

Viikkoraportti 1

Työn edistyminen

Tällä viikolla olen aloittanut projektin kanssa tyoskentelyn. Aiheenani on koneoppimisella toteutettava nimigeneraattori, jonka kirjoitan Java-kielellä. Olen valinnut algoritmit ja tietojärjestelmät, kuten määrittelydokumentissa sanotaan. En ole vielä kirjoittanut riviäkään koodia, mutta olen luonut projektin, lisännyt sen githubiin ja tehnyt suunnitelmia projektin etenemiseksi.

Opin

Olen käyttänyt gittiä aikaisemminkin, mutta siitä on jo useita kuukausia, joten kesti hetken palautella kaikki mieleen. Opin kuitenkin, kuinka tärkeä datan määrä koneoppimisessa on, joten päädyin muuttamaan alkuperäisen ideani (motivaatiolausegenerointi) joksikin vähän yksinkertaisemmaksi, josta on helpompi löytää riittävästi dataa.

Epäselvää

Ei vielä mitään. Muutamia kysymyksiä on, mutta lähinnä ne liittyvät algoritmeihin ja tietojärjestelmiin ja tiedän, että saan niihin vastauksen kun jähden tutkiskelemaan syvemmin

Seuraavaksi

Ensi viikolla aion käydä läpi materiaalia valittuihin algoritmeihin ja tietojärjestelmiin liittyen, ja ryhtyä soveltamaan sitä aiheeseeni. Haluaisin saada jonkinikäisen rungon/kokonaiskuvan valmiiksi.

Viikkoraportti 2

Työn edistyminen

Tällä viikolla sain aloitettua projektin virallisesti. Olen saanut aikaiseksi ne luokat, joita uskon tarvitsevani:

Kirjain: säilyttää kirjaimen tiedot ja puun tapaukselle, jossa kyseinen kirjain on “ensimmäinen”

Kirjaimet: Taulukko, joka huolehtii kaikista kirjaimista ja siitä, mikä kirjain on milloinkin ensimmäinen tai sen hetkisen tutkimuksen alainen kirjain. Myös hoitaa tallentamisen ja tekstitiedoston lukemisen.

Nimeäjä: Luokka, joka loppupelissä luo nimen. Kyselee kirjaimet- luokalta tietoa todennäköisyyksistä ja arpoo ne. Arpomista ei ole vielä painotettu

Opettaja: Luokka, jossa käydään läpi data ja lasketaan todennäköisyydet kullekin kirjaimelle ja kirjainparin jälkeiselle kirjaimelle. Luokilla on perus metodit tehtynä (Get, Set, Tallenna, Lue jne). Alla näkyy muutama alustava kaavio ja pseudokoodi projektiin liittyen

Opin

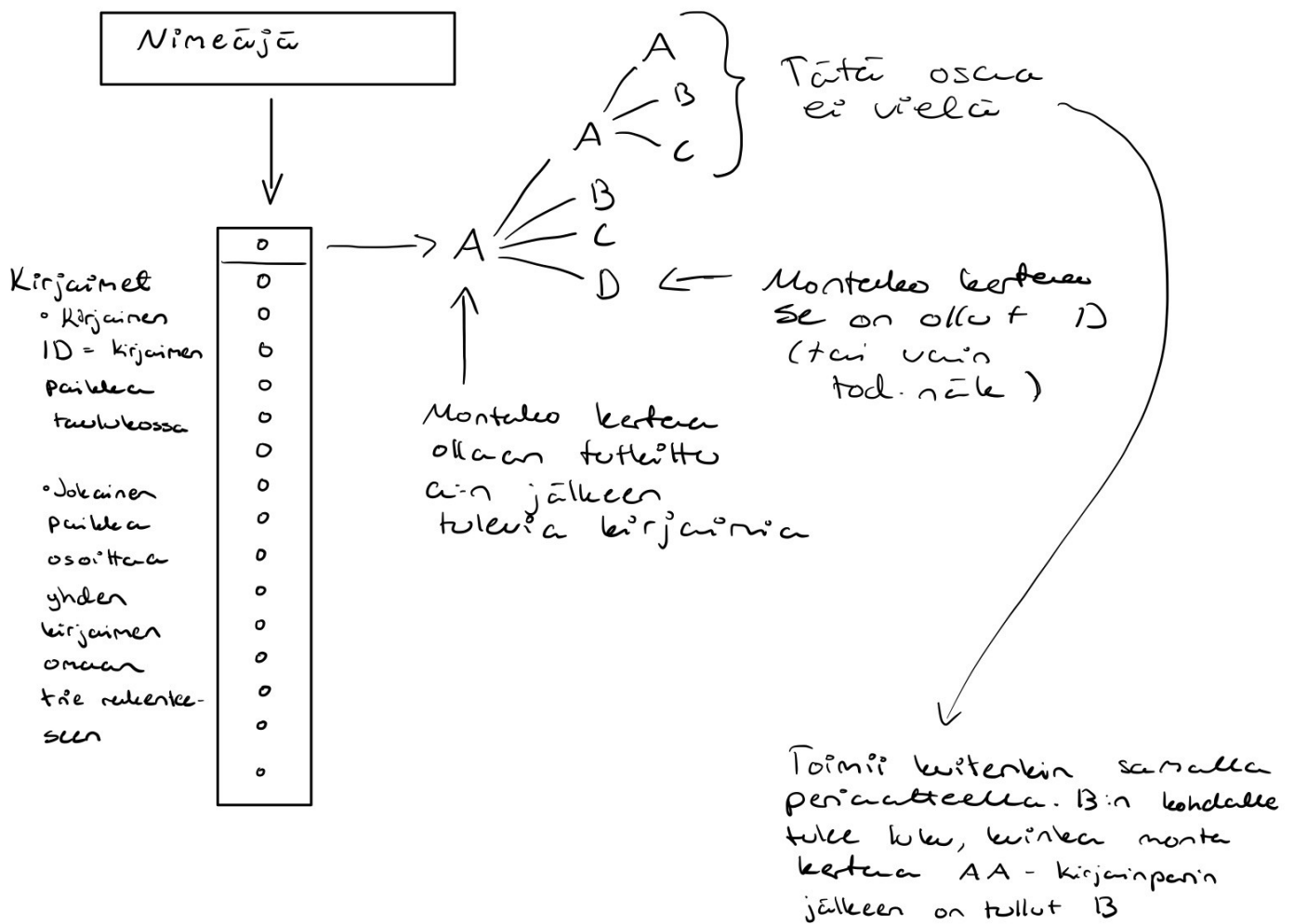
Opin, että Puu-rakennetta sanotaan myös trie-rakenteeksi. En ollut koskaan aikaisemmin kuullut termiä, joten se hämmensi hetken. Opin myös, että puurakennetta voi käyttää melko moneen asiaan.

Epäselvää

Ei ole suuria kysymyksiä tällä hetkellä. Yleisesti mietityttää, että olenko lähdössä ratkaisemaan tätä oikealla ja kaikista tehoikkaimmalla tavalla.

Seuraavaksi

Ensi viikolla työstän Opetus-luokkaa ja testailen kaikki muut luokat toimiviksi.



Pseudokoodi - runko

1. Ota kirjain (todennäköisyys ensimmäisessä kirjaimessa)
2. Mene kirjaimen puuhun ja arvo kirjain (todennäköisyydet painotettu)
3. Lisää kirjain string-jonoon
4. Aseta kirjain uudeksi puukirjaimeksi
5. Onko string yli 15 merkkiä pitkä (jos on, lopeta)
6. Katsotaan, lopetettuanko (todennäköisyydet väärä kirjain ja nimen pituus)
7. Jatka kohdasta 2.

↳ Kun tarjotun nimen tarjotaan "Hyväksyntä, kuulostaa suomalaiselta" ja valmis nimi pyydetään jonkin hyväksymis-protokollan läpi

Pseudokoodi opetus

1. Ota sana
2. +1 alkukirjain pinoon alkukirjaimelle
3. Seuraava kirjain +1 seuraavalle kirjaimelle kirjaimen puussa, +1 kokonaiskirjainten (tämän kirjaimen määrän)
4. Onko seuraava kirjainta, jos ei +1 viimeisen kirjain
ja +1 sanan pituudelle
5. Siirry seuraavaan ja jatka kohdasta 2

Viikkoraportti 3

Työn edistyminen

Tällä viikolla olin heiman jumissa trie-rakenteen kanssa, joten projekti eteni melko hitaasti.

Joiduinmyös tekemään uudelleen muutamia luokkia ja oliaota.

Nyt minulla on kuitenkin melko

selvät sävelet siitä, minne ollaan menossa ja miten. Tällä

viikolla keräsin myös opetusdatan kasaan

ja muokkasin sen luettavaan muotoon

Opin

Opin trie-raknteesta ja siitä, miten sen avulla voidaan

toteuttaa tämä projekti. Sain mahtavan ahaa-

elämyksen.

Epäselvää

Ei mitää, tällä hetkellä kaikki on melkoisen selvää.

Seuraavaksi

Ensi viikolla testailen koodia ja haluaisin saada nimen luojan

toimimaa. Sen lisäksi tallenna- ja lue-

metodit olisi hyvä saada toimimaan

Viikkoraportti 5

Työn edistyminen

Tällä viikolla sain tiedostoihin tallentamisen ja lukemisen aikaiseksi, avan kuten halusinkin. Kerkesin myös aloittaa opetus-luokaa ja sitä, mitä kaikkea siihen tarvitsee.

Opin

Palautin mieleeni, miten kirjoittaminen ja lukeinen toimivat, sekä miten saada järkevästi toString ja parse metodit toimimaan.

Epäselvää

Ei mitään epäselvyyksiä

Seuraavaksi

Ensi viikon iso urakka on saada opetuskoodi toimimaan halutulla tavalla ja käydä vihdoinkin koko nimiläjä läpi

Viikkoraportti 6

Työn edistyminen

Olen saanut algoritmin toimimaan hyvin. Se osaa luoda suomenkielen kuuloisia nimiä, kuten Erttari, Änämö ja Kasuliini. Lisäsin trie-olion joka osaa lisätä puuhunsa halutun pituisia merkkijonoja. Työ on siis periaatteessa valmis, mutta testit ja dokumentointi puuttuvat vielä. Lisäksi haluan tehdä pieniä hienosäätöjä, kuten kysyä käyttäjältä nimen pituutta, ensimmäistä alkukirjainta yms, sekä muokata niin, että ensimmäiset tai viimeiset kaksi kirjainta eivät voi olla kaksi kaksoiskonsonanttia. Haluaisin myös mahdollisesti testata eri kielillä, esim Englanniksi samaa, minkä pitäisi olla melko helppoa, vaihtaa vain nimien dataa siihen, mikä haluttu kieli on.

Opin

Opin tällä viikolla trie-rakenteesta vielä enemmän, ja pitkän debuggauksen jälkeen sain sen toimimaan halutulla tavalla.

Epäselvää

Ei epäselvyyksiä

Seuraavaksi

Seuraavaksi edessä hienosäätöä ja testailua

Viikko 7

Työn edistyminen

Tällä viikolla lisäsin toisen kielivaihtoehdon, joten nyt käyttäjä voi valita kahden kielen välillä (englanti tai suomi). Tällä viikolla ten myös joitakin testejä, en saanut vielä kaikkia testejä valmiiksi, joten se jatkuu ensi viikolla. Ensi viikolla pitäisi saada helposti kaikki valmiiksi, varsinkin kun muutama muu kurssi loppuu alkuviiikosta.

Opin

En oikeastaan mitään uutta, tämä viikko on ollut vain lisäystä ja täydentelyä, ei niinkään uusia konsepteja

Epäselvää

Ensiviikolla pitäisi palauttaa mielutiten jar-tiedostona, mutta en ole ihan varma, miten tämä toimii. Olisin myös kiinnostunut siitä, miten tämän saisi toimimaan niin, että ihmisellä, jolla ei ole eclipseä (jolla minä tämän tein) voisi vain avata sen. Se ei välttämättä kuulu kurssin osa-alueisiin, mutta olisin silti kiinnostunut siitä, enkä ole varma, mistä sitä tietoa voisi lähteä hakemaan

Seuraavaksi

Ensi viikolla pitää valmistella kaikki tarvittavat dokumentit, viimeistellä testaus ja jos jää aikaa, hienosäätää koodia. Haluaisin esimerkiksi muokata niin, että tietyille kielille pätee tiettyjä sääntöjä, kuten että Suomenkielinen nimi ei voi alkaa tai loppua kahdella samalla konsonantilla (esim Kkonnanu on nimi, jonka algoritmi antoi, enkä haluaisi kahta K-kirjainta alkuun). Mutta tämä tehdään vain, jos jää aikaa. Tämä kurssi on ollut erittäin mielenkiintoinen, ja saattaa olla, että jatkan algoritmin kanssa leikkimistä myös kurssin jälkeen