

NodePieSpy

<https://github.com/tuhoojabotti/NodePieSpy>

1. Sisällys

- [1. Sisällys](#)
- [2. Johdanto](#)
- [3. Ominaisuudet](#)
 - [3.1. Päättelyheuristiikat](#)
 - [3.1.1. Vierekkäisyys](#)
 - [3.1.2. Suora puhuttelu](#)
 - [3.1.3. Epäsuora maininta](#)
 - [3.2. Historia](#)
 - [3.3. Ajallinen rappeutuminen](#)
 - [3.4. Monen kanavan tuki](#)
 - [3.5. Analysointialgoritmit](#)
 - [3.5.1. BFS-puu](#)
 - [3.5.2. Ryvästäminen](#)
- [4. Ongelmia](#)
 - [4.1. Nimimerkkien vaihtelu](#)
 - [4.2. Yksityisviestit](#)

2. Johdanto

NodePieSpy on työkalu, jonka avulla voi analysoida sosiaalisia verkkoja. Ohjelmisto toimii tällä hetkellä vain irssin (<http://irssi.org/>) lokien perusteella, mutta se on tulevaisuudessa laajennettavissa esimerkiksi Facebook- tai Github-suhteiden kanssa toimivaksi.

Tällä hetkellä ohjelmisto koostuu kahdesta komponentista. NodePieSpy on palvelinohjelma, joka hoitaa tiedon lataamisen ja käsittelyn, ja tarjoaa HTTP:n yli rajapinnan, jolla tietoja voi pyytää JSON-formaatissa. NodePieSpy-Client on HTML-sivu, joka käyttää kyseistä rajapintaa ja visualisoi tulokset. Tämä käyttöliittymä ei ole harjoitustyön osana, vaikka sitä tullaankin käyttämään projektin testauksessa ja esittelyssä.

Projektin inspiraationa toimii PieSpy: <http://www.jibble.org/piespy/>

3. Ominaisuudet

Tässä osiossa määritellään kaikki erilaiset ominaisuudet, mitä projektiin olen suunnitellut.

3.1. Päättelyheuristiikat

Tärkein osa koko ohjelmistossa on itse tietojen jäsentäminen sosiaaliseksi verkoksi. Tiedot tulevat Irssi IRC-asiakasohjelman lokitiedostoista.

3.1.1. Vierekkäisyys

Peräkkäiset viestit irkissä kuuluvat yleensä yhteen, joskin on myös todennäköistä, että keskustelu vaihtuu, joten tämän heuristiikan painoarvon tulee olla aika pieni.

Parannuksena tähän voisi ottaa huomioon myös viestien ajallisen merkityksen.

3.1.2. Suora puhuttelu

Mikäli viestin alussa on jonkun nimimerkki, kyseessä on hänelle tarkoitettu viesti.

3.1.3. Epäsuora maininta

Mikäli viestissä mainitaan joku nimimerkki, mutta se ei ole viestin alussa, on kyseessä epäsuora maininta.

3.2. Historia

Olisi kiva kerätä historiaa, miten sosiaalinen verkosto muuttuu ajan myötä ja tehdä siitä vaikka animaatio.

3.3. Ajallinen rappeutuminen

Suhteiden tulisi rappeutua ajan myötä, jotta verkosta ei tule liian iso ja sekava.

3.4. Monen kanavan tuki

Olisi kiva nähdä yhtä aikaa monen kanavan verkko. Esimerkiksi Helsingin yliopiston kurssien irc kanavien muodostama verkko.

3.5. Analysointialgoritmit

Minkälaista tietoa tästä verkosta saa sitten irti.

3.5.1. BFS-puu

Generoimalla BFS-puu jostain käyttäjästä, selviää monen henkilön kautta jokaisella irkkaajalla on yhteys muihin kanavalla irkkaaviin. BFS:n generoiminen onnistunee ajassa $O(|V| + |E|)$.

3.5.2. Ryvästäminen

Olisi mukava nähdä millaisia ryhmiä irkkaajat muodostava keskenään. Nämä ongelmat ovat NP-täydellisiä, mutta jokin yksinkertaisempi algoritmi lienee riittävän hyvä, sillä verkon rakennekin perustuu olettamuksiin.

Markov Cluster -algoritmillä ryvästäminen onnistuu ajassa $O(|V|^3)$.

4. Ongelmia

Mitä erilaisia ongelmia voi tulla vastaan, mitä voi ottaa huomioon—tai sitten ei.

4.1. Nimimerkkien vaihtelu

Irkissä nimimerkkiä voi vaihtaa todella helposti ja se on aika yleistä. Olisi hyvä seurata näitä muutoksia. Lisäksi käyttäjällä voi olla useita yleisiä nimimerkkejä, joten olisi hyvä olla lista eri vaihtoehtoista, jotka kuitenkin lasketaan samaksi. Esimerkiksi yhteysvirheen yhteydessä nimimerkin perään voi joutua laittamaan jotain, koska vanha nimimerkki ei ole vielä vapautunut.

4.2. Yksityisviestit

Yksityisviestien analysointi olisi todella hyvä tiedonlähde, mutta näitä tietoja ei oikein voi saada ja se olisi yksityisyydelle vähän ikävää.