niin varmaan jonkinlaista pinoa joutuu käyttämään selvittämään mitkä kohdat katsotaan ensin.

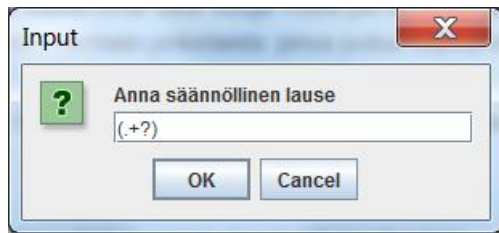
Taulukoita varmasti tarvitsee merkitsemään mitä eri merkit tarkoittivat ja taulukoita varmasti tarvitsee myös tietojen tulostamista varten.

Mitä ongelmaa ratkaisit ja miksi valitsit kyseiset algoritmit/tietorakenteet

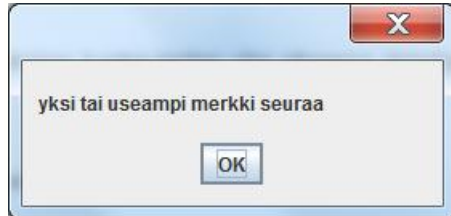
Pino toimii hyvin sellaisten ongelmien kanssa jossa pitää tietää missä järjestyksessä asiat pitää käydä läpi eli jos sulkulauseita.

Taulukoita tarvitsee jotta voi tallentaa kunkin tiedon ylös oikeassa järjestyksessä ja sitten tulostaa vastaukset.

Mitä syötteitä ohjelma saa ja miten näitä käytetään



Käyttäjää antaa säännöllisen lauseen Inputtina ja ohjelma palauttaa selityksen outputtina



Tavoitteena olevat aika- ja tilavaativuudet (m.m. O-analyysi):

Pinoa kuvaavaa taulukkoa täytyy varmaan aina aika-ajoin kasvattaa kaksinkertaiseksi kun jono loppuu tulee aikavaatimukseksi  $O(n)$ . Yksittäisen push- tai pop-kutsun arvoksi tulee lähes  $O(1)$ , paitsi tietty sille joka kasvattaa taulukkoa.

Tilaa pino vie syötteen verran eli  $O(n)$

Säännölliset lauseet eivät ole itselleni vielä aivan tuttuja joten voi olla että muitakin tietorankenteita tarvitaan. Tämä selviää projektin vanhetessa. Samoin voi olla että lauseita joutuu kelaamaan useampaan suuntaan joten aika- ja tilavaatimus voi olla paljon pahempi.

Kaikenkaikkiaan algoritmin pitäisi siis selvittää operaatioista  $O(n)$  aika- ja tilavaatimuksilla.

Lähteet:

<http://www.regular-expressions.info/tutorial.html>

<http://www.regular-expressions.info/reference.html>

<http://perldoc.perl.org/perlre.html#Escape-sequences>