

Ensimmäinen viikkoraportti

Aiheen valinta

Valitsin aiheekseni suosittelualgoritmin. Ensimmäisellä viikolla olen pääasiassa suunnitellut ja valmistellut tätä algoritmia. Ideanani on suositella asioita käyttäjälle useamman eri muuttujan perusteella. Ensimmäinen näistä muuttujista on ns. *collaborative filtering*, jossa käyttäjien tekemiä arvosteluja vertaillaan toisiinsa ja tällöin voidaan päätellä mistä tuotteesta käyttäjä A tykkää käyttäjän B perusteella. Käyttäjä B löydetään vertailemalla kaikkia käyttäjiä ja selvittämällä mikä on käyttäjän A lähin naapuri arvostelujen perusteella. Toisena muuttujana käytetään asioiden samankaltaisuutta, eli sitä liittyvätkö eri asiat toisiinsa - samankaltainen lähin naapuri ratkaisu. Tässä vertailu tehdään painoarvojen perusteella, jokaisella asialla on "*quality*" eli jokin ominaisuus, tälle ominaisuudelle on annettu painoarvo, joka määrittää kuinka kriittinen tämä ominaisuus on sen arvostelun kannalta. Esimerkiksi ruuassa "suolainen" on painoarvoltaan suurempi ominaisuus kuin esimerkiksi "porkkana".

Laajennettavuus

Lisäksi algoritmiin on upotettu kaksi vaihtoehtoista ominaisuutta. Ensimmäinen on kategoriat, joka mahdollistaa tulosten suodatuksen eri kategorioiden sisällä. Toinen ominaisuus on arvostelujen painoarvon laskeminen niiden iän mukaan, mutta nämä ominaisuudet ovat toissijaisia verrtuna ensimmäisen kappaleen toimintoihin.

Huolen aiheet suunnittelussa

Ainut merkittävä huoli tässä vaiheessa on algoritmin tehokkuus. Mikäli käyttäjien määrä on pilvissä ja jokaisella on satoja arvosteluja voi tuloksen tuottaminen olla hidasta. Tähän tulen pureutumaan myöhemmin tutkiessani mahdollisia pullonkauloja.

Toteutus

Aloittelin algoritmin toteuttamista luomalla muutaman perustason luokan: Item, Quality, User, Rating, Category, Value ja abstrakti luokka nimeltä AbstractElement sekä kirjoitin JavaDocit niille. Lisäksi loin Junit-testit: CategoryTest, ItemTest, QualityTest, RatingTest, StarTest, UserTest ja ValueTest. En kirjoittanut testeille JavaDocia, koska en usko, että niitä tarvitaan.

Suunnitelmat

Toteutan seuraavaksi muutaman yksinkertaisen generaattorin testidatalle, jonka avulla voin alkaa myöhemmin toteuttaa algoritmini laskennallista puolta. Lisäksi tarvitsen luokan, joka muistaa mitä dataa on tällä hetkellä tallennettu ja se onkin prioriteettini ensiviikolle. Tulevaisuuden suunnitelmana on myös tarkastella mahdollisuuksia ArrayListin korvaajaksi tietorakenteena.