

1. JOHDANTO: TRAAPIJA

Järjestelmän tarkoitus

Traapijan tarkoitus on kerätä twiittejä reaaliaikaisesti ja historiallisesti, mutta [Twitterin API:n](#) rajoitukset huomioiden käytännössä "ryöpyissä".

Tiivis kuvaus siitä mistä on kyse

Verkkoselaimesta käytettävä Twitter-ryömiä "Traapija", jolle voi antaa tiettyjä hashtageja kerättäväksi joko tiettyyn aikarajaukseen perustuen tai reaaliaikaisesti. Sovellus ei ainakaan beta-versiona tule olemaan avoimessa verkossa, vaan se on tarkoitettu suljettuun käyttöön pienille tutkimustiimeille.

Millaisen toiminnan tukemiseen järjestelmä on tarkoitettu.

Datajournalistien ja tutkijoiden käyttöön.

Mitkä ovat järjestelmän tavoitteet

Kerätä twiittejä tietokantaan.

Toteutus-/toimintaympäristö

Itse ryömiä kirjoitetaan hyödyntäen [muiden esimerkkejä](#) ja Pythonin [Beautiful Soup -kirjastoa](#) ja [Cronjobsia](#) "ryöppyjen" ajastamiseksi. Verkkosivu toteutetaan käyttäen HTML:ää ja CSS:ää [Bootstrapilla](#) sekä mahdollisesti [Jekyllillä](#) (HTML) tai [Twigillä](#) (PHP). Mallit, näkymät ja kontrolleri kirjoitetaan PHP:llä kurssin ohjeiden mukaisesti.

Missä ympäristössä työ toteutetaan (yleensä laitoksen users-palvelimella Tomcat- tai Apache- palvelimen alla)

Laitoksen users-palvelimella.

Täytyykö web-sovelluksen alustajärjestelmän tukea jotain tiettyä ohjelmointikieltä. (esim. Java, Ruby, PHP..?)

Mahdollisesti Pythonia ja PHP:ta, en ole vielä varma missä määrin sovellus tulee vaatimaan selaintukea näille kielille vaikka käytän niitä.

Täytyykö käyttäjän selaimen tukea jotain tiettyä ohjelmointikieltä (esim. javascript?)

En ole vielä varma missä itse ryömiä toimii, käyttäjän koneella vai sovelluksen käytössä olevilla palvelimilla. Jos käyttäjän koneella niin koneen on tuettava Pythonia. Muuten mahdollisesti javascriptia ja jQuerya, jos päädyn tuunaamaan graafista käyttöliittymää näillä.

Edellyttääkö ohjelmisto jonkun tietyn tietokannan käyttöä vai voiko sitä vaihtaa helposti. Useimmat työt toimivat vain yhdellä kannalla

Tarkoitus olisi, että ohjelmisto voidaan siirtä toimimaan myös toisilla tietokannoilla. Esimerkiksi datajournalismiin erikoistunut [Knightlabs ohjeistaa](#) säilömään Twitter-ryömiöiden keräämää dataa [MongoDB](#)-tietokannassa.

2. KÄYTTÖTAPAUKSET

A. Käyttäjät

Tilaa työ

Käyttäjä voi tehdä sovellukselle tilauksen työstä, eli pyytää ryömiä noutamaan dataa tietyin rajauksin.

Kirjaudu sisään

Sovelluksen on oltava suojassa sisäänkirjauksen takana monesta syystä. Ensimmäkin sillä kerätty data on arvokasta eikä ole mitään syytä laskea sitä avoimeen verkkoon ilman saatesanoja tai seulaamista. Toiseksi Twitterin API:ssa on nykyään varsin rajalliset mahdollisuudet hakea dataa: käyttäjä saa vain haetusta objektista riippuen tietyn määrän dataa joka 15 minuutti. Siksi on selvää, että Traapija

Rajaa hakua

Rajauksiin sisältyvät toistaiseksi vain hashtagit. Lisäksi käyttäjä voi määritellä miltä ajalta twittejä haetaan. Jos twittejä haetaan "strömaten" niin takarajaa ei tarvita, vaan työ keskeytetään käsin. Jos twittejä haetaan tietyltä ajanjaksolta on ajanjaksolle määriteltävä alku- ja loppuajankohta.

Lataa tietokanta, tilaa "näkymä"

Käyttäjät voivat tilata valmiista tietokannasta "näkymiä" tiettyjen rajausten perusteella tai ladata tietokannan kokonaisuudessaan CSV- ja XLS-muodoissa.

Poista tilaus

Käyttäjät voivat itse poistaa aloittamansa tilaukset ja vastaanottamansa valmiit tietokannat.

B. Ylläpito

"Traapijalla" ei välttämättä ole ylläpitoa, sillä käyttäjäkunta tulee olemaan aluksi hyvin pieni (yksi tai muutama henkilö). Käyttötapauksissa on silti eritelty miten ylläpito toimisi jos sellainen sisällytettäisiin järjestelmään.

Luo uusi käyttäjä

Traapijaan ei voi rekisteröityä ulkopuolisena, vaan ylläpito luo käyttäjät omasta takaa ja jakaa käyttöoikeuksia valitsemilleen henkilöille tai yhteisöille.

Vastaanota ilmoitus uudesta työstä sähköpostiin

Jotta järjestelmä ei ylikuormitu, ylläpidon on hyväksyttävä kaikki tilaukset. Uusista tilauksista tulee ylläpidolle sähköposti-ilmoitus. Tilaukset hyväksytään Traapijan sisällä.

Hyväksy työ

Ylläpidon on hyväksyttävä kaikki työtilaukset.

